Engineering Chemistry

Second Edition

About the Author

O G Palanna obtained his PhD in solid-state chemistry from IIT Bombay, in 1975. He has been working as Professor and Head in the Department of Chemistry, RNS Institute of Technology, Bangalore, since 2004.

He has 47 years of teaching experience in engineering chemistry, during that period he was a professor and HOD of the Department of Chemistry, St Philomena College, Puttur, Karnataka, where he taught chemistry for 35 years. He was the Director (Academics), Vice-Principal, Professor and HOD of chemistry, AMC Engineering College, Bangalore, for 4 years. He was also a professor and HOD of chemistry, East-West College of Engineering, Bangalore for a year.

His fields of research are materials science, electrochemistry, rechargeable lithium ion cells, fuel cells, super-capacitors and corrosion science. He has a number of research papers to his credit in national and international journals. He has also authored a few books in chemistry.

Engineering Chemistry

Second Edition

O G Palanna

Former Professor and R & D Head
Department of Chemistry
R N S Institute of Technology
Bangalore, Karnataka

McGraw Hill Education (India) Private Limited
CHENNAI

McGraw Hill Education Offices
Chennai New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

McGraw Hill Education (India) Private Limited

Published by McGraw Hill Education (India) Private Limited
444/1, Sri Ekambara Naicker Industrial Estate, Alapakkam, Porur, Chennai - 600 116

Engineering Chemistry, 2e

This edition can be exported from India only by the publishers,
McGraw Hill Education (India) Private Limited

Print Edition:
ISBN-13: 978-93-5260-577-4
ISBN-10: 93-5260-577-2

E-Book Edition:
ISBN-13: 978-93-5260-609-2
ISBN-10: 93-5260-609-4

Managing Director: *Kaushik Bellani*

Director—Science & Engineering Portfolio: *Vibha Mahajan*
Lead—Science & Engineering Portfolio: *Hemant Jha*

Content Development Lead: *Shalini Jha*
Content Developer: *Shehla Mirza*

Production Head: *Satinder S Baveja*
Sr. Manager—Production: *Piyaray Pandita*

General Manager—Production: *Rajender P Ghansela*
Manager—Production: *Reji Kumar*

Typeset at The Composers, 260, C.A. Apt., Paschim Vihar, New Delhi 110 063 and printed at

Cover Printer:

Visit us at: www.mheducation.co.in

Dedicated to
my loving wife Prema Sarojini Palanna for her inspiration, support
and encouragement throughout my academic excellence
and for what I am today.

Contents

Preface to the Second Edition

As a result of good feedback and encouragement from students and teaching faculty from various technical universities, I am delighted to present the second edition of 'Engineering Chemistry' book with the addition of few more topics of prime interest. The previous edition was highly appreciated and used as a text/reference book to most of the Technical universities of India. The book is widely distributed among the popular universities such M. G. University, Kerala; CUSAT, Cochin; Dr. Abdul Rahman University, Calicut; VTU, Karnataka; Pune University, Pune; Karunya University, TN; Osmania University, Hyderabad; Chitkara University; Mahamaya Technical University, Noida; Chandigarh University, Gharaun; Dehradun Institute Technology (DIT), Dehradun; Dr. Mahalingam University; Guru Govind Singh University, Delhi; I.I.T, Mandi; Indraprastha University, Delhi; Jamia Millia Islamia University, Delhi; K L University, Andhra Pradesh; N.I.T, Goa; Tripura University, Tripura; Sri Sai University, Himachal Pradesh and host of other Engineering colleges in India.

I am very much delighted and wrote few more chapters which are added to this new edition in addition to the publication of its first edition in (2009). The chapters included in the second edition of the book are: (1) Nanoscience and Technology, (2) Lubrication, (3) Thermal Analysis, (4) Dynamics of Electro-analytical Methods, and (5) Supercapacitors.

This book is written in a simple language with logical flow of topics along with relevant concepts, theories, explanations, illustrations and applications. I look forward to receiving the feedback from both—the teaching faculty and students for the further improvement of the book. I express my wholehearted thanks to the staff of McGraw Hill Education, especially to M/S Vibha Mahajan, Shalini Jha and Koyal Ghosh for their sustained interest in the publication of this book. My deep sense of appreciation is for Ms Piyali Chatterjee and Piyaray Pandita for their constant support, help and co-operation.

May 2017 **O G Palanna**

Preface

Introduction to the Course

It is desirable that every student of Engineering need to acquire profound knowledge of the basics and applications of Engineering Chemistry. The Engineering Chemistry book is designed to fulfil the requirements of learning of the basic concepts, theory and technical applications. The book is exclusively best suited for I/II year semester students of B.E/B.Tech of any technical universities of India. The entire 23 chapters written in the book cover almost all the topics related to engineering aspects of Chemistry.

Target Audience

Since, the book is widely accepted for I/II year semester students of B.E/B.Tech of any technical universities of India, the book emphasizes on the basic concepts of chemistry and lays stress on the fundamental principles and motivates every students to acquire profound knowledge for real applications.

Objective of the Revision

The main objective of revision is to incorporate prime interest topics like 'Nanoscience and Nanotechnology', 'Thermal Analysis' and 'Supercapacitor'; which are added to the first edition. In addition, two new chapters—'Lubrication' and 'Dynamical Analytical Methods' are included in the second edition of the book. Minor changes are effected in Chapter 1 and Chapter 8 to update the syllabi of technical universities.

Roadmap to Target Course

The content of the second edition of the book contains 23 chapters which will cater to the needs of student community of any technical universities in India and abroad. The chapters included in the book have logical flow, comprehensive coverage of topics, exhaustive, illustrative, simple and clear to understand for any student. The book is ideally suited for I/II semester students of B.E/B.Tech of any technical university of India.

Salient Features

Five new chapters are included in the second edition of book. The following below mentioned reviewers have stressed their report on the style of writing. The new chapters have been reviewed by professors from various states of India.

We would like to thank the following reviewers who went through the text and provided valuable suggestions and comments. Their names are given as below:

Dr. (Mrs.) A. A. Kittur	*SDM College of Engineering & Technology, Dharwad*
Dr. Priyanka Sharma	*GD Goenka University, Gurugram*
Dr. Shilpi Agarwal	*University of Petroleum and Energy Studies, Dehradun*
Dr. Sarita Khaturia	*Mody University, Sikar*
Dr. Manoj kumar sahani	*Kailash Institute of Pharmacy and Management, Gida*
Dr. Sunil Verma	*Rawal Institute of Engineering and Technology, Faridabad*
Dr. Sharmila Pokharna	*Anand International College of Engineering, Jaipur*
Dr. Deeksha Ranjan	*Rama University, Kanpur*
Dr. Hema Singh	*Allenhouse Institute of Technology, Kanpur*

Dr. Atul Kumar Srivastava	*Pranveer Singh Institute of Technology, Kanpur*
Dr. Kushal Qanungo	*Mody University, Sikar*
Dr. Hiremath Suresha Babu	*Rao Bahadur Y Mahabaleshwarappa Engineering College, Cantonment Bellary*
Dr. Mohd. Shahnawaz Khan	*J K Lakshmipat University, Mahapura*

The comments of reviewers are summarized as follows:

1. **Written Language:** A clear and simple lucid language is used in writing the chapters of the book so that an average student can understand all the topics easily.
2. **Rigour:** The author gives the student clear and concise information to understand the various concepts of Chemistry to Engineering applications. The author did sincere efforts to make chapters more informative and the concept and applications were described in detailed.
3. **Pedagogy:** The book is written with a sound pedagogical features. A proper balance between theoretical and applied aspects has been well maintained. The book contains a wide range of topics, comprehensive, rich in fundamental/conceptual based problems and illustrative with solved examples.
4. **Approach:** The style of writing of the basic concepts of topics of the book is precise and good. A simple and holistic contextual approach of writing was executed with a large number of illustrations, flow charts, tables and diagrams.

What is New in the Second Edition

A few minor changes are made in Chapter 1. Various types of renewable (nonconventional) energies are highlighted in Chapter 1. In addition, 'Boiler Water', Boiler Feed Water, Causes of Corrosion and Softening of Boiler Feed Water are discussed in Chapter 8. Topics related to DTA, TGA, DSC and Cyclic Voltametry are deleted from Chapter 9 since, 'Thermogravimetry' and 'Coulometry' are dealt in different chapters.

Apart from these, all other topics of Chapters 1–18 are retained without a change. In the second edition of the book, five new chapters are added to the book. Chapter 19 deals with Nanoscience and Nano-technology. This chapter encompasses entire nanosized materials, their synthesis and applications to various fields. The topics such as carbon nanotubes, graphene, fullerenes, nanoclusters, nanohorns, nanofibres, nanowires, nanocomposites, etc. are discussed in detail along with their applications. Chapter 20 deals with topics related to lubrication of liquids, solids and semisolids and their applications to industry. Three important techniques of thermal analysis such as TGA, DTA and DSC are discussed in Chapter 21. Detailed theory, working and applications of these three methods of thermal analysis are discussed in Chapter 21. Chapter 22 deals with two important dynamical analytical methods and their applications. The basic working principle and applications of 'Electrogravimetry' and 'Coulometry' are discussed in detail. A brief discussion of supercapacitors is presented in Chapter 23.

Organization of the Book

Chapter 1 introduces the different sources of energy and types of fuels. Concept of calorific value and knocking are also discussed in this chapter. In **Chapter 2,** electrochemical energy systems are treated exhaustively. It discusses types of electrochemical cells, types of electrodes, and determination of pH, using different electrodes. **Chapter 3** deals mainly with battery and fuel cell technology. Basic concepts of batteries, classification of batteries, along with battery characteristics are explained. The portion on fuel cell technology covers the classification and types of fuel cells. Then follows **Chapter 4** on corrosion science in which types of corrosion, electrochemical theory of corrosion, and galvanic series are discussed. This chapter also deals with corrosion control in detail. **Chapter 5** introduces to metal finishing and its technological importance. The concept of electroplating is treated thoroughly in this chapter. **Chapter 6**, on liquid crystals, deals with the classification of liquid crystals, their chemical constitution and behavior. The technology of liquid crystal display is also explained in the chapter.

Then follows **Chapter 7** on high polymers in which mechanism of different types of polymerization reactions is illustrated. It also discusses the synthesis, properties and applications of various commercial polymers. **Chapter 8** introduces to sources of drinking water and water analysis. Concepts of hardness of water, BOD, COD and different methods of water purification are discussed in detail. **Chapter 9** focusses on instrumental methods of analysis—colorimetry, potentiometry, conductometry, and different spectroscopic methods. **Chapter 10** covers the major topics of solid state chemistry. It mainly includes classification of solids, crystal structure of solids, and basic symmetry operations. It also introduces the reader to the very important concept of powder X-ray diffraction. Further, **Chapter 11** takes the reader through the electrical properties of solids. New trends, scope and importance of biotechnology are dealt in **Chapter 12**.

Chapter 13 explains environmental pollution, its types and causes. Environmental hazards such as acid rain, ozone depletion, global warming, water and land pollution are also discussed in great details. **Chapter 14** takes up the phase rule with phase diagrams of one-component and two-component systems. **Chapter 15** thoroughly covers the concept of chemical thermodynamics—terminology, types of systems, thermodynamic properties, and laws of thermodynamics with their applications. Further, **Chapter 16** deals with chemical kinetics—rate of reaction, factors affecting reaction rate, order and molecularity of reaction, and collision model. **Chapter 17** on refractories explains the classification and properties of refractories. Various refractory materials are also discussed in the chapter. **Chapter 18** on abrasives covers the types and classification of abrasives, various abrasive products and their applications.

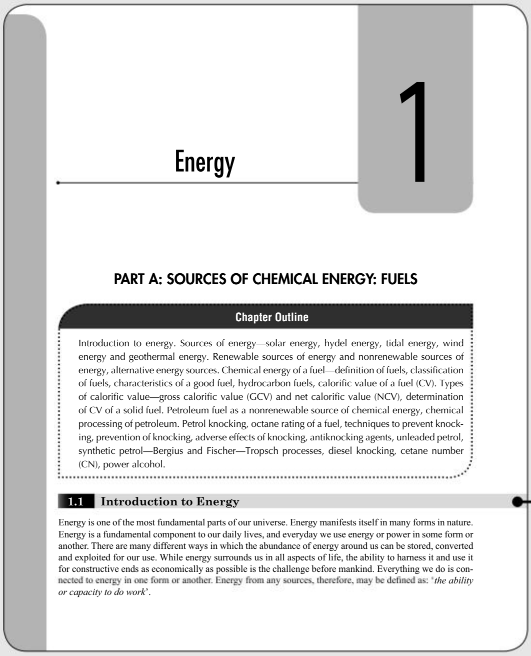
Chapter 19 deals with the important topic of nanotechnology. It majorly covers nanomaterials, their classification and types; carbon nanotubes; nanostructured bulk materials; quantum dots, their fabrication and applications; organic nanoparticles; and nanocomposite. The topic of lubrication is discussed next in **Chapter 20**. Lubricants, their requirement, types and functions; lubrication mechanism; solid and gaseous lubricants; different classifications of lubricants are covered well in the chapter. **Chapter 21** explains various thermal analysis techniques—TGA, DTA, and DSC. The basic principle, instrumentation, and applications of these techniques are dealt. **Chapter 22** deals with dynamic electroanaytical methods. It extensively covers electrogravimetry and coulometry. Finally, the last chapter of the book, **Chapter 23**, deals with the types of supercapacitors, their applications, pseudocapacitors, and hybrid supercapacitors.

Acknowledgements

I express my wholehearted thanks to all the staff of McGraw Hill Education, especially to M/S Vibha Mahajan, Shalini Jha and Koyal Ghosh for their sustained interest in the publication of this book. My deep sense of appreciation is for Ms Piyali Chatterjee and Piyaray Pandita for their constant support, help and co-operation.

May 2017 **O G Palanna**

Visual Walkthrough



1

Energy

PART A: SOURCES OF CHEMICAL ENERGY: FUELS

Chapter Outline

Introduction to energy. Sources of energy—solar energy, hydel energy, tidal energy, wind energy and geothermal energy. Renewable sources of energy and nonrenewable sources of energy, alternative energy sources. Chemical energy of a fuel—definition of fuels, classification of fuels, characteristics of a good fuel, hydrocarbon fuels, calorific value of a fuel (CV). Types of calorific value—gross calorific value (GCV) and net calorific value (NCV), determination of CV of a solid fuel. Petroleum fuel as a nonrenewable source of chemical energy, chemical processing of petroleum. Petrol knocking, octane rating of a fuel, techniques to prevent knocking, prevention of knocking, adverse effects of knocking, antiknocking agents, unleaded petrol, synthetic petrol—Bergius and Fischer—Tropsch processes, diesel knocking, cetane number (CN), power alcohol.

1.1 Introduction to Energy

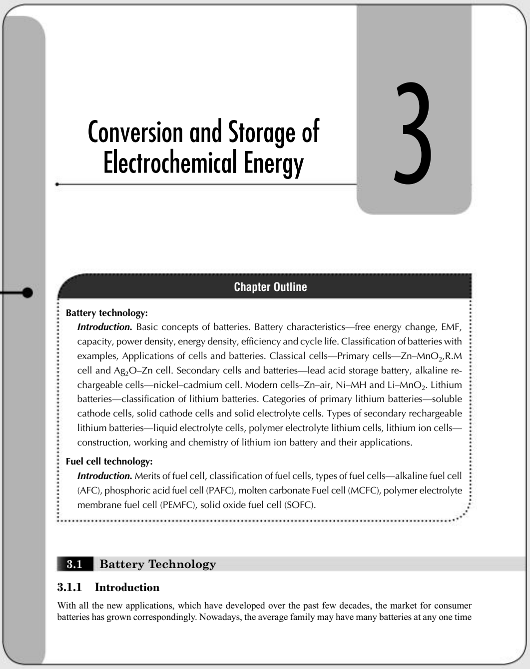
Energy is one of the most fundamental parts of our universe. Energy manifests itself in many forms in nature. Energy is a fundamental component to our daily lives, and everyday we use energy or power in some form or another. There are many different ways in which the abundance of energy around us can be stored, converted and exploited for our use. While energy surrounds us in all aspects of life, the ability to harness it and use it for constructive ends as economically as possible is the challenge before mankind. Everything we do is connected to energy in one form or another. Energy from any sources, therefore, may be defined as: '*the ability or capacity to do work*'.

Introduction

Each chapter begins with an introduction which highlights the text which is presented in a lucid manner.

Chapter Outline

The depth of topics covered is highlighted in the chapter before the introduction and discussion of the contents.



3

Conversion and Storage of Electrochemical Energy

Chapter Outline

Battery technology:

Introduction. Basic concepts of batteries. Battery characteristics—free energy change, EMF, capacity, power density, energy density, efficiency and cycle life. Classification of batteries with examples, Applications of cells and batteries. Classical cells—Primary cells—Zn–MnO_2,R.M cell and Ag_2O–Zn cell. Secondary cells and batteries—lead acid storage battery, alkaline rechargeable cells—nickel–cadmium cell. Modern cells–Zn–air, Ni–MH and Li–MnO_2. Lithium batteries—classification of lithium batteries. Categories of primary lithium batteries—soluble cathode cells, solid cathode cells and solid electrolyte cells. Types of secondary rechargeable lithium batteries—liquid electrolyte cells, polymer electrolyte lithium cells, lithium ion cells—construction, working and chemistry of lithium ion battery and their applications.

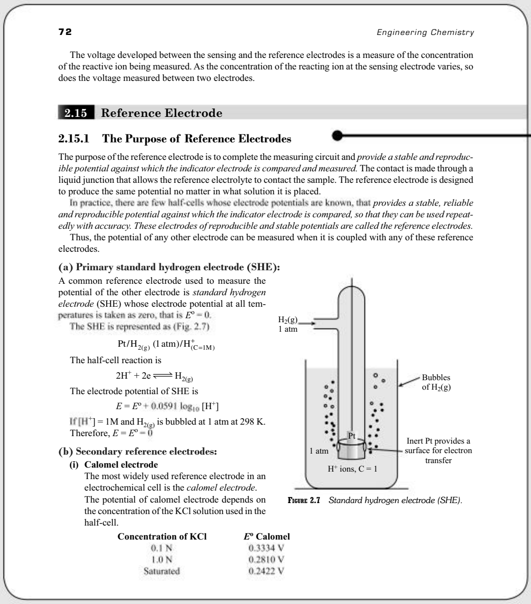
Fuel cell technology:

Introduction. Merits of fuel cell, classification of fuel cells, types of fuel cells—alkaline fuel cell (AFC), phosphoric acid fuel cell (PAFC), molten carbonate Fuel cell (MCFC), polymer electrolyte membrane fuel cell (PEMFC), solid oxide fuel cell (SOFC).

3.1 Battery Technology

3.1.1 Introduction

With all the new applications, which have developed over the past few decades, the market for consumer batteries has grown correspondingly. Nowadays, the average family may have many batteries at any one time



72 *Engineering Chemistry*

The voltage developed between the sensing and the reference electrodes is a measure of the concentration of the reactive ion being measured. As the concentration of the reacting ion at the sensing electrode varies, so does the voltage measured between two electrodes.

2.15 Reference Electrode

2.15.1 The Purpose of Reference Electrodes

The purpose of the reference electrode is to complete the measuring circuit and *provide a stable and reproducible potential against which the indicator electrode is compared and measured.* The contact is made through a liquid junction that allows the reference electrolyte to contact the sample. The reference electrode is designed to produce the same potential no matter in what solution it is placed.

In practice, there are few half-cells whose electrode potentials are known, that *provides a stable, reliable and reproducible potential against which the indicator electrode is compared, so that they can be used repeatedly with accuracy. These electrodes of reproducible and stable potentials are called the reference electrodes.*

Thus, the potential of any other electrode can be measured when it is coupled with any of these reference electrodes.

(a) Primary standard hydrogen electrode (SHE):

A common reference electrode used to measure the potential of the other electrode is *standard hydrogen electrode* (SHE) whose electrode potential at all temperatures is taken as zero, that is $E^o = 0$.

The SHE is represented as (Fig. 2.7)

$$\mathrm{Pt/H_{2(g)}\ (1\,atm)/H^+_{(C=1M)}}$$

The half-cell reaction is

$$2H^+ + 2e \rightleftharpoons H_{2(g)}$$

The electrode potential of SHE is

$$E = E^o + 0.0591 \log_{10} [H^+]$$

If $[H^+] = 1M$ and $H_{2(g)}$ is bubbled at 1 atm at 298 K.
Therefore, $E = E^o = 0$

FIGURE 2.7 *Standard hydrogen electrode (SHE).*

(b) Secondary reference electrodes:

(i) Calomel electrode

The most widely used reference electrode in an electrochemical cell is the *calomel electrode*. The potential of calomel electrode depends on the concentration of the KCl solution used in the half-cell.

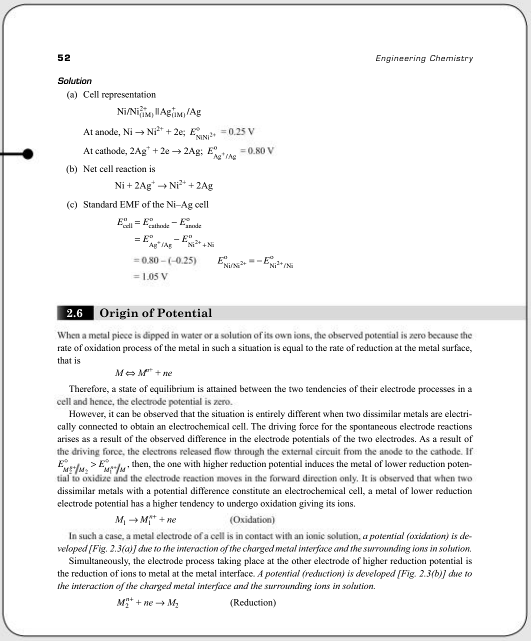
Concentration of KCl	E^o Calomel
0.1 N	0.3334 V
1.0 N	0.2810 V
Saturated	0.2422 V

Sections and Sub-sections

A logical flow of basics, concepts, theories etc are presented by introducing sections and subsections of topics covered in the chapters.

Solutions

Clear logical reasoning is stressed to explain the topics with exhaustive illustrations in all the chapters.



52 *Engineering Chemistry*

Solution

(a) Cell representation

$$\mathrm{Ni/Ni^{2+}_{(1M)} \| Ag^+_{(1M)}/Ag}$$

At anode, $Ni \rightarrow Ni^{2+} + 2e$; $E^o_{NiNi^{2+}} = 0.25$ V

At cathode, $2Ag^+ + 2e \rightarrow 2Ag$; $E^o_{Ag^+/Ag} = 0.80$ V

(b) Net cell reaction is

$$Ni + 2Ag^+ \rightarrow Ni^{2+} + 2Ag$$

(c) Standard EMF of the Ni–Ag cell

$$\begin{aligned} E^o_{cell} &= E^o_{cathode} - E^o_{anode} \\ &= E^o_{Ag^+/Ag} - E^o_{Ni^{2+}+Ni} \\ &= 0.80 - (-0.25) \qquad E^o_{Ni/Ni^{2+}} = -E^o_{Ni^{2+}/Ni} \\ &= 1.05\ V \end{aligned}$$

2.6 Origin of Potential

When a metal piece is dipped in water or a solution of its own ions, the observed potential is zero because the rate of oxidation process of the metal in such a situation is equal to the rate of reduction at the metal surface, that is

$$M \Leftrightarrow M^{n+} + ne$$

Therefore, a state of equilibrium is attained between the two tendencies of their electrode processes in a cell and hence, the electrode potential is zero.

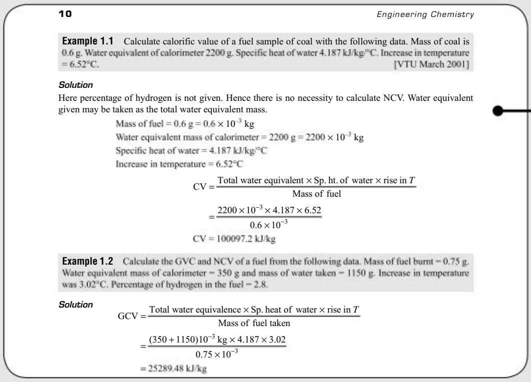
However, it can be observed that the situation is entirely different when two dissimilar metals are electrically connected to obtain an electrochemical cell. The driving force for the spontaneous electrode reactions arises as a result of the observed difference in the electrode potentials of the two electrodes. As a result of the driving force, the electrons released flow through the external circuit from the anode to the cathode. If $E^\circ_{M_2^{n+}/M_2} > E^\circ_{M_1^{n+}/M}$, then, the one with higher reduction potential induces the metal of lower reduction potential to oxidize and the electrode reaction moves in the forward direction only. It is observed that when two dissimilar metals with a potential difference constitute an electrochemical cell, a metal of lower reduction electrode potential has a higher tendency to undergo oxidation giving its ions.

$$M_1 \rightarrow M_1^{n+} + ne \qquad \text{(Oxidation)}$$

In such a case, a metal electrode of a cell is in contact with an ionic solution, *a potential (oxidation) is developed [Fig. 2.3(a)] due to the interaction of the charged metal interface and the surrounding ions in solution.*

Simultaneously, the electrode process taking place at the other electrode of higher reduction potential is the reduction of ions to metal at the metal interface. *A potential (reduction) is developed [Fig. 2.3(b)] due to the interaction of the charged metal interface and the surrounding ions in solution.*

$$M_2^{n+} + ne \rightarrow M_2 \qquad \text{(Reduction)}$$



10 Engineering Chemistry

Example 1.1 Calculate calorific value of a fuel sample of coal with the following data. Mass of coal is 0.6 g. Water equivalent of calorimeter 2200 g. Specific heat of water 4.187 kJ/kg/°C. Increase in temperature = 6.52°C. [VTU March 2001]

Solution

Here percentage of hydrogen is not given. Hence there is no necessity to calculate NCV. Water equivalent given may be taken as the total water equivalent mass.

Mass of fuel = 0.6 g = 0.6×10^{-3} kg

Water equivalent mass of calorimeter = 2200 g = 2200×10^{-3} kg

Specific heat of water = 4.187 kJ/kg/°C

Increase in temperature = 6.52°C

$$\text{CV} = \frac{\text{Total water equivalent} \times \text{Sp. ht. of water} \times \text{rise in } T}{\text{Mass of fuel}}$$

$$= \frac{2200 \times 10^{-3} \times 4.187 \times 6.52}{0.6 \times 10^{-3}}$$

$$\text{CV} = 100097.2 \text{ kJ/kg}$$

Example 1.2 Calculate the GVC and NCV of a fuel from the following data. Mass of fuel burnt = 0.75 g. Water equivalent mass of calorimeter = 350 g and mass of water taken = 1150 g. Increase in temperature was 3.02°C. Percentage of hydrogen in the fuel = 2.8.

Solution

$$\text{GCV} = \frac{\text{Total water equivalence} \times \text{Sp. heat of water} \times \text{rise in } T}{\text{Mass of fuel taken}}$$

$$= \frac{(350 + 1150)10^{-3} \text{ kg} \times 4.187 \times 3.02}{0.75 \times 10^{-3}}$$

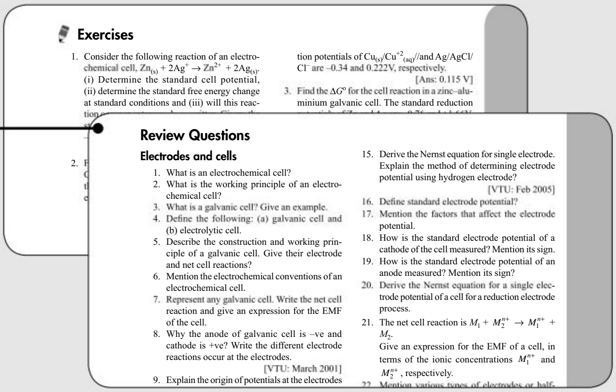
$$= 25289.48 \text{ kJ/kg}$$

Solved Examples

Uniqueness of the chapters are the presentation of numerous worked problems based on the concepts and principles of topics, which are solved and given in the book for proper comprehension of the topics.

Exercise Problems

The topics covered in the chapters are very exhaustive and is enriched with worked problems and exercises.



Exercises

1. Consider the following reaction of an electrochemical cell, $Zn_{(s)} + 2Ag^{+} \rightarrow Zn^{2+} + 2Ag_{(s)}$. (i) Determine the standard cell potential, (ii) determine the standard free energy change at standard conditions and (iii) will this reac-

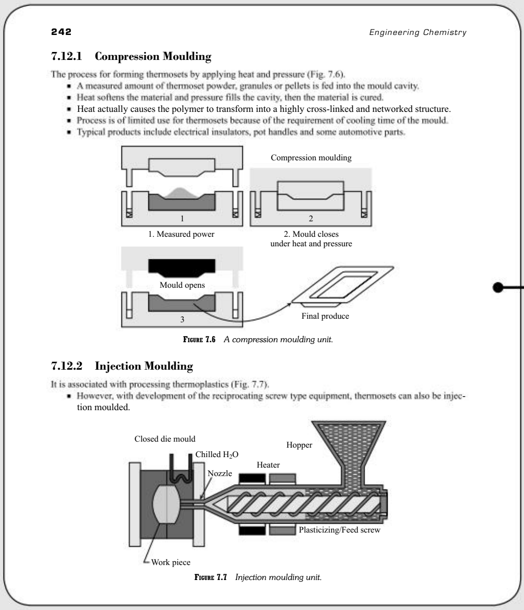
tion potentials of $Cu_{(s)}/Cu^{+2}_{(aq)}$//and Ag/AgCl/ Cl^{-} are –0.34 and 0.222V, respectively. [Ans: 0.115 V]

3. Find the ΔG^{o} for the cell reaction in a zinc–aluminium galvanic cell. The standard reduction

Review Questions

Electrodes and cells

1. What is an electrochemical cell?
2. What is the working principle of an electrochemical cell?
3. What is a galvanic cell? Give an example.
4. Define the following: (a) galvanic cell and (b) electrolytic cell.
5. Describe the construction and working principle of a galvanic cell. Give their electrode and net cell reactions?
6. Mention the electrochemical conventions of an electrochemical cell.
7. Represent any galvanic cell. Write the net cell reaction and give an expression for the EMF of the cell.
8. Why the anode of galvanic cell is –ve and cathode is +ve? Write the different electrode reactions occur at the electrodes. [VTU: March 2001]
9. Explain the origin of potentials at the electrodes
15. Derive the Nernst equation for single electrode. Explain the method of determining electrode potential using hydrogen electrode? [VTU: Feb 2005]
16. Define standard electrode potential?
17. Mention the factors that affect the electrode potential.
18. How is the standard electrode potential of a cathode of the cell measured? Mention its sign.
19. How is the standard electrode potential of an anode measured? Mention its sign?
20. Derive the Nernst equation for a single electrode potential of a cell for a reduction electrode process.
21. The net cell reaction is $M_1 + M_2^{n+} \rightarrow M_1^{n+} + M_2$.
Give an expression for the EMF of a cell, in terms of the ionic concentrations M_1^{n+} and M_2^{n+}, respectively.
22. Mention various types of electrodes or half-



242 Engineering Chemistry

7.12.1 Compression Moulding

The process for forming thermosets by applying heat and pressure (Fig. 7.6).

- A measured amount of thermoset powder, granules or pellets is fed into the mould cavity.
- Heat softens the material and pressure fills the cavity, then the material is cured.
- Heat actually causes the polymer to transform into a highly cross-linked and networked structure.
- Process is of limited use for thermosets because of the requirement of cooling time of the mould.
- Typical products include electrical insulators, pot handles and some automotive parts.

FIGURE 7.6 *A compression moulding unit.*

7.12.2 Injection Moulding

It is associated with processing thermoplastics (Fig. 7.7).

- However, with development of the reciprocating screw type equipment, thermosets can also be injection moulded.

FIGURE 7.7 *Injection moulding unit.*

Illustrations

Beautiful illustrative figures are included in all the chapters to understand the basic concepts of topics. There are nearly 150 such figures in the book.

1

Energy

PART A: SOURCES OF CHEMICAL ENERGY: FUELS

Chapter Outline

Introduction to energy. Sources of energy—solar energy, hydel energy, tidal energy, wind energy and geothermal energy. Renewable sources of energy and nonrenewable sources of energy, alternative energy sources. Chemical energy of a fuel—definition of fuels, classification of fuels, characteristics of a good fuel, hydrocarbon fuels, calorific value of a fuel (CV). Types of calorific value—gross calorific value (GCV) and net calorific value (NCV), determination of CV of a solid fuel. Petroleum fuel as a nonrenewable source of chemical energy, chemical processing of petroleum. Petrol knocking, octane rating of a fuel, techniques to prevent knocking, prevention of knocking, adverse effects of knocking, antiknocking agents, unleaded petrol, synthetic petrol—Bergius and Fischer—Tropsch processes, diesel knocking, cetane number (CN), power alcohol.

1.1 Introduction to Energy

Energy is one of the most fundamental parts of our universe. Energy manifests itself in many forms in nature. Energy is a fundamental component to our daily lives, and everyday we use energy or power in some form or another. There are many different ways in which the abundance of energy around us can be stored, converted and exploited for our use. While energy surrounds us in all aspects of life, the ability to harness it and use it for constructive ends as economically as possible is the challenge before mankind. Everything we do is connected to energy in one form or another. Energy from any sources, therefore, may be defined as: '*the ability or capacity to do work*'.

1.2 Sources of Energy

Energy source—a supply of energy available for use (sometimes referred to as resources). Primary energy sources are sources that are naturally available on the earth. Energy resources can be described as follows:

- Renewable energy
- Nonrenewable energy

1.2.1 Renewable Energy (or Nonconventional Energy)

Renewable energy is a source of energy that can never be exhausted. A resource which is generated from sources which are not finite or exhaustible. Renewable energy sources are recurring or continuous. Under the category of renewable energy or nonconventional energy are sources such as wave power, tides, wind power, solar power or geothermal energy, agricultural residue, firewood and animal dung.

(a) Solar Energy

Sun generates energy and is known as *solar energy*. This energy can be collected and converted in a few different ways. The range is from solar water heating with solar collectors or attic cooling with solar attic fans for domestic use to the complex technologies of direct conversion of sunlight to electrical energy using photovoltaic cells.

Solar energy now powers everything from portable radios to homes, stores and neighbourhood and is inexhaustible and cheaper.

Solar energy is radiant light and heat from the sun harnessed using a range of ever-evolving technologies such as solar heating, photovoltaics, solar thermal energy, solar architecture and artificial photosynthesis.

Photovoltaic (PV) cells convert sunlight to direct current (DC) electricity. The inverter converts DC into alternating current (AC) electricity. The electrical panel sends power to your lights and appliances.

Advantages of solar:

Solar energy is a resource that is not only sustainable for energy consumption, it is indefinitely renewable. Solar power can be used to generate electricity and it is also used in relatively simple technology to heat water (solar water heaters). Solar energy is a clean and renewable energy source. Once a solar panel is installed, solar energy can be produced free of charge. Solar energy causes no pollution. Solar cells make absolutely no noise at all.

(b) Hydel Energy

Flowing water generates energy that can be captured and transformed into electricity. This is called *hydroelectric power* or *hydropower*.

The most common type of hydroelectric power plant uses a dam on a river to store water in a reservoir. Water released from the reservoir flows through a turbine, spinning it, which in turn activates a generator to produce electricity. Energy from moving water is the largest source of renewable electricity. Hydroelectricity is the term referring to electricity generated by hydropower; the production of electrical power through the use of the gravitational force of falling or flowing water. Hydropower is one of the best, cheapest and cleanest sources of energy. It is a technology with enormous potential, which could exploit the water resources to supply energy to remote rural areas with little access to conventional energy sources. It also eliminates most of the negative environmental effects associated with large hydro projects.

(c) Tidal Energy

Tidal power, which is also known as *tidal energy*, is a form of hydropower that converts the energy obtained from tides into useful forms of power, mainly electricity. Although not yet widely used, tidal power has potential for future electricity generation. Tides are more predictable than wind energy and solar power.

Tidal energy is a form of hydropower that converts the energy of the tides into electricity or other useful forms of power. The tide is created by the gravitational effect of the sun and the moon on the earth causing cyclical movement of the seas. Tidal energy is therefore an entirely predictable form of renewable energy, which can be harnessed in two forms.

Tidal range is the vertical difference in height between the high tide and the succeeding low tide.

Artificial tidal barrages or lagoons may be constructed to capture the tide. Turbines in the barrier or lagoon generate electricity as the tide floods into the reservoir; water thus retained can then be released through turbines, again generating electricity once the tide outside the barrier has receded.

Among sources of renewable energy, tidal power has traditionally suffered from relatively high cost and limited availability of sites with sufficiently high tidal ranges or flow velocities, thus constricting its total availability.

Ocean thermal is energy derived from waves and also from tidal waves. Through the method of cogeneration, a cleaner and less polluting form of energy is being generated.

(d) Wind Energy

Wind energy is a source of renewable power which comes from air current flowing across the earth's surface. Wind turbines harvest this kinetic energy and convert it into usable power which can provide electricity for home, farm, school or business applications on small or large scales.

Wind energy is one of the fastest growing sources of electricity and one of the fastest growing markets in the world today. These growth trends can be linked to the multidimensional benefits associated with wind energy.

Today, a windmill (*wind turbine)* can be used to convert wind's energy to generate electricity. Wind turbines, like windmills, are mounted on a tower to capture the most energy. At 100 feet (30 m) or more above ground, they can take advantage of the faster and less turbulent wind. Turbines catch the wind's energy with their propeller-like blades. Usually, two or three blades are mounted on a shaft to form a rotor.

A blade acts much like an airplane wing. When the wind blows, a pocket of low-pressure air forms on the downwind side of the blade. The low-pressure air pocket then pulls the blade towards it, causing the rotor to turn. This is called lift. The force of the lift is actually much stronger than the wind's force against the front side of the blade, which is called drag. The combination of lift and drag causes the rotor to spin like a propeller, and the turning shaft spins a generator to make electricity.

Wind turbines can be used as standalone applications, or they can be connected to a utility power grid or even combined with a photovoltaic (solar cell) system. For utility-scale sources of wind energy, a large number of wind turbines are usually built close together to form a wind plant. Several electricity providers today use wind plants to supply power to their customers.

Standalone wind turbines are typically used for water pumping or communications. However, homeowners, farmers and ranchers in windy areas can also use wind turbines as a way to cut their electric bills.

Small wind systems also have potential as distributed energy resources. Distributed energy resources refer to a variety of small, modular power-generating technologies that can be combined to improve the operation of the electricity delivery system.

(e) Geothermal energy

Geothermal energy is derived from hot dry rocks, magma, hot water springs, natural geysers, etc. Geothermal energy taps the earth's internal heat for a variety of uses, including electric power production and the heating and cooling of buildings. Reservoirs of steam and hot water beneath the earth's surface hold enormous potential as a renewable energy resource.

Geothermal energy is created by harnessing geothermal energy from the earth. Heat from the centre of the earth conducts outwards and heats up the outer layers of rock called the mantle. When this type of rock melts and becomes molten it is called magma. Magma can reach just below the earth's surface. Rain water sometimes seeps down through geological fault lines and cracks becoming super-heated by the hot rocks below. Some of this super-heated water rises back to the surface of the earth where it emerges as hot springs or even geysers. Sometimes the hot water becomes trapped below the surface as a geothermal reservoir. One way of producing electricity from geothermal energy is by drilling wells into the geothermal reservoirs. The hot water that rises emerges at the surface as steam. The steam is used to drive turbines producing electricity. If the water is not hot enough to produce steam, it can still be used to heat homes and businesses, saving gas/electricity.

Advantages

1. Geothermal energy is relatively environmental friendly. Pollution in the form of fumes is not produced although usually drilling of the earth's surface takes place. The surrounding environment is not harmed with the exception of the land required for the power plant and transport links.
2. Unlike wind power, geothermal power can be relied on as it provides constant power.
3. The use of conventional polluting fuels such as oil and coal can be reduced if geothermal and other alternative energy forms are used (reducing pollution).
4. Geothermal power can take different forms. For instance, it can be used to produce electricity or the hot water can be used directly to heat homes and businesses.

(f) Bioenergy

Biomass—firewood, animal dung, biodegradable waste from cities and crop residues—is a source of energy when it is burnt.

We have used biomass energy or *bioenergy*—the energy from organic matter—for thousands of years, ever since people started burning wood to cook food or to keep warm.

And today, wood is still our largest biomass energy resource. But many other sources of biomass can now be used, including plants, residues from agriculture or forestry and the organic component of municipal and industrial wastes. Even the fumes from landfills can be used as a biomass energy source.

The use of biomass energy has the potential to greatly reduce our greenhouse gas emissions. Biomass generates about the same amount of carbon dioxide as fossil fuels, but every time a new plant grows, carbon dioxide is actually removed from the atmosphere. The net emission of carbon dioxide will be zero as long as plants continue to be replenished for biomass energy purposes. These energy crops, such as fast-growing trees and grasses, are called *biomass feedstocks.* The use of biomass feedstock can also help increase profits for the agricultural industry.

1.2.2 Nonrenewable Chemical Energy Sources (or Conventional Energy)

Nonrenewable energy comes from sources that cannot be replenished during a human's life. Some nonrenewable sources include fossil fuels (coal, oil, natural gas, propane) and uranium and plutonium used in nuclear reactors. Coal and petroleum are called fossil fuels, since they are supposed to have been formed over a million years by the decay of vegetable and animal matters buried under the earth due to the heat, pressure and biological action.

Fuels are one of the primary requisites of human beings. Earlier fuels were used mainly for cooking, heating and lighting. But modern civilization needs fuels for illumination, transportation, manufacturing and other applications. The main source of fuel was wood in earlier times. The discovery of the use of fossil fuels such as coal and petroleum actually led to industrial revolution.

1.3 Alternative Energy Sources

In the past century, it has been seen that the consumption of nonrenewable sources of energy has caused more environmental damage than any other human activity. Electricity generated from fossil fuels such as coal and crude oil has led to high concentrations of harmful gases in the atmosphere. This has in turn led to many problems being faced today such as ozone depletion and global warming. Vehicular pollution has also been a major problem.

Therefore, alternative sources of energy have become very important and relevant in today's world. These sources, such as the sun and wind, can never be exhausted and, therefore, are called renewable. They cause less emission and are available locally. Their use can, to a large extent, reduce chemical, radioactive and thermal pollution. They stand out as a viable source of clean and limitless energy. These are also known as nonconventional sources of energy. Most of the renewable sources of energy are fairly nonpolluting and considered clean though biomass, a renewable source, is a major polluter indoors.

1.4 What is Chemical Energy of a Fuel?

The combustion of any compound (fuel) containing carbon and other atoms in the presence of oxygen liberates enormous amounts of energy in the form of heat and this energy is referred to as chemical energy liberated from a fuel.

$$\text{Organic (carbonaceous) matter} + O_2 \rightarrow \text{product} + \text{heat}$$

Many people do not realize that of all the various kinds of energy and power that humanity has created over the past several centuries; chemical energy is the form of energy that is used most often. Chemical energy is possibly the easiest source of energy to obtain, and it is certainly the most efficient form of energy to use and to store. Chemical energy is known to be a source of life and source of power for many species for several billions of years. Recently, technology has advanced so that we can use chemical energy towards renewable energy sources that will affect our global future.

1.5 Definition of Fuels

The urge for better living and the growing industrialization has created a demand for energy as a fuel. Fuel may be defined as a naturally occurring or artificially prepared combustible carbonaceous material used mainly as a source of heat and light and in few cases as a source of raw material.

1.6 Classification of Fuels

Depending on the origin, fuels are classified as primary and secondary fuels (Fig. 1.1). They may be solid, liquid and gaseous depending on their physical state of existence.

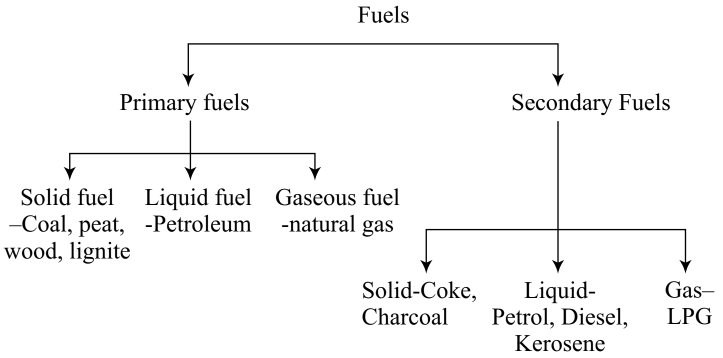


FIGURE 1.1 *Types of fuel.*

1.7 Characteristics of a Good Fuel

A good fuel has the following characteristics:

1. It should be easily available in the natural and or in any other form and cheap.
2. Storage, transportation and handling of the fuel should be easy.
3. A fuel should have low ignition temperature and burn efficiently. The ignition temperature of a hydrocarbon fuel depends on the percentage of hydrogen in it. Higher the percentage of hydrogen lower will be the ignition temperature.
4. There should be control for combustion rate which evidently depends on the chemical composition of the fuel. The rate of combustion also depends upon the availability of oxygen.
5. The level of emission of smoke should be low. Formation of smoke depends upon the percentage of carbon. Higher the percentage of carbon higher will be the smoke. Thus, aromatic hydrocarbons with high carbon contents produce more smoke.
6. There should be minimum of waste after the combustion of a fuel. Solid fuels leave behind ash as an undesirable waste. Ash affects the burning characteristics of the fuel and may also corrode the furnace part.
7. The fuel should have high calorific value. Heat energy produced by the combustion of unit mass of the fuel must be high.
8. The fuel should have low moisture content. High percentage of moisture increases the ignition temperature and also reduces calorific value.

1.8 Hydrocarbon Fuels

The main hydrocarbon fuels are coal, petroleum and natural gas. Naturally occurring coal is available in many forms in the solid state. Petroleum is the only primary liquid fuel and natural gas is the primary gaseous fuel. Coal is the main source of aromatic compounds while petroleum is the principal source of aliphatic compounds. *Petroleum* along with hydrocarbons contains organic compounds containing nitrogen, sulphur and oxygen. Hydrocarbon fuels contain various types of hydrocarbons which may be represented as C_xH_y where x and y are simple integers and x is called carbon number. Hydrocarbons are classified into various homologous series (see Table 1.1).

Strictly speaking, petroleum consists entirely of aliphatic hydrocarbons, those composed of hydrogen and carbon.

Name of the homologous series	General molecular formula	Examples
Alkanes	C_nH_{2n+2} where $n = 1, 2, 3, 4$	CH_4 Methane C_2H_6 Ethane
Alkenes	C_nH_{2n} where $n = 2, 3, 4$	C_2H_4 Ethene C_3H_6 Propene
Alkynes	C_nH_{2n-2} where $n = 2, 3, 4$, etc.	C_2H_2 Ethyne C_2H_4 Propyne
Cycloalkanes	C_nH_{2n} where $n = 3, 4$	C_3H_6 Cyclopropane C_4H_8 Cyclobutane
Aromatic hydrocarbons	C_nH_{2n-6} where $n = 6, 7, 8…..$	C_6H_6 Benzene C_7H_8 Toluene

Table 1.1

Hydrocarbons

- The four lightest hydrocarbons—CH_4 (methane), C_2H_6 (ethane), C_3H_8 (propane) and C_4H_{10} (butane)—are all gases, boiling at −107, –67, −43 and –18°C, respectively.
- The chains in the C_{5-7} range are all light, easily vaporized, clear liquid called naphthas. They are used as solvents, dry cleaning fluids and other quick-drying products.
- The chains from C_6H_{14} to $C_{12}H_{26}$ are blended together and used as petrol for cars, airplanes, etc. Kerosene is made up of chains in the C_{10} to C_{15} range, followed by diesel fuel/heating oil (C_{10}–C_{20}) and heavier fuel oils as the ones used in trucks and ship engines. These petroleum compounds are all liquid at room temperature.
- Lubricating oils and semisolid greases (including Vaseline®) range from C_{16} to C_{20}.
- Chains above C_{20} form solids, starting with paraffin wax, then tar and asphaltic bitumen.

1.9 Calorific Value of a Fuel (CV)

The *calorific value* of a fuel is measured in terms of the *heating efficiency* or the *performance* during its ignition in the presence of oxygen.

Calorific value of a fuel is defined as the quantity of heat evolved by the complete combustion of unit quantity (mass or volume) of the fuel in air or oxygen.

The unit of calorific value of a solid or liquid is cal/g, or kcal/kg or J/kg. When reporting the CV for a gaseous fuel, it is customary to refer the temperature and pressure of the gas along with the heat quantity per unit volume, since volume of a gas varies with temperature and pressure. Hence, the unit of CV of a gaseous fuel is expressed as kJ/m^3 at the given temperature and pressure.

Generally, calorific value of fuels is expressed in two ways:

(i) Gross/higher calorific value represented as GCV or HCV
(ii) Net (lower) calorific value represented as NCV or LCV

Gross/Higher Calorific Value (GCV)

Gross calorific value is defined as the quantity of heat evolved by the complete combustion of unit mass (a units volume) of the fuel in air or oxygen and the products of combustion are brought down to room temperature.

Explanation: Most of the fuels used by us are hydrocarbon fuels. During combustion, carbon becomes oxidized to CO_2 and hydrogen to steam. When the products of combustion are cooled to room temperature, the steam present condenses releasing its latent heat of vaporization. Hence, GVC is made up of the heat of combustion of the fuel plus the latent heat of condensation of steam.

Net Calorific Value (NCV)

Net calorific value is defined as the quantity of heat evolved by the complete combustion of unit mass (or unit volume) of the fuel and the products of combustion are let off into the atmosphere.

Explanation: If the combustion products, like steam and CO_2 of fuels, are allowed to escape into the atmosphere, the CV measured under these conditions does not include the latent heat of condensation of steam. The heat quantity so measured is called *net calorific value*.

Net calorific value = [GVC–latent heat of condensation of steam]

1.10 Determination of Calorific Value of a Solid Fuel

Bomb Calorimeter

A known mass of the solid or liquid fuel is completely burnt in excess of oxygen in a bomb calorimeter (Fig. 1.2).

The heat liberated is absorbed by surrounding water and brass calorimeter. The increase in temperature of the calorimeter and water is noted using a sensitive thermometer. Calorific value of the fuel is calculated using the principle that the heat lost by the burning fuel is equal to the heat gained by water and calorimeter.

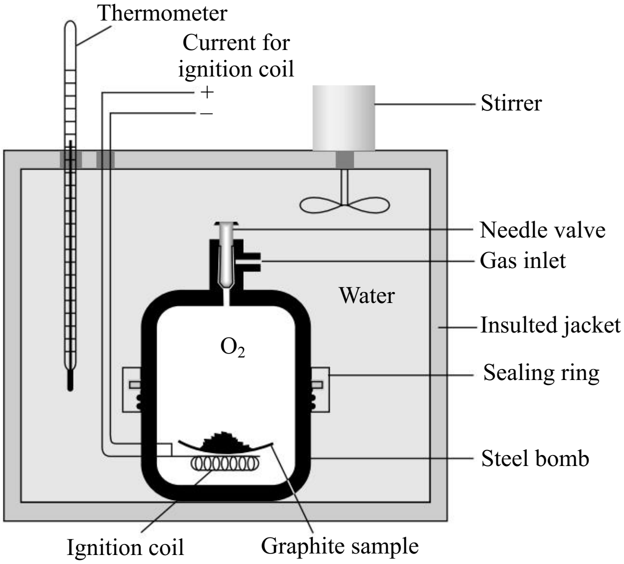


Figure 1.2 *A bomb calorimeter.*

Apparatus

A bomb calorimeter, shown in Fig. 1.2, consists of a thick steel cylinder or bomb, which can be closed with a needle valve. There is a gas inlet valve for oxygen. Two platinum wires are fixed in the cap for sending current for ignition coil. One of them carries a graphite/platinum cup. An ignition coil (magnesium wire) connects the cup and the other platinum wire. The bomb calorimeter is kept in a known quantity of water contained in another outer brass calorimeter. The brass calorimeter is kept inside a nonconducting jacket over ebonite (insulated) blocks. A sensitive thermometer and a stirrer are also provided and they are kept dipping in the water.

Working

A known weight of the fuel is kept in the platinum/graphite cup. The bomb is charged with oxygen up to 25 atmospheres using the oxygen valve. The apparatus is arranged as shown in Fig 1.2. The initial temperature of water is noted. The platinum wires are connected to a source of current and ignition of the fuel is initiated by passing a current through the ignition coil. The fuel completely burns and the heat evolved is absorbed by water in the outer calorimeter, which is constantly stirred. The final temperature of water is noted.

Calculation of GCV

Mass of the solid or liquid fuel = m kg

Mass of water taken = $w_1$kg

Water equivalent mass of calorimeter and other accessories = w_2 kg

Total water equivalent mass = $(w_1 + w_2)$ kg

Initial temperature = T_1K

Final temperature = T_2K

Increase in temperature = $(T_2 - T_1)$ K

Heat lost by 'm' kg of a fuel = [Heat gained by calorimeter + water]

= mass × Sp. heat × increase in temperature

$= (w_1 + w_2) \times 4.187 \times (T_2 - T_1)$ kJ

$$\text{Gross calorific value} = \frac{(w_1 + w_2) \times 4.187 \times (T_2 - T_1)}{m}$$

Note: The heat quantity calculated above is the gross calorific value because the products of combustion remain inside the calorimeter and brought down to ambient temperature.

Calculation of NCV

If the percentage of hydrogen in the fuel is known, the *net calorific value* of a fuel can be calculated. Let the percentage of hydrogen be x.

$$\underset{2\text{ kg}}{H_2} + \frac{1}{2}O_2 \rightarrow \underset{18\text{ kg}}{H_2O}$$

Since 2 kg of hydrogen produces 18 kg steam, x kg of hydrogen will produce $= \frac{18}{2} \times x$ kg of steam. That is, 100 kg of fuel which contains x kg of hydrogen will produce $9x$ kg of steam.

$$1 \text{ kg fuel produces} = \frac{9x}{100} = 0.09x \text{ kg steam}$$

Latent heat of steam = 587 × 4.187 kJ/kg

Heat liberated by condensation of $0.09x$ kg steam $= 0.09x \times 587 \times 4.187$ kJ

Therefore, NCV = GCV – heat of condensation of steam formed from 1 kg fuel

$= [\text{GCV} - 0.09x \times 587 \times 4.187]$ kJ/kg

Example 1.1 Calculate calorific value of a fuel sample of coal with the following data. Mass of coal is 0.6 g. Water equivalent of calorimeter 2200 g. Specific heat of water 4.187 kJ/kg/°C. Increase in temperature = 6.52°C. [VTU March 2001]

Solution

Here percentage of hydrogen is not given. Hence there is no necessity to calculate NCV. Water equivalent given may be taken as the total water equivalent mass.

Mass of fuel = 0.6 g = 0.6×10^{-3} kg

Water equivalent mass of calorimeter = 2200 g = 2200×10^{-3} kg

Specific heat of water = 4.187 kJ/kg/°C

Increase in temperature = 6.52°C

$$\text{CV} = \frac{\text{Total water equivalent} \times \text{Sp. ht. of water} \times \text{rise in } T}{\text{Mass of fuel}}$$

$$= \frac{2200 \times 10^{-3} \times 4.187 \times 6.52}{0.6 \times 10^{-3}}$$

$$\text{CV} = 100097.2 \text{ kJ/kg}$$

Example 1.2 Calculate the GVC and NCV of a fuel from the following data. Mass of fuel burnt = 0.75 g. Water equivalent mass of calorimeter = 350 g and mass of water taken = 1150 g. Increase in temperature was 3.02°C. Percentage of hydrogen in the fuel = 2.8.

Solution

$$\text{GCV} = \frac{\text{Total water equivalence} \times \text{Sp. heat of water} \times \text{rise in } T}{\text{Mass of fuel taken}}$$

$$= \frac{(350 + 1150)10^{-3} \text{ kg} \times 4.187 \times 3.02}{0.75 \times 10^{-3}}$$

$$= 25289.48 \text{ kJ/kg}$$

Heat evolved by condensation of water vapour = $0.09x \times 587 \times 4.187$ kJ/kg
where x is percentage of hydrogen higher.

$$= 0.09 \times 2.8 \times 587 \times 4.187$$
$$= 627.83 \text{ kJ/kg}$$

Therefore, NCV = GCV – heat of condensation of steam
= 25289.48 – 627.83
= 24,661.65 kJ/kg

Example 1.3 Calculate the gross calorific value of a coal sample from the following data:

Weight of coal sample taken = 5.5×10^{-3} kg
Weight of water taken in the calorimeter = 2.5 kg
Water equivalent of calorimeter = 0.5 kg
Initial temperature of water = 24°C
Final temperature of water = 28°C [VTU Feb 2005]

Solution

$$\text{Gross calorific value of coal sample} = \frac{\text{Total water equivalent} \times \text{sp. heat} \times \text{rise in temp}}{\text{Mass of coal}}$$

$$\text{CCV} = \frac{[0.5 + 2.5] \times 4.187 \times 4}{5.5 \times 10^{-3}}$$

$$= 9135.27 \text{ kJ/kg}$$

Example 1.4 0.6 g of a coal sample with 92 percent C, 5 percent hydrogen and 3 percent ash caused a rise in the temperature of 2000 g water by 3.2°C in a bomb calorimeter experiment. Calculate the gross and net calorific value of coal. Given water equivalent = 200 g, specific heat of water = 4.187 kJ/kg/°C, latent heat of steam = 587 cal/g (1 cal = 4.18 J). [VTU July 2005]

Solution

Given

Mass of coal taken = 0.60 g = 0.6×10^{-3} kg.

Hydrogen in coal sample = 5%

Rise in temperature of water = 3.2°C

Mass of water taken = 2000 g = 2.0 kg

Water equivalent of calorimeter = 200 g = 0.2 kg

Specific heat of water = 4.187 kJ/kg/°C

Latent heat of steam = 587 cal/g

= 587×4.187 kJ/kg

$$\text{Gross calorific value of coal} = \frac{(W_1 + W_2) \times \text{sp. heat} \times \text{rise in temperature}}{\text{Mass of coal}}$$

$$\text{GCV} = \frac{2.2 \times 4.187 \times 3.2}{0.6 \times 10^{-3}} = 49127.46 \text{ kJ/kg}$$

Heat of condensation of steam = $0.09x \times 587 \times 4.187$ kJ/kg

where x is the % of hydrogen contained in the sample.

= $0.09 \times 5 \times 587 \times 4.187 = 1104.147$ kJ/kg

Therefore net calorific value = 49127.46 – 1105.99

= 48021.47 kJ/kg

Example 1.5 The gross calorific value of a bituminous coal is 36,000 kJ/kg in an experiment, 0.83 g of this coal burnt under 1.2 kg of water in a bomb calorimeter. Due to combustion, the temperature of water increased by 3.92°C. Calculate the water equivalent of the calorimeter. Specific heat of water = 4.2 kJ/kg. [VTU Jan 2006]

Solution

Gross calorific value of bituminous sample = 36,000 kJ/kg

Mass of coal taken = 0.83 g = 0.83×10^{-3} kg

Mass of water = 1.2 kg

Rise in temperature = 3.92°C

Specific heat of water = 4.2 kJ/kg/°C

$$\text{Gross calorific value} = \frac{(W_1 + W_2) \times \text{sp. heat} \times \text{rise in temperature}}{\text{Mass of coal}}$$

$$= \frac{[1.2 + W_2] \times \text{sp. heat} \times \text{rise in temp.}}{\text{Mass of the sample}}$$

Therefore,

$$36{,}000 = \frac{[1.2 + W_2] \times 4.187 \times 3.92}{0.83 \times 10^{-3}} = [1.2 + 22{,}483] + [W_2 \times 22{,}483] = 26980 + [W_2 \times 22{,}483]$$

$$W_2 = \frac{[\text{GCV} \times \text{mass of sample taken} - W_1]}{\text{Sp. heat} \times \text{rise in temperature}} \;;\; W_2 = \frac{36{,}000 - 26980}{22{,}483} = 0.409 \text{ kg}$$

Example 1.6 On burning 0.75×10^{-3} kg of a solid fuel in a bomb calorimeter, the temperature of 2.5 kg water is increased from 24 to 28°C. The water equivalent of calorimeter and the latent heat of steam are 0.485 kg and 4.2×587 kJ/kg, respectively. Specific heat of water is 4.187 kJ/kg/°C. If the fuel contains 2.5 percent hydrogen, calculate its gross and net calorific values. [VTU July 2006]

Solution

Mass of fuel = 0.75×10^{-3} kg
Mass of water taken = 2.5 kg
Water equivalent of calorimeter = 0.485 kg
Latent heat of steam = 4.2×587 kJ/kg
Increase in temperature = 4°C

$$\text{Gross calorific value} = \frac{(W_1 + W_2) \times \text{sp. heat} \times \text{rise in temperature}}{\text{Mass of coal}}$$

$$= \frac{[2.5 + 0.485] \times 4.187 \times 4}{0.75 \times 10^{-3}}$$

$$= 66657.04 \text{ kJ/kg}$$

Heat of condensation of water vapour $= 0.09x \times 587 \times 4.187$

$$= 0.09 \times 2.5 \times 587 \times 4.187$$

$$= 552.998 \text{ kJ/kg}$$

Therefore, the net calorific value = GCV −heat of condensation of steam.

$$= 66657.0 - 553 = 66104.00 \text{ kJ/kg}$$

1.11 Petroleum Fuel as a Nonrenewable Source of Chemical Energy

Petroleum is called a *non-renewable* energy source because it takes millions of years to form. The term petroleum comes from the Latin term petra means 'rock' and oleum means 'oil'. It is used to describe a broad range of hydrocarbons that are found indifferent forms such as gases, liquids or solids beneath the surface of the earth. The two most common forms are natural gas and crude oil. Petroleum is a naturally occurring complex mixture made up predominantly of carbon and hydrogen compounds, but also frequently containing significant amounts of nitrogen, sulphur and oxygen together with smaller amounts of nickel, vanadium and other elements.

The systematic diagram of petroleum mining is shown in Fig. 1.3(a) and (b).

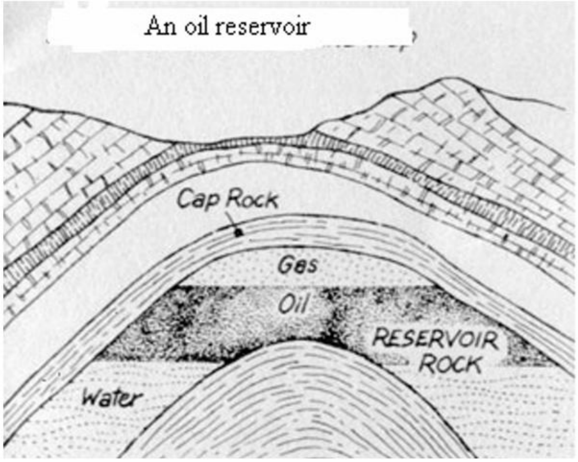


FIGURE 1.3 *(a) Petroleum reservoir.*

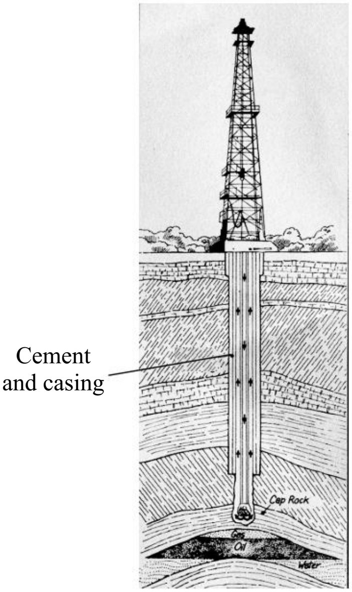


FIGURE 1.3 *(b) Petroleum rig work station.*

Crude oil is a complex mixture that is between 50 and 95 percent hydrocarbon by weight. The first step in refining crude oil involves separating the oil into different hydrocarbon fractions by distillation. Crude oil obtained from a refinery is a mixture of aliphatic hydrocarbons with dark colour and unpleasant odour. Impure petroleum is first freed from dirt and water. The expulsion of dissolved gases is carried out by heating the crude. Separation of the mixture of hydrocarbons is carried out by fractional distillation of crude using a fractionating column [Fig. 1.4(a) and (b)].

Fractionating column consists of a tall steel cylinder filled with trays (a) at various heights. Each tray contains a number of holes closed with bubble caps (B) with toothed edges. Purified crude oil is vaporized and sent up the fractionating column.

The temperature at the bottom of the column is around 600°C and as the height increases temperature becomes less and less. As the vapours rise up the column, the heavier and less volatile components condense and return to the lower tray. The heat released during the condensation vaporizes the more volatile components, which rise up the tower. As the vapours pass through the bubble cap, they blow through the already condensed liquid leading to further vaporization of the liquid and condensation of heavier fractions. Thus, repeated condensation and vaporization is achieved in the column. Crude oil thus gets separated into various fractions and collects at the trays at different heights in the column. Any excess liquid in a tray overflows through the downspout C into the lower tray. Each fraction is drained off through the respective outlet D. Each fraction actually consists of a mixture of hydrocarbons with a range of boiling points and often the boiling point ranges of successive fraction overlap. Each fraction may be further fractionated to get products of desired properties. The important fractions, their boiling point ranges, carbon content and applications are given in Table 1.2.

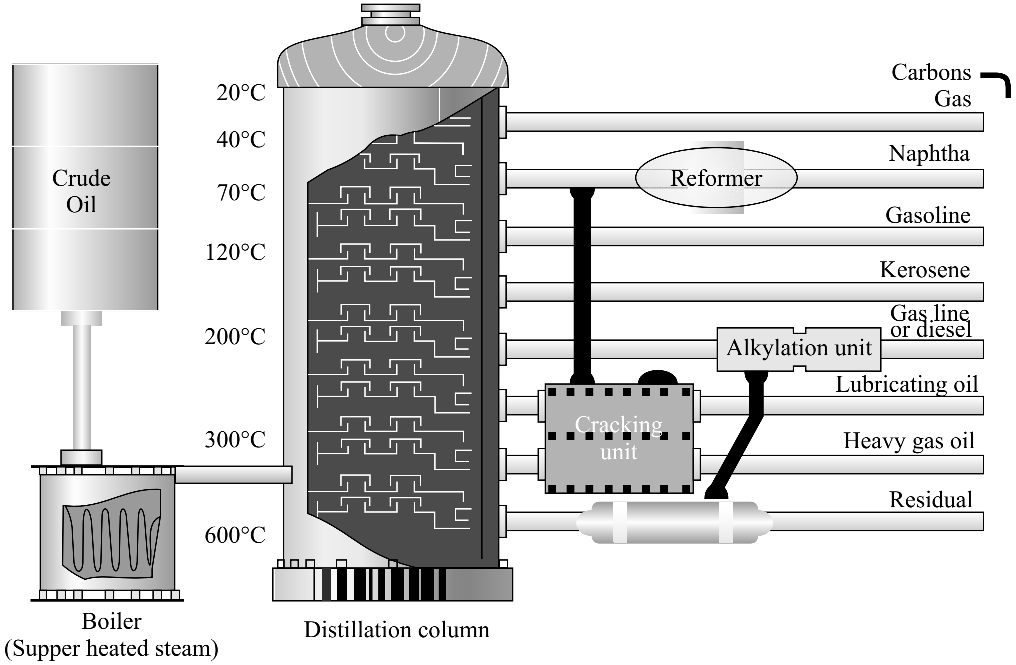


FIGURE 1.4 *(a) Fractionating columns with distillates of crude oil.*

Fraction	Boiling point range °C	Carbon number	Uses
Gas	Up to 20°C	C_1–C_4	LPG
Petroleum ether	20–70°C	C_5–C_6	Solvent in pharmaceuticals
Gasoline (petrol)	40–100°C	C_5–C_{10}	Fuel for petrol engines, aviation petrol
Naphtha	100–160°C	C_5–C_9	as solvent, dry cleaning liquid, etc.
Kerosene	160–230°C	C_{11}–C_{12}	Domestic fuel for illumination
Light gas oil	300–400°C	C_{18}–C_{25}	Lubrication, Cracking stock
Paraffin wax	Low melting	C_{25}–C_{40}	Wax, candles, matches, etc.
Asphalt (Solid residue)	–	–	Surfacing roads

Table 1.2

Products of primary distillation of crude oil

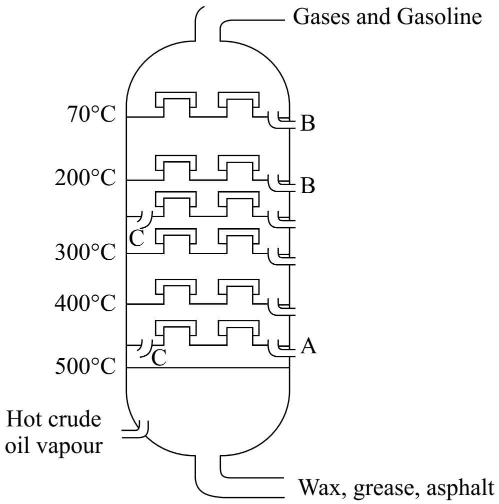


FIGURE 1.4 *(b) Fractionating column.*

1.12 Chemical Processing of Petroleum

Primary distillation of crude oil yields just about 20 percent of straight run petrol only. This quantity is not enough to meet the increasing need for petrol.

Another significant portion of refining is converting individual molecules into different sizes and shapes. There are three main conversion processes in the refinery:

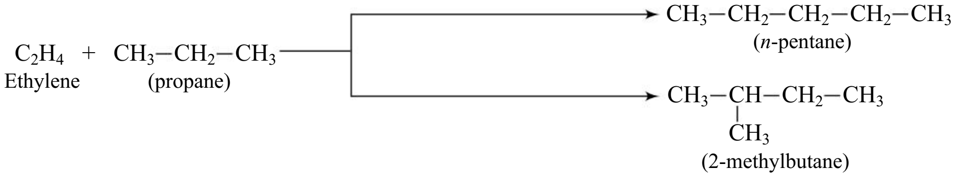
1. *Cracking*—The heavier fractions of petroleum refining may be converted to more useful lighter fractions by a process called *cracking* and this involves change in molar mass.
2. *Combining*—Combining takes place in three units: the alkylation unit, the dimersol and the polymerization unit. In all three of these units, small LPG molecules are combined into larger, gasoline molecules. FHR uses sulphuric acid, nickel and phosphoric acid to combine different combinations of LPG into gasoline.
3. *Reforming*—Sometimes a change in molecular structure may be produced without much change in molar mass and this is called *reforming*.

(1) Petrol Cracking

Cracking may be defined as the process of breaking down hydrocarbons of higher molar mass (high boiling) into lighter hydrocarbons (low boiling).

$$C_{12}H_{26} \xrightarrow[\text{in the absence of air}]{\text{Heat + Pressure}} C_6H_{14} + C_6H_{12}$$

Sometimes the product of cracking may recombine to form branched alkanes, alkenes, etc.



$$\underset{\text{Ethylene}}{C_2H_4} + \underset{\text{(propane)}}{CH_3-CH_2-CH_3} \rightarrow \underset{(n\text{-pentane})}{CH_3-CH_2-CH_2-CH_2-CH_3}$$

$$\underset{\text{Ethylene}}{C_2H_4} + \underset{\text{(propane)}}{CH_3-CH_2-CH_3} \rightarrow \underset{\text{(2-methylbutane)}}{CH_3-\underset{\displaystyle CH_3}{\underset{|}{CH}}-CH_2-CH_3}$$

Three types of cracking are known (a) thermal cracking, (b) catalytic cracking and (c) Hydrocracking.

(a) Thermal cracking

Heavy oil feedstock is subjected to high pressure and temperature (700°C) in the absence of air and catalyst. The cracked products are cooled and fractionated to get gasoline and other products of low molar mass. The method is old and seldom used now since there is no control over the products of cracking and the efficiency is low.

Heating petroleum at high temperatures (>350°C) favours the breakdown of higher molecular weight components into smaller (lower molecular weight) fragments. For example, thermal cracking of dodecane ($C_{12}H_{26}$) yields hexane and 1-hexene both of which are viable gasoline components. Although olefins (alkenes) such as 1-hexene are rarely found in unrefined petroleum, they are abundant in processed petroleum due to the cracking phenomenon.

$$\underset{\text{dodecane}}{CH_3-(CH_2)_{10}-CH_3} \rightarrow \underset{\text{hexane}}{CH_3-(CH_2)_4-CH_3} + \underset{\text{1-hexene}}{H_2C{=}CH-(CH_2)_3-CH_3}$$

(b) Catalytic cracking

The vaporized feedstock is made to undergo cracking by coming into contact with a solid catalyst in finely powdered state. Cracking takes place at low temperatures and pressures. By controlling the conditions, it is possible to control the nature of the end products. Catalytic cracking is essentially thermal decomposition of hydrocarbons in the presence of catalysts such as a crystalline aluminosilicates (zeolites) or molecular sieves. In petroleum refining, catalytic cracking has largely superseded thermal cracking because it produces gasoline richer in branched paraffins, cycloparaffins and aromatics.

Two types of catalytic cracking methods are known as:

(i) Fixed bed catalytic cracking

Fixed bed catalytic cracking is used to convert heavy hydrocarbon fractions obtained by vacuum distillation into a mixture of more useful products such as petrol and light fuel oil. In this process, the feedstock undergoes a chemical breakdown, under controlled heat (450–500°C) and pressure, in the presence of a catalyst—a substance which promotes the reaction without itself being chemically changed. Small pellets of silica-alumina or silica-magnesia have proved to be the most effective catalysts.

The cracking reaction yields petrol, LPG, unsaturated olefin compounds, cracked gas oils, a liquid residue called cycle oil, light gases and a solid coke residue. Cycle oil is recycled to cause further breakdown and the coke, which forms a layer on the catalyst, is removed by burning. The other products are passed through a fractionator to be separated and separately processed.

(ii) Fluidized bed catalytic cracking

Principle: In the fluidized bed catalytic cracking, the finely powdered catalyst is kept agitated by a gas stream of the vaporized heavy oil feedstock. Thus, the catalyst can be handled like a fluid system and can be pumped like a liquid. This brings about a close contact between the catalyst and the reactant and results in a more efficient cracking.

Method: A schematic representation of the fluidized bed catalytic cracking is given in Fig. 1.5. Catalyst (Y type of zeolite with rare earth oxide or a zeolite called ZSM5) in the finely divided state is kept in the catalyst chamber. It is kept in a fluid state by the upcoming feedstock vapours under pressure. Cracking takes place efficiently since there is very close contact between the catalyst and reactant. The cracked products are withdrawn from the top of the cracking chamber and fed directly into a fractionating column for further separation. The uncracked oil is sent for a second stage of cracking.

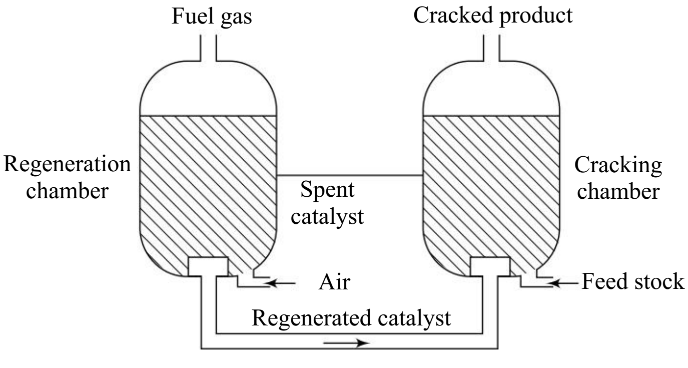


FIGURE 1.5 *Fluidized catalytic bed.*

A spent catalyst is continuously pumped from the catalyst chamber into a regeneration chamber. Hot air is blown through it to burn the carbon deposit on the catalyst surface. The regenerated catalyst is sent back to the catalyst chamber. Thus, the process of cracking is made continuous.

(iii) Hydrocracking

Similar to fluid catalytic cracking, but uses a different catalyst, lower temperatures, higher pressure and hydrogen gas. It takes heavy oil and cracks it into gasoline and kerosene (jet fuel). After various hydrocarbons are cracked into smaller hydrocarbons, the products go through another fractional distillation column to separate them.

Hydrocracking can increase the yield of petrol components, as well as being used to produce light distillates. It produces no residues, only light oils. Hydrocracking is catalytic cracking in the presence of hydrogen. The extra hydrogen saturates, or hydrogenates, the chemical bonds of the cracked hydrocarbons and creates isomers with the desired characteristics. Hydrocracking is also a treating process, because the hydrogen combines with contaminants such as sulphur and nitrogen, allowing them to be removed.

Gas oil feed is mixed with hydrogen, heated and sent to a reactor vessel with a fixed bed catalyst, where cracking and hydrogenation take place. Products are sent to a fractionator to be separated. The hydrogen is recycled. Residue from this reaction is mixed again with hydrogen, reheated and sent to a second reactor for further cracking under higher temperatures and pressures. In addition to cracked naphtha for making petrol, hydrocracking yields light gases useful for refinery fuel, or alkylation as well as components for high quality fuel oils, lube oils and petrochemical feedstocks.

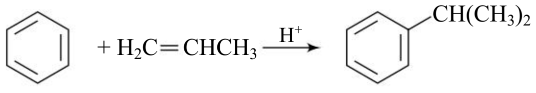
(2) Combining or Unification Process

Sometimes, you need to combine smaller hydrocarbons to make larger ones—this process is called *unification.* The major unification process is called *catalytic reforming* and uses a catalyst (platinum, platinum-rhenium mix) to combine low weight naphtha into aromatics, which are used in making chemicals and in blending gasoline. A significant by-product of this reaction is hydrogen gas, which is then used for hydrocracking.

Alkylation

The structures of molecules in one fraction are rearranged to produce another. Commonly, this is done using a process called *alkylation*.

Acid catalysts such as sulphuric acid or aluminium trichloride promote reaction of either a highly-branched paraffin such as isobutane or an aromatic compound with an olefin such as ethylene or propene. In below equation, we observe that benzene can react with propene to yield isopropylbenzene (cumene).



benzene cumene (isopropylbenzene)

(6)

Olefins such as propylene and butylene are produced by catalytic and thermal cracking. In alkylation, low molecular weight compounds, such as propylene and butylene, are mixed in the presence of a catalyst such as hydrofluoric acid or sulphuric acid. *Alkylation* refers to the chemical bonding of these light molecules with isobutane to form larger branched-chain molecules (isoparaffins) that make high octane petrol. The products of alkylation are *high octane hydrocarbons*, which are used in gasoline blends to reduce knocking**.**

e.g. C_4H_{10} + C_4H_8 → C_8H_{18}
(Isobutene) (butylene) (isooctane)

(3) Reforming of petrol

Reforming is a process involving modification of the structure of molecules; which uses heat, pressure and a catalyst (usually containing platinum) to produce high octane petrol and petrochemical feedstock. This is carried out to convert straight chain hydrocarbons with low octane number to branched chain or cyclic compounds with high octane number (upgradation of gasoline).

Reforming converts a portion of these compounds to isoparaffins and aromatics, which are used to blend higher octane petrol

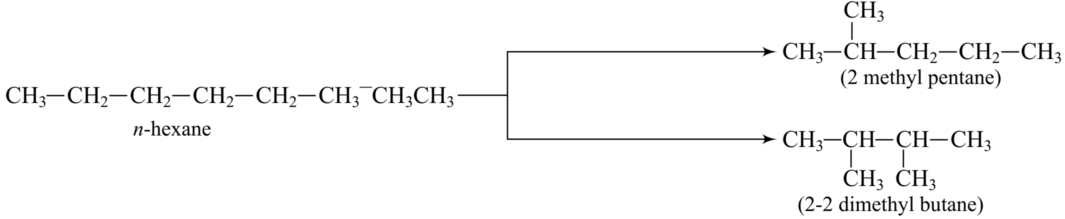
- paraffins are converted to isoparaffins
- paraffins are converted to naphthenes
- naphthenes are converted to aromatics.

$$\begin{array}{ccccc} C_7H_{16(\text{catalyst})} & \rightarrow & C_7H_8 & + & 4H_2 \\ \text{(Heptane)} & & \text{(Toluene)} & & \text{(Hydrogen)} \\ C_6H_{12} & \rightarrow & C_6H_6 & + & 3H_2 \\ \text{(Cyclohexane)} & \rightarrow & \text{(Benzene)} & & \text{(Hydrogen)} \end{array}$$

Reforming is usually carried out by passing the vapours of the petroleum over platinum supported on alumina in presence of hydrogen. A number of reforming reactions are known. A few of the important reforming reactions are given below.

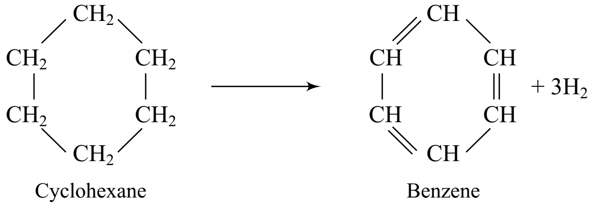
(a) ***Isomerization*** refers *to chemical rearrangement of straight-chain hydrocarbons* (*paraffins*), *so that they contain branches attached to the main chain* (*isoparaffins*). This is done due to two reasons:

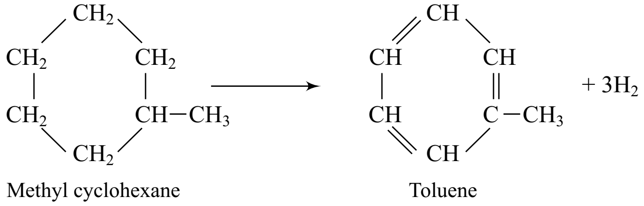
1. they create extra isobutane feed for alkylation
2. they improve the octane of straight run pentanes and hexanes and hence make them into better petrol blending components.



Pentanes and hexanes are the lighter components of petrol. Isomerization can be used for the conversion of a straight chain compound into a branched chain hydrocarbon.

(b) ***Dehydrogenation.*** Cyclic alkanes undergo dehydrogenation to form cyclic alkanes or aromatic compounds.

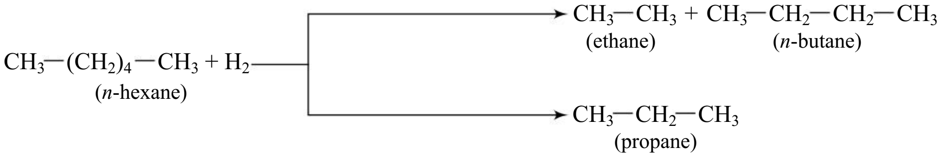




(c) ***Cyclization and dehydrogenation.*** *n*-alkanes become cycloalkanes, which undergo dehydrogenation and form aromatic compounds.

$$\underset{n\text{-hexane}}{C_6H_{14}} \longrightarrow \underset{\text{cyclohexane}}{C_6H_{12}} + H_2 \longrightarrow \underset{\text{benzene}}{C_6H_6} + 3H_2$$

Hydrocracking: *n*-paraffins are mixed with hydrogen and passed over a hot catalyst. Hydrocracking takes place resulting in two lighter hydrocarbons.



$$\underset{(n\text{-hexane})}{CH_3{-}(CH_2)_4{-}CH_3} + H_2 \longrightarrow \underset{(\text{ethane})}{CH_3{-}CH_3} + \underset{(n\text{-butane})}{CH_3{-}CH_2{-}CH_2{-}CH_3}$$

$$\longrightarrow \underset{(\text{propane})}{CH_3{-}CH_2{-}CH_3}$$

1.13 Petrol Knocking

Petrol is a mixture of different hydrocarbons. Diesel is the fifth fraction, released after naphtha and kerosene. Diesel is harder to vaporize than petrol. This is why the two are not interchangeable in car engines.

Mechanism of Knock in Petrol Engines

In a petrol engine, combustion is triggered by a spark. *However, under certain temperature and pressure conditions, the fuel can ignite spontaneously*. When this happens there is a very rapid reaction, of some or all of the fuel-air mixture in an engine. The flame speed is many times greater than that which follows normal spark ignition. This is called *knocking*, *auto-ignition (or pinking)*. The noise associated with it is called *knock*. Knocking is influenced by several factors—for example, the build-up of deposits, hard acceleration in a high gear, high-load driving (uphill or towing) and an out-of-tune engine (wrong spark timing or fuel/air ratio) can lead to increased *knocking*.

Knock Resistance

Engine knocking is compression detonation of fuel in the *power stroke* of the engine (Fig. 1.6). Knocking occurs when the air–fuel mixture auto-ignites all at once (or sometimes perhaps when the flame front goes very fast because of early ignition timing), before the flame front from spark plug ignition can reach it. The explosive reaction causes combustion to stop before the optimum timing, causing a decrease in performance. A fuel with a high auto-ignition temperature that burns reasonably fast and thus does not need early ignition timing will most often have high practical value knock resistance. Ethanol is such a fuel.

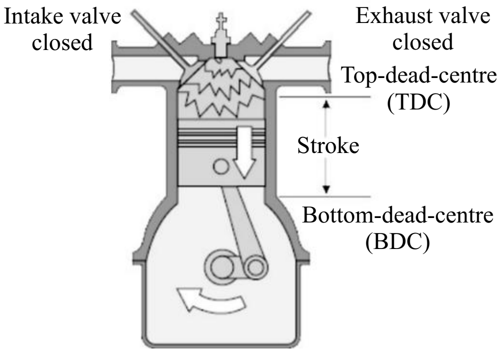


Figure 1.6 *Power stroke of a four stroke engine with a (air+ fuel) compressed mixture.*

1.14 Octane Rating of a Fuel

Octane Number of a Fuel

The octane number of a fuel is a measure of its ability to resist knocking. A fuel with a high octane number has a greater resistance to spontaneous combustion. Obviously then, the lower the octane number the higher the chance of auto-ignition.

For over a century, the internal combustion engine and its fuel, petrol, have had a major effect on our lives. Despite controversies about traffic congestion, pollution caused by lead compounds amongst others, and large-scale CO_2 emissions contributing to global warming, the importance of petrol is unlikely to diminish for a long time.

(a) Petrol Composition

Petrol as a mixture of hydrocarbons is being considered as fuels or as sources of fuels. Petrol is a complex mixture of at least 100 different compounds, mostly hydrocarbons. Most of these are branched-chain alkanes, and some are aromatic compounds.

(b) Petrol and the Internal Combustion Engine

A number of things happen to the petrol in the internal combustion engine, including:

- Petrol is vaporized
- The vapour is mixed with air
- The petrol-air mixture is compressed
- The mixture is ignited by a spark from the spark plug and burned
- The gases produced by the combustion reaction expand
- Expansion causes the piston to move that is kinetic energy is produced.

(c) Premature Ignition

The greater the extent to which gases are compressed the more they tend to heat up. Sometimes this causes ignition before the spark is produced. This is intended in a diesel engine, where there is no spark plug, but in a petrol engine the occurrence is called *auto-ignition* or *knocking* or pinking. This is quite a problem as it can cause loss of power, with obvious danger, or damage to the engine.

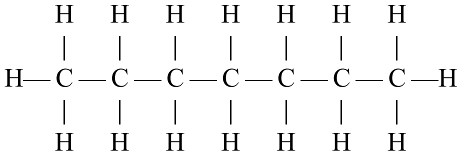
It can be prevented in two ways during petrol manufacture:

1. Use of additives.
2. Use of a suitable mixture of high-octane compounds.

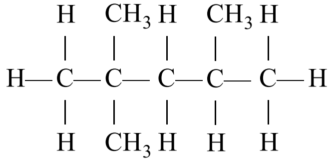
(d) Octane Rating

The octane rating is a measure of the tendency of a fuel to auto-ignite. The lower the octane rating the more likely it is that auto-ignition will occur. Clearly, high-octane fuels are more desirable.

The scale is an arbitrary one. Two compounds were chosen, *heptane* (C_7H_{16}) and 2,2,4-*trimethylpentane* ($CH_3{-}C(CH_3)_2{-}CH_2{-}CH(CH_3){-}CH_3$).



(heptane)



(2,2,4-trimethylpentane)

Heptane has a high tendency to auto-ignite, so it was given an octane number of 0. On the other hand, 2,2,4-trimethylpentane has a low tendency to auto-ignite, so it was given a rating of 100.

A mixture of these two compounds containing 95 percent of 2,2,4-trimethylpentane (isooctane) is said to have an octane number of 95. A mixture of compounds with an identical tendency to auto-ignite, under the same conditions of compression, would thus also be given an octane rating of 95. A compound that is less likely to auto-ignite than pure 2,2,4-trimethylpentane would have an octane rating of more than 100.

(e) Addition of Oxygenates, e.g. Alcohols or Ethers

The compounds such as alcohols and ethers work by raising the octane number of the fuel. They cause less pollution, because apart from not containing lead, they produce lower levels of carbon monoxide when they burn. The most commonly used oxygenates are MTBE (methyl tertiary butyl ether) ethyl alcohol. The systematic name is 2-methoxy-2-methylpropane. Its octane rating is 118.

(f) Measures to Obtain High Octane Compounds

Apart from the use of additives, knocking may also be prevented by using a mixture of high-octane compounds in petrol manufacture. Certain molecular features are desirable in ensuring that compounds have high-octane ratings.

These are
(a) a high degree of branching
(b) short chain length
(c) the existence of rings.

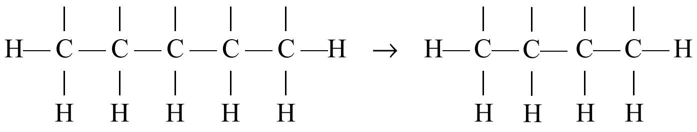
High-octane compounds can be obtained from low octane compounds by three processes, each involving the use of catalysts:
(a) Isomerisation
(b) Dehydrocyclisation
(c) Catalytic cracking.

These processes will be explained in the succeeding paragraphs.

Isomerization

When certain compounds are heated in the presence of a suitable catalyst, a different structural isomer of the particular compound is formed, e.g.

A straight chain alkane such as pentane (C_5H_{12}), which has an octane number of 62, is heated in the presence of a suitable catalyst. The chain breaks, the fragments rejoin to form a branched compound, 2 methylbutane, ($CH_3CH(CH_3)CH_2CH_3$), which has an octane number of 93.



$$\begin{array}{ccccccccccc} & H & & H & & H & & H & & H & \\ & | & & | & & | & & | & & | & \\ H— & C & — & C & — & C & — & C & — & C & —H \\ & | & & | & & | & & | & & | & \\ & H & & H & & H & & H & & H & \end{array} \rightarrow \begin{array}{ccccccccc} & H & & CH_3 & & H & & H & \\ & | & & | & & | & & | & \\ H— & C & — & C & — & C & — & C & —H \\ & | & & | & & | & & | & \\ & H & & H & & H & & H & \end{array}$$

heptane 2-methylbutane

Clearly, the product would be a much more suitable component of petrol than the original pentane.

Dehydrocyclization

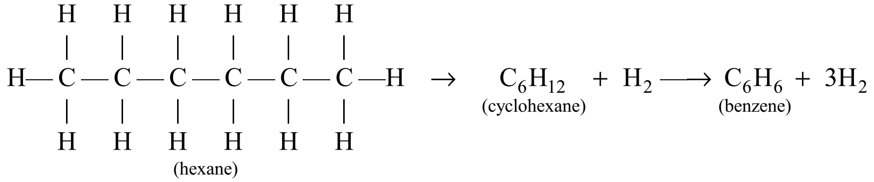
This process involves the formation of a ring compound, accompanied by the removal of a valuable by-product, hydrogen gas, e.g.,

- A straight chain alkane such as hexane (C_6H_{14}), which has an octane number of 25, is heated in the presence of a suitable catalyst.
- The catalyst causes the alkane to change to a cycloalkane, e.g. cyclohexane, of octane number 83

$$C_6H_{14} \rightarrow (CH_2)_6 + H_2$$

- The catalyst causes the cycloalkane to further change to an aromatic compound, e.g. benzene, of octane number greater than 100.

$$(CH_2)_6 \rightarrow C_6H_6 + 3H_2$$

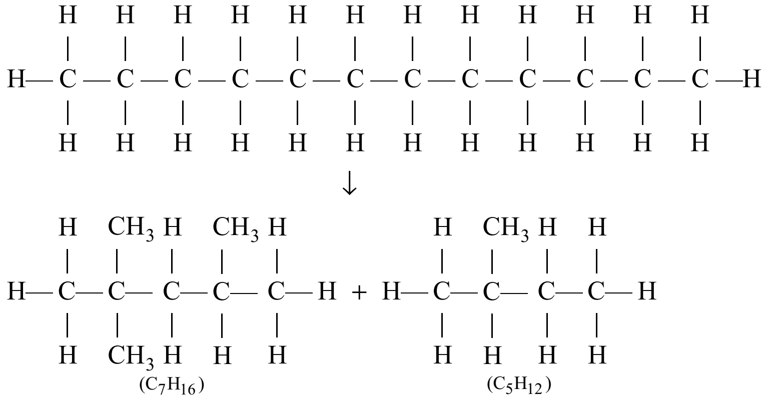


$$\underset{\text{(hexane)}}{H-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-H} \rightarrow \underset{\text{(cyclohexane)}}{C_6H_{12}} + H_2 \longrightarrow \underset{\text{(benzene)}}{C_6H_6} + 3H_2$$

In terms of octane ratings, the overall effect is to increase from 25 to a number greater than 100. The same catalyst, platinum dispersed on aluminium oxide, is used for both stages.

Catalytic Cracking

This process involves taking heavy oil such as kerosene or diesel and heating it to a high temperature in the presence of a catalyst. The large molecule breaks down into several smaller ones, some saturated, some unsaturated, e.g.



$$H-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-H$$

$$\downarrow$$

$$\underset{(C_7H_{16})}{H-\overset{H}{\underset{H}{C}}-\overset{CH_3}{\underset{CH_3}{C}}-\overset{H}{\underset{H}{C}}-\overset{CH_3}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-H} + \underset{(C_5H_{12})}{H-\overset{H}{\underset{H}{C}}-\overset{CH_3}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-H}$$

The unsaturated products are used as feedstocks for the polymer industry. The saturated products are usually high-octane branched chain alkanes suitable for making petrol.

Good petrol has an octane number of 97 or more. Petrol with an octane number of 98 has the same anti-knocking characteristics as a mixture of 98 percent iso-octane and 2 percent heptane.

A high tendency to auto-ignite, or low octane rating, is undesirable in a gasoline engine but desirable in a diesel engine.

1.15 Techniques to Prevent Knocking

If petrol were used as it is produced, straight from the fractionating column, it would have a very *low* octane number and therefore would be unsuitable for its purpose. To prevent knocking or self-ignition, additives must be added to the petrol. In the past, small amounts of tetraethyl lead $Pb(C_2H_5)_4$ were used. However, this was found to be extremely damaging to the environment and so it is advisable to use unleaded petrol.

To increase the octane number of the fuel, other techniques had to be introduced. Such reformation of petrol techniques are as follows:

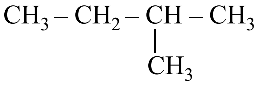
- Isomerization,
- Catalytic cracking
- Reforming and
- Adding oxygenates to petrol.

Isomerization

Isomers are chemical compounds which have the same number and type of each atom but a different arrangement or structure of atoms. *Isomerization* involves altering the chemical structure of a compound. It usually occurs when the substance is heated in the presence of a suitable catalyst. The shorter and more branched the alkane chain the higher the octane number, so creating isomers with these properties increases their octane number making them more suitable to be used as petrol.

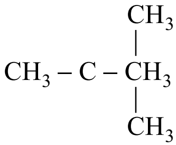
$CH_3 - CH_2 - CH_2 - CH_2 - CH_3$

Pentane (C_5H_{12})



$$CH_3 - CH_2 - \underset{\displaystyle CH_3}{\underset{|}{CH}} - CH_3$$

2-Methyl butane (C_5H_{12})



$$CH_3 - \overset{\displaystyle CH_3}{\overset{|}{\underset{\displaystyle CH_3}{\underset{|}{C}}}} - CH_3$$

2, 2-Dimethylpropane (C_5H_{12})

Catalytic Cracking

Catalytic cracking is the breaking down of long-chain hydrocarbons into shorter chain hydrocarbons. This involves taking heavy oil such as kerosene and heating it to a high temperature in the presence of a catalyst. The process was developed by scientists when they found there was a surplus of heavier fractions (kerosene) from the fractional distillation of crude oil.

Dehydrocyclisation

Dehydrocyclization, also known as reforming, involves the use of a catalyst to form ring compounds from straight chain alkanes. Hydrogen is formed as a very valuable by-product. An alkane is converted to a cycloalkane and then to an aromatic compound. Aromatic compounds have far greater octane numbers than alkanes and therefore are more suitable to be used for petrol.

Oxygenation

An oxygenate is any fuel that contains oxygen. Adding oxygenates, such as methanol (CH_3OH), ethanol (C_2H_5OH) or MTBE (methyl tertiary butyl ether), to a fuel is a more environmentally friendly way of increasing the octane number.

In Brazil, ethanol made from sugarcane residues is added to petrol and is called *gasohol*. Up to 20 percent alcohol can be added without any major modifications to the engine. This saves importing fuel and will help preserve existing reserves. It may also be an alternative to petrol. A huge advantage is that it is a renewable resource and is more environmentally friendly.

1.16 Prevention of Knocking

Knocking in petrol engine may be minimized by the following methods:

- A suitable change in engine design may be made to have optimum CR
- High octane number fuel may be used
- Antiknocking agents may be used

1.17 Adverse Effects of Knocking

- Knocking produces undesirable rattling sound.
- It results in decrease in efficiency.
- Driving and travelling become unpleasant.
- It increases fuel consumption.
- It causes mechanical damage to the engine parts and spark plug.

1.18 Antiknocking Agents: TEL and TML

Knocking of petrol may be reduced by the addition of some organolead compounds into it. The substance added to control knocking is called antiknocking agent. (i) *Tetra ethyl lead* (TEL), $Pb(C_2H_5)_4$, (ii) tetra methyl lead (TML), $Pb(CH_3)_4$ or (iii) a mixture of TEL and TML is used as an antiknocking agent. They are used along with ethylene dichloride or dibromide.

Explanation: Antiknocking agents TEL and TML produce ethyl and methyl free radical by dissociation. These radicals combine with the free radicals formed by branching chain reaction during the combustion of gasoline and form stable hydrocarbons. Thus, the populations of chain carriers are kept under check. Thus knocking is reduced greatly even by the use of small quantities of TEL or TML.

1.18.1 Effects of TEL and TML

TEL and TML are converted to Pb or PbO and deposited on the engine parts or the exhaust pipe causing damages. But if they are used along with ethylene dichloride or dibromide, Pb + PbO are converted to volatile $PbCl_2$ or $PbBr_2$ which escape as gases into atmosphere. But again these lead halides are poisonous to human beings and animals. Hence the use of organolead compounds should be discouraged.

1.19 Unleaded Petrol

Knocking of gasoline can be minimized by addition of antiknocking agents such as TEL or TML. But TEL and TML form Pb or PbO deposits and cause damage to engine parts. Therefore, TEL + TML are used along with $C_2H_4Cl_2$ or $C_2H_4Br_2$. These convert Pb or PbO into volatile $PbCl_2$ or $PbBr_2$, which escape into air through the exhaust pipe. Lead salts are poisonous to animals and human beings. Hence, usage of leaded petrol should be discouraged. Unleaded petrol does not contain TEL or TML. To avoid knocking in them various measures are taken.

- They are formed by mixing compounds of higher Octane number.
- They are mixed with compounds like methyl t-butyl ether or Ethyl t-butyl ether which also serves as antiknocking agent, but are not harmful.

All cars are fitted with *catalytic convertors* which convert the straight chain compounds [low O.N.] in gasoline to branched chain compounds (higher O.N.).

1.20 Synthetic Petrol

The synthetic petrol can be obtained by any of the following methods:

(1) Bergius Process

An intimate paste of finely powdered bituminous coal and heavy oil is fed into the converter along with catalyst Sn or nickel oleate. The temperature of the reactor is maintained at 450°C and the conversion is carried out at a pressure of 250–300 atms for about 2 hours. Under these conditions, hydrogen combines with coal to form saturated hydrocarbons, which decompose at the maintained conditions to yield low boiling liquid hydrocarbons. When the issuing gases from the reacting vessel are condensed, a liquid resembling crude oil is obtained. This is further subjected to fractional distillation gives the following fractions:

- gasoline,
- middle oil, and
- heavy oil.

The middle oil and the heavy oils obtained are further recycled to get more gasoline. The yield of gasoline from this method may be about 60 percent.

(2) Fischer–Tropsch Process

The raw material, water gas ($CO + H_2$), obtained from passing steam over coke mixed with H_2 and is passed through Fe_2O_3 (and/Na_2CO_3) to remove any sulphide contents in the mixture. The mixture of gases are compressed to 20 atms and then led through converter containing catalysts such as cobalt, thoria, magnesia and mostly kieselgar earth at 250°C. The products are mainly saturated and unsaturated hydrocarbons.

$$nCO + 2nH_2 \rightarrow C_nH_{2n} + nH_2O$$

$$nCO + (2n + 1)\,H_2 \rightarrow C_nH_{2n+2} + nH_2O$$

The hot gaseous products are cooled by passing through a condenser. The condensate is a crude oil which on fractionation yields (i) gasoline and (ii) high boiling heavy oil. The heavy oil is again recycled to get more gasoline.

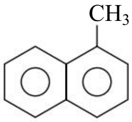
1.21 Diesel Knocking

Diesel engines are compression ignition engines. The fuel is made up of straight chain hydrocarbons with a boiling point around 100–360°C. In a diesel engine there is no spark plug. In the suction stroke, only air is drawn into the cylinder. In the compression stroke, air is highly compressed till its temperature reaches 500°C. Now the fuel is injected as a spray into the engine.

The diesel vaporizes, attains self-ignition temperature and burns. If vaporization and combustion are instantaneous, fuel burns smoothly and efficiently. But usually there is a time lag between vaporization and combustion called diesel lag (ignition delay). This causes the accumulation of diesel in the vapour state and the whole mixture of diesel vapour and air gets ignited with explosion. This is called *diesel knocking*.

1.22 Cetane Number (CN)

Cetane number is an indicator of *the readiness with which a given diesel undergoes compression ignition*. Straight chain compounds undergo easy compression ignition and *n*-cetane is chosen as the upper limit of cetane number = 100. Branched chain and cyclic compounds do not undergo compression ignition readily and 1-methyl naphthalene is chosen as the lower limit of cetane number = 0.



1 Methyl naphthalene
CN = 0

$CH_3—(CH_2)_{14}—CH_3$
n-cetane
CN = 100

Definition: *Cetane number of a diesel is the percentage by volume of n-cetane in a mixture of n-cetane and 1 methyl naphthalene that gives the same knocking as the diesel under consideration.*

Prevention of Diesel Knocking

Diesel knocking can be minimized as

1. by using diesel with high cetane number—containing more straight chain hydrocarbons.
2. certain organic additives such as ethyl nitrate $C_2H_5–O–NO_2$, amyl nitrate $CH_3\text{-}(CH_2)_3–CH_2–O–NO_2$ increase the cetane number of a diesel.

1.23 Power Alcohol

Alcohol is an excellent alternative motor fuel for gasoline engines. The alcohols are fuels of the family of the ‘oxygenates’ having one or more oxygen, which contributes to the combustion. Only two of the alcohols are technically and economically suitable as fuels for internal combustion engines that is Methanol and Ethanol.

If about 20–25% ethyl alcohol is blended with petrol and used as a fuel, it is known as power alcohol for internal combustion engine.

Gasohol is a mixture of 90 percent unleaded gasoline and 10 percent ethyl alcohol (ethanol). Its performance as a motor vehicle fuel is comparable to that of 100 percent unleaded gasoline, with the added benefit of superior antiknock properties (no premature fuel ignition). No engine modifications are needed for the use of gasohol, which has in recent years gained some acceptance as an alternative to pure gasoline.

Methanol is produced by a variety of process, the most common are as follows: distillation of wood; distillation of coal; natural gas and petroleum gas. Ethanol is produced mainly from biomass transformation, or bioconversion. It can also be produced by synthesis from petroleum or mineral coal.

Combustion Characteristics of Alcohols as a Fuel

There are some important differences in the combustion characteristics of alcohols and hydrocarbons.

- Alcohols have higher *flame speeds* and extended flammability limits. Also, alcohols produce a *great number of product moles per mole of fuel burnt, therefore, higher pressure are achieved.*
- Alcohols burn with no luminous flame and produce almost *no soot*, especially methanol. The tendency to soot increases with molecular weight.
- The onset of ignition (due to heat and pressure during compression stroke) leads to the formation of a flame front which propagates at high speed throughout the whole mixture. This leads to increased volumetric efficiency and reduced compression temperatures. Together with the low level of combustion temperature, these effects also improve the thermal efficiency by 10 percent.
- The higher flame speed, giving earlier energy release in the power stroke, results in a *power increase of 11 percent* at normal conditions and up to 20 percent at the higher levels of a compression ratio (14:1).
- The oxygen contents of alcohols *depress the heating value* of the fuel in comparison with hydrocarbon fuels.
- The heat of *combustion per unit volume* of alcohol is approximately half that of isooctane. However, the stoichiometric fuel-air mass ratios are such big that the quantity of energy content based on unit mass of stoichiometric mixture become comparable with that of hydrocarbons.

Engine Response to Alcohol as Fuel

Some properties of alcohols can be turned into advantages by designing the engine to adequately take advantage of such properties.

The *mass caloric value,* or the *energy available per unit of* mass of the fuel in a liquid state, is the most important characteristic of a fuel.

The *heat of combustion*, or the energy available per unit of volume of a stoichiometric mixture of fuel and air.

The fuel consumption is inversely proportional to the *caloric value*.

Other properties favourable to the increase of power and reduction of fuel consumption are:

- Number of molecules or products is more than that of reactants
- Extended limits of flammability
- High octane number.

Other properties favourable to the increase of power and reduction of fuel consumption are:

- Number of molecules or products is more than that of reactants
- Extended limits of flammability
- High octane number.

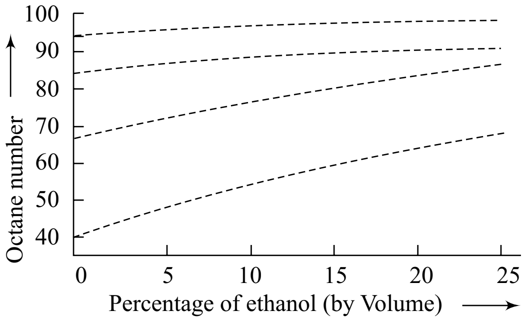


Figure 1.7 *Octane rating versus % of ethanol blend.*

Octane Ratings

The graph (Fig. 1.7) illustrates another major advantage of alcohol blends, namely the *ability of alcohol to raise* *the antiknock quality of the gasoline with which it is mixed*. This means, of course, that lower, cheaper grades of gasoline can be used to obtain a fuel with the desired octane.

Advantages of Alcohol Power Alcohol

Alcohol has characteristics which makes it a natural engine fuel:

- It has a high 'octane' rating, which prevents engine detonation (knock) under load.
- It burns clean, in fact, that not only are noxious emissions drastically reduced, but the internal parts of the engine are purged of carbon and gum deposits which, of course, do not build up as long as alcohol is used as fuel.
- An alcohol burning engine tends to run cooler than its gasoline-powered counterpart, thus extending engine life and reducing the chance of overheating.
- Less starting problems.
- Alcohol removes traces of moisture in the petrol.

Disadvantages of Blended Alcohol

- Alcohol lowers the calorific value of petrol.
- Due to considerable surface tension, alcohol do not atomize at low temperature.
- Due to oxidation of alcohol in storage tanks, corrosion can occur.

PART-B: RENEWABLE SOURCE OF ENERGY (SOLAR ENERGY)

Chapter Outline

Introduction, importance of solar (PV) cells. Photovoltaic cell, the *p/n* junction of a solar cell. How photovoltaic cell converts light into electricity. Doping of silicon for photovoltaic cell, and chemical properties grade solar grade silicon. Production of solar grade silicon from quartz, chemical vapour deposition technique. Manufacture of photovoltaic cell.

1.24 Photovoltaic Cells

Solar energy is the source of all life on earth. Without it, we would not be here today living on such a rich and diverse planet. The sun radiates most of the available energy to us. It provides us with food energy through plant photosynthesis and provides the heat that we need to survive. Trapped solar energy is released when we burn fossil fuel reserves and the sun drives the earth's weather systems which provide renewable forms of energy like wind, solar and wave power. It is now widely recognized that utilizing the sun's natural energy can offer real alternatives to burning finite resources of fossil fuels or endangering future generations by relying on dangerous technologies such as nuclear power. Solar energy has long been used for space heating utilizing passive solar design. One of the most exciting areas of development has come in the form of the *photovoltaic cell; in which the 'photovoltaic effect' is the basic physical process through which a PV cell converts sunlight into electricity.* The photovoltaic effect will continue as long as light strikes the photovoltaic cell: every time the photons will create new electron/hole-pairs. This implies that no materials are being wasted; it is a *renewable process*.

Solar cells are composed of various semiconducting materials. Semiconductors are materials, which become electrically conductive when supplied with light or heat, but which operate as insulators at low temperatures. Over 95 percent of all the solar cells produced worldwide are composed of the semiconductor material silicon (Si).

Crystalline solar cells are made from crystalline silicon and are fragile and break easily. The crystalline structure of silicon is shown in Fig. 1.8.

As the second most abundant element in the earth's crust, silicon has the advantage, of being available in sufficient quantities, and additionally processing the material does not burden the environment. To produce a solar cell, the semiconductor is contaminated or 'doped'. 'Doping' is the intentional introduction of chemical elements, with which one can obtain a surplus of either positive charge carriers (*p*-conducting semiconductor layer) or negative charge carriers (*n*-conducting semiconductor layer) from the semiconductor material. If two differently contaminated semiconductor layers are combined, then a so-called *p*–*n*-junction results on the boundary of the layers.

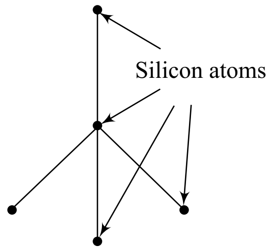


Figure 1.8 *Structure of silicon.*

Crystalline silicon has a bandgap energy of 1.1 electron-volts (eV). The bandgap energies of other effective PV semiconductors range from 1.0 to 1.6 eV. In this range, electrons can be freed without creating extra heat (Fig. 1.9).

The photon energy of light varies according to the different wavelengths of the light. The entire spectrum of sunlight, from infrared to ultraviolet, covers a range of about 0.5 eV to about 2.9 eV. For example, red light has an energy of about 1.7 eV, and blue light has an energy of about 2.7 eV. Most PV cells cannot use about 55 percent of the energy of sunlight, because this energy is either below the bandgap of the material or carries excess energy.

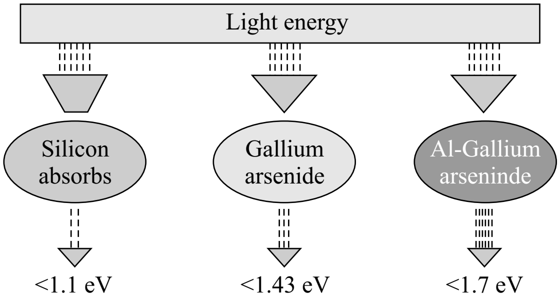


Figure 1.9 *Bandgap energies of other PV semiconductors.*

Different PV materials have different energy bandgaps. Photons with energy equal to the bandgap energy are *absorbed to create free electrons*.

1.25 Importance of Solar Cells

The photovoltaic cell offers a limitless and environmentally friendly source of electricity. Solar cells have many applications. They are particularly well suited to, and historically used in situations where electrical power from the *grid* is unavailable, such as in remote area power systems, *earth* orbiting *satellites*, handheld *calculators*, remote radiotelephones, *water* pumping applications, etc. Solar cells (in the form of modules or *solar panels*) are appearing on building roofs where they are connected through an *inverter* to the electricity grid.

The main advantages of silicon as a PV material are because of the following:

- Thin wafers of silicon
- Much bigger
- Much cheaper (delete)
- Silicon is abundant (sand)
- Nontoxic, safe
- Light carries energy into cell
- Cells convert sunlight energy into electric current.

1.26 Photovoltaic Cell

A *solar cell,* or *photovoltaic cell*, is a semiconductor device consisting of a large-area *p–n junction diode*, which in the presence of *sunlight* is capable of generating usable *electrical energy*. This conversion is called the *photoelectric effect*. The field of research related to solar cells is known as *photovoltaics*.

A typical silicon PV cell (Fig. 1.10) is composed of a thin wafer consisting of an ultra-thin layer of phosphorus-doped (*n*-type) silicon on top of a thicker layer of boron-doped (*p*-type) silicon. An electrical field is created near the top surface of the cell where these two materials are in contact, called the *p–n* junction. When sunlight strikes the surface of a PV cell, this *electrical field provides momentum and direction* to light-stimulated electrons, resulting in a flow of current when the solar cell is connected to an electrical load.

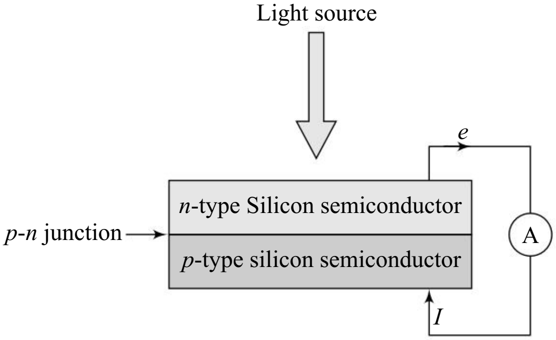


Figure 1.10 *Diagram of photovoltaic cell.*

Regardless of size, a typical silicon PV cell produces about 0.5–0.6 V DC under open-circuit, no-load conditions. The *current* (and *power*) output of a PV cell depends on its *efficiency* and size (*surface area*), and is proportional to the *intensity* of sunlight striking the surface of the cell. The electron flow provides the current (I), and the cell's electric field causes a voltage (V).

With both current and voltage, we have *power* (P), which is just the product of the two. Therefore, when an external load (such as an electric bulb) is connected between the front and back contacts, electricity flows in the cell, working for us along the way.

For example, under peak sunlight conditions a typical commercial PV cell with a surface area of 160 cm^2 will produce about 2-W peak power.

1.27 The *p–n* Junction of a Solar Cell

A solar cell is a large-area semiconductor *p–n junction*. To understand the working of a *p–n* junction it is convenient to imagine *what happens when a piece of n-type silicon is brought into contact with a piece of p-type silicon*. In practice, however, the *p–n* junctions of solar cells are not made in this way, but rather, usually, by diffusing an *n-type* dopant into one side of a *p-type* wafer.

A *p–n junction* is formed by combining *n*-type and *n*- and *p*-type semiconductors together in very close contact. The term junction refers to the region where the two types of semiconductor meet. *Extra valence electrons in the n-type layer move into the p-type layer filling the holes in the p-type layer forming what is* *called a depletion zone.* The depletion zone does not contain any mobile positive or negative charges. Moreover, *this zone keeps other charges from the p- and n-type layers from moving across it.*

A region depleted of carriers is left around the junction, and a small electrical imbalance exists inside the solar cell. This electrical imbalance amounts to about 0.6–0.7 V. So due to the *p*/*n*-junction, a built-in *electric field* is always present across the solar cell.

It can be thought of as the border region between the *p*-type and *n*-type blocks as shown in Fig. 1.11.

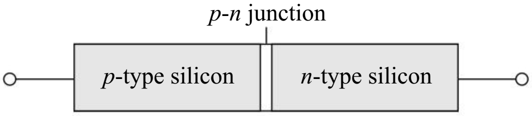


FIGURE 1.11 *A silicon p–n junction with no applied voltage.*

The *p–n* junction possesses some interesting properties which have useful applications in modern electronics. *p*-doped semiconductor is relatively *conductive*. The same is true for *n*-doped semiconductor, but the junction between them is a *nonconductor*.

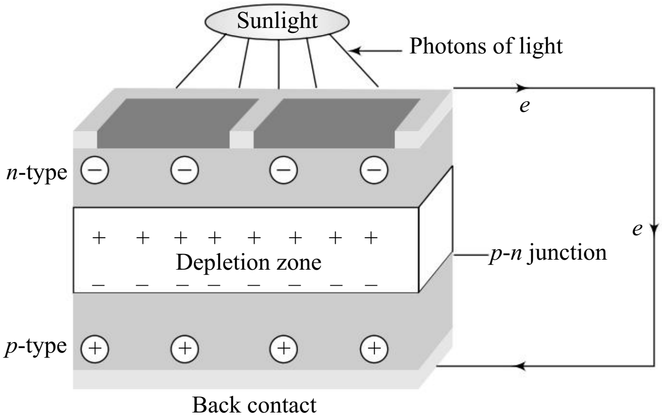


FIGURE 1.12 *A working solar cell indicating the formation of p/n junction.*

This nonconducting layer, called the *depletion zone or p/n junction*, occurs because the electrical *charge carriers* in *doped n-type* and *p-type* silicon (*electrons* and *holes,* respectively) attract and eliminate each other in a process called *recombination* (Fig. 1.12). By manipulating this nonconductive layer, *p–n* junctions are commonly used as *diodes*: electrical switches that allow a flow of *electricity* in one direction but not in the other (opposite) direction. This property is explained in terms of the *forward-bias* and *reverse-bias* effects, where the term *bias* refers to an application of electric voltage to the *p–n* junction.

1.28 How Photovoltaic Cell Converts Light into Electricity

The photovoltaic cell is the component responsible for converting light to electricity. When sunlight strikes a photovoltaic cell, part of the light particles (*photons*), which contain energy, is absorbed by the cell. By the absorption of a photon a (negative) *electron* is knocked loose from a silicon atom, and a positive '*hole*' remains. The freed electron and the positive hole together are neutral. Therefore, in order to be able to generate electricity, the electron and the hole need to be separated from each other. Therefore, a photovoltaics cell has an artificial junction layer, also called the *p/n-layer*, as shown in Fig. 1.13.

When photons hit the solar cell, freed electrons (-) attempt to unite with holes on the *p*-type layer. *The p/n-junction, a one-way road, only allows the electrons to move in one direction. If we provide an external conductive path, electrons will flow through this path to their original (p-type) side to unite with holes.*

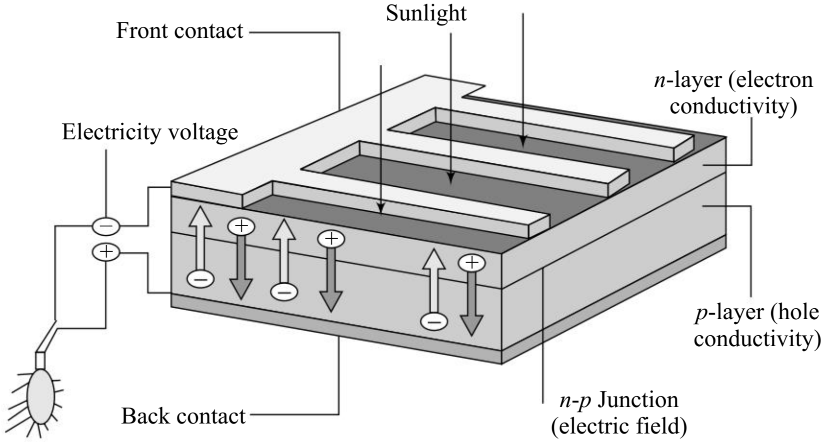


FIGURE 1.13 *Working of a solar cell with a p/n junction.*

Therefore, when the electric contacts on the front and rear are being connected through an external circuit, *the freed electrons can only return to the positively charged holes by flowing through this external circuit*, thus generating current.

The photovoltaic effect will continue as long as light strikes the photovoltaic cell: every time the photons will create new electron/hole-pairs. This implies that no materials are being wasted; it is a renewable process.

The *electrical power* that can be extracted from a photovoltaic cell is proportional to its *area* and to the *intensity* of the sunlight that hits the area, and is measured in watt (W).

1.29 Doping of Silicon for Photovoltaics

In semiconductor production, *doping refers to the process of intentionally introducing very small amount of impurities into an extremely pure (also referred to as intrinsic) semiconductor in order to change its electrical properties*. The impurities are dependent upon the type of semiconductor. Lightly and moderately doped semiconductor is referred to as *extrinsic*.

Boron, arsenic, phosphorus and occasionally gallium are used to dope silicon. Boron is the *p-type* dopant of choice for silicon integrated circuit production, since it diffuses at a rate which makes junction depths easily controllable. Phosphorus is typically used for bulk doping of silicon wafers, whilst arsenic is used to diffuse junctions, since it diffuses more slowly than phosphorus and is thus more controllable.

(a) Doping of Phosphorous: *n*-Type Silicon Semiconductor

The process of 'doping' introduces an atom of another element into the silicon crystal to alter its electrical properties. The dopant has either three or five valence electrons, as opposed to silicon's four. Phosphorus atoms, which have five valence electrons, are used for doping *n-type* silicon (because phosphorus provides its fifth, free, electron). A phosphorus atom occupies the same place in the crystal lattice that was occupied formerly by the silicon atom it replaced. Four of valence electrons of phosphorus combines with four silicon valence electrons. But the fifth valence electron remains free, without bonding responsibilities. When numerous phosphorus atoms are substituted for silicon in a crystal, many free electrons become available.

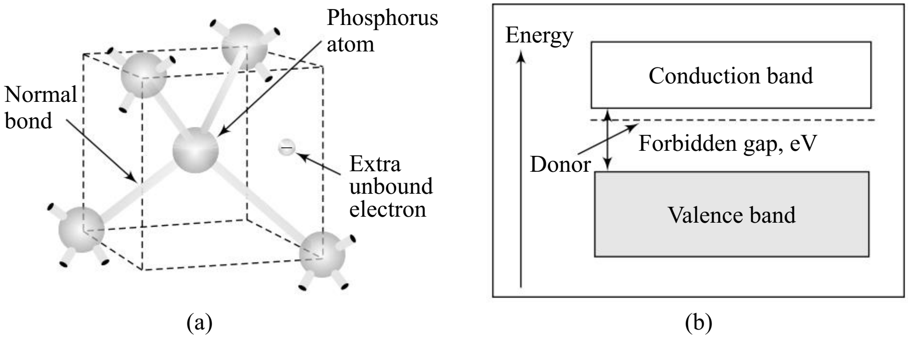


FIGURE 1.14 *(a) Phosphorus doping. (b) Donor energy levels.*

Substituting a phosphorus atom (with five valence electrons) for a silicon atom in a silicon crystal leaves an extra, unbonded electron [as shown in Fig. 1.14(a)] that is relatively free to move around the crystal and its band conduction is depicted in Fig 1.14(b).

The most common method of doping is to coat the top of a layer of silicon with phosphorus and then heat the surface. This allows the phosphorus atoms to diffuse into the silicon. The temperature is then lowered so that the rate of diffusion drops to zero. Other methods of introducing phosphorus into silicon include gaseous diffusion, a liquid dopant spray-on process and a technique in which phosphorus ions are driven precisely into the surface of the silicon.

(b) Doping of Boron: *p*-Type Silicon Semiconductor

Of course, *n*-type silicon cannot form the electric field by itself; it is also necessary to have some silicon altered to have the opposite electrical properties. So, boron, which has three valence electrons, is used for doping *p*-type silicon. Boron is introduced during silicon processing, where silicon is purified for use in PV devices. When a boron atom assumes a position in the crystal lattice formerly occupied by a silicon atom, there is a bond missing an electron (in other words, an extra hole) as shown in Fig. 1.15(a).

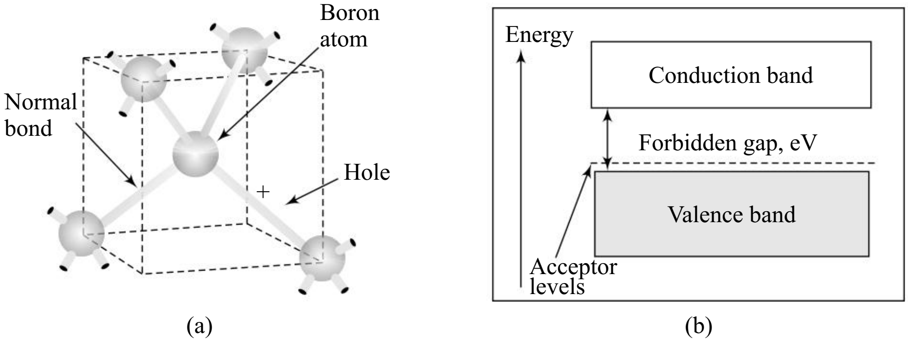


FIGURE 1.15 *(a) Doping boron. (b) Acceptor energy levels.*

Substituting a boron atom (with three valence electrons) for a silicon atom in a silicon crystal leaves a hole (a bond missing an electron) that is relatively free to move around the crystal (Fig 1.15(b).

1.30 Physical and Chemical Properties of Silicon

Silicon is the most abundant electropositive element in the earth's crust. It's a metalloid with a marked metallic luster and very brittle. It is usually tetravalent in its compounds, although sometimes its bivalent, and it's purely electropositive in its chemical behaviour. Moreover, penta-coordinated and hexa-coordinated silicon compounds are also known.

Elemental silicon has the physical properties of metalloids, similar to *Germanium*, situated under it in the group IV of the *periodic table*. Silicon is an intrinsic semiconductor in its purest form, although the intensity of its semiconduction is highly increased by introducing small quantities of impurities. Silicon is similar to metals in its chemical behaviour.

It is almost as electropositive as tin and much more positive than germanium or lead. According to this metallic character, it forms tetrapositive ions and various covalent compounds; it appears as a negative ion only in a few silicides and as a positive constituent of oxyacids or complex anions.

It forms various series of hydrides, various halides (many of which contain silicon-silicon bonds) and many series of compounds which contain oxygen, which can have ionic or covalent properties.

Applications

Silicon is the principal component of glass, cement, ceramics, most semiconductor devices and silicones, the latter a plastic substance often confused with silicon. Silicon is also an important constituent of some steels and a major ingredient in bricks. It is a refractory material used in making enamels and pottery.

- Elemental raw silicon and its intermetallic compounds are used as alloy integrals to provide more resistance to *aluminium, magnesium, copper* and other metals.
- Metallurgic silicon with 98–99 percent purity is used as raw material in the manufacture of organosilicic and silicon resins, seals and oils.
- Silicon chips are used in integrated circuits. Photovoltaic cells for direct conversion of solar energy use thin cut slices of simple silicon crystals of electronic grade.
- Silicon dioxide is used as a raw material to produce elemental silicon and silicon carbide. Big silicon crystals are used for piezoelectric glasses.
- Melted quartz sands are transformed in silicon glasses which are used in laboratories and chemical plants, as well as in electric insulators.
- A colloidal dispersion of silicon in water is used as a coating agent and as ingredient for certain enamels.

It is known that silicon forms compounds with 64 out of the 96 stable elements and possibly form silicides with other 18 elements. Apart from metallic silicides, which are used in big quantities in metallurgy, it forms important commonly used compounds with *hydrogen, carbon*, halogens, *nitrogen, oxygen* and sulphur. Moreover, many useful organosilicic by-products.

Silicon in the Environment

Silicon is found in many dioxide forms and in uncountable variations from the natural silicates.

The silicon is much more abundant than any other element, apart from the oxygen. It constitutes 27.72 percent of the solid earth's crust, while the oxygen constitutes 46.6 percent, and the next element after silicon, aluminium, is found to be 8.13 percent.

Sand is used as source of the silicon produced commercially. A few silicate minerals are mined, e.g., talc and mica. Other mined silicates are feldspars, nephenile, olivine, vermiculite, perlite, kaolinite, etc. At the other extreme there are forms of silica so rare that they are desirable for this reason alone: gemstone opal, agate and rhinestone.

Health Effects of Silicon

Silicon concentrates in no particular organ of the body but is found mainly in connective tissues and skin.

Silicon is nontoxic as an element; and in all its natural forms, namely silica and silicates, which are most abundant in nature. Elemental silicon is an inert material, which appears to lack the property of causing fibrosis in lung tissue. However, slight pulmonary lesions have been reported in laboratory animals from intratracheal injections of silicon dust. Silicon dust has little adverse effect on the lungs and does not appear to produce significant organic disease or toxic effects when exposures are kept beneath exposure limits.

Lung cancer is associated with occupational exposures to crystalline silica specifically quartz and cristobalite. Several epidemiological studies have reported statistically significant numbers of excess deaths or cases of immunologic disorders and autoimmune diseases in silica-exposed workers. Crystalline silica may affect the immune system, leading to mycobacterial infections (tuberculous and nontuberculous) or fungal, especially in workers with silicosis.

1.31 Production of Solar Grade Silicon from Quartz

Chemical Methods

The silica is reduced (oxygen removed) through a reaction with carbon in the form of coal, charcoal and heating to 1500–2000°C in an electrode arc furnace.

$$\underset{\text{(Quartz)}}{SiO_2} + \underset{\text{(coke)}}{C} \longrightarrow Si + CO_2$$

Liquid silicon collects at the bottom of the furnace, and is then drained and cooled. The silicon produced via this process is called *metallurgical grade silicon. The resulting silicon is metallugical grade silicon (MG-Si*). It is 98 percent pure silicon.

The use of silicon in semiconductor devices demands a much greater purity than afforded by metallurgical grade silicon. A number of methods have been used to produce high-purity silicon.

A small amount of the metallurgical grade silicon is further refined for the semiconductor industry. Powdered MG-Si is reacted with anhydrous HCl at 300°C in a fluidised bed reactor to form $SiHCl_3$

$$Si + 3HCl \rightarrow SiHCl_3 + H_2\uparrow$$

During this reaction impurities such as Fe, Al and B react to form their halides (e.g. $FeCl_3$, $AlCl_3$ and BCl_3). The $SiHCl_3$ has a low boiling point of 31.8°C and distillation is used to purify the $SiHCl_3$ from the impurity halides. The resulting $SiHCl_3$ now has electrically active impurities (such as Al, P, B, Fe, Cu or Au) of less than 1 ppb.

Finally, the pure $SiHCl_3$ is reacted with hydrogen at 1100°C for ~200–300 hours to produce a very pure form of silicon.

$$SiHCl_3 + H_2 \rightarrow Si + 3HCl$$

The reaction takes place inside large vacuum chambers and the silicon is deposited onto thin polysilicon rods (small grain size silicon) to produce high-purity polysilicon rods of diameter 150–200 mm. The process is often referred to as the *Siemens process*.

Silicon produced from this and similar process is called *polycrystalline silicon*. Polycrystalline silicon typically has impurity levels of 1 part per billion or less.

The resulting rods of semiconductor grade silicon are broken up to form the feedstock for the crystallization process. The production of semiconductor grade silicon requires a lot of energy. Solar cells can tolerate higher levels of impurity than integrated circuit fabrication and there are proposals for alternative processes to create a 'solar-grade' silicon.

Physical Methods

1. In *zone melting*, the first silicon purification method to be widely used industrially, rods of metallurgical grade silicon were heated to melt at one end. Then, the heater was slowly moved down the length of the rod, keeping a small length of the rod molten as the silicon cooled and resolidified behind it. Since most impurities tend to remain in the molten region rather than resolidify, when the process is complete, most of the impurities in the rod had been moved into end that is the last to be melted. This end is then cut off and discarded, and the process repeated if a still higher purity is desired.
2. The majority of silicon crystals grown for *device production* are produced by the *Czochralski process*, since it is the cheapest method available. However, silicon single-crystals grown by the Czochralski method contain impurities since the crucible which contains the melt dissolves. For certain electronic devices, particularly those required for high power applications, silicon grown by the Czochralski method is not pure enough.

(1) Single-Crystal Silicon

The most widely used technique for making single-crystal silicon is the Czochralski process, in which a seed of single-crystal silicon contacts the top of molten silicon. As the seed is slowly raised, atoms of the molten silicon solidify in the pattern of the seed and extend the single-crystal structure.

After growing the silicon ingot, we must saw it into thin *wafers* for further processing into PV cells.

(a) Czochralski Process

In the Czochralski process, a seed crystal is dipped into a crucible of molten silicon and withdrawn slowly, pulling a cylindrical single crystal as the silicon crystallizes on the seed (Fig. 1.16).

For these applications, float-zone silicon (FZ-Si) can be used instead.

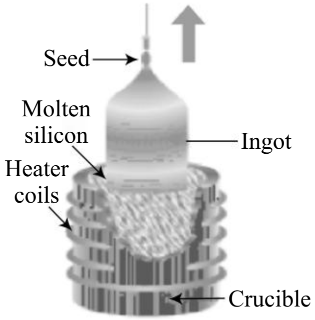


Figure 1.16 *Silicon crystal from melt.*

(b) Float-zone silicon

The float-zone process produces purer crystals than the Czochralski method, because they are not contaminated by the crucible used in growing Czochralski crystals. In the float-zone process, a silicon rod is set atop a seed crystal and then lowered through an electromagnetic coil. The coil's magnetic field induces an electric field in the rod, heating and melting the interface between the rod and the seed. Single-crystal silicon forms at the interface, growing upward as the coils are slowly raised.

Once the single-crystal rods are produced, by either the Cz or Fz method, they must be sliced or sawn to form thin wafers. The resulting thin wafers are then doped to produce the necessary electric field. They are then treated with a coating to reduce reflection, and coated with electrical contacts to form functioning PV cells.

(2) Multicrystalline Silicon

Multicrystalline silicon devices are generally less efficient than those of single-crystal silicon, but they can be less expensive to produce. The multicrystalline silicon can be produced in a variety of ways. The most popular commercial methods involve a casting process in which molten silicon is directly cast into a mould and allowed to solidify into an ingot Fig. 1.17.

The starting material can be a refined lower-grade silicon, rather than the higher semiconductor grade required for single-crystal material. The cooling rate is one factor that determines the final size of crystals in the ingot and the distribution of impurities. The mould is usually square, producing an ingot that can be cut and sliced into square cells that fit more compactly into a PV module.

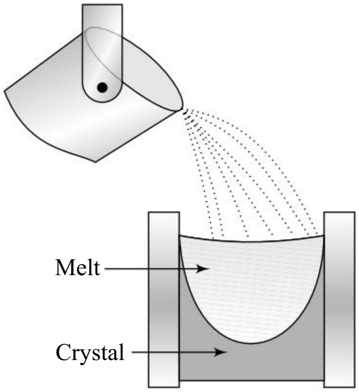


Figure 1.17 *Silicon crystal from melt.*

(3) Amorphous Silicon

Amorphous solids, like common glass, are materials whose atoms are not arranged in any particular order. They do not form crystalline structures at all, and they contain large numbers of structural and bonding defects. But they have some economic advantages over other materials that make them appealing for use in solar electric, or photovoltaic (PV), systems.

Amorphous silicon is common in solar-powered consumer devices that have *low power* requirements, such as wristwatches and calculators.

Amorphous silicon absorbs solar radiation 40 times more efficiently than does single-crystal silicon, so a film only about 1 μm thick can absorb 90 percent of the usable light energy shining on it. This is one of the chief reasons that amorphous silicon could reduce the cost of photovoltaics. Other economic advantages are that it can be produced at lower temperatures and can be deposited on low-cost substrates such as plastic, glass and metal. This makes amorphous silicon ideal for building-integrated PV products like the one shown in the photo. And these characteristics make amorphous silicon the leading thin-film PV material.

1.32 Chemical Vapour Deposition Technique

The general technique of deposition is known as *chemical vapour deposition* (CVD). CVD is commonly used to deposit layers of polycrystalline silicon, silicon dioxide and silicon nitride on the substrate.

CVD is accomplished by placing the *substrate wafers* in a reactor chamber and heating them to a certain temperature. Controlled amounts of silicon or nitride source gases, usually carried by either nitrogen and / or hydrogen, are added to the reactor. Dopant gases may also be added if desired. A reaction between the source gases and the wafer occurs, thereby depositing the desired layer. Reaction temperatures between 500 and 1100°C and pressures ranging from atmospheric to low pressure are used, depending on the specific deposition performed. Heating is usually accomplished with radio frequency, infrared, or thermal resistance heating. Common source gases include silane, silicon tetrachloride, ammonia and nitrous oxide. Some dopant gases that are used include arsine, phosphine and diborane.

Chemical vapour deposition (CVD) is a very versatile process (Fig. 1.18) used in the production of coatings, powders, fibres and monolithic parts. With CVD, it is possible to produce almost any metallic or nonmetallic element, including carbon and silicon, as well as compounds such as carbides, nitrides, borides, oxides, intermetallics and many others. In addition to being able to penetrate porous bodies, blind holes, large L/D tubes, etc.

CVD offers many advantages over other deposition processes.

These include:

- Versatile—can deposit any element or compound
- High purity—typically 99.99 percent
- High density—nearly 100 percent of theoretical
- Material formation is well below the vapour generation
- Coatings deposited by CVD are conformal and near net shape
- Economical in production, since many parts can be coated at the same time.

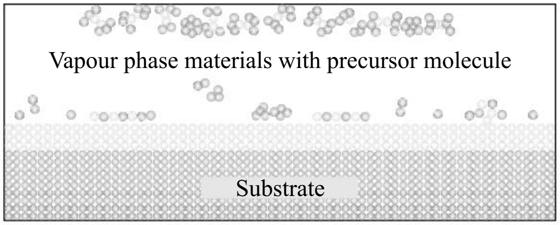


Figure 1.18 *Chemical vapour deposition process of a material with a precursor.*

Coatings, the most common application of CVD, generally fall into one of two categories (electronic materials or protective coatings), and are applied either as conversion coatings or as deposited coatings. Conversion coatings involve the surface formation of a compound where one of the elemental components is already present on the surface. Oxidizing a silicon wafer to form silicon dioxide (SiO_2) is an example of a conversion coating. On the other hand, all of the elemental constituents of a deposited coating can come from the vapour phase. For example, the deposition of silicon from silane would be considered a deposited coating.

1.33 Manufacture of Photovoltaic Cell

The process of fabricating conventional single—and polycrystalline silicon PV cells need very pure semiconductor-grade polysilicon—a material processed from quartz and used extensively throughout the electronics industry. The polysilicon is then heated to melting temperature, and trace amounts of boron are added to the melt to create a *p*-type semiconductor material.

Next, an ingot, or block of silicon is formed, commonly using one of two methods:

1. by growing a pure crystalline silicon ingot from a seed crystal drawn from the molten polysilicon or
2. by casting the molten polysilicon in a block, creating a polycrystalline silicon material.

Individual wafers are then sliced from the ingots using wire saws and then subjected to a surface etching process. After the wafers are cleaned, they are placed in a phosphorus diffusion furnace, creating a thin *n*-type semiconductor layer around the entire outer surface of the cell. Next, an antireflective coating is applied to the top surface of the cell, and electrical contacts are imprinted on the top (negative) surface of the cell. An aluminized conductive material is deposited on the back (positive) surface of each cell, restoring the *p*-type properties of the back surface by displacing the diffused phosphorus layer. Each cell is then electrically tested, sorted based on current output, and electrically connected to other cells to form cell circuits for assembly in PV modules.

Exercises

1. A coal has the following composition by weight C = 90 percent, S = 0.5 percent, N = 0.5 percent, and ash = 2.5 percent. Net calorific value of the fuel was found to be 8490.5 Kcal/kg. Calculate the percentage of H and GCV of coal. (Bombay, B.E 1997). [Ans: 8731.8 kcal/kg].

2. 0.6 g of coal sample with 92 percent of carbon, 5 percent of hydrogen and 3 percent of ash caused a rise in temperature of 1500 g of water by 4.2°C in a bomb calorimeter experiment. Calculate the gross and net calorific value of coal given water equivalent of calorimeter = 200 g. [Ans: GCV= 49,825.3 kJ/kg].
3. A solid fuel contains 80 percent carbon,10 percent hydrogen, 2 percent sulphur, 3 percent nitrogen 1 percent oxygen and 4 percent ash. Calculate the GCV and NCV value of solid fuel. [Ans: GCV=9913.9 Kcal/kg].
4. Calculate the GCV and NCV of a fuel from the following data. Mass of fuel burnt = 0.55 g. Water equivalent mass of calorimeter = 500 g and mass of water taken = 1000 g. Increase in temperature was 2.8°C. Percentage of hydrogen in the fuel = 3.2.
   [Ans: GCV =17,585.4 kJ/kg, Heat of condensation of steam = 707.84 kJ/kg, NCV= 16,877.56 kJ/kg⁻]
5. A sample of coal contains C = 93 percent, H = 6 percent and ash = 1 percent. Calculate the gross and net calorific value of the coal from the following data:
   Weight of coal burnt =0.92 g
   Weight of water taken = 550 g
   Water equivalent of calorimeter = 2200 g
   Rise in temperature = 2.42°C
   [Ans: GCV = 30,287.48 kJ/kg, NCV = 28,960.29 kJ/kg]
6. 0.72 g of a fuel containing 80% carbon, when burnt in bomb calorimeter increased the temperature of water from 27.3°C – 29.1°C, if the calorimeter contains 250 g of water and its water equivalent is 150g. Calculate the GCV of the fuel.
   (Mangalore, BE, 1997) [Ans: 4187 kJ/kg].
7. 0.94g of a fuel on combustion with oxygen increased the temperature of water in a calorimeter containing 1400 g of water by 2.5°C. The water equivalent of calorimeter is 150 g. Calculate the GCV of the fuel.
   [Ans: 4122.3 cal/g].
8. A coal has the following analysis data: Carbon = 84 percent, sulphur = 1.5 percent, Nitrogen = 0.6 percent hydrogen 5.5 percent, and oxygen 8.4 percent Find the gross and net calorific values of the fuel.
   [Ans: GCV = 8356 kcal/kg and NCV = 7912.2 kcal/kg]
9. Calculate the Gross calorific value of a coal from the following data:
   Mass of coal = 0.5 g.
   Water equivalent of calorimeter = 2000 g
   Specific heat of water = 4.187 kJ/kg/°C
   Rise in temperature of water = 5.4°C.
   [Ans: 90,439.2 kJ/kg].
10. 0.134 kg of a solid fuel was burnt in a bomb calorimeter. Calculate the gross and net calorific values if weight of water taken is 2.1 kg. water equivalent of calorimeter is 1.1 kg, initial and final temperatures of water are 24 and 35°C respectively and the percentage of hydrogen in the fuel is 2. Specific heat of water = 4.18 kJ/kg/°C. [Ans: GCV = 1098.03 kJ/kg, NCV = 656.37 kJ/kg].

Review Questions

Fuels

1. Write a note on energy systems.
2. What do you mean by nonconventional source of energy? Mention few examples.
3. What are conventional energy sources? Give examples.
4. A note on solar energy.
5. A note on hydel energy.
6. A note on Tidal energy.
7. A note on wind energy
8. A note on geothermal energy.
9. What are fuels? How are they classified? Explain with suitable examples.
10. Define calorific value of a fuel
11. Define (i) GCV and (ii) NCV. How are they related to each other?

12. Describe bomb calorimetric method of determination of the calorific value of a solid or liquid fuel. [VTU: Aug 2000, Feb 2000, Aug 2001]
13. Write a brief account of hydrocarbon fuels.
14. Describe Bomb calorimetric method for the determination of calorific value of a gaseous fuel.
15. Write a short note on fractional distillation of petroleum. Explain the different methods of cracking.
16. What is cracking? [March 1999]
17. Describe the fluidized bed catalytic cracking. What are its advantages? [VTU: Aug 1999, Feb. 2000]
18. What is reforming? Explain with examples any four reforming reactions. How does reforming increase octane number? [VTU: Aug. 2001]
19. Explain the term petrol knocking. How is it minimized? Explain the mechanism of knocking in chemical. [VTU: Aug 2000, March 1999]
20. Write a note on antiknocking agents.
21. Explain the term octane number.
22. Explain how octane number can be modified.
23. Write a note on unleaded petrol
24. What is meant by diesel knocking? How can it be minimized?
25. Give reasons for the following.
    (a) GCV is always higher than NCV
    (b) In Bomb calorimeter what we measure is Gross Calorific Value of the fuel.
    (c) Fluidized bed catalytic cracking is superior to fixed bed catalytic cracking.
    (d) Unleaded petrol is better for IC engines
    (e) Reforming reaction increase octane number of fuel.
26. 0.945 g of a fuel on complete combustion in excess of oxygen increased temperature of water is a calorimeter from 13.25°C to 19.2°C. The mass of water in the calorimeter was 1458 g. Calculate GCV if water equivalent of calorimeter is 144 g. [Ans: GCV: 42,232.67 kJ/kg]
27. Write a note on power alcohol.
28. Mention few characteristics of alcohol blended petrol for internal combustion engines as fuel.
29. (a) Describe the experimental determination of calorific value of a solid fuel using Bomb calorimeter.
    (b) What is synthetic petrol? Describe the Bergius method of synthetic of petrol
    (c) Write a note on reforming of petrol.
    (d) Explain construction and working of silicon photovoltaic cell.
30. (a) Define (i) Gross calorific value (ii) Net calorific value of a fuel. How are they related?
    (b) Calculate the gross and net calorific value of a coal sample from the following data obtained from a bomb calorimeter experiment:
        (i) Weight of coal = 0.65 kg
        (ii) weight of water taken in calorimeter = 1200 kg
        (iii) Water equivalent of calorimeter = 400 kg
        (iv) Latent heat of steam = 587 × 4.2 kJ/kg
        (v) Hydrogen in the coal sample = 2 percent
        (vi) Rise in temperature = 1.8°C
        (vii) Specific heat of water = 4.187 kJ/kg/°C

        [Ans.: GCV:18551.63 kJ/kg^{-}, NCV = 18107.86 kJ/kg]
31. What is knocking in IC engines? Explain the mechanism of knocking. How can it be prevented?
32. (a) Describe the Bomb calorimetric method of determination of calorific value of solid fuel.
    (b) What is reforming of petroleum? Give any three reactions involved in reforming.
    (c) What is power alcohol? Give its advantages as a fuel.
    (d) What is a photo voltaic cell? Explain its working.
33. (a) Define gross and net calorific value of a fuel. Describe how the calorific value of a gaseous fuel is determined by using Bomb calorimeter.
    (b) Write a note on 'Cetane number'.

34. (a) Distinguish between gross and net calorific value of a fuel.
(b) What is meant by cracking of petroleum? Explain fluidized bed catalytic cracking.
(c) On burning 0.96 g of a solid fuel in Bomb calorimeter, the temperature of 3500 g of water increased by 2.7°C Water equivalent of calorimeter, and latent heat of steam are 385 g and 587 cal/g, respectively. If the fuel contains 5 percent H_2 calculate its gross and net calorific values.
[Ans: GCV = 45749.52 kJ/kg, NCV = 44643.52 kJ/kg]
(d) Write a note on power alcohol

35. (a) What is cracking of petroleum? Explain the process of fluidized bed catalytic cracking of petroleum?
(b) Define gross and net calorific value of a fuel. Explain bomb calorimetric method of determination of calorific value of a solid fuel.
(c) On burning 0.75×10^{-3} kg of a solid fuel in a bomb calorimeter, the temperature of 2.5 kg of water is increased from 240° to 280°C. The water equivalent of calorimeter and latent heat of steam are 0.485 kg and 4.2×587 kJ/kg, respectively. Specific heat of water is 4.2 kJ/kg/°C. If the fuel contains 2.5 percent hydrogen, calculate its gross and net calorific values.
[Ans: GCV: 6,68,640 kJ/kg, NCV: 6,67,708.08 kJ/kg]

36. (a) Define calorific value of fuel. Describe the Boy's calorimetric method of determining the calorific value of a gaseous fuel.
(b) (i) A 0.6 g coal sample with 92 percent C, 5 percent H_2 and 3 percent ash, caused a rise in the temperature of 2000 g of water by 3.20 C in a bomb calorimeter experiment. Calculate the gross and net calorific value of a coal , given water equivalent = 200 g
(ii) Specific heat of water = 4.187 kJ/kg/°C
(iii) Latent heat of steam = 580 cal/g
(1 calorie = 4.18 J)
[Ans: GCV: 44,661.33 kJ/kg, NCV: 43,570.35 kJ/kg]
(c) Write a note on knocking in IC engines.

37. (a) Define the following; gross calorific value, net calorific value.
(b) The gross calorific value of a sample of bituminous coal is 36,000 kJ/kg. In an experiment, 0.83 g of this coal burnt under 1.2 kg of water equivalent of the calorimeter. Due to combustion, the temperature of water rose by 3.92°C. Calculate the water equivalent of the calorimeter. Specific heat of water = 4.2 kJ/kg/°C.
[Ans: 0.6149 kg]
(c) Explain the process of fluidized bed catalytic cracking of petroleum products

38. (a) Describe the experimental determination of calorific value of a solid fuel using Bomb calorimeter.
(b) What is synthetic petrol? Describe the Bergius method of synthetic of petrol
(c) Write a note on reforming of petrol.
(d) Explain construction and working of silicon photovoltaic cell.

39. (a) Define (i) Gross calorific value (ii) Net calorific value of a fuel. How are they related?
(b) Calculate the gross and net calorific value of a coal sample from the following data obtained from a bomb calorimeter experiment:
(i) Weight of coal = 0.65 kg
(ii) weight of water taken in calorimeter = 1200 kg
(iii) Water equivalent of calorimeter = 400 kg
(iv) Latent heat of steam = 587×4.2 kJ/kg
(v) Hydrogen in the coal sample = 2%
(vi) Rise in temperature = 1.8°C
(vii) Specific heat of water = 4.187 kJ/kg/°C [Ans: GCV:18551.63 kJ/kg, NCV = 18107.86 kJ/kg]

(c) What is knocking in IC engines? Explain the mechanism of knocking. How can it be prevented?

40. (a) Describe the Bomb calorimetric method of determination of calorific value of solid fuel.
(b) What is reforming of petroleum? Give any three reactions involved in reforming.
(c) What is power alcohol? Give its advantages as a fuel.
(d) What is a photovoltaic cell? Explain its working.

41. (a) Define gross and net calorific value of a fuel. Describe how the calorific value of a gaseous fuel is determined by using Bomb calorimeter.
(b) Write a note on 'Cetane number'.

42. (a) Distinguish between gross and net calorific value of a fuel
(b) What is meant by cracking of petroleum? Explain fluidized bed catalytic cracking
(c) On burning 0.96 g of a solid fuel in Bomb calorimeter, the temperature of 3500 g of water increased by 2.7°C Water equivalent of calorimeter, and latent heat of steam are 385 g and 587 cal/g, respectively. If the fuel contains 5 percent H_2, calculate its gross and net calorific values.
[Ans: GCV = 45,749.52 kJ/kg, NCV = 44,643.52 kJ/kg]
(d) Write a note on power alcohol

43. (a) What is cracking of petroleum? Explain the process of fluidized bed catalytic cracking of petroleum.
(b) Define gross and net calorific value of a fuel. Explain bomb calorimetric method of determination of calorific value of a solid fuel.
(c) On burning 0.75×10^{-3} kg of a solid fuel in a bomb calorimeter, the temperature of 2.5 kg of water is increased from 240–280°C. The water equivalent of calorimeter and latent heat of steam are 0.485 kg and 4.2 $\times$ 587 kJ/kg, respectively. Specific heat of water is 4.2 kJ/kg/°C. If the fuel contains 2.5 percent hydrogen, calculate its gross and net calorific values.
[Ans: GCV: 6,68,640 kJ/kg, NCV: 6,67,708.08 kJ/kg]

44. (a) Define calorific value of fuel. Describe the Boy's calorimetric method of determining the calorific value of a gaseous fuel.
(b) A 0.6 g coal sample with 92 percent C, 5 percent H_2 and 3 percent ash, caused a rise in the temperature of 2000 g of water by 3.20 C in a bomb calorimeter experiment. Calculate the gross and net calorific value of a coal, given water equivalent = 200 g
Specific heat of water = 4.187 kJ/kg/°C
Latent heat of steam = 580 cal/g
(1 calorie = 4.18 J) delete
[Ans: GCV: 44661.33 kJ/kg, NCV: 43570.35 kJ/kg]
(c) Write a note on knocking in IC engines

45. (a) Define the following; gross calorific value, net calorific value.
(b) The gross calorific value of a sample of bituminous coal is 36,000 kJ/kg. In an experiment, 0.83 g of this coal burnt under 1.2 kg of water equivalent of the calorimeter. Due to combustion, the temperature of water rose by 3.92°C. Calculate the water equivalent of the calorimeter. Specific heat of water = 4.2 kJ/kg /°C.
[Ans: 0.6149 kg]
(c) Explain the process of fluidized bed catalytic cracking of petroleum products.

46. (a) Define calorific value? What are gross and net calorific values?
[UP Technical University, 2014–15]
(b) Calculate the net and gross calorific value of a coal sample containing 84 percent of carbon, 1.5 percent of sulphur, 6 percent nitrogen, and 8.4 percent of oxygen. Calorific value of carbon, hydrogen, sulphur are 8080 kJ/kg, 34,500 kcal/kg and 2240 kJ/kg, respectively, latent heat of steam is 587cal/g.
[UP Technical University, 2012–2013]

47. On burning 0.83 g of a solid fuel in a Bomb calorimeter, the temperature of 3500 g of water increased from 26.5–29.2°C. Water equivalent of calorimeter and latent heat of steam are 385 g and 587 g, respectively. If the fuel contains 0.7 percent hydrogen, calculate its gross and net calorific values. [UP Technical University, 2012–2013]
48. A coal sample has the following composition: C = 70 percent, Hydrogen 10 percent, Nitrogen 3 percent, Oxygen 2 percent, and Ash 12 percent. If 20 percent excess air is required for complete combustion, then calculate the amount of air required for complete combustion of 1 kg of coal. [Pune University, 2015]
49. Define GRV and CRV. Give the justification how they are related. [Pune University, 2015]
50. Define Gross/higher calorific value and justify the relationship between GCV and NCV of the fuel, if its H% (Hydrogen). [Pune University, 2013]
51. Draw neat labelled diagram and give the construction and working of bomb calorimeter to determine GCV of a fuel. State formula with corrections to calculate GCV. [Pune University, 2016].
52. Differentiate HCV and LCV. [JNTU, 2014]
53. What is knocking of petrol engine? Define Octane number and explain effect of chemical structure of hydrocarbons present in petrol on knocking. [Pune University, 2013]

Fuels-Photovoltaic Cells

1. What is a photovoltaic cell?
2. How do you synthesize a *p*/*n* junction material for photovoltaic cell?
3. What is a *p*–*n* junction? Mention its importance in PV cells.
4. How does a photovoltaic cell work? Explain.
5. Explain the doping concept of a silicon semiconductor.
6. How do you get an *n*-type silicon semiconductor for PV cells
7. How do you obtain *p*-type silicon?
8. What are the physical and chemical properties of silicon?
9. Mention few applications of silicon.
10. How do you get photovoltaic grade silicon?
11. Mention the types of PV grade silicon.
12. How to get poly crystalline PV grade silicon?
13. Give any one method to obtain PV grade silicon.
14. Explain zone refining technique for purification of silicon and diffusion method. [VTU: 2016]
15. Give the construction and working of a photovoltaic cell. [VTU: June/July 2016]
16. Explain the production of solar grade silicon by Union Carbide process. [VTU: June/July 2015]
17. What is doping? Explain doping of silicon by diffusion method. [VTU: June/July 2015]

2

Electrochemical Energy Systems

Chapter Outline

Introduction

Electrochemical cell—classification of electrochemical cells, differences of galvanic and electrolytic cells, galvanic cell—electrochemical conventions, origin of electrode potential, electrode potentials, standard electrode potentials and electrochemical series. Measurement of electrode potentials and IUPAC sign. Nernst equation for single electrode potential, EMF of a cell. Types of electrodes or half cells, types of ion-selective electrodes, applications of ion-selective electrodes, how an ion-selective electrode works. Reference electrodes, factors affecting the potential of a reference electrode, limitations of primary reference electrode and advantages of secondary reference electrode. Ion-selective glass electrode, determination of pH of a solution using a glass electrode and a calomel electrode, determination of pH of a solution using a glass and a common Ag/AgCl electrode, concentration cells.

2.1 Introduction to Electrode Potentials and Cells

Chemical reactions involving transfer of electrons are called oxidation and reduction reactions or redox reactions. A spontaneous electrode redox process is always accompanied by a decrease in free energy. The spontaneous redox reaction at the electrode occurs as a result of the different abilities of metals to lose their electrons and the ability of the electrons to flow through the circuit. It is observed that when two dissimilar metals, one with higher electrode potential and the other with lower potential, and are connected and are immersed in their own respective metal ion solutions, the metals spontaneously develop a tendency to lose or gain electrons at their respective electrodes. In other words, *spontaneous redox reactions of a voltaic cell take place at the two electrodes with a decrease of free energy*.

2.2 Electrochemical Cell

An electrochemical cell is a device consisting of two electrodes, each in contact with a solution of its own ions, and transforms the free energy change of the redox reaction at the electrodes into electrical energy.

Energy is the driving force for chemical reactions in electrodes. In a redox reaction, the free energy released in a reaction due to movement of charged particles gives rise to a 'potential difference'. *The maximum potential difference between the two electrodes is called the 'electromotive force', E, in volts for a galvanic cell*. Therefore, the Gibbs free energy, ΔG, is the negative value of the maximum available electrical work. The electrical work $W_{\max}$ is the product of the charge and the potential of the cell.

$$\text{EMF (V)} = \frac{\text{Work (J)}}{\text{Charge (C)}}$$

$$E = \frac{-w}{q}$$

$$w = -qE \text{ and } q = nF \quad \& \quad w_{\max} = \Delta G$$

Therefore $\quad \Delta G = -nFE$

Standard condition:

$$\Delta G^\circ = -nFE^\circ$$

where 'q' is the charge in Coulombs and 'E' is the potential in volts. Note also that $q = nF$, and $F = 96{,}485$ C is the Faraday constant, whereas n is the number of moles of electrons in the reaction equation.

The working principle of an electrochemical cell is the transformation of the change of free energy of the redox reactions of the electrodes into electrical energy,

That is, $\quad \Delta G = -nFE \quad$ (if, the cell potential (E) is +ve)

In a redox reaction, the energy released in a reaction due to the movement of charged particles gives rise to a potential difference. The maximum potential difference is called the electromotive force (EMF), E, in volts. From the above, it should be apparent that the potential difference between the electrodes of a cell is a measure of the tendency for the cell reaction to take place: the more positive the cell potential, the greater the tendency for the reaction to proceed to spontaneity.

The electrochemical cells are classified into the following types.

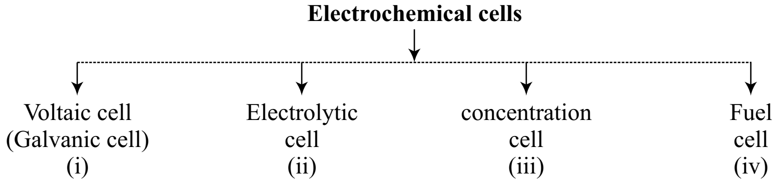


FIGURE 2.1 *(i)–(iv) depicts the various types of electrochemical cells.*

(i) Galvanic cells

The free energy change in a galvanic cell, or in a spontaneous cell reaction, must be negative and the potential difference of the electrodes (cell voltage) must be positive.

- Energy is released from spontaneous redox reaction
- System does work on load/surroundings
- Anode (–ve): oxidation
- Cathode (+ve): reduction

In galvanic cell, the free energy change of electrode redox reactions is converted into electric energy; that is, $\Delta G = -nFE$.

The common examples of galvanic cells are primary irreversible cells (dry cells) and secondary cells (lead storage battery, Ni–Cd battery, etc.).

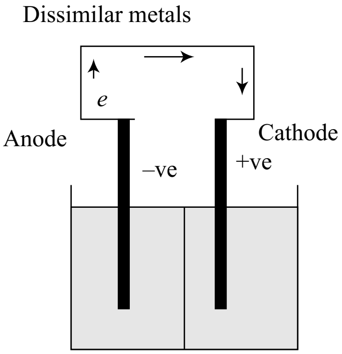


FIGURE 2.1 *(i): A galvanic cell.*

(ii) Electrolytic cells

It is a device in which electrical energy from an external source can be used to produce chemical reactions.

- Energy is absorbed to drive non-spontaneous redox reaction.
- Surroundings(power supply) do work on system (cell).
- Anode (+ve): oxidation
- Cathode (–ve): reduction.

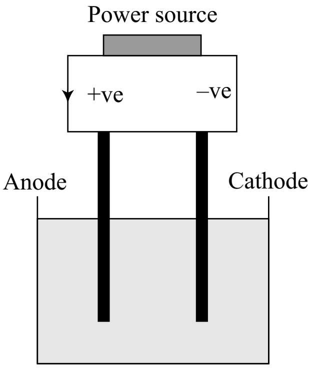


FIGURE 2.1 *(ii): An electrolytic cell.*

In electrolytic cells, the reaction is being driven in the non-spontaneous direction by external electrical force and the free energy change is positive, that is, $-nFE = \Delta G$.

(iii) Concentration cell:

A concentration cell is also an electrochemical device that generates electrical energy when two electrodes of the same metal are in contact with solutions of its ions at different concentrations.

In concentration cell, too, the free energy change of electrode reactions is converted into electric energy; that is, $\Delta G = -nFE$.

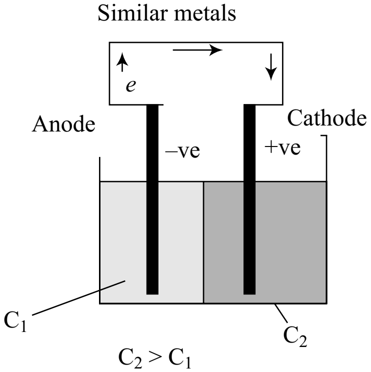


FIGURE 2.1 *(iii): Concentration cell.*

(iv) Fuel cells:

Contrast to the above, a fuel cell is also an electrochemical device, which operates with continuous replenishment of the fuel at the electrode and no charging is required.

In a fuel cell, the free energy change of electrode redox reactions is converted into electric energy.

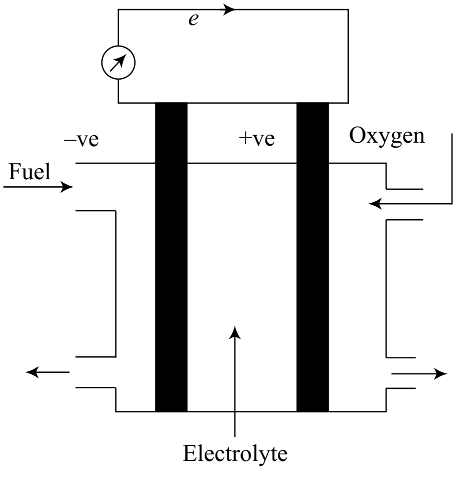


FIGURE 2.1 *(iv): A fuel cell.*

In a fuel cell, the free energy change of electrode redox reactions is converted into electric energy; that is, $\Delta G = -nFE$.

2.3 Difference between Galvanic and Electrolytic Cell

	Galvanic cell		Electrolytic cell
1	It is a device to convert chemical energy into electrical energy.	1	It is a device to convert electrical energy to chemical energy.
2	The redox reactions occurring at the electrodes are spontaneous.	2	The redox reactions occurring at the electrodes are non-spontaneous and take place only when electrical energy is supplied.
3	Electrodes are of dissimilar metals.	3	Electrodes used may be of dissimilar or same metals.
4	Each metal is dipped in its own ions	4	Only one electrolyte is taken.
5	Each electrode is immersed in its own ions and both have separate compartments.	5	Both the electrodes are immersed in the same electrolyte solution.
6	The two electrolyte solutions are connected by a salt bridge.	6	No salt bridge is needed.
7	–ve electrode of the cell is anode. +ve electrode of the cell is cathode.	7	+ve electrode of the cell is anode. –ve electrode of the cell is cathode.

2.4 A Galvanic Cell

Daniel cell is the best example of a galvanic cell. It consists of two electrodes of dissimilar metals, Zn and Cu; each electrode is in contact with a solution of its own ions; $ZnSO_4$ and $CuSO_4$, respectively.

In other words, each electrode of a galvanic cell may be regarded as a 'half-cell'. The two solutions are connected by a salt bridge. The redox electrode reactions of the Daniel cell can be represented as two half-cell reactions. They are

At anode,

$$Zn \rightarrow Zn^{2+} + 2e \qquad \text{(Oxidation)}$$

At cathode,

$$Cu^{2+} + 2e \rightarrow Cu \qquad \text{(Reduction)}$$

The overall cell redox reaction is

$$Zn + Cu^{2+} \rightarrow Zn^{2+} + Cu$$

If $E^{o}_{Cu^{2+}/Cu} > E^{o}_{Zn^{2+}/Zn}$, electrode reactions of the cell take place spontaneously and the cell potential (E) is +ve.

The electrode potential is a quantitative measure of the driving force or tendency of the half-cell reaction to occur. In a Daniel cell (Fig. 2.2), when the Zn electrode is coupled with the Cu electrode, the tendency of Zn (low reduction potential) to dissolve and form Zn^{2+} ions is greater than the tendency of Zn^{2+} ions to get deposited, and therefore, Zn half-cell acquires *negative charge*. In contrast, the tendency of Cu (high reduction potential) to dissolve in solution as Cu^{2+} is lesser than the tendency of Cu^{2+} ions to get deposited as Cu at the copper half-cell. Thus, the copper half-cell becomes *positively* charged.

By IUPAC convention, *the half-cell at which the oxidation electrode reaction takes place is called as the anode and at which the reduction electrode reaction takes place is called as the cathode.*

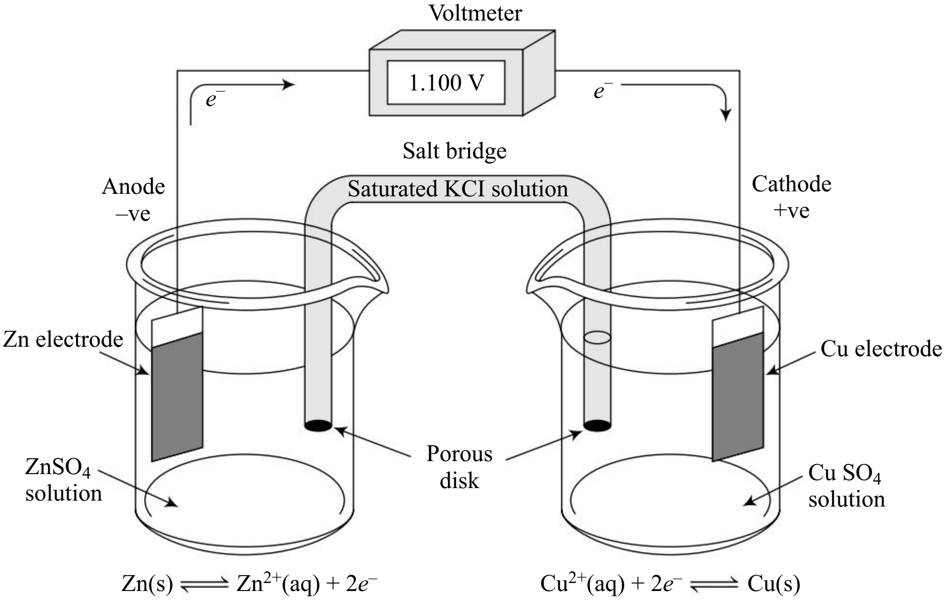


Figure 2.2 *A Daniel cell.*

Therefore, in the Daniel cell, Zn is the negative electrode (anode) and copper is the positive electrode (cathode). The electron flows from the zinc half-cell to the copper half-cell of the Daniel cell. The free energy change of the above electrode reactions is converted to electric energy.

2.5 Electrochemical Conventions

In the study of various types of electrochemical cells, the following conventions have to be strictly adopted. They are as follows:

(1) Sign of the electrode potential

If reduction reaction takes place at the cathode of the cell, then the sign of the electrode potential is +ve. The sign of the electrode potential of the anode in terms of reduction potential of the cell is –ve. That is

At cathode,

$$M^{n+} + ne \rightarrow M; \qquad E^{\circ}_{\mathrm{M}^{n+}/M} = +\mathrm{ve}$$

At anode,

$$M^{n+} + ne \rightarrow M; \qquad E^{\circ}_{\mathrm{M}^{n+}/M} = -\mathrm{ve}$$

Not necessarily as stated above, it is observed in actual practice that an electrochemical cell can be constructed with both either +ve or –ve reduction potentials; however, in each case, one of the electrodes should have higher reduction potential than the other.

(2) Representation of an electrode

If oxidation reaction takes place at the anode half-cell, then the electrode is conveniently represented as:

$$M/M^{n+}_{(C_1)}, \qquad \text{Example: } \mathrm{Zn/Zn^{2+}} \qquad \text{(Oxidation)}$$

If reduction reaction takes place at the cathode, then the electrode is represented as:

$M^{n+}_{(C_2)}/M,$ Example: Cu^{2+}/Cu (Reduction)

The vertical line indicates the interface between the metal and their ions.

(3) Representation of a cell

The following points have to be noted to represent an electrochemical cell.

(a) 'Anode' is always written to the left side.

$M/M^{n+}_{(C_1)}$ (oxidation)

(b) 'Cathode' of the cell is written to the right side of anode.

$M^{n+}_{(C_2)}/M$ (Reduction)

(c) Two vertical (||) lines denote the interface between the two salt solutions.
Therefore, a complete cell can be represented as:

$$M/M^{n+}_{(C_1)} \| M^{n+}_{(C_2)}/M$$

Example: $Zn/Zn^{2+}_{(C_1)} \| Cu^{2+}_{(C_2)}/Cu$

(4) EMF of a cell

An electrochemical cell is obtained by coupling two half-cells, in which one of the electrodes has a higher reduction potential than the other, that is

$$E^{o}_{\text{cathode}} > E^{o}_{\text{anode}}$$

EMF of a cell is defined '*as the algebraic difference between the reduction potentials of the cathode and the anode'*.

The measured EMF of a cell is

$$E_{\text{cell}} = E^{o}_{\text{cathode}} - E^{o}_{\text{anode}}$$

where E^{o}_{cathode} is the standard reduction electrode potential of the cathode and E^{o}_{anode} is the standard reduction electrode potential of the anode.

(5) Electrode and cell reactions

The net cell reaction is the sum of the two half-cell reactions that takes place in the cell.

For any electrochemical cell, the two electrode reactions are

(a) At anode,

$M_1 \rightarrow M_1^{n+} + ne$ (Oxidation)

(b) At cathode,

$M_2^{n+} + ne \rightarrow M_2$ (Reduction)

Net cell reaction, $M_1 + M_2^{n+} \rightarrow M_1^{n+} + M_2$

The electrical energy (nFE) is always produced at the expense of the decrease of free energy ($-\Delta G$) of the redox electrode processes of the cell.

$$\Delta G = -nFE$$

and hence, the EMF of the cell is +ve; if $E^{o}_{\text{cathode}} > E^{o}_{\text{anode}}$, the cell reaction is spontaneous.

Example 2.1 The electrodes reduction potentials of Zn and Cu electrodes are –0.76 and +0.34 V, respectively. Calculate the EMF of the Daniel cell.

Solution

Since $E^{o}_{Cu^{2+}/Cu} > E^{o}_{Zn^{2+}/Zn}$, the two half-cell reactions are as follows:

At anode,

$$Zn \rightarrow Zn^{2+} + 2e \qquad \text{(Oxidation);} \qquad E_{Zn/Zn^{2+}} = +0.76\ \text{V}$$

At cathode,

$$Cu^{2+} + 2e \rightarrow Cu \qquad \text{(Reduction);} \qquad E^{o}_{Cu^{2+}/Cu} = 0.34\ \text{V}$$

By convention, we have

$$\begin{aligned} E_{cell} &= E^{o}_{Cu^{2+}/Cu} - E^{o}_{Zn^{2+}/Zn} \\ &= 0.34 - (-0.76) = 1.10\ \text{V} \end{aligned}$$

Example 2.2 Represent a cadmium–copper cell. Give the electrode and net cell reactions. If the standard reduction potentials of Cd and Cu are –0.40 and 0.34 V, respectively, calculate the standard EMF of the cell.

Solution

In this problem, $E^{o}_{Cu^{2+}/Cu} > E^{o}_{Cd^{2+}/Cd}$

Therefore, Cd acts as the anode and Cu behaves as the cathode.

(i) Cell representation

$$Cd/Cd^{2+} \,||\, Cu^{2+}/Cu$$

(ii) Electrode reactions

$$\text{At anode, } Cd \rightarrow Cd^{2+} + 2e; \quad E^{o}_{Cd/Cd^{2+}} = 0.40\ \text{V}$$

$$\text{At cathode, } Cu^{2+} + 2e \rightarrow Cu; \quad E^{o}_{Cu^{2+}/Cu} = 0.34\ \text{V}$$

(iii) The net cell reaction

$$Cd + Cu^{2+} \rightarrow Cd^{2+} + Cu$$

(iv) Calculation of standard EMF of the cell.

$$\begin{aligned} E^{o}_{cell} &= E^{o}_{cathode} - E^{o}_{anode} \\ &= E^{o}_{Cu^{2+}/Cu} - E^{o}_{Cd^{2+}/Cd} \end{aligned}$$

Since, $E^{o}_{Cd/Cd^{2+}} = -E^{o}_{Cd^{2+}/Cd}$

$$= 0.34 - (-0.40) = 0.74\ \text{volt}$$

Example 2.3 The oxidation potential of the nickel half-cell is 0.25 V and the reduction electrode potential of silver half-cell is 0.80 V. (a) The cell representation, (b) the cell reaction and (c) calculate the standard EMF of the Ni–Ag cell.

Hint: $E^{o}_{Ag^{+}/Ag} > E^{o}_{Ni^{2+}/Ni}$

Solution

(a) Cell representation

$$\mathrm{Ni/Ni^{2+}_{(1M)}\,||\,Ag^{+}_{(1M)}/Ag}$$

At anode, $\mathrm{Ni \rightarrow Ni^{2+} + 2e}$; $E^{\mathrm{o}}_{\mathrm{NiNi^{2+}}} = 0.25$ V

At cathode, $\mathrm{2Ag^{+} + 2e \rightarrow 2Ag}$; $E^{\mathrm{o}}_{\mathrm{Ag^{+}/Ag}} = 0.80$ V

(b) Net cell reaction is

$$\mathrm{Ni + 2Ag^{+} \rightarrow Ni^{2+} + 2Ag}$$

(c) Standard EMF of the Ni–Ag cell

$$E^{\mathrm{o}}_{\mathrm{cell}} = E^{\mathrm{o}}_{\mathrm{cathode}} - E^{\mathrm{o}}_{\mathrm{anode}}$$
$$= E^{\mathrm{o}}_{\mathrm{Ag^{+}/Ag}} - E^{\mathrm{o}}_{\mathrm{Ni^{2+}+Ni}}$$
$$= 0.80 - (-0.25) \qquad E^{\mathrm{o}}_{\mathrm{Ni/Ni^{2+}}} = -E^{\mathrm{o}}_{\mathrm{Ni^{2+}/Ni}}$$
$$= 1.05 \text{ V}$$

2.6 Origin of Potential

When a metal piece is dipped in water or a solution of its own ions, the observed potential is zero because the rate of oxidation process of the metal in such a situation is equal to the rate of reduction at the metal surface, that is

$$M \Leftrightarrow M^{n+} + ne$$

Therefore, a state of equilibrium is attained between the two tendencies of their electrode processes in a cell and hence, the electrode potential is zero.

However, it can be observed that the situation is entirely different when two dissimilar metals are electrically connected to obtain an electrochemical cell. The driving force for the spontaneous electrode reactions arises as a result of the observed difference in the electrode potentials of the two electrodes. As a result of the driving force, the electrons released flow through the external circuit from the anode to the cathode. If $E^{\circ}_{M_2^{n+}/M_2} > E^{\circ}_{M_1^{n+}/M}$, then, the one with higher reduction potential induces the metal of lower reduction potential to oxidize and the electrode reaction moves in the forward direction only. It is observed that when two dissimilar metals with a potential difference constitute an electrochemical cell, a metal of lower reduction electrode potential has a higher tendency to undergo oxidation giving its ions.

$$M_1 \rightarrow M_1^{n+} + ne \qquad \text{(Oxidation)}$$

In such a case, a metal electrode of a cell is in contact with an ionic solution, *a potential (oxidation) is developed [Fig. 2.3(a)] due to the interaction of the charged metal interface and the surrounding ions in solution.*

Simultaneously, the electrode process taking place at the other electrode of higher reduction potential is the reduction of ions to metal at the metal interface. *A potential (reduction) is developed [Fig. 2.3(b)] due to the interaction of the charged metal interface and the surrounding ions in solution.*

$$M_2^{n+} + ne \rightarrow M_2 \qquad \text{(Reduction)}$$

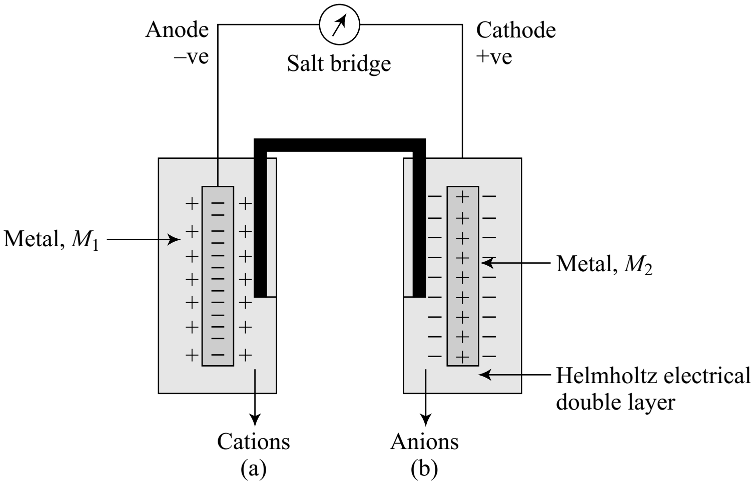


FIGURE 2.3 *Origin of electrode potential.*

In the cell shown in Fig. 2.3, M_1 and M_2 are dissimilar metal electrodes (half-cells) dipped in their own ionic solutions, respectively. The redox reactions take place at the two half-cells and ultimately a dynamic equilibrium is established between the +ve/–ve charges at the electrode interface. Therefore, due to the interaction of opposite charges at the respective electrodes, +ve (a) or –ve (b) ions remain close to the metal. This charged layer around an electrode is called *Helmholtz electrical double layer (Fig. 2.4).*

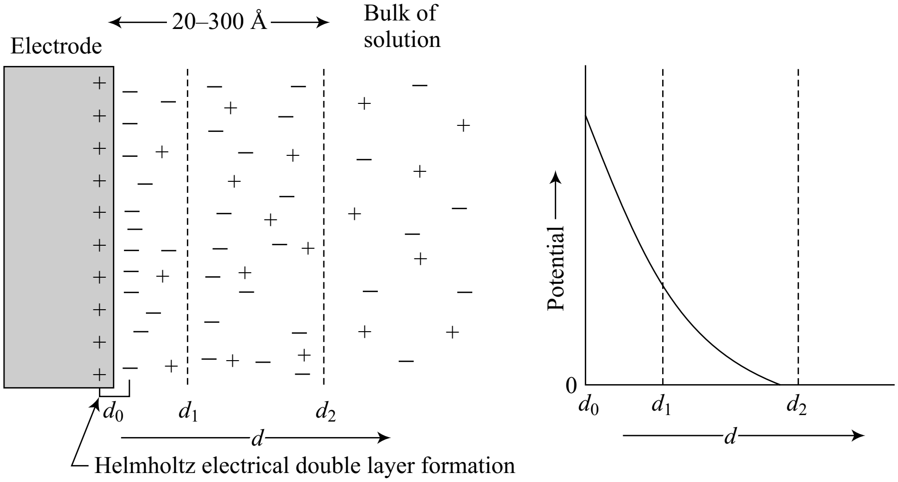


FIGURE 2.4 *Illustration of the formation of Helmholtz electrical double layer at the metal and ionic solution.*

As a result, a potential exists between the interface of a metal electrode (M) and its ionic solution (M^{n+}) at a particular temperature. This is called its electrode potential.

2.7 Electrode Potentials

(1) Single electrode potential

An electrochemical cell can be obtained by coupling two electrodes of metals having a difference in their standard electrode potentials, that is $E^{o}_{\text{cathode}} > E^{o}_{\text{anode}}$. When two dissimilar metals are in contact with their own ions in a cell, the metals have a tendency to lose or gain electrons during their electrode processes. Therefore, the *electrode potential* or *single electrodes potential* of a metal electrode of a cell is dependent on *the measure of the metal to lose or gain electrons, when it is in contact with a solution of its own ions*.

Single electrode potential can, thus, be defined as the potential established at the interface between the metal (M) and the ionic solution (M^{n+}) when it is in contact with a solution of its own ions. It is denoted as $E^{o}_{M^{n+}/M}$ (or $E^{o}_{M/M^{n+}}$).

When two electrodes are coupled to form a cell, the one with lower reduction electrode potential value acts as the '*anode*' due to the oxidation electrode process and the electrode with higher reduction electrode potential value acts as the '*cathode*'.

It is observed that the value of single electrode potential, *E*, in a cell depends on

(a) the nature of the metal electrode, *M*
(b) the concentration of metal ions, M^{n+} and
(c) the temperature.

(2) Standard electrode potentials (electrochemical series)

The standard electrode potentials of a large number of half-cells have been measured using a standard hydrogen electrode (SHE) as the reference electrode ($E^{o} = 0$).

The electrode potentials of electrode/half-cells can be measured in their standard states.

The standard electrode potential is defined as the potential of an electrode measured at 298K and unit (1M) metal in concentration.

The standard electrode potentials of two half-cells are denoted as $E^{o}_{M^{n+}/M}$ or $E^{o}_{M/M^{n+}}$

Cathode (reduction) Half-cell reaction	Standard potential E° (V)
$Li^{+}_{(aq)} + e^{-} \rightarrow Li_{(s)}$	–3.04
$K^{+}_{(aq)} + e^{-} \rightarrow K_{(s)}$	–2.92
$Ca^{2+}_{(aq)} + 2e^{-} \rightarrow Ca_{(s)}$	–2.76
$Na^{+}_{(aq)} + e^{-} \rightarrow Na_{(s)}$	–2.71
$Mg^{2+}_{(aq)} + 2e^{-} \rightarrow Mg_{(s)}$	–2.38
$Al^{3+}_{(aq)} + 2e^{-} \rightarrow Al_{(s)}$	–1.66
$Zn^{2+}_{(aq)} + 2e^{-} \rightarrow Zn_{(s)}$	–0.76
$Cr^{3+}_{(aq)} + 3e^{-} \rightarrow Cr_{(s)}$	–0.74
$Fe^{2+}_{(aq)} + 2e^{-} \rightarrow Fe_{(s)}$	–0.41
$Cd^{2+}_{(aq)} + 2e^{-} \rightarrow Cd_{(s)}$	–0.40

(*Contd.*)

(*Contd.*)

Cathode (reduction) Half-cell reaction	**Standard potential E° (V)**
$Ni^{2+}_{(aq)} + 2e^- \rightarrow Ni_{(s)}$	–0.23
$Sn^{2+}_{(aq)} + 2e^- \rightarrow Sn_{(s)}$	–0.14
$Pb^{2+}_{(aq)} + 2e^- \rightarrow Pb_{(s)}$	–0.13
$Fe^{3+}_{(aq)} + 3e^- \rightarrow Fe_{(s)}$	–0.04
$2H^{+}_{(aq)} + 2e^- \rightarrow H_{2(g)}$	0.00
$Sn^{4+}_{(aq)} + 2e^- \rightarrow Sn^{2+}_{(aq)}$	0.15
$Cu^{2+}_{(aq)} + e^- \rightarrow Cu^{+}_{(aq)}$	0.16
$AgCl_{(s)} + e^- \rightarrow Ag_{(s)} + Cl^{-}_{(aq)}$	0.22
$Cu^{2+}_{(aq)} + 2e^- \rightarrow Cu_{(s)}$	0.34
$ClO^{-}_{3(aq)} + H_2O_{(l)} + 2e^- \rightarrow ClO^{-}_{2(aq)} + 2OH^{-}_{(aq)}$	0.35
$Cu^{+}_{(aq)} + e^- \rightarrow Cu_{(s)}$	0.52
$I_{2(s)} + 2e^- \rightarrow 2I^{-}_{(aq)}$	0.54
$Fe^{3+}_{(aq)} + e^- \rightarrow Fe^{2+}_{(aq)}$	0.77
$Hg^{2+}_{2(aq)} + 2e^- \rightarrow Hg_{(l)}$	0.80
$Ag^{+}_{(aq)} + e^- \rightarrow Ag_{(s)}$	0.80
$Hg^{2+}_{(aq)} + 2e^- \rightarrow Hg_{(l)}$	0.85
$ClO^{-}_{(aq)} + H_2O_{(l)} + 2e^- \rightarrow Cl^{-}_{(aq)} + 2OH^{-}_{(aq)}$	0.90
$2Hg^{2+}_{(aq)} + 2e^- \rightarrow Hg^{2+}_{2(aq)}$	0.90
$Br_{2(l)} + 2e^- \rightarrow 2Br^{-}_{(aq)}$	1.07
$Cr_2O^{2-}_{7(aq)} + 14H^{+}_{(aq)} + 6e^- \rightarrow 2Cr^{3+}_{(aq)} + 7H_2O_{(l)}$	1.33
$Ce^{4+}_{(aq)} + e^- \rightarrow Ce^{3+}_{(aq)}$	1.44
$MnO^{-}_{4(aq)} + 8H^{+}_{(aq)} + 5e^- - Mn^{2+}_{(aq)} + 4H_2O_{(l)}$	1.49
$Co^{3+}_{(aq)} + e^- \rightarrow Co^{2+}_{(aq)}$	1.82
$F_2 + 2e \rightarrow 2F^-$	2.87

Table 2.1

Standard electrode potentials in aqueous solution at 298K

(3) Applications of electrochemical series

(i) Comparison of the relative oxidizing and reducing powers:

It can be remarked from Table 2.1 that the greater the reduction potential, greater the tendency of element/ion for reduction. Fe^{2+} has the highest reduction potential (2.87) and Li^+ (–3.04) has the lowest reduction potential. Therefore, it is evident that Fe^{2+} can be reduced most easily whereas Li^+ ions are reduced with great difficulty. In other words, Li^+ ions will be oxidized most easily and hence is the best reducing agent and vice versa.

(ii) Relative activities of the metals:

The relative activities of few metals appear in the following order: Mg > Zn > Fe > Cu > Ag

This depicts that greater the oxidation potential of a metal, more easily it can lose electrons and hence greater is its reactivity. As result, a metal with greater oxidation potential can displace metals with lower oxidation potentials from their salts from their salt solution.

(iii) To calculate the standard EMF of any electrochemical cell:

In an electrochemical cell, oxidation and reduction reactions occur at the electrodes.

The EMF of the cell is the algebraic difference of their reduction potentials.

(iv) To predict the spontaneity of redox reaction:

The free energy change in a galvanic cell, or in a spontaneous reaction, must be negative and the EMF of the cell must be positive. In other words, if the EMF of the cell is negative, the redox reaction is non-spontaneous.

2.8 Measurement of Electrode Potentials

The single electrode potential of a half-cell can be determined always by coupling this with a reference electrode, *standard hydrogen electrode* (SHE). In accordance with the IUPAC[1] convention, the sign of the electrode potentials can be obtained thus:

(1) Determination of standard electrode potential of the cathode:

If the electrode process of a half-cell is reduction when the referred electrode is coupled with a standard hydrogen electrode (SHE) as shown in Fig. 2.4. The electrode potential is assigned a +ve value.

$$M^{n+} + ne \rightarrow \mathrm{M}; \qquad E^{o}_{\mathrm{M}^{n+}/M} = +\mathrm{ve}.$$

Example:

Consider a cell (Fig. 2.5) as represented

$$\mathrm{Pt/H_2(1\,atm)/H^{+}_{(1\mathit{M})}||Cu^{2+}_{(1\mathit{M})}/Cu}$$

When a copper electrode is connected with the standard hydrogen electrode, reduction takes place at the cathode. The electrode reactions are

$$\text{At anode, } \mathrm{H_2 \rightarrow 2H^+ + 2e} \qquad \text{(Oxidation)}$$

$$\text{At cathode, } \mathrm{Cu^{2+} + 2e \rightarrow Cu} \qquad \text{(Reduction)}$$

$$E^{o}_{\mathrm{Cu^{2+}/Cu}} > E^{o}_{\mathrm{SHE}}$$

---

[1] IUPAC: International Union of Pure & Applied Chemistry.

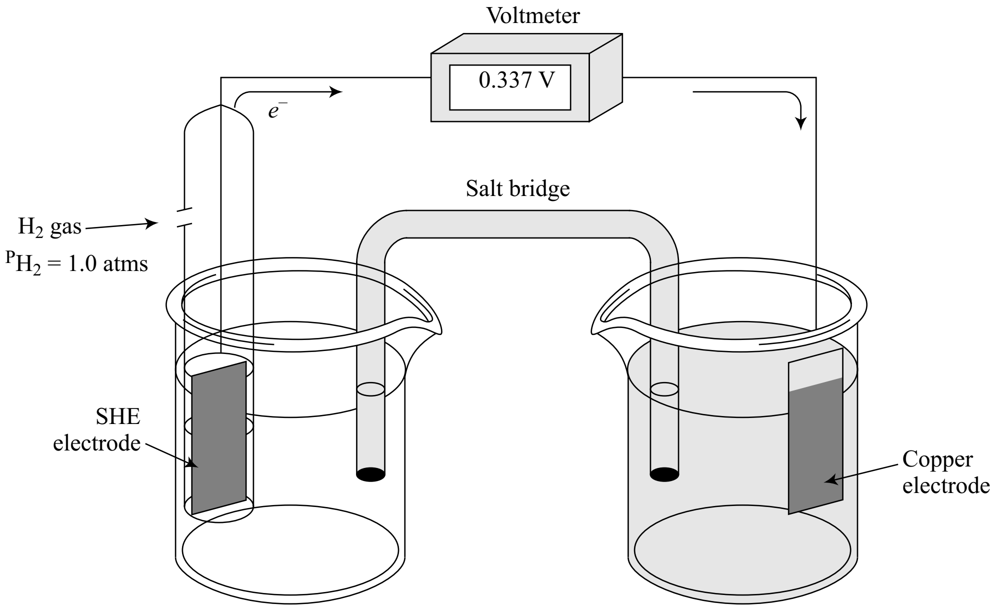


FIGURE 2.5 *A cell to determine potential of a cathode.*

Hence, according to the electrochemical convention, 'the potential of copper electrode is taken as +ve'.

$$E^o_{cell} = E^o_{Cu^{2+}/Cu} - (0) = E^o_{Cu^{2+}/Cu}$$

Thus, $E^o_{Cu^{2+}/Cu}$ is positive.

That is, $E^o_{cell} = E^o_{Cu^{2+}/Cu} - E^o_{H^2}$

$= 0.34 - (0) = 0.34$ V

(2) Determination of standard electrode potential of the anode:

A cell is constructed (Fig. 2.6) by assuming the zinc electrode as the anode, which is coupled with a standard hydrogen electrode (SHE) as shown in Fig. 2.6.

$$Zn/Zn^{2+}_{(C_1)} || H^+_{(C=1M)}/H_{2(g)}(1\ atm)/Pt$$

$$E^o_{SHE} > E^o_{Zn^{2+}/Zn}$$

The electrode reactions of the cell are

At anode, $Zn \rightarrow Zn^{2+} + 2e$; $\quad E^o_{Zn/Zn^{2+}} = 0.76$ V

At cathode, $2H^+ + 2e \rightarrow H_{2(g)}$; $\quad E^o_{2H^+/H2} = 0$

The EMF of the cell is

$$E^o_{cell} = E^o_{cathode} - E^o_{anode} = E^o_{H^+/H_2} - E^o_{Zn^{2+}/Zn}$$

$$E^o = 0 - E^o_{Zn^{2+}/Zn} = -[E^o_{Red}] = -ve$$

Hence, the electrode potential of the anode is 'negative'.

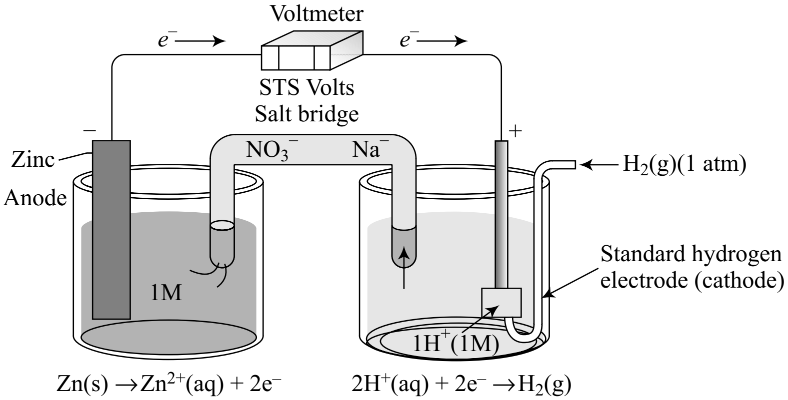


FIGURE 2.6 *A cell to determine potential of an anode.*

Example:

$$\begin{aligned} E^o_{Zn/Zn^{2+}} &= -(-0.76) = 0.76\text{ V (Oxidation potential)} \\ &= -0.76\text{ V (reduction potential)} \end{aligned}$$

In brief, it can be remarked that

$$E^o_{\text{oxidation}} = -E^o_{\text{reduction}}$$

2.9 Derivation of Nernst Equation for Single Electrode Potential

In electrochemical cell, the potential of an electrode depends on the following factors:

1. Nature of the electrode
2. Concentration of metal ions
3. Temperature

According to thermodynamics, Nernst showed that the free energy change of any one of the electrode processes is related to the potential developed (E) at the electrode with a concentration of M^{n+} ions at the interface at T^o absolute.

In an electrochemical cell, if reduction electrode process takes place at the cathode then,

$$M^{n+} + 2e \rightleftharpoons M \tag{2.1}$$

The decrease of free energy change ($-\Delta G$) accompanying the reduction reaction at the cathode is given as

$$\Delta G = \Delta G^o + RT \ln K \tag{2.2}$$

where ΔG^o is the decrease of free energy accompanying the above electrode reaction when all the reactants and products are in their standard states (298 K and 1M) and 'K' denotes the ratio of the concentration of products and reactants at the equilibrium electrode reaction.

$$K = \left[\frac{M}{M^{n+}}\right]$$

The free energy change of the reduction electrode reaction (Eq. (2.1)) is

$$\Delta G = \Delta G^o + 2.303\, RT \log_{10} \frac{[M]}{[M^{n+}]} \tag{2.3}$$

The electrical energy of a reversible cell is given by the decrease of free energy ($-\Delta G$) of the electrode reaction occurring in the cell. This free energy change is equal to the *electrical network done* (W_{net}) by the cell, that is

$$-W_{\text{net}} = \Delta G$$

Thus, the potential is developed at the electrode at the expense of the decrease in free energy of the electrode process.

$$\Delta G = -nFE$$

where 'E' denotes the single electrode potential in volts and nF denotes the quantity of electricity produced in *coulombs*, n is the number of electrons liberated at one electrode.

And similarly, $\Delta G^{\text{o}} = -nFE^{\text{o}}$ at standard conditions.

Standard free energy change (ΔG^{o}) is defined as the change in free energy which takes place when the reactants in the standard state (1 atm, 298 K) are converted into the products in their standard state.

Hence, $$-nFE = -nFE^{\text{o}} + 2.303\ RT \log_{10}\left[\frac{1}{M^{n+}}\right] \qquad (2.4)$$

dividing by $-nF$, we have

$$E = E^{\text{o}} - \frac{2.303RT}{nF}\log_{10}\left[\frac{1}{M^{n+}}\right] \qquad (2.5)$$

$$E = E^{\text{o}} + \frac{2.303RT}{nF}\log_{10}[M^{n+}] \qquad (2.6)$$

or this relationship can be written as

$$E = E^{\text{o}} + \frac{0.0591}{n}\log_{10}[M^{n+}] \qquad (2.7)$$

where R, n and F have their usual significance. Equation (2.6) is known as the *Nernst equation* for the single electrode potential.

In general, for any redox electrode processes of the cell

$$\text{Oxidized state} + ne \rightleftharpoons \text{Reduced state,}$$

The potentials of the electrodes are given as

(i) At cathode,

$$E = E^{\text{o}} + \frac{0.0591}{n}\ \log_{10}\,[M^{n+}]\ \text{(Reduction)}$$

(ii) At anode,

$$E = E^{\text{o}} - \frac{0.0591}{n}\ \log_{10}\,[M^{n+}]\ \text{(Oxidation)}$$

2.10 EMF of a Cell

Let us consider a cell,

If, $$E^{\text{o}}_{\text{cathode}} > E^{\text{o}}_{\text{anode}}$$

$$M_1|M^{n+}_{(C_1)}||M^{n+}_{(C_2)}|M_2$$

which is obtained by connecting two half-cells or electrodes.

The net cell reaction is as given

$$M_1 + M_2^{n+} \rightleftharpoons M_1^{n+} + M_2$$

The EMF of the cell is

$$E_{\text{cell}} = E^{\circ}_{\text{cell}} - \frac{2.303RT}{nF}\log_{10}\frac{[M_1^{n+}]}{[M_2^{n+}]}$$

$$E_{\text{cell}} = E^{\circ}_{\text{cell}} - \frac{2.303RT}{nF}\log_{10}\frac{[M_1^{n+}]}{[M_2^{n+}]}$$

$$E^{\circ}_{\text{cell}} = E^{\circ}_{M_2^{n+}/M_2} - E^{\circ}_{M_1^{n+}/M_1} = E^{\circ}_{\text{cathode}} - E^{\circ}_{\text{anode}}$$

where, M_1^{n+} = oxidized state of metal M_1 and M_2^{n+} = oxidized state of metal M_2.

Example 2.4 Write an expression for the EMF of a Daniel cell.

Solution

If, $E^{\circ}_{\text{cathode}} > E^{\circ}_{\text{anode}}$, the net cell reaction of a Daniel cell is

$$\text{Zn} + \text{Cu}^{2+} \rightleftharpoons \text{Zn}^{2+} + \text{Cu}$$

Therefore, the EMF of the Daniel cell is given as

$$E_{\text{cell}} = E^{\circ}_{\text{cell}} - \frac{0.0591}{2}\log_{10}\frac{[\text{Zn}^{2+}]}{[\text{Cu}^{2+}]} \text{ at } 25°\text{C}$$

where $E^{\circ}_{\text{cell}} = E^{\circ}_{\text{cathode}} - E^{\circ}_{\text{anode}}$

$$= E^{\circ}_{\text{Cu}^{2+}/\text{Cu}} - E^{\circ}_{\text{Zn}^{2+}/\text{Zn}}$$

$$= 0.34 - (-0.76)$$

$$= 1.10 \text{ V.}$$

Hence, $E_{\text{cell}} = 1.1 - \frac{0.0591}{2}\log_{10}\frac{[\text{Zn}^{2+}]}{[\text{Cu}^{2+}]}$

Example 2.5 The standard reduction electrode potential, E°, of copper is 0.34V and the concentration of Cu^{2+} ion is 0.015 M. Find the (i) reduction electrode potential, E, of copper and (ii) free energy change of electrode reaction.

Solution

(i) The reduction electrode reaction is

$$\underset{\text{(mol ions.)}}{\text{Cu}^{2+}} \quad + \quad \underset{\text{(2 mol electrons)}}{2\text{e}} \quad \rightarrow \quad \underset{\text{(mol atom)}}{\text{Cu}}$$

The Nernst equation for single reduction electrodes potential is

$$E = E^{\text{o}} + \frac{0.0591}{n}\log_{10}[\text{Cu}^{2+}]$$

$$E = 0.34 + \frac{0.0591}{2}\log_{10}[0.015]$$

$E_{\text{cell}} = 0.296$ V.

(ii) $\Delta G = -nFE$

$n = 2$ mol electrons, $E = 0.296$ V, $F = 96{,}500$ C/mol

$-\Delta G = -[2\ (\text{mol}) \times 96{,}500(\text{C/mol})] \times 0.296(\text{V})$

$= -57{,}128$ Volt-Coulomb $= -57.13$ KJ/mol. [Volt-Coulomb = Joule].

Example 2.6 The standard reduction electrode potential of copper and zinc are 0.34 and –0.76 V, respectively. If the cell develops an EMF of 1.10 V, find the free energy change of the overall cell reaction.

Solution

$$\text{Zn} + \text{Cu}^{2+} \rightarrow \text{Zn}^{2+} + \text{Cu}$$

Data: $n = 2$ mol electrons, $E^o = 1.10$ V, and $F = 96{,}500$ C/mol
$\Delta G = -nFE$
$\Delta G = -[2 \times 96{,}500 \times 1.1 = -2{,}12{,}300$ J $= -212.3$ KJ/mol
$\Delta G = -212.3$ KJ/mol

The cell reaction is spontaneous because of the decrease of free energy (–212.3 KJ/mol).

Example 2.7 The EMF of cell, $\text{Cd}_{(s)}/\text{Cd}^{2+}$ (0.01 M) || Cu^{2+} (0.5 M)/Cu, is 0.79 V. Determine the standard reduction potential of Cd electrode, if the standard electrode potential of copper is 0.34 V.

Solution

The EMF of the cell is

$$E_{\text{cell}} = E^{\text{o}}_{\text{cell}} - \frac{0.0591}{2}\log_{10}\frac{[\text{Cd}^{2+}]}{[\text{Cu}^{2+}]}$$

Therefore,

$$0.79 = E^{\text{o}}_{\text{cell}} - \frac{0.0591}{2}\log_{10}\frac{[0.01]}{[0.5]}$$

$$E^{\text{o}}_{\text{cell}} = 0.79 + \frac{0.0591}{2}\log_{10}\frac{[0.01]}{[0.5]}$$

$$= 0.79 - 0.05 = 0.74 \text{ V}$$

If

$$E^{\text{o}}_{\text{cell}} = E^{\text{o}}_{\text{Cu}^{2+}/\text{Cu}} - E^{\text{o}}_{\text{Cd}^{2+}/\text{Cd}}$$

$$E^{\text{o}}_{\text{Cd}^{2+}/\text{Cd}} = 0.34 - 0.74 = -0.40 \text{ V}$$

Example 2.8 Calculate the standard electrodes potential $E^{\text{o}}_{\text{Cu}^{2+}/\text{Cu}}$ of copper, if its electrode potential, E, at 25°C is 0.296V when the concentration of Cu^{2+} ion is 0.015 M.

Solution

The Nernst equation for single electrodes potential is

$$E = E^{\mathrm{o}} + \frac{0.0591}{n}\log_{10}[\mathrm{Cu}^{2+}]$$

Given: $E = 0.296$ V.

$$E = E^{\mathrm{o}}_{\mathrm{Cu}^{2+}/\mathrm{Cu}} + \frac{0.0591}{2}\log_{10}[0.015]$$

Therefore,

$$E^{\mathrm{o}}_{\mathrm{Cu}^{2+}/\mathrm{Cu}} = 0.296 - \frac{0.0591}{2}\log_{10}[0.015]$$

$$= 0.296 + 0.0540$$

$$= 0.35 \text{ volt}$$

Example 2.9 Calculate the reduction electrode potential of copper when it is in contact with 0.5M copper sulphate solution at 298K. The E^{o} value of copper is 0.34 V.

Solution ***The reduction electrode reaction of copper half-cell is given as,***

$$\mathrm{Cu}^{2+} + 2\mathrm{e} \rightleftharpoons \mathrm{Cu}$$

According to the Nernst equation,

$$E = E^{\mathrm{o}}_{\mathrm{Cu}^{2+}/\mathrm{Cu}} + \frac{0.0591}{2}\log_{10}[\mathrm{Cu}^{2+}]$$

$$= 0.34 + \frac{0.0591}{2}\log_{10}[0.5]$$

$$= 0.34 + 0.002$$

$$= 0.342 \text{ Vat } 298 \text{ K}$$

Example 2.10 Consider the following cell, Ni/Ni^{2+} (0.01M) || Cu^{2+} (0.5M)/Cu. The standard reduction potentials of Ni and Cu are –0.25 and 0.34 V, respectively. Write the electrode reactions and calculate the EMF of the cell at 298 K.

Solution

If, $E^{\mathrm{o}}_{\mathrm{Cu}^{2+}/\mathrm{Cu}} > E^{\mathrm{o}}_{\mathrm{Ni}^{2+}/\mathrm{Ni}}$

$$\mathrm{Ni/Ni}^{2+}\ (0.01\mathrm{M})||\mathrm{Cu}^{2+}\ (0.5\mathrm{M})/\mathrm{Cu}$$

The electrode reactions of the above cell are
At anode,

$$\mathrm{Ni} \rightarrow \mathrm{Ni}^{2+} + 2\mathrm{e}; \qquad E^{\mathrm{o}}_{\mathrm{Ni/Ni}^{2+}} = 0.25 \text{ V}$$

At cathode,

$$\mathrm{Cu}^{2+} + 2\mathrm{e} \rightarrow \mathrm{Cu}; \qquad E^{\mathrm{o}}_{\mathrm{Cu}^{2+}/\mathrm{Cu}} = 0.34 \text{ V}$$

The net cell reaction is

$$\underset{}{\mathrm{Ni}} + \underset{(0.5\ \mathrm{M})}{\mathrm{Cu^{2+}}} \rightarrow \underset{(0.01\mathrm{M})}{\mathrm{Ni^{2+}}} + \mathrm{Cu}$$

Therefore, $E^{o}_{cell} = [E^{o}_{Cu^{2+}/Cu} - E^{o}_{Ni^{2+}/Ni}] = [0.34 - (-0.25)]$

Since $E^{o}_{red} = -E^{o}_{oxid}$

$= 0.59$ V

EMF of the cell is written as

$$E_{cell} = E^{o}_{cell} - \frac{0.0591}{2}\log_{10}\frac{[\mathrm{Ni}^{2+}]}{[\mathrm{Cu}^{2+}]}$$

$$= 0.59 - \frac{0.0591}{2}\log_{10}\frac{[0.01]}{[0.5]}$$

$$= 0.59 + 0.05 = 0.64 \text{ V}$$

Example 2.11 Consider an electrochemical cell.

$\mathrm{Fe/Fe^{2+}}$ (0.02M) || $\mathrm{Cd^{2+}}$ (1M)/Cd

(i) write the cell reaction and (ii) calculate the EMF of the cell and determine the direction of the spontaneous reaction. Given the standard reduction potentials of iron and cadmium are –0.44 and –0.40 V, respectively.

Solution

If, $E^{o}_{Cd^{2+}/Cd} > E^{o}_{Fe^{2+}/Fe}$

$\mathrm{Fe/Fe^{2+}}$ (0.02M) || $\mathrm{Cd^{2+}}$ (1M)/Cd

(i) The two half-cell reactions of the cell are
At anode,

$\mathrm{Fe \rightarrow Fe^{2+} + 2e}$; $E^{o}_{Fe/Fe^{2+}} = 0.44$ V

At cathode,

$\mathrm{Cd^{2+} + 2e \rightarrow Cd}$; $E^{o}_{Cd^{2+}/Cd} = -0.40$ V

(ii) The net cell reaction is

$\mathrm{Fe + Cd^{2+} \rightarrow Fe^{2+} + Cd}$

(iii) The EMF of the cell is

$$E_{cell} = E^{o}_{cell} - \frac{0.0591}{2}\log_{10}\frac{[\mathrm{Fe}^{2+}]}{[\mathrm{Cd}^{2+}]}$$

where $E^{o}_{cell} = E^{o}_{Cd^{2+}/Cd} - E^{o}_{Fe^{2+}/Fe}$

$= -0.40 - (-0.44)$

$= 0.04$ V

$$\text{Therefore, } E_{\text{cell}} = 0.04 - \frac{0.0591}{2}\log_{10}\frac{[0.02]}{[1]}$$

$$= 0.04 + 0.0503 = 0.09 \text{ V}$$

The EMF of the cell is +ve and the cell reaction $Fe + Cd^{2+} \rightarrow Fe^{2+} + Cd$ proceeds spontaneously.

Example 2.12 Write the half-cell reaction and net cell reaction of the given cell

$$Cd_{(s)}/Cd^{2+}\ (0.01\ M)\ \|\ Cu^{2+}\ (0.5\ M)/Cu$$

The standard reduction potentials are – 0.40 and 0.34 V, respectively. [VTU: Aug/Sept. 1999]

Solution

If, $E^{o}_{Cu^{2+}/Cu} > E^{o}_{Cd^{2+}/Cd}$

$$Cd/Cd^{2+}\ (0.01M)\ \|\ Cu^{2+}\ (0.5M)/Cu$$

(a) The half-cell reactions are
At anode,

$$Cd \rightarrow Cd^{2+} + 2e; \quad E^{o}_{Cd/Cd^{2+}} = 0.40\ V$$

At cathode,

$$Cu^{2+} + 2e \rightarrow Cu; \quad E^{o}_{Cu^{2+}/Cu} = +\ 0.34\ V$$

(b) The net cell reaction is

$$\underset{}{Cd} + \underset{(0.5M)}{Cu^{2+}} \rightarrow \underset{(0.01M)}{Cd^{2+}} + Cu$$

(c) The EMF of the cell is

$$E_{\text{cell}} = E^{o}_{\text{cell}} - \frac{0.0591}{2}\log_{10}\frac{[Cd^{2+}]}{[Cu^{2+}]}$$

$$\text{where} \quad E^{o}_{\text{cell}} = E^{o}_{Cu^{2+}/Cu} - E^{o}_{Cd^{2+}/Cd}$$

$$= +\ 0.34 - (-0.40)$$

$$= 0.74\ V$$

$$\text{Therefore, } E_{\text{cell}} = 0.74 - \frac{0.0591}{2}\log_{10}\frac{[0.01]}{[0.5]}$$

$$= 0.74 + 0.05$$

$$= 0.79\ V$$

Example 2.13 Calculate the voltage of the cell, $Mg/\ Mg^{2+}_{(aq)}\ \|\ Cd^{2+}_{(aq)}\ /Cd$, at 25°C when $[Cd^{2+}] = 7 \times 10^{-11}$ m; $[Mg^{2+}] = 1.0$ M and $E^{o}_{\text{cell}} = 1.97$ V. [VTU: Aug/Sept 2000]

Solution

The electrode reactions of the cell are

At anode,

$$\mathrm{Mg \rightarrow Mg^{2+} + 2e}$$

At cathode,

$$\mathrm{Cd^{2+} + 2e \rightarrow Cd}$$

The net cell reaction is

$$\mathrm{Mg + Cd^{2+} \rightarrow Mg^{2+} + Cd}$$

The EMF of the cell is

$$\begin{aligned} E &= E^{\mathrm{o}} - \frac{0.0591}{2}\log_{10}\frac{[\mathrm{Mg}^{2+}]}{[\mathrm{Cd}^{2+}]} \\ &= E^{\mathrm{o}} - \frac{0.0591}{2}\log_{10}\frac{[1.0]}{[7\times10^{-11}]} \\ &= 1.97 - 0.0296\log_{10}\frac{[1.0]}{[7\times10^{-11}]} \\ &= 1.67\ \mathrm{V} \end{aligned}$$

Example 2.14 Calculate the voltage generated in the following cell Fe/Fe^{2+} || Mn^{2+}/Mn when an iron rod is immersed in 1.0 M FeSO$_4$ solution and an Mn rod is immersed in 0.1 M MnSO$_4$ solution.
Given $E^{\mathrm{o}}_{\mathrm{Fe^{2+}/Fe}} = -0.40$ V and $E^{\mathrm{o}}_{\mathrm{Mn^{2+}/Mn}} = -1.18$ V [VTU: Feb/March 2001]

Solution

If, $E^{\mathrm{o}}_{\mathrm{Fe^{2+}/Fe}} > E^{\mathrm{o}}_{\mathrm{Mn^{2+}/Mn}}$

That is, $-0.40 > -1.18$

(a) Representation of the cell is

$$\mathrm{Mn/Mn^{2+}\ (0.01\ M)\ ||\ Fe^{2+}\ (1.0\ M)/Fe}$$

Since, $E^{\mathrm{o}}_{\mathrm{Fe^{2+}/Fe}} > E^{\mathrm{o}}_{\mathrm{Mn^{2+}/Mn}}$

The two electrode reactions are
At anode,

$$\underset{(0.01\mathrm{M})}{\mathrm{Mn \rightarrow Mn^{2+} + 2e}}\ ; \qquad E^{\mathrm{o}}_{\mathrm{Mn/Mn^{2+}}} = 1.18\ \mathrm{V}$$

At cathode,

$$\underset{(0.01\mathrm{M})}{\mathrm{Fe^{2+} + 2e \rightarrow Fe}}; \qquad E^{\mathrm{o}}_{\mathrm{Fe^{2+}/Fe}} = -0.40\ \mathrm{V}$$

(b) The net cell reactions is

$$\mathrm{Mn + Fe^{2+} \rightarrow Mn^{2+} + Fe}$$

(c) The EMF of the cell is

$$E_{\mathrm{cell}} = E^{\mathrm{o}}_{\mathrm{cell}} - \frac{0.0591}{2}\log_{10}\frac{[\mathrm{Mn}^{2+}]}{[\mathrm{Fe}^{2+}]}$$

where $E^{o}_{cell} = E^{o}_{Fe^{2+}/Fe} - E^{o}_{Mn^{2+}/Mn}$

$$= -0.40 -(-1.18)$$
$$= 0.78 \text{ V}$$

Therefore, the EMF of the cell is

$$E_{cell} = 0.78 - \frac{0.0591}{2}\log_{10}\frac{0.1}{1.0}$$
$$= 0.78 + 0.02960$$
$$= 0.8096 \text{ V}$$

Example 2.15 A cell is constructed by coupling a zinc electrode dipped in 0.5M $ZnSO_4$ and a nickel electrode dipped in 0.05M $NiSO_4$. Write the cell representation, cell reaction. Calculate EMF of the cell, given that standard reduction potential of Zn and Ni as –0.76 and –0.25 V, respectively.

[VTU: July/August 2002].

Solution

If $E^{o}_{Ni^{2+}/Ni} > E^{o}_{Zn^{2+}/Zn}$

The cell is represented as

$$Zn/Zn^{2+}\ (0.5M)//Ni^{2+}\ (0.05M)/Ni$$

The electrode reactions of the above cell are
At anode,

$$Zn \rightarrow Zn^{2+} + 2e; \qquad E^{o}_{Zn/Zn^{2+}} = +0.76 \text{ V}$$

At cathode,

$$Ni^{2+} + 2e \rightarrow Ni; \qquad E^{o}_{Ni/Ni^{2+}} = -0.25 \text{ V}$$

The net cell reaction is

$$\underset{}{Zn} + \underset{(0.05M)}{Ni^{2+}} \rightarrow \underset{(0.5M)}{Zn^{2+}} + Ni$$

Therefore,

$$E^{o}_{cell} = [-0.25 - (-0.76)] \qquad \text{since } (E^{o}_{red} = -E^{o}_{oxid})$$
$$= 0.51 \text{ V}$$

EMF of the cell is written as

$$E_{cell} = E^{o}_{cell} - \frac{0.0591}{2}\log_{10}\frac{[0.5]}{[0.05]}$$
$$= 0.51 - 0.02955 \log_{10} [10]$$
$$= 0.51 - 0.02955$$
$$= 0.48 \text{ V.}$$

Example 2.16 A cell is formed by dipping a nickel rod in 0.01M Ni^{2+} solution and a lead rod in 0.5M Pb^{2+} solution. The standard electrode potentials of Ni and Pb are –0.24 and – 0.13V, respectively. Write the cell representation, cell reaction and calculate the EMF of the cell. [VTU: July/August 2003]

Solution

If, $E^{o}_{Pb^{2+}/Pb} > E^{o}_{Ni^{2+}/Ni}$

The cell is represented as

$$Ni/Ni^{2+}\ (0.0.01M)//Pb^{2+}\ (0.5M)/Pb$$

The electrode reactions of the above cell are
At anode,

$$Ni \rightarrow Ni^{2+} + 2e; \qquad E^{o}_{Ni/Ni^{2+}} = -0.24\ V$$

At cathode,

$$Pb^{2+} + 2e \rightarrow Pb; \qquad E^{o}_{Pb^{2}/Pb} = -0.13\ V$$

The net cell reaction is

$$\underset{}{Ni} + \underset{(0.5M)}{Pb^{2+}} \rightarrow \underset{(0.01M)}{Ni^{2+}} + Pb$$

Therefore,

$$\begin{aligned} E^{o}_{cell} &= [-0.13 - (-0.24)] \qquad \text{since } (E^{o}_{red} = -E^{o}_{oxid}) \\ &= 0.11\ V \end{aligned}$$

The EMF of the cell is written as

$$\begin{aligned} E_{cell} &= E^{o}_{cell} - \frac{0.0591}{2}\log_{10}[0.01/0.5] \\ &= 0.11 - 0.02955\ \log_{10}[0.02] \\ &= 0.11 - [-0.00922] \\ &= 0.12\ V. \end{aligned}$$

Example 2.17 Write the electrode reaction and calculate the EMF of the following cell at 298K. Given $E^{\circ} = 0.46$ V.

Solution

$$Cu_{(s)}/Cu^{2+}\ (1 \times 10^{-2}M)//Ag^{+}\ (1 \times 10^{-1}M)/Ag_{(s)}.$$

At anode,

$$Cu_{(s)} \rightarrow Cu^{2+} + 2e$$

At cathode,

$$2Ag^{+} + 2e \rightarrow 2Ag_{(s)}$$

$$Cu_{(s)} + 2Ag^{+} \rightarrow Cu^{2+} + 2Ag_{(s)};\ \text{given: } E^{o}_{cell} = E^{o}_{c} - E^{o}_{a} = 0.46\ V.$$

This states explicitly that the *cell potential is a measure of the free energy change per mole of electrons transferred*; this explains why we do not have to multiply the $E°$s of half-cell reactions by stoichiometric factors when we are finding the $E°$ of a complete cell.

$$E_{\text{cell}} = 0.46 - \frac{0.0591}{2}\log_{10}\frac{[1\times10^{-2}]}{[1\times10^{-1}]^2}$$

$$= 0.46 \text{ V}$$

Example 2.18 The E^{o} values of Li/Li$^+$, Zn/Zn^{+2}, Cu/Cu^{+2} and Ag/Ag$^+$ are – 3.0, –0.77, +0.33 and +0.80 V, respectively. Which combination of the electrodes you use to construct a cell of the highest EMF if the ionic concentrations are 0.1,1.0, 10 and 0.01M, in the same order. Justify your answer.

[VTU: July/August 2005]

Solution

The potential of a half-cell is

$$E = E^{\text{o}} + \frac{0.0591}{n}\log_{10}[M^{+n}] \qquad \text{Nernst equation}$$

(i) For Li/Li$^+$, the electrode potential is

$$E = -3.0 + 0.0591 \log_{10} [0.10]$$
$$= -3.0 - 0.0591 = -2.419 \text{ V.}$$

(ii) For Zn/Zn^{+2}, the electrode potential is

$$E = -0.77 + \frac{0.0591}{2}\log_{10}[1.0]$$
$$= -0.77 \text{ V}$$

(iii) For Cu/Cu^{+2}, the electrode potential is

$$E = 0.34 + \frac{0.0591}{2}\log_{10}[10.0] = 0.34 + 0.0295$$
$$= 0.3695 \text{ V.}$$

(iv) For Ag/Ag$^+$, the electrode potential is

$$E = 0.80 + 0.0591 \log_{10} [0.010]$$
$$= 0.80 - 0.0591 \times 0.3010$$
$$= 0.80 - 0.1773 = 0.7227 \text{ V.}$$

From the above values, the highest value of EMF of a cell can be obtained by coupling Ag/Ag$^+$ as the cathode and Li/Li$^+$ as the anode of the cell.

$E_{\text{cell}} = E_{\text{cathode}} - E_{\text{anode}}$ (both potentials are reduction potentials).

$$E_{\text{Ag/Ag}^+} > E_{\text{Li/Li}^+}, \text{ that is, } 0.7227 > -2.419$$

$$E_{\text{cell}} = E_{\text{Ag/Ag}^+} - E_{\text{Li/Li}^+} = 0.7227 - [-2.419]$$
$$= 0.7227 + 2.419 = 3.1417 \text{ V.}$$

Example 2.19 Calculate the potential of Ag–Zn cell at 298K, if the concentration of Ag^+ and Zn^{2+} are 5.2×10^{-6} and 1.3×10^{-3} M, respectively. E^o of the cell at 298K is 1.56V. Calculate the change in free energy ΔG for the reduction of 1mole of Ag^+.

Solution

We know that, $E^o_{Ag^+/Ag} > E^o_{Zn^{+2}/Zn}$

At anode,

$$Zn \rightarrow Zn^{2+} + 2e$$

At cathode, $n = 2$

$$2Ag^+ + 2e \rightarrow 2Ag$$

$$Zn + 2Ag^+ \rightarrow 2Ag + Zn^{2+}; \; E^o_{cell} = E^o_{cathode} - E^o_{anode} = 1.56 \text{ V}$$

This states explicitly that the *cell potential is a measure of the free energy change per mole of electrons transferred*; this explains why we do not have to multiply the $E°$ of half-cell reactions by stoichiometric factors when we are finding the $E°$ of a complete cell.

Therefore, $E_{cell} = E^o_{cell} - \frac{0.0592}{2}\log_{10}\frac{[1.3\times10^{-3}]}{[5.2\times10^{-6}]^2}$

$$= 1.56 - \frac{0.0592}{2}\log_{10}\frac{[1.3\times10^{-3}]}{[5.2\times10^{-6}]^2}$$

$$\Delta G = -nEF = -2 \times 1.25 \times 96{,}500 \text{ for 2 moles } Ag^+ \text{ ions}$$

$$= -246 \text{ kJ}$$

Therefore, 1 mole Ag^+ ions = 123.0 KJ/mol

2.11 Types of Electrodes or Half Cells

An electrochemical cell can be conveniently constructed by selecting suitably any of the following types of electrodes or half-cells, each of which has its own developed potential values.

(1) Metal–metal ion electrode

A half-cell of this kind consists of a metal dipped in a solution containing its own ions.

The half-cell is represented as

M/M^{n+} (Oxidation)

The redox reaction at the electrodes

$$M \rightleftharpoons M^{n+} + ne$$

Example: Zn/Zn^{2+}, Cu/Cu^{2+}

(2) Metal–metal salt–ion electrodes

In this type of half-cell, *a metal is in contact with a sparingly soluble salt of the same metal dipped in a solution containing anion of the salt.*

Calomel and silver–silver chloride electrodes are the common electrodes generally used in cells.

(a) Calomel electrode, $Hg/Hg_2Cl_2/Cl^-$

(b) Silver–silver chloride electrode, $Ag/AgCl_{(s)}/Cl^-$

(3) Gas electrode

A gas electrode consists of a particular gas flushed around an inert electrode (Pt), which is dipped in a solution containing ions to which it is reversible. At the metal interface, there is equilibrium between the gas and its ions.

Consider standard hydrogen electrode (SHE). The half-cell is represented as

$$\mathrm{Pt/H_{2(g)}\ (1\,atm)/H^{+}_{(C=1M)}}$$

The electrode reaction is reduction, if the half-cell is made as the cathode.

(4) Oxidation–reduction electrodes

In this type of half-cell, *electrode potential arises due to the presence of both oxidized and reduced species of the same metal at an inert electrode like Pt.*

The developed potential could be described due to the tendency of one form to change into other more stable form.

The potential is sensed by the inert electrode, Pt.

Examples: $\mathrm{Pt/Fe^{2+}}$; $\mathrm{Fe^{3+}}$, $\mathrm{Pt/Sn^{2+}}$; $\mathrm{Sn^{4+}}$

(5) Ion-selective electrodes (ISE)

A membrane of a half-cell is sensitive to a particular ion in a solution; the ion exchange process takes place between the membrane electrode and the solution containing the specific ions and develops a potential, which depends on the concentration of that particular ionic species.

For examples glass ISE, calcium ISE and ammonium ISE.

(6) Ion-selective electrodes (ISEs)

Ion-sensitive electrodes (or chemical sensors) are miniaturized analytical devices that are *used to interface between chemical species and electronic devices*, which can provide the information about the presence of specific ions (or compounds) in samples. *The potential developed at the ion-sensitive sensor is a measure of the concentration of the ionic species of interest.* Different types of ion-selective electrodes or chemical sensors are available, which are sensitive selectively to a particular ionic species. *The electrodes not only can detect but also measure the concentration of a specific chemical species directly in a given solution are referred to as ion-selective electrodes.* These electrodes are frequently used as potentiometric electrodes during laboratory analysis as well as in industry, process control, physiological measurements and environmental monitoring.

An *ion-selective membrane* is the key component of all potentiometric ion sensors. The voltage developed between the sensing and the reference electrodes is a measure of the concentration of the reactive ion being measured. As the concentration of the reacting ion at the sensing electrode varies, so does the voltage measured between two electrodes.

2.12 Types of Ion Selective Electrodes

(a) Glass membrane electrodes

Glass membrane electrodes are formed by the doping of the silicon dioxide glass matrix with various chemicals. The most common of the glass membrane electrodes is the pH electrode. Glass membrane electrodes are also available for the measurement of sodium ions. The glass membrane electrode exchanges $\mathrm{H^+}$ ions from an acid solution.

$$\mathrm{Gl^{-}\,Na^{+} + H^{+} \rightleftharpoons Gl^{-}\,H^{+} + Na^{+}}$$

(b) Polymer membrane electrodes

Polymer membrane electrodes consist of various ion exchange materials in an inert matrix such as PVC, polythene or silicone rubber. After the membrane is formed, it is sealed to the end of a PVC tube. The potential developed at the membrane surface is related to the concentration of the species of interest. Potassium, calcium and nitrate ion-sensitive electrodes are the examples.

(c) Solid state electrodes

Solid state electrodes utilize relatively insoluble inorganic salts in a membrane. Solid state electrodes exist in homogeneous or heterogeneous form. In both types, potentials are developed at the membrane surface due to the ion exchange process. Examples include silver/sulphide, chloride and fluoride.

(d) Gas-sensing electrodes

Gas-sensing electrodes are available for the measurement of ammonia, carbon dioxide, nitrogen oxide and sulphur dioxide. These electrodes have a gas permeable membrane and an internal buffer solution. The pH of the buffer solution changes as the gas reacts with it. The change is detected by a combination pH sensor within the housing. Due to the construction, gas-sensing electrodes do not require an external reference electrode.

2.13 Applications of Ion Selective Electrodes (ISE's)

The ISE can be used in a wide variety of experiments, including the following:

Water hardness (calcium ISE)

Calcium is very useful as an indicator to detect the hardness of water in fresh water samples. Even though water hardness is also due to other ions (Mg^{2+} and Fe^{3+}), because calcium normally predominate hardness as Ca^{2+} ion, which is a reliable indicator of water hardness.

Nitrate monitoring (nitrate ISE)

This ISE can be used to qualitatively determine the concentration of the nitrate ion (NO_3^-) in fresh water samples. This ion may appear in samples due to waste or fertilizer runoff. Increased levels of nitrate may also be present in water samples in waste from plant and animal decay.

*Salinity and chloride monitoring in sea water (*chloride ISE*)-*

This ISE can be used to determine the concentration of chloride ions (Cl–) or salinity levels in sea water samples. The wide range of this electrode allows us to measure salinity (as sodium chloride) or chloride concentration of sea water.

Monitoring steams for ammonium (ammonium ISE)

Ammonium (NH^{4+}) is often present as a result of water runoff from fertilized fields. This ISE gives you an easy way to check for this environmentally important ion.

Calcium in milk (calcium ISE)

This ISE can be used to determine calcium levels in skim milk.

2.14 How an Ion Selective Electrode Works

Ion-selective electrodes are membrane-based electrodes that measure a specific ion (e.g. Na^+, Ca^{2+}) in aqueous solutions. *When the membrane of the electrode is in contact with a solution containing the specific ion, a voltage, dependent on the level of that ion in solution, develops at the membrane.* All ISEs measure the specific ion concentration directly. Samples need to be aqueous to avoid contaminating or dissolving the membrane.

The voltage developed between the sensing and the reference electrodes is a measure of the concentration of the reactive ion being measured. As the concentration of the reacting ion at the sensing electrode varies, so does the voltage measured between two electrodes.

2.15 Reference Electrode

2.15.1 The Purpose of Reference Electrodes

The purpose of the reference electrode is to complete the measuring circuit and *provide a stable and reproducible potential against which the indicator electrode is compared and measured.* The contact is made through a liquid junction that allows the reference electrolyte to contact the sample. The reference electrode is designed to produce the same potential no matter in what solution it is placed.

In practice, there are few half-cells whose electrode potentials are known, that *provides a stable, reliable and reproducible potential against which the indicator electrode is compared, so that they can be used repeatedly with accuracy. These electrodes of reproducible and stable potentials are called the reference electrodes.*

Thus, the potential of any other electrode can be measured when it is coupled with any of these reference electrodes.

(a) Primary standard hydrogen electrode (SHE):

A common reference electrode used to measure the potential of the other electrode is *standard hydrogen electrode* (SHE) whose electrode potential at all temperatures is taken as zero, that is $E^o = 0$.

The SHE is represented as (Fig. 2.7)

$$Pt/H_{2(g)}\ (1\,atm)/H^{+}_{(C=1M)}$$

The half-cell reaction is

$$2H^{+} + 2e \rightleftharpoons H_{2(g)}$$

The electrode potential of SHE is

$$E = E^o + 0.0591 \log_{10} [H^{+}]$$

If $[H^{+}] = 1M$ and $H_{2(g)}$ is bubbled at 1 atm at 298 K.
Therefore, $E = E^o = 0$

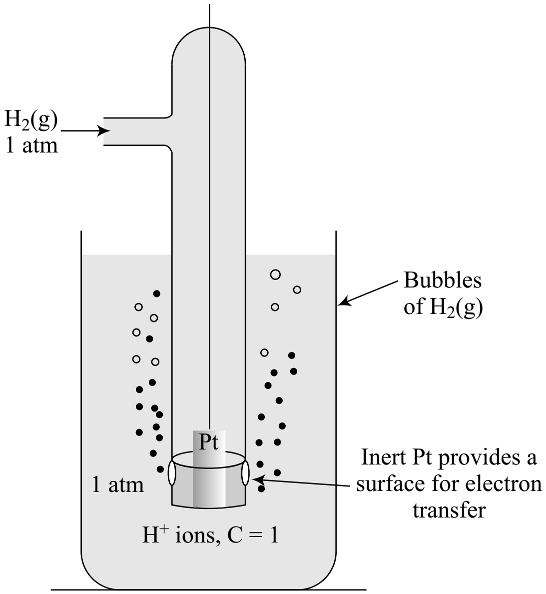


Figure 2.7 *Standard hydrogen electrode (SHE).*

(b) Secondary reference electrodes:

(i) Calomel electrode

The most widely used reference electrode in an electrochemical cell is the *calomel electrode*. The potential of calomel electrode depends on the concentration of the KCl solution used in the half-cell.

Concentration of KCl	E^o Calomel
0.1 N	0.3334 V
1.0 N	0.2810 V
Saturated	0.2422 V

Generally, the calomel electrode with saturated KCl solution is used in electrochemical cells.

Construction:

A calomel electrode consists of a narrow glass tube, the bottom of which has a layer of Hg and followed by a paste of Hg_2Cl_2 and Hg. The remaining portion of the glass tube is filled with saturated KCl solution (Fig. 2.8).

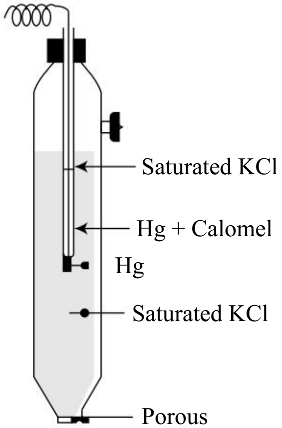


Figure 2.8 *Dipping type of calomel electrode.*

A Pt wire is dipped in the Hg layer for electrical contact. This is placed inside an outer glass jacket with a tapering porous end and is filled with saturated KCl solution. The calomel electrode is represented as

$$Hg|Hg_2Cl_2|\,Cl^-_{sat}$$

If the cathode of the cell is calomel electrode, then the electrode reaction is

$$Hg_2Cl_2 + 2e \rightleftharpoons 2Hg_{(s)} + 2Cl^-$$

The reduction electrode potential, E, is given as

$$E_{cal} = E^o - 0.0591\log_{10}[Cl^-].$$

Demerits of calomel electrode:

Calomel electrodes should not be used above 50°C because the mercurous chloride breaks down yielding unstable readings. It has similar stability to silver/silver chloride and a better thermal hysteresis. It is less prone to contamination because the mercury–mercurous chloride interface is protected inside a tube not in direct contact with the electrolyte. Reproducibility of a calomel electrode is much better than a silver/silver chloride. Recently, calomel has disfavoured because of its toxicity.

(ii) Silver–silver chloride electrode

This electrode is an example of a metal–metal salt ion reference electrode. It consists of a silver electrode coated with a sparingly soluble AgCl and is immersed in a solution containing Cl^- ions (saturated KCl) (Fig. 2.9).

Commercial reference electrodes consist of a plastic tube electrode body. The electrode is a silver wire that is coated with a thin layer of silver chloride, either by electroplating or by dipping the wire in molten silver chloride. The silver–silver chloride electrode can be represented as

$$Ag/AgCl_{(s)}/Cl^-$$

The electrode or half-cell reaction is

$$AgCl_{(s)} + e \rightleftharpoons Ag_{(s)} + Cl^-$$

The electrode potential, E, is given by

$$E_{Ag/AgCl} = E^o - 0.0591 \log_{10} [Cl^-]$$

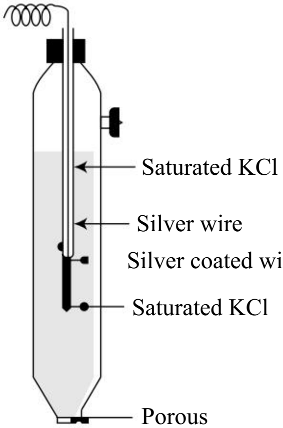


FIGURE 2.9 *A silver–silver chloride electrode.*

(i) Merits of silver–silver chloride electrode:

- The reduction electrode potential of this electrode is dependent on the concentration of Cl^- ions.
- Its ease of manufacture and its superior temperature range, actually usable even above 130°C.
- The potential of the electrode remains constant as long as the chloride concentration remains constant.
- It is used today by many electrochemists because it is simple, inexpensive, very stable and non-toxic.

(ii) Applications

- Silver chloride is used in
- Reference electrodes, e.g. in pH meters.
- Photography films and plate
- Infrared windows where it can be used in contact with many aqueous solutions, e.g. in FTIR spectrometers, in which case it can be an alternative to KBr, which is attacked by water.

2.16 What Factors can Effect a Reference Electrode?

The potential of a reference electrode varies with temperature (typically 0.5–1.0 mV/°C). Therefore, precise measurement of redox potentials requires the use of a constant temperature bath for the cell. The potentials of the silver/silver chloride and calomel reference electrode are also affected by the concentration of chloride in the electrode solution, which must therefore be maintained at a constant value by proper storage.

2.17 Limitations of Primary Reference Electrode and Advantages of Secondary Reference Electrode

The standard hydrogen electrode(SHE) has its own limitation for working during electrochemical measurements, for reasons stated below:

- It is difficult to maintain H^+ ion concentration as IM.
- It is difficult to maintain a flow of 1 atm of hydrogen at the interface of the Pt electrode
- Impurities present in hydrogen flow may poison the Pt electrode which alters the potential value.

To overcome these difficulties, secondary reference electrodes of the types (i) calomel and silver–silver chloride have been successfully designed. They have these advantages over the primary reference electrodes.

- They have constant and stable potential values at a particular temperature.
- They are calibrated with respect to SHE
- They have the least potential gradient with temperature.
- The potential values are reliable and measurements can be made with great accuracy.

2.18 Ion-selective Electrode

Glass Electrode

A glass membrane of a silicate glass is found to be sensitive to hydrogen $[H^+]$ ions of an acidic solution. The membrane undergoes ion exchange reaction, the Na^+ ions of glass are exchanged for H^+ ions.

$$\underset{\text{membrane}}{[H^+] + Na^+Gl^-} \rightarrow \underset{\text{membrane}}{Na^+ + H^+Gl^-}$$

When a thin-walled glass bulb containing $\frac{N}{10}$ acid HCl solution, as shown in Fig. 2.10, is immersed in an acidic solution of different concentrations $[C_1 > C_2]$, a 'boundary potential, E_b' is developed across gel layers of the glass membrane.

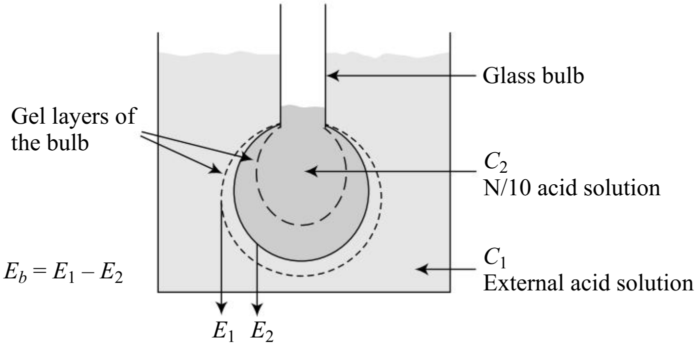


FIGURE 2.10 *A glass electrode.*

This potential, E_b, arises due to a difference in the H^+ ion concentration inside and outside the glass bulb, that is, 'E_b' $= E_1 - E_2$.

Thus,

$$E_b = [E_1 - E_2] = \frac{2.303\,RT}{nF}\log_{10}\frac{C_1}{C_2}$$

where C_1 is the concentration of H^+ ions of the solution into which the glass membrane is dipped.

The concentration of H^+ ions (C_2) inside the bulb is constant.

Therefore, $$E_b = \text{Constant} + \frac{0.0591}{n}\log_{10}[C_1]$$

$$= \text{Constant} + \frac{0.0591}{n}\log_{10}[H^+]$$

Since, $\text{pH} = -\log_{10}[H^+]$

Therefore, $E_b = K - 0.0591\ {}_{\text{p}}\text{H}$

Construction of glass electrode:

A glass membrane electrode is made up of a thin long glass tube with a thin-walled glass membrane at the bottom end (Fig. 2.11). A low melting point and high electrical conductance silicate glass may, specifically, be selected for the purpose to act as a glass membrane.

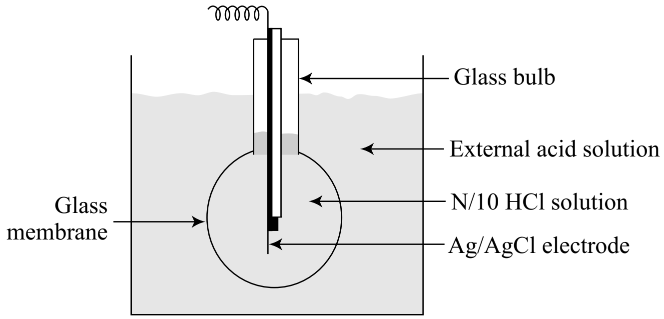


FIGURE 2.11 *A combined glass electrode coupled with Ag/AgCl electrode.*

Ag/AgCl electrode is dipped in a $\frac{N}{10}$ solution of HCl which is taken inside the bulb of the electrode and is connected to the Pt wire for external contact. The glass electrode is represented as

$Ag/AgCl_{(s)}/N/10$ HCl solution/glass

The combined glass electrode is immersed in an acidic (external) solution, then the potential of the glass electrode, E_G is given by

$$E_G = E_b + E_{Ag/AgCl}$$

Theoretically, $E_b = 0$, when $C_1 = C_2$. However, a small additional, E_{asy}, potential contribution exists, even if $C_1 = C_{2.}$

Thus, $E_G = E_b + E_{Ag/AgCl} + E_{asy}$

That is,

$$E_G = [K - 0.0591\ \text{pH}] + E_{Ag/Cl} + E_{asy}$$

The potential of glass electrode is given by

$$E_G = E_G^{o} - 0.0591\ \mathrm{P^H}$$

where $E_G^{o} = [K + E_{\mathrm{Ag/AgCl}} + E_{\mathrm{asy}}]$ is constant for a particular glass material.

The above expression indicates that the potential of a glass electrode, E_G, varies with the pH of the acidic solution.

2.19 Determination of P^H of a Solution using a Glass and a Calomel Reference Electrode

It is clear that 'E_G' of the glass electrode depends on the pH of the solution in which the glass electrode is dipped.

$$E_G = E^{o} - 0.0591\ \mathrm{P^H}$$

where E_G^{o} is the standard electrode potential of the glass electrode.

A glass electrode is coupled with a saturated calomel electrode (SCE) to determine the $\mathrm{P^H}$ of an acidic solution.

The cell is represented as

$$\mathrm{Hg/Hg_2Cl_2/Cl^-}\ \|\text{(unknown) }\mathrm{H^+}\text{ solution/glass electrode}$$

The potential established at the indicator glass electrode is higher than the reference saturated calomel electrode. Therefore, we have

$$E_G > E_{\mathrm{SCE}}.$$

The coupled two half-cells, SCE and the glass electrodes are then connected to a potentiometer to measure the E_{cell}. Knowing the E_G^{o} of the glass and E_{SCE} of calomel electrodes, the $\mathrm{P^H}$ of unknown solution is calculated thus:

$$\begin{aligned} E_{\mathrm{cell}} &= E_{\mathrm{cathode}} - E_{\mathrm{anode}} = E_G - E_{\mathrm{SCE}} \\ &= [E_G^{o} - 0.0591\ \mathrm{P^H}] - E_{\mathrm{SCE}} \end{aligned}$$

where, $E_G^{o} = [K + E_{\mathrm{Ag/AgCl}} + E_{\mathrm{asy}}]$,

Therefore, $0.0591\ \mathrm{pH} = E_G^{o} - E_{\mathrm{SCE}} - E_{\mathrm{cell}}$

The pH of the given solution is

$$\mathrm{pH} = \frac{E_G^{o} - E_{\mathrm{SCE}} - E_{\mathrm{cell}}}{0.0591}$$

2.20 Determination of P^H of a Solution using Glass and a Common Silver–Silver Chloride Reference Electrode

The Combined Glass Electrode

The heart of the glass electrode is a thin glass membrane, specially fabricated to preferentially exchange $\mathrm{H^+}$. The outside of the membrane is in contact with the analyte solution containing the unknown $[\mathrm{H^+}]$. The inside of the membrane contacts a hydrochloric acid solution of fixed (*N*/10) concentration. A silver wire coated with

AgCl dips into this solution; the other end of the wire is connected to the measuring device. A combination glass electrode with silver/silver chloride reference may be represented as shown in Fig. 2.10.

The Construction and the Mechanism of the Response of pH Electrode

A change in H^+ ion concentration causes a change in composition of the glass membrane due to an ion exchange process involving the solution and the membrane. A corresponding change in membrane potential, proportional to pH, is measured. All other potentials are constant. In effect, the membrane potential (variable) is measured against two fixed potentials, the external reference and the internal reference, both Ag/AgCl reference electrodes.

A combined glass electrode is made up of a thin long glass tube with a thin-walled glass membrane bulb at the bottom end. A low melting point and high electrical conductance silicate glass may be selected for the purpose to act as a membrane (Fig. 2.10). The $Ag/AgCl/Cl^-$ electrode is immersed inside the bulb containing N/10 HCl (C_1) solution.The outer interface of the glass bulb is in contact with the H^+ ions of unknown (C_2) solution as shown in Fig. 2.12. The $Ag/AgCl/Cl^-$ electrode in the combined glass cell behaves as a common reference electrode for both the inner and outer H^+ ions.

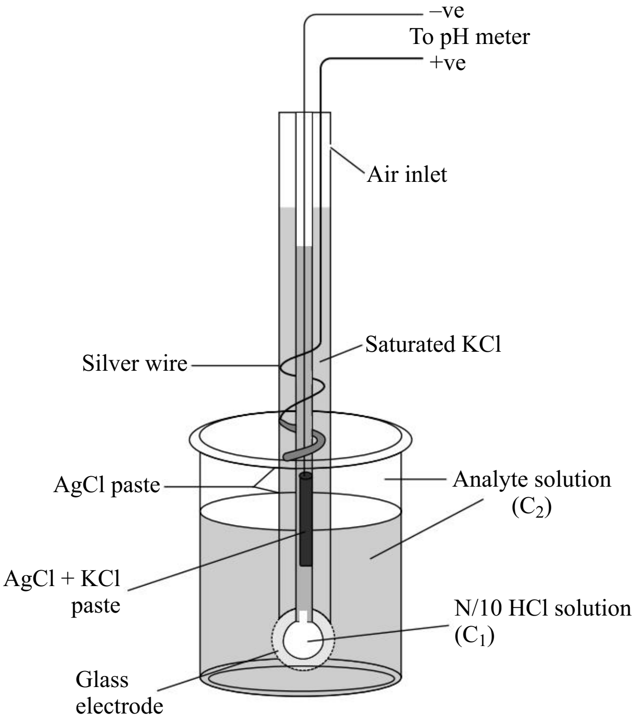


FIGURE 2.12 *A typical combined glass electrode coupled with the common Ag/AgCl electrode.*

If $C_2 > C_1$, the cell can be represented as

$$\underset{\text{[Combined Glass electrode]}}{\downarrow}$$

$$Ag_{(s)}/AgCl_{(s)}/Cl_{(aq)}//H^+\text{(aq, inside)}\|\ H^+\text{(aq, outside)}/Cl^-_{(aq)}/AgCl_{(s)}/Ag_{(s)}$$

When a thin-walled glass bulb containing N/10 HCl solution is immersed in an acidic solution (unknown) concentration and if $C_2 > C_1$, a boundary potential, E_b, is developed across the two gel layers of the glass.

The ion-sensitive glass electrode (ISE) measures the potential, E_b, of a specific H^+ ions of solution in which the cell is immersed. This potential is measured against a stable internal as well the outer reference electrode, $Ag/AgCl_{(s)}/Cl^-$.

Therefore, E_{cell} is given thus:

$$\begin{aligned} E_{cell} &= E^o_{(cathode),ISE} - E^o_{Ag/AgCl(anode)} \\ &= [E_b + E_{asy}] - E^0_{Ag/AgCl(anode)} \\ &= [K - 0.0591p^H + E_{asy}] - E^o_{Ag/AgCl(anode)} \end{aligned}$$

where $E^o_G = [K + E_{asy}]$ is constant for a particular glass material and E_{asy} the potential contribution, exists even if $C_1 = C_2$

The potential of the cell is given by

$$E_{cell} = [E^o_G - 0.0591 \text{ pH}] - E^o_{Ag/AgCl}$$

That is, $0.0591pH = E^o_G - E_{cell} - E^o_{Ag/AgCl}$

Therefore, $pH = E^o_G - E_{cell} - E^o_{Ag/AgCl}$
$= 0.0591$

2.21 Concentration Cells

The electrode potential of a half-cell varies with the concentration of the metal ions around the interface of the electrode in solution. Unlike a galvanic cell, in concentration cell two similar metal electrodes are in contact with the solutions of its ions at different concentrations, resulting in difference in their electrode potentials (see Fig. 2.13).

A concentration cell is also an electrochemical device that generates electrical energy when two electrodes of the same metal are in contact with solutions of its ions at different concentrations.

Consider a concentration cell of the type represented

$$Cu/Cu^{2+}{}_{(C_1)} \,||\, Cu^{2+}{}_{(C_2)}/Cu$$

where C_1 and C_2 are the molar concentrations of the Cu^{2+} ions in contact with two copper electrodes.

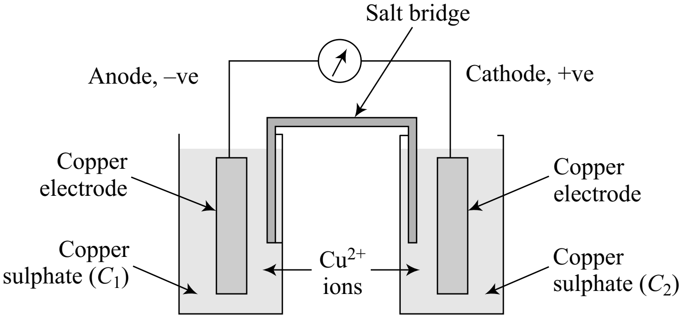


FIGURE 2.13 *A concentration cell of two similar metals.*

By electrochemical convention, if $C_2 > C_1$, then, we have the following reactions.

At anode,

$$\text{Cu} \rightarrow \text{Cu}^{2+} + 2e; \qquad E^{o}_{\text{Cu/Cu}^{2+}} = -\text{ve volt}, \qquad \text{(Oxidation)}$$

At cathode

$$\text{Cu}^{2+} + 2e \rightarrow \text{Cu}; \qquad E^{o}_{\text{Cu}^{2+}/\text{Cu}} = +\text{ve volt}, \qquad \text{(Reduction)}$$

Thus,

The net cell reaction of the concentration cell, $\underset{(C2)}{\text{Cu}^{2+}} \rightarrow \underset{(C1)}{\text{Cu}^{2+}}$.

The voltage of the cell exists as long as $C_2 > C_1$.

The EMF of the concentration cell will be the algebraic difference of the two reduction electrode potentials.

$$E_{\text{cell}} = E^{o}_{\text{cathode}} - E^{o}_{\text{anode}}$$

$$E_{\text{cell}} = \left[E^{o}_{\text{Cu}^{2+}/\text{Cu}} + \frac{0.0591}{2}\log_{10} C_2\right] - \left[E^{o}_{\text{Cu}^{2+}/\text{Cu}} + \frac{0.0591}{2}\log_{10} C_1\right]$$

where $E^{o}_{\text{cell}} = [E^{o}_{\text{Cu}^{2+}/\text{Cu}} - E^{o}_{\text{Cu}^{2+}/\text{Cu}}] = 0$

$$E^{o}_{\text{cell}} = 0$$

Since both the electrodes are of the same material, copper,

Thus, $E_{\text{cell}} = \dfrac{0.0591}{2}\log_{10}\dfrac{C_2}{C_1}$

The EMF of the cell is +ve only if $C_2 > C_1$.

In general, the EMF of a concentration cell is given by

$$E_{\text{cell}} = \frac{0.0591}{2}\log_{10}\frac{C_2}{C_1}$$

The following characteristics of concentration cell can be noted:

1. When $C_2 = C_1$, the concentration cell does not generate electrical energy.
2. When $C_2 = C_1$, log C_2/C_1 is positive, and therefore, E_{cell} is positive.
3. Higher the ratio of C_2/C_1, greater the cell potential.

Example 2.20 Consider a concentration cell of silver electrodes as represented

$$\text{Ag/Ag}^{+}\ (0.00475\ \text{M}) \parallel \text{Ag}^{+}\ (0.043\ \text{M})/\text{Ag}$$

Hint: An electrode of higher concentration of Ag^{+} (0.043 M) acts as the cathode, while the other silver electrode with Ag^{+} ion concentration (0.00475 M) acts as the anode.

Solution

At anode,

$$\text{Ag}_{(s)} \rightarrow \underset{(0.00475\ \text{M})}{\text{Ag}^{+}} + e$$

At cathode,

$$\underset{(0.043\ \text{M})}{\text{Ag}^{+}} + e \rightarrow \text{Ag}_{(s)}$$

The net reaction of the concentration cell is

$$\underset{(0.043\ \text{M})}{Ag^+} \rightarrow \underset{(0.00475\ \text{M})}{Ag^+}$$

The EMF of the concentration cell is

$$E_{\text{cell}} = \frac{0.0591}{1}\log_{10}\frac{[0.043]}{[0.00475]}$$
$$= 0.056\ \text{V}$$

Example 2.21 A concentration cell is constructed by dipping two copper electrodes in 0.001 and 0.1 M $CuSO_4$ and the two solutions are connected by a salt bridge. Calculate the EMF of the cell at 298 K. Since 0.1 > 0.001 M, the electrode that is in contact with higher Cu^{2+} ion concentration acts as the cathode.

Solution:

(i) Representation of concentration cell

$$Cu/Cu^{2+}(0.001\text{M}) \,||\, Cu^{2+}\ (0.1\text{M})/Cu$$

(ii) The two electrode reactions are as follows:
At anode,

$$Cu \rightarrow \underset{(0.001\text{M})}{Cu^{2+}}$$

At cathode,

$$\underset{(0.1\text{M})}{Cu^{2+}} + 2e \rightarrow Cu$$

(iii) The net reaction of concentration cell

$$\underset{(0.1\text{M})}{Cu^{2+}} \rightarrow \underset{(0.001\text{M})}{Cu^{2+}}$$

The EMF of the concentration cell is

$$E_{\text{cell}} = \frac{0.0591}{n}\log_{10}\frac{C_2}{C_1}$$
$$E_{\text{cell}} = \frac{0.0591}{2}\log_{10}\frac{0.1}{0.001}$$
$$= \frac{0.0591}{2}\log_{10}100$$
$$= \frac{0.0591}{2}\times 2$$
$$= 0.0591\ \text{V}$$

Example 2.22 Calculate the EMF of the concentration cell

$$Cu/Cu^{2+}\ (0.072\ \text{M}) \,||\, Cu^{2+}\ (0.32\ \text{M})/Cu$$

Comment on the spontaneity of the cell reaction. The cathode of the concentration cell is the one having higher Cu^{+2} (0.32 M) ion concentration.

Solution

The two half-cell reactions are as follows:
At anode,

$$\underset{}{Cu} \rightarrow \underset{(0.072M)}{Cu^{2+}} + 2e$$

At cathode,

$$\underset{(0.032M)}{Cu^{2+}} + 2e \rightarrow Cu$$

Cell reaction

$$\underset{(0.032\ M)}{Cu^{2+}} \rightarrow \underset{(0.072\ M)}{Cu^{2+}}$$

The EMF of the concentration cell is

$$E_{cell} = \frac{0.0591}{2}\log_{10}\frac{0.032}{0.072}$$
$$= 0.0726\ V$$

Since, E_{cell} is +ve, electron flows from the anode to the cathode of cell.

Example 2.23 A concentration cell was constructed by immersing two silver electrodes in 0.01 and 10 M $AgNO_3$ solutions. Write the cell reactions and calculate the EMF of the concentration cell.

[VTU: 1999]

Solution

The half-cell Ag/Ag^+ (10 M) acts as the cathode since Ag^+ (10 M) > Ag^+ (0.01 M).

(i) Representation of concentration cell,

$$Ag/Ag^+\ (0.01\ M)\ ||\ Ag^+\ (10\ M)/Ag$$

(ii) The electrode reactions are as follows:
At anode,

$$Ag_{(s)} \rightarrow \underset{(0.01\ M)}{Ag^+}$$

At cathode,

$$\underset{(10.0\ M)}{Ag^+} + e \rightarrow Ag_{(s)}$$

(iii) The net cell reaction is

$$\underset{(10.0\ M)}{Ag^+} \rightarrow \underset{(0.01\ M)}{Ag_{(s)}}$$

(iv) The EMF of the concentration cell is

$$E_{cell} = \frac{0.0591}{1}\log_{10}\frac{10.0}{0.7}$$

$$= 0.0591 \log_{10} 1000 = 0.0591 \times 3$$
$$= 0.1773 \text{ V}$$

Example 2.24 A concentration cell was constructed by immersing two silver electrodes in 0.05 and 0.1 M $AgNO_3$ solutions. Write cell representation, cell reactions and calculate the EMF of the concentration cell. [VTU July 2006].

Solution

If $0.1 \text{ M} > 0.05 \text{ M},\ E^{o}_{Ag+(0.1M)} > E^{o}_{Ag+(0.05M)}$

The cell representation is

$$Ag/AgNO_3\ (0.05M)//AgNO_3\ (0.1M)/Ag$$

At anode,

$$Ag \rightarrow Ag^{+}\ (0.05M) + e$$

At cathode,

$$Ag^{+}\ (0.1\ M) + e \rightarrow Ag$$

Net cell reaction,

$$Ag^{+}\ (0.1\ M) \rightarrow Ag^{+}\ (0.05M)$$

The EMF of a concentration cell is

$$E_{\text{cell}} = \frac{0.0591}{n} \log_{10} \frac{[C_2]}{[C_1]}$$
$$= \frac{0.0591}{1} \log_{10} \frac{[0.1]}{[0.05]} = 0.0591 \log_{10} 0.20 \times 10$$
$$= 0.0591 \times 2.030 = 0.12 \text{ V}$$

Exercises

1. Consider the following reaction of an electrochemical cell, $Zn_{(s)} + 2Ag^{+} \rightarrow Zn^{2+} + 2Ag_{(s)}$. (i) Determine the standard cell potential, (ii) determine the standard free energy change at standard conditions and (iii) will this reaction occur spontaneously as written. Given: the standard reduction potentials of Zn and Ag are –0.76 and 0.80 V, respectively. [Ans: $\Delta G^{o} = -30.106$ KJ/mole].
2. Find the standard potential (EMF) of the cell $Cu_{(s)}/Cu^{+2}{}_{(aq)}//Cl^{-}/AgCl_{(s)}/Ag_{(s)}$ and predict the direction of electron flow when the two electrodes are connected. The standard reduction potentials of $Cu_{(s)}/Cu^{+2}{}_{(aq)}//$and Ag/AgCl/$Cl^{-}$ are –0.34 and 0.222V, respectively. [Ans: 0.115 V]
3. Find the ΔG^{o} for the cell reaction in a zinc–aluminium galvanic cell. The standard reduction potentials of Zn and Ag are –0.76 and +1.66V, respectively. [Ans: –521 KJ].
4. Consider an electrochemical cell $Mg_{(s)}/\ Mg^{2+}_{(aq)}\ //\ Ag^{2+}_{(aq)}\ /Ag_{(s)}$ show that the cell voltage is unaffected by multiplying the reaction equation by a positive number. The standard reduction potentials of Mg and Ag are –2.38 and +0.80 V, respectively.

5. The standard cell potential E^o for the reaction $Fe + Zn^{2+} \rightarrow Fe^{2+} + Zn$ is –0.353 V. If a piece of iron is placed in a 1M Zn^{2+} solution, what is the equilibrium concentration of Fe^{2+}.
   [Ans: 1.2×10^{-12} M].

6. Calculate the EMF for the cell $Sn_{(s)}/Sn^{2+}$ (0.15M)//Ag^+(0.30M)/$Ag_{(s)}$ at 25°C.
   The standard reduction potentials of Sn and Ag are –0.14 and +0.80V, respectively.
   [Ans: 0.95 V].

7. Consider the given cell, $Al_{(s)}/Al^{3+}_{(aq)}//Mg^{2+}_{(aq)}/Mg_{(s)}$. Write the electrode and net cell reactions and calculate the E^o for the cell. The standard reduction potentials of Al and Mg are –1.66 and –2.38V, respectively. [Ans: 0.70 V].

8. A copper wire is dipped in $AgNO_3$ solution place in a beaker A. A silver wire is dipped in a solution of copper sulphate placed in a beaker B. If standard reduction potentials for copper and silver are 0.34 and 0.80 V, predict in which beaker the ions present will get reduced?
   [Ans: Ag^+ will get reduced to Ag in beaker A].

9. Using standard electrode potentials, predict the reaction, if any, that occurs between $Fe^{3+}_{(aq)}$ and I^-. Given: $E^o_{Fe^{3+}/Fe^{2+}} = 0.77$ V; $E^o_{I2/I^-} = 0.54$ V.
   [Ans: $2Fe^{3+} + 2I^- \rightarrow 2Fe^{2+} + I_2$].

10. The EMF of the following cell is found to be 0.20 V at 298 K
    $Cd_{(s)}/Cd^{2+}_{(aq)}$ (?)// $Ni^{2+}_{(aq)}$ (2.0M)/$Ni_{(s)}$
    What is the molar concentration of Cd^{2+} ions in solution? [Ans: 0.0409 M].

11. Calculate the EMF of the cell in which the reaction is
    $$Mg_{(s)} + 2Ag^+_{(aq)} \rightarrow Mg^{2+}_{(aq)} + 2Ag_{(s)}$$
    When [$Mg^{2+}_{(aq)}$] = 0.130M and [$Ag^+_{(aq)}$] = 1×10^{-4} M. [Ans: 2.96 V].

12. Calculate the equilibrium constant for the reaction
    $$Zn_{(s)} + Cd^{2+}_{(aq)} \Leftrightarrow Zn^{2+}_{(aq)} + Cd_{(s)}$$
    Given: $E^o_{cell} = 0.36$ V. [Ans: 1.52×10^{12}]

13. Calculate the standard free energy change and maximum work obtainable for the reaction occurring in the Daniel cell.
    $$Zn + Cu^{2+} \Leftrightarrow Zn^{2+} + Cu$$
    Given: $E^o_{Zn^{2+}/Zn} = -0.76$ V; $E^o_{Cu^{2+}/Cu} = 0.34$ V and $F = 96{,}500$ C/mol.
    How is it related to the equilibrium constant for the reaction?
    [Ans: $\Delta G^o = -212.3$ KJ; $K = 1.6 \times 10^{37}$].

14. Calculate the equilibrium constant for the following reaction
    $$Cu_{(s)} + 2Ag^+ \rightarrow Cu^{2+}_{(aq)} + 2Ag_{(s)}$$
    Given: $E^o_{cell} = 0.46$ V. [Ans: 3.69×10^{15}].

15. Consider the cell, $Mg_{(s)}/Mg^{2+}_{(aq)}//Ag^+_{(aq)}/Ag_{(s)}$. The standard reduction potentials of Mg and Ag are –2.38 and 0.80 V, respectively. Calculate the equilibrium constant at 298 K and also the maximum work that can be obtained using the cell. [Ans: $K = 189 \times 10^{107}$].

16. Consider an electrochemical cell,
    $Al_{(s)}/Al^{3+}$(0.01M)//Fe^{2+} (0.20M)/$Fe_{(s)}$.
    The standard reduction potentials of Al and Mg are –1.66 and –0.44 V, respectively.
    Calculate the W_{max} that can be obtained by the cell.
    [Ans: $E^o_{cell} = 1.22$ V; $W_{max} = 705.8$ KJ].

17. Using the Nernst equation calculate the electrode potential of $Sn^{2+} + 2e \rightarrow Sn$ when [Sn^{2+}] = 0.5M. ($E_{Sn2+/Sn} = -0.14$ V).

18. Compute the cell potential of the Ag+/Ag couple with respect to Cu^{2+}/Cu if the concentration of Ag^+ and Cu^{2+} are 4.2×10^{-6} and 1.3×10^{-3} M, respectively. E^o of Ag and Cu are 0.8 and 0.34V, respectively.

19. Calculate the voltage of the cell Mg/ $Mg^{2+}_{(aq)}$ // $Cd^{2+}_{(aq)}$ /Cd at 25°C, when [Cd^{2+}] = 7.0×10^{-11} M, [Mg^{2+}] = 1.0 M and $E^o_{cell} = 1.97$ V
    [VTU: August 2000]

20. Write the half-cell and net cell reactions for the cell, $Cd_{(s)}/Cd^{2+}$(0.01M)//Cu^{2+} (0.5M)/$Cu_{(s)}$ the standard reduction potentials of cadmium and copper are –0.4 and 0.34 V, respectively. Calculate the EMF of the cell.

21. An electrochemical cell consists of iron electrode dipped in 0.1M $FeSO_4$ and a silver

electrode is dipped in 0.05M $AgNO_3$. Write the cell representation, cell reaction and calculate the EMF of the cell at 298 K. Given the standard reduction potentials of iron and silver electrodes are –0.44 and +0.80 V, respectively. [VTU July 2004].

Review Questions

Electrodes and cells

1. What is an electrochemical cell?
2. What is the working principle of an electrochemical cell?
3. What is a galvanic cell? Give an example.
4. Define the following: (a) galvanic cell and (b) electrolytic cell.
5. Describe the construction and working principle of a galvanic cell. Give their electrode and net cell reactions?
6. Mention the electrochemical conventions of an electrochemical cell.
7. Represent any galvanic cell. Write the net cell reaction and give an expression for the EMF of the cell.
8. Why the anode of galvanic cell is –ve and cathode is +ve? Write the different electrode reactions occur at the electrodes. [VTU: March 2001]
9. Explain the origin of potentials at the electrodes of a galvanic cell.
10. What is the criterion that decides anode and cathode of a galvanic cell consisting of two dissimilar metals that are immersed in their own salt solutions?
11. Discuss the origin of electrode potential? Derive the Nernst equation for single electrode potential. [VTU: July 2004]
12. Define single electrode potential? How is it denoted conventionally for both oxidation and reduction processes.
13. Define the terms
    (a) Standard electrode potential
    (b) EMF of a cell.
14. Define single electrode potential and standard electrode potential? Derive the Nernst equation for electrode potential. [VTU: Jan 2003]
15. Derive the Nernst equation for single electrode. Explain the method of determining electrode potential using hydrogen electrode? [VTU: Feb 2005]
16. Define standard electrode potential?
17. Mention the factors that affect the electrode potential.
18. How is the standard electrode potential of a cathode of the cell measured? Mention its sign.
19. How is the standard electrode potential of an anode measured? Mention its sign?
20. Derive the Nernst equation for a single electrode potential of a cell for a reduction electrode process.
21. The net cell reaction is $M_1 + M_2^{n+} \rightarrow M_1^{n+} + M_2$.
    Give an expression for the EMF of a cell, in terms of the ionic concentrations M_1^{n+} and M_2^{n+}, respectively.
22. Mention various types of electrodes or half-cells.
23. Write a note on ion-selective electrode. [VTU: Jan 2003]
24. What is a metal–metal ion electrode? Give an example.
25. What is a metal–metal salution electrode? Give an example.
26. What is a gas electrode? Mention an example.
27. What is an oxidation–reduction electrode? Give an example.
28. What is an ion-selective electrode? Mention an example.
29. What do you mean by a reference electrode? Mention the different types of reference electrodes.
30. Describe standard hydrogen electrode (SHE)?

31. Explain the construction and working of calomel electrode. Mention its advantages. [VTU: Jan 2003].
32. How is potential of an electrode measured using a calomel electrode?
33. Write brief notes on:
    (i) Calomel electrode
    (ii) Glass electrode. [VTU: Feb 2005]
34. Describe the construction of calomel electrode? Why is it called a secondary reference electrode?
35. Construct and describe a calomel electrode.
36. Describe a silver–silver chloride electrode.
37. Explain the construction and working of Ag/AgCl electrode. [VTU: Aug 1999]
38. What is a glass electrode? Describe the construction and working principle of a glass electrode.
39. Give the principle of glass electrode. How is it constructed?
40. What are ion-selective electrodes? Give the construction of glass electrode and explain the experimental method of determining pH using glass electrode. [VTU: July 2003, 2004]
41. What is glass electrode? How is it constructed? Describe the experimental determination of pH of a solution using glass electrode. [VTU: July 2005]
42. Describe a method by which the pH of an unknown solution is determined using a glass electrode.
43. Define a concentration cell.
44. Distinguish a galvanic cell and a concentration cell. Illustrate with examples.
45. Derive an expression for the EMF of a concentration cell.
46. What are concentration cells? Derive an expression for the EMF of a concentration cell. [VTU: March 2000)
47. What are concentration cells? Explain with an example, calculate the EMF of the given cell at 298 K. $Ag_{(s)}/AgNO_3(0.018\ M)//AgNO_3(1.2M)/Ag_{(s)}$. [VTU: Aug 2000, Feb 2002]
48. Derive the Nernst equation for the potential of a single electrode from thermodynamic principle. From this, deduce an expression for the EMF of a copper concentration cell in which the copper ions ratio is 10. Calculate the EMF of this cell at 25°C. [VTU: July2005]
49. Derive the Nernst equation for the potential of a single electrode from thermodynamic principle. From this, deduce an expression for the EMF of a copper concentration cell in which the copper ions ratio is 10. Calculate the EMF of this cell at 25°C.
50. What is glass electrode ? How is it constructed? Describe the experimental determination of pH of a solution using glass electrode.
51. The E^o values of Li/Li^+, Zn/Zn^{++}, Cu/Cu^{++} and Ag/Ag^{++} are –3.0, –0.77, +0.33 and +0.80 V, respectively. Which combination of the electrodes you will use to construct a cell of highest EMF if the ionic concentrations are 0.1, 1.0, 10 and 0.01 M in the same order. Justify your answer.
52. Write a note on calomel electrode. [VTU: Jan 2005]
53. (a) Explain the construction of calomel electrode. Explain how this electrode is used to determine the potential unknown electrode.
    (b) The spontaneous galvanic cell tin/tin ion (0.024 m) 11 tin ion (0.064 m)/tin develops an EMF of 1.126 25°C. Calculate the valency of tin.
    (c) Explain the principles of a membrane electrode. Mention the different types of membranes available. [VTU: Jan 2006]
54. (a) Derive the Nernst equation for single electrode. Explain the determination of single electrode potential using standard hydrogen electrode.
    (b) Explain the construction and working of the calomel electrode.
    (c) A concentration cell was constructed by immersing two silver electrodes in 0.05 M and 0.2 M $AgNO_3$ solution. Write cell representation, cell reactions and calculate the EMF of the concentration cell. [VTU: July 2006]
55. (a) What are reference electrodes? Explain the construction and working of Ag/AgCl electrode.

(b) Write the half-cell and net cell reactions and also calculate the voltage generated in the following cell $Mn/Mn^{2+}//Fe^{2+}/Fe$ when iron rod is immersed in 6.9×10^{-4} m $FeSO_4$ and Mn rod is immersed in 2.6×10^{-6} m $MnSO_4$ solution. Give E^o for Fe^{2+}/Fe is –0.4 V and Mn/Mn^{2+} is –1.18 V.

(c) What are ion-selective electrodes? Explain the construction and application of glass electrodes in P^H determination.

[VTU: July 2007]

56. Derive Nernst equation on electrode potential?

[VTU: July 2008)

57. (a) Define (i) Signal electrode potential and (ii) Standard electrode potential.

(b) An electrochemical galvanic cell is obtained by coupling silver [$E^o_{Ag^+/Ag} = 0.80$ V] with standard hydrogen electrode [$E^o_{SHE} = 0$] at 298 K. How would you determine the potential of silver electrode? Represent the electrochemical cell and write the cell reactions.

(c) Consider an electrochemical cell as given, Zn/Zn^{2+} (0.005M)//Ag^+(0.1 M)/Ag. The cell reaction is spontaneous at 298 K. The standard reduction potential of zinc and silver are –0.76 and 0.80 V, respectively. Write the redox electrode reactions with their respective electrode potentials, net cell reaction and calculate the EMF of the cell.

(d) What are reference electrodes? Explain the construction and working of a calomel electrode. [VTU: January 2008.)

58. (a) What are ion-selective electrodes? Explain the measurement of pH of a solution using glass electrode?

(b) Explain the origin of single electrode potential.

(c) Describe the construction and working of a calomel electrode.

(d) Calculate the EMF of the cell Fe/Fe^{2+} (0.01 M)//Ag+ (0.1M)/Ag at 298 K, if standard electrode potential of Fe and Ag electrodes are – 0.42 and 0.8 V, respectively.

[VTU: 2008].

59. (a) Define reference electrode. Explain the measurement of single electrode potential using calomel. Electrode. Derive Nernst's equation for single electrode potential.

(b) What are types of electrodes? Explain briefly with an example for each.

[VTU: June/July 2016]

60. (a) Explain the construction and working of silver–silver chloride electrode.

(b) What are ion-selective electrodes? Discuss the construction and working of glass electrode.

[VTU: CBSE, June/July 2016]

61. (a) Define standard electrode potential. Derive Nernst's equation for single electrode potential.

(b) What are concentration cells? Give an example. Derive an expression for the concentration of EMF of concentration cell.

(c) Mention any four advantages of calomel electrode. An electrochemical cell of a nickel electrode dipped in 0.05M $NiSO_4$ and lead electrode dipped in 0.5M $PbSO_4$ solution. Write the cell scheme, cell reaction and calculate the EMF of the cell at 298K. Given the standard electrode potentials of Ni and Pb electrodes are –0.24 and – 0.13 V, respectively.

[VTU: June/July 2016]

63. (a) Explain the construction and working of glass electrode.

(b) Give the construction of calomel electrode. Justify that it is reversible.

[VTU: June/July 2015]

3

Conversion and Storage of Electrochemical Energy

Chapter Outline

Battery technology:

Introduction. Basic concepts of batteries. Battery characteristics—free energy change, EMF, capacity, power density, energy density, efficiency and cycle life. Classification of batteries with examples, Applications of cells and batteries. Classical cells—Primary cells—Zn–MnO_2,R.M cell and Ag_2O–Zn cell. Secondary cells and batteries—lead acid storage battery, alkaline rechargeable cells—nickel–cadmium cell. Modern cells–Zn–air, Ni–MH and Li–MnO_2. Lithium batteries—classification of lithium batteries. Categories of primary lithium batteries—soluble cathode cells, solid cathode cells and solid electrolyte cells. Types of secondary rechargeable lithium batteries—liquid electrolyte cells, polymer electrolyte lithium cells, lithium ion cells—construction, working and chemistry of lithium ion battery and their applications.

Fuel cell technology:

Introduction. Merits of fuel cell, classification of fuel cells, types of fuel cells—alkaline fuel cell (AFC), phosphoric acid fuel cell (PAFC), molten carbonate Fuel cell (MCFC), polymer electrolyte membrane fuel cell (PEMFC), solid oxide fuel cell (SOFC).

3.1 Battery Technology

3.1.1 Introduction

With all the new applications, which have developed over the past few decades, the market for consumer batteries has grown correspondingly. Nowadays, the average family may have many batteries at any one time in and around the home. Many of these batteries are of advanced design and construction, giving greatly improved performance as a result of developments in materials science and technology. Although most small consumer batteries are still of the primary variety, there is growing trend to adopt secondary (rechargeable) batteries as being more economical.

A cell has only two electrodes (or half cells) that generate electrical energy and the EMF thus, generated depend on the magnitude of the electrode potentials of the two electrodes. Depending on the specific purpose, higher voltage can be achieved by coupling a number of cells in series. The arrangement of two or more cells coupled in series is called a battery. *The working principle of a battery is the transformation of free energy change of redox reactions of the electrode active materials of cells into electrical energy.*

Batteries are the indispensable source of portable energy. Nowadays, wide varieties of batteries are available in the market; and these find extensive applications in modern technology. The batteries are constructed as per the desired requirements and also to suit specific applications.

Depending on the type of battery, they are used in electronic gadgets, pacemakers, calculators, power supplies, telecommunication equipment, and so on. High power batteries are also being tested and used in small cars, to minimize the air pollution problem. Nowadays, batteries of long shelf life, power, recharging capacity, tolerance to extreme conditions and reliability are desirable, and have wide applications too.

3.1.2 Basic Concepts of Batteries

A galvanic cell is a device that generates electrical energy at the expense of decrease of free energy of electrode reactions of a cell.

A ***battery*** *is a device that consists of one or more of cells connected in series or parallel or both and converts the chemical energy by means of an electrochemical oxidation–reduction reaction depending on their desired output voltage and capacity* (Fig. 3.1).

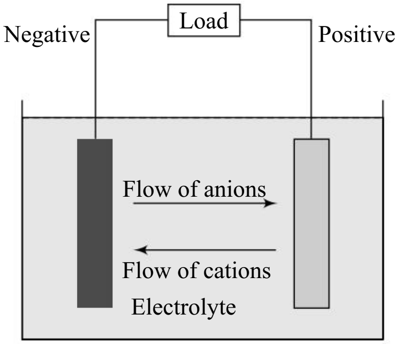


FIGURE 3.1 *An electrochemical cell.*

The cell consists of three major aspects:

1. The anode (or the –ve electrode) is oxidised during the electrochemical reaction and liberates electrons to the external circuit.

   $$M \rightarrow M^{n+} + ne$$

   The materials having the following properties are preferable as an anodic material. Efficiency as a reducing agent, high coulombic output (Ah/g), good conductivity, stability, ease of fabrication and low cost.
2. The cathode (or positive electrode) is reduced during the electrochemical reaction, which accepts electrons from the external circuit.

   $$M^{n+} + ne \rightarrow M$$

   The cathode must be an efficient oxidizing agent, be stable when in contact with the electrolyte, and have useful working voltage.
3. The electrolyte provides the medium for transfer of ions inside the cell between the anode and the cathode. The electrolyte must have good ionic conductivity.

The battery itself can be built in many shapes and configurations—cylindrical, button and flat.

3.1.3 Battery Characteristics

A battery may be specifically designed, constructed and used based on its characteristic properties. The few important characteristics of batteries are mentioned below.

1. Free energy change

Whenever electrode reactions occur in the presence of active materials in a battery, there occurs a decrease in free energy of the redox system.

$$-\Delta G = nEF$$

where F = Faraday (96500 C or 26.8 Ah)
n = number of electrons involved in electrode reactions.
E = potential, Volts.

2. EMF of a battery

The net voltage (or EMF) of battery depends on the total number of cells, which constitute a battery. The overall reaction of a cell of a battery is given as

$$M_1 + M_2^{n+} \rightleftharpoons M_1^{n+} + M_2$$

Thus, the EMF of a cell in a battery, E_B is

$$E_B = E^o_{cell} - \frac{0.0591}{n}\log_{10}\frac{[M_1^{n+}]}{[M_2^{n+}]}$$

where $E^o_{cell} = E^o_{cathode} - E^o_{anode}$ and $\frac{[M_1^{n+}]}{[M_2^{n+}]}$ is the ratio of the ionic concentrations of M_1^{n+} and M_2^{n+}.

It is evident, therefore, that the EMF of the cell of a battery is dependent on.

(i) p.d between the cathode and anode,
(ii) the ratio of the ionic concentration of M_1^{n+} and M_2^{n+}, and
(iii) the temperature.

Summarizing, it can be remarked that

- EMF of the battery is higher, if the electrode potential difference between the two electrodes is more.
- EMF of the cell decreases with the increasing molar concentration of $[M_1^{n+}]$ in the numerator of the expression.

If the temperature of the battery increases, the EMF of the cell gets reduced marginally.

3. Amps/Ampere-hour

Also known as *Amperes*. This is the rate at which electrons flow in a wire. The units are coulombs per second, or since an electron has a charge of 1.602×10^{-19} coulombs, an amp is $6.24 \times 10^{+18}$ electrons per second.

One ampere-hour is equal to a current of one ampere flowing for one hour. A unit quantity of electricity used as a measure of the amount of electrical charge that may be obtained from a storage battery before it requires recharging.

4. Capacity

The capacity of a battery is expressed as the total quantity of electricity involved in the electrochemical reaction and is defined in terms of coulombs or amperes-hours (Ah) or the total number of ampere hour or watt-hours that can be withdrawn from a fully charged cell or battery under specified conditions of discharge is termed as the capacity of a battery. The theoretical capacity of a battery is the quantity of electricity involved in the electro-chemical reaction. It is denoted Q and is given by

$$Q = xnF$$

where x = the number of moles of reaction, n = the number of electrons transferred per mole of reaction and F = Faraday's constant.

The capacity is usually given in terms of mass, not the number of moles:

$$Q = \frac{nF}{M_r}$$

where M_r = molecular mass.

This gives the capacity in units of ampere-hours per gram (Ah/g). The ampere-hour capacity of a battery on discharge is determined by a number of factors, among which the following are the most important: final limiting voltage; quantity of electrolyte; discharge rate; density of electrolyte; design of separators; temperature, age and life history of the battery; and number, design and dimensions of electrodes.

The total capacity, Ah or Wh, that will be obtained from a cell or battery at defined discharge rates and other specified discharge or operating conditions is known as its 'available capacity'.

5. Power

The *power* generated by a battery can be calculated as

$W = V.I$, where V = cell voltage and I = cell current (or rate)

since $V = IR$; $w = I^2R$ and $w = V^2/R$

The *energy* generated by a battery is power × time or

$$E = V.I.\ t = qV$$

where q = charge = rate × time

6. Power density

The *power density* is usually discussed in terms of the cell mass:

Power density = Energy (E)/time (t)/mass (kg)

= Energy (qV)/time (t)/mass (kg) of cell

= Power/mass (units are W/kg)

The ratio of the power delivered by a cell or a battery to its weight, w/kg, is also known as the power density of a battery. During the discharge of a battery, the power density decreases.

This is related to the *energy density* at a given discharge rate and indicates how rapidly the cell can be discharged and how much power generated. A cell with high energy density may exhibit a significant voltage and capacity drop at higher discharge rates and power density, therefore, has a low power density.

7. Energy density

The energy density or capacity—determined by the voltage of the cell and the amount of charge that can be stored, $E = qV$, this parameter is usually evaluated on a weight or volume basis:

Theoretical. weight. capacity = qV/mass (units are W h/kg)

Theoretical. volume. capacity = qV/volume (units are Watt h/L)

The energy density of a cell or a battery is also described as the ratio of the energy output of a cell or battery to its weight, Wh/kg.

8. Efficiency

The ratio of the output of a battery on the discharge to the input required to restore it to the initial state of charge under specified conditions

(i) 'Voltage efficiency' is described as the ratio of average voltage during discharge to average voltage during recharge under specified conditions.
(ii) 'Watt-hour efficiency' is known as the ratio of watt-hours delivered on discharge of a battery to the watt-hour needed to restore it to its original state under specified conditions of charge and discharge.
(iii) Ampere-hour efficiency—The ratio of the output of a secondary cell or a battery, measured in ampere-hours, to the input required to restore the initial state of charge, under specified conditions (also coulombic efficiency).

9. Cycle life

For rechargeable batteries, the duration of satisfactory performance, measured in years or in the *number of charge/discharge cycles*. In practice, end of life is usually considered to be reached when the cell or battery delivers approximately 80 percent of rated ampere-hour capacity.

Shelf Life—*The duration of storage under specified conditions at the end of which a cell or a battery still retains the ability to give a specified performance.*

For a dry cell, the period of time (measured from date of manufacture), at a storage temperature of 21°C (69°F), after which the cell retains a specified percentage (usually 90 percent) of its original energy content.

10. Tolerance to service conditions

The battery duty may require that it provides power continuously, intermittently or at an irregular rate.

It may also be expected to perform its duty or be stored under a range of conditions (e.g.−40 to +55°C to be successful in the extremes of winter and summer).

The battery may also have to be tolerant to various types of misuse, including occasional shorting, vibration and shock.

3.1.4 Commercial Cell or Battery

An assembly of electrochemical cells is called a commercial battery, provided it satisfies the following requisites.

1. Long shelf life
2. Recharging capacity
3. Easily portable
4. Compact and lightweight and
5. Low priced

3.1.5 Classification of Batteries

Electrochemical cells or batteries are identified as primary (non-rechargeable) or secondary (rechargeable), depending on their capability of being electrically recharged. The batteries are classified as (i) primary, (ii) secondary, and (iii) reserve batteries.

1. Primary batteries

The working principle of a primary battery is the conversion of the free energy change of the active materials during electrode processes into the electrical energy. *A battery which is not intended to be recharged and discarded when the battery has delivered all its electrical energy is known as a primary battery.*

The net cell reaction of a primary battery is irreversible and as long as the active materials are present in a battery, the cell generates electrical energy. In other words, primary batteries cannot be recharged.

Example: Zn–MnO_2 dry cell.

2. Secondary or rechargeable batteries

A secondary battery is known as a galvanic battery, which after discharge, may be restored to the fully charged state by the passage of an electric current through the cell in the opposite direction to that of the discharge.

In other words, the net cell reactions of battery can be reversed. They are storage devices for electrical energy and are known as 'storage batteries'.

Examples: Lead–acid battery and Ni–Cd battery.

The secondary batteries have advantages over the other primary batteries that the net cell reactions can be reversed during the charging process and the current can be drawn during the discharge process. The secondary batteries have better cycle life and capacity, so that it can be used over and over again. The secondary batteries are classified into two types:

1. Acid storage battery—lead–acid battery.
2. Alkaline storage battery—Ni–Cd battery.

3. Reserve batteries

In these reserve types of batteries, vital component is separated from the rest of the battery prior to activation. Under this condition, chemical deterioration or self-discharge is essentially eliminated, and the battery is capable of long-term storage. Usually, an electrolyte is the component that is isolated.

These batteries are used, for example, to deliver high power for relatively short periods of time, in missiles, torpedoes and other weapon systems.

Example: LiV_2O_5 cell.

3.1.6 Important Applications of Batteries

Batteries have a wide variety of applications and their market is very potential

- Car batteries
  High specific power is required for the large mechanical load, low operating temps can arise, safety and environmental concerns an important issue. The Pb–acid cell is currently used exclusively.
- Electrically powered vehicles use batteries to replace the combustion engine. The batteries will need sufficient energy and power densities, and environmental safety, and cost issues become especially important for batteries produced on a large scale.
- Secondary cells for portable high-power apps such as portable computers (a rapidly growing market—currently these use NiCd, NiMH or lithium ion cells), power tools, flashlights, etc.
- Primary cells for portable low-power applications such as watches, metres, cameras, calculators. In these devices, low power densities are often acceptable, but low self-discharge rates and high energy densities are desirable to provide long service life. Lithium batteries are most often used.
- Military applications: Examples include power for missile or torpedo guidance, drive or activation (these can obviously be primary cells), and communication devices. Fuel cells are currently used extensively.
- Pacemakers
- Cost is not a primary issue, but concerns focus on a long service life, safety and discharge profile. A lithium–iodine cell is typically used.
- Hearing aids: Usually use mercury, silver oxide or zinc–air batteries because the high volumetric energy densities allow for a smaller battery.

3.1.7 Classical Batteries

The important classes of classical batteries, whose construction, working and applications are discussed below:

3.1.7.1 Primary cells

1. Zn–MnO_2 dry cell

Anode—Zn cylinder
Cathode—MnO_2/C
Electrolyte—$ZnCl_2 + NH_4Cl$

This is a primary battery, popularly known as the *dry cell.* It consists of a graphite rod (with a metal cap) at the centre, which acts as the cathode of the cell. The cathode is surrounded by a thick moist paste of graphite and MnO_2. The outer layer consists of a paste of $ZnCl_2$ + NH_4Cl. The entire assembly is placed inside a thin zinc cylinder container, which acts as an anode of the cell as shown Fig. 3.2. The whole assembly is wrapped with polypropylene or cardboard materials, which prevent leakage, if any. The electrochemical representation of Zn–MnO_2 dry cell is

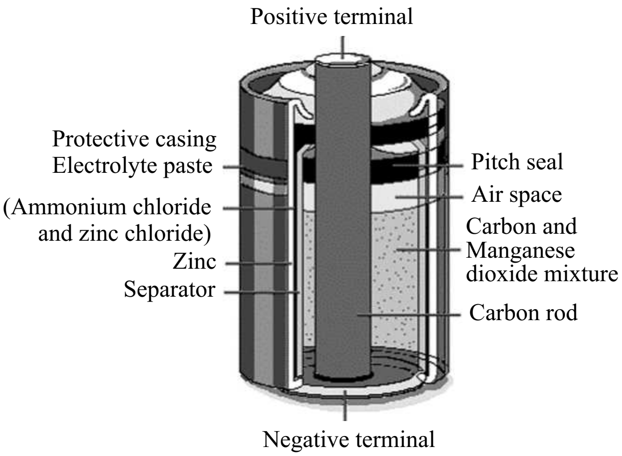


FIGURE 3.2 *Zn–MnO_2 cell (dry cell)*

$$Zn/Zn^{2+} \| NH_4^+/MnO_2/C$$

The electrode reaction of the primary cell is
At anode,

$$Zn \rightarrow Zn^{2+} + 2e$$

At cathode,

$$2MnO_2 + 2H_2O + MnO(OH) + 2OH^-$$

The net cell reaction of the primary cell is

$$Zn + 2MnO_2 + 2H_2O \rightarrow 2MnO(OH) + Zn^{2+} + 2OH^-$$

In the above electrode process, MnO_2 is reduced to Mn_2O_3.

Secondary reactions:

Though, the secondary reactions also take place inside the dry cell, the free energy change of these reactions does not contribute significantly to the EMF of the dry cell. The following reactions occur.

$$2NH_4Cl + 2OH^- \rightarrow 2NH_3 + 2H_2O + 2Cl^-$$
$$Zn^{2+} + 2NH_3 + 2Cl^- \rightarrow Zn(NH_3)_2Cl_2$$

The hydroxyl ions produced during the electrode reaction react with NH_4Cl and formed NH_3. The overall secondary reactions taking place inside the cell is the formation of $Zn(NH_3)_2Cl_2$ complex.

Limitations:

1. The capacity of the battery is low. They are not suitable for high drain applications (such as driving electric motors) as they readily polarize and the available capacity falls sharply with the increased discharge rate.
2. The batteries are not chargeable.
3. The shelf life of the battery is not especially long.

4. The optimum temperature range of operation is 20–40°C; outside this temperature range the performance deteriorates markedly.

Advantages and uses:

- The battery/cell is a convenient source of power for portable electric and electronic devices, lighting, photographic equipment, computers, communication equipment, watches and calculators and a variety of other applications.
- The other general advantages are good shelf life, reasonable power density, low cost, reliable performance and ready availability.

Cell performance:

- The primary cells generate almost constant voltage for small discharge current at a reasonable moderate loads. For reasons, natural deterioration is observed, though these cells have long shelf life.
- The efficiency of the cell is high, if the battery operates within the optimum temperature, 40°C.

2. Alkaline primary cells

One of the materials used in the dry cell is slightly acidic and corrodes the zinc container. The dry cell, therefore, have a limited shelf life. In order to overcome this limitation of dry cell, alkaline electrolyte (KOH) is used in primary cells to obtain better shelf life.

There are two types of alkaline primary cells.

(i) HgO–Zn cell and

(ii) Ag_2–Zn cell

(i) HgO–Zn cell

This alkaline primary cell is known as Reuben–Mallori (RM) cell.

In RM cell, an amalgamated zinc rod is at the centre which acts as the anode of the cell. An absorbent material (cellulose) containing 40 percent KOH saturated with ZnO surrounds the anode. This is surrounded by a paste of 5–10 percent graphite in HgO which is introduced into the steel case as shown in Fig. 3.3. The anode and the cathode materials are separated by a layer of insulating material.

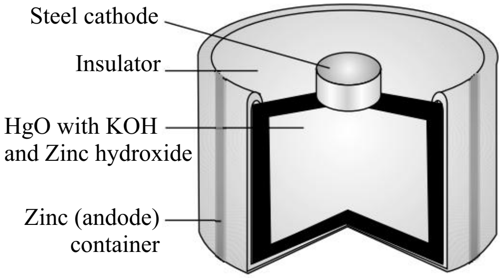


Figure 3.3 *A Reuben–Mallori (RM) cell.*

The redox reactions of RM cell are

At anode,

$$Zn + 2OH^- \rightarrow ZnO_{(s)} + H_2O + 2e$$

At cathode,

$$HgO_{(s)} + H_2O + 2e \rightarrow Hg_{(s)} + 2OH^-$$

The overall cell reaction is

$$Zn + HgO_{(s)} \rightarrow ZnO_{(s)} + Hg_{(s)}$$

The voltage of the RM cell is 1.35 V. It has high capacity and the potential remains constant till its shelf life. It is used extensively in electronic watches, electronic gadgets, etc.

(ii) $Ag_2O_{(s)}$–Zn cell

It is sold in the market as a button-type alkaline primary cell.

A typical button type Ag_2O–Zn cell is shown in Fig. 3.4. In this cell, the Ag_2O material is pressed into the thin button type metal case, which acts as the cathode. An absorbent material soaked in KOH is placed in between the anodic and cathodic compartments. A zinc metal is at the centre of the cell which acts as the anode of the cell. The electrolyte is KOH solution.

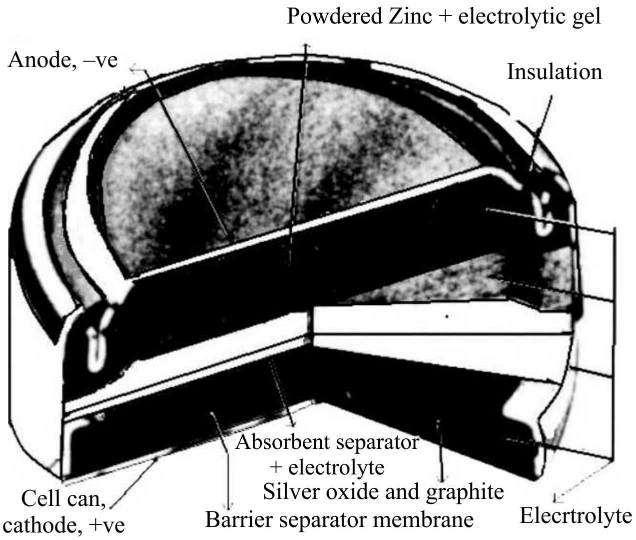


FIGURE 3.4 *A Ag_2O–Zn cell.*

The redox reactions of the cell are
At anode,

$$Zn + 2OH^- \rightarrow ZnO_{(s)} + H_2O + 2e$$

At cathode,

$$Ag_2O + H_2O + 2e \rightarrow 2Ag_{(s)} + 2OH^-$$

The overall cell reaction is

$$Zn + Ag_2O \rightarrow ZnO_{(s)} + 2Ag_{(s)}$$

The voltage generated by the cell is 1.5 V. This cell offers better cell performance than others because of its high voltage, long life, more reliable and nontoxic. These button-type cells are used in electronic watches, calculators and small electronic goods.

3.7.2 Secondary Battery/Cell.

(1) Lead–Acid Battery

(a) Discharging electrode reactions

This is a secondary storage battery. A lead–acid battery in the charged condition can be represented as

$$Pb/PbSO_{4(s)}H_2SO_{4(aq)} \parallel PbSO_{4(s)}PbO_{2(s)}/Pb$$

The net cell reactions of the lead–acid battery are reversible and hence, the reactions of the cell can be reversed, or in other words, the cell can be charged.

The lead–acid battery has a number of cells coupled in series to obtain higher voltage. A typical cell of a battery is shown in Fig. 3.5(a).

(a) Anode—Pb grid with Spongy lead
(b) Cathode—Pb–Sb with lead dioxide
(c) Electrolyte—28–30 percent H_2SO_4 acid.

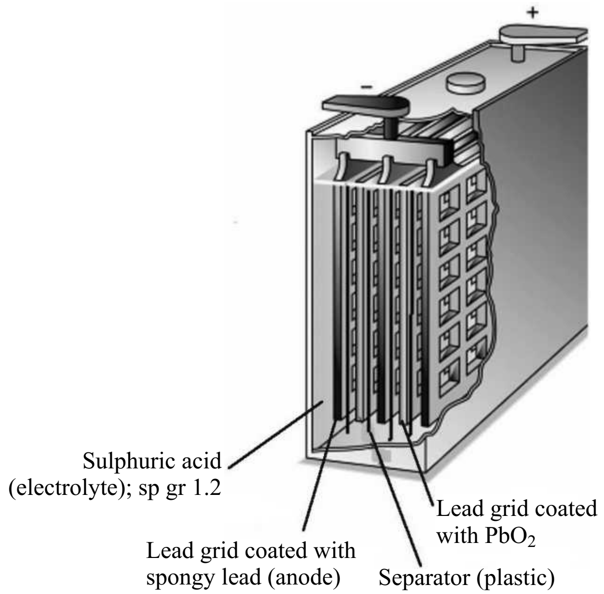


Figure 3.5(a) *A lead–acid battery*

The anode of the cell is a lead grid filled with spongy lead, while the cathode of the cell is also a lead–antimony grid containing PbO_2. The anode and the cathode of the cell of a battery is separated by inserting an inert porous material in between them. The electrolyte is about 28–30 percent H_2SO_4 (sp.gr.1.2. at 25 percent C). The whole set-up of number of such cells are enclosed inside an ebonite case, as shown in Fig. 3.5(b).

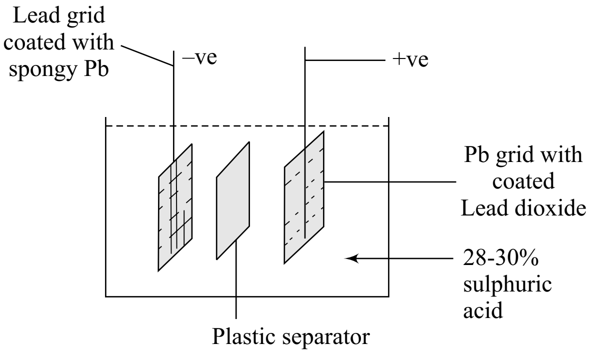


Figure 3.5(b) *A typical cell of a battery*

(i) The discharge reactions of a battery

During the discharge process of a battery, the free energy change of the redox reaction is converted into electrical work.

That is, $-\Delta G = nEF.$

The electrode reactions of a cell (or battery) can be as follows:

At anode,

$$\mathrm{Pb} \rightarrow \mathrm{Pb}^{2+} + 2e$$

$$\begin{array}{l} \mathrm{Pb}^{2+} + \mathrm{SO}_4^{2-} \rightarrow \mathrm{PbSO}_4 \\ \hline \underset{(s)}{\mathrm{Pb}} + \mathrm{SO}_4^{2-} \rightarrow \mathrm{PbSO}_4 + 2\mathrm{e} \end{array} \quad E^{\circ}_{\mathrm{Pb/Pb^{2+}}} = -0.36\ \mathrm{V}$$

At cathode,

$$\begin{array}{l} \mathrm{PbO}_2 + 4\mathrm{H}^+ + 2e \rightarrow \mathrm{Pb}^{2+} + 2\mathrm{H}_2\mathrm{O} \\ \mathrm{Pb}^{2+} + \mathrm{SO}_4^{2-} \rightarrow \mathrm{PbSO}_4 \\ \hline \mathrm{PbO}_2 + 4\mathrm{H}^+ + \mathrm{SO}_4^{2-} + 2e \rightarrow \mathrm{PbSO}_4 + 2\mathrm{H}_2\mathrm{O} \end{array} \quad E^{\circ}_{\mathrm{PbO_2/PbSO_4}} = 1.69\ \mathrm{V}$$

The net cell reaction is

$$Pb + PbO_2 + 2H_2SO_4 \rightarrow 2PbSO_4 + 2H_2O;\ E_{cell} = 2.00\ V$$

The voltage produced during the discharge process is 2.0 V. The $PbSO_4$ formed during the discharge process gets adhered to the electrodes. The $PbSO_4$, thus, deposited on the electrode interface is utilized again during the charging process of the battery.

Since, the EMF of the cell is dependent on the concentration of H_2SO_4, a 28–30 percent (sp. gr. 1.2) sulphuric acid is used as an electrolyte.

(ii) Charging reactions of a battery

The chemical energy of a battery can be restored by utilizing the electrical energy from an external source that is, $neF = -\Delta G$.

A battery needs charging when the specific gravity of H_2SO_4 acid falls below 1.2. The cells of a battery can be charged by reversing the electrode reactions. This can be achieved by applying a higher EMF of an external source. During the charging process, each cell of a battery acts as an electrolytic cell. Lead and lead dioxides are deposited on their respective electrodes.

The electrode reactions taking place during charging are

$$PbSO_4 + 2e \rightarrow Pb + SO_4^{2-}$$
$$PbSO_4 + 2H_2O \rightarrow PbO_2 + SO_4^{2-} + 4H^+ + 2e$$

The overall reaction is

$$2PbSO_4 + 2H_2O \rightarrow Pb + PbO_2 + 2H_2SO_4$$

The sp. gr of H_2SO_4 inside the battery increases to about 1.2.

In general, if the lead–acid battery is reversible with a better cycle life, then the net battery cell reaction during discharging charging processes for a cycle can be represented as

$$Pb + PbO_2 + 2H_2SO_4 \underset{\text{charge}}{\overset{\text{discharge}}{\rightleftarrows}} 2PbSO_4 + 2H_2O$$

Lead-storage batteries are used in automobiles, UPS, telephone exchanges, trains, laboratories, etc.

Example 3.1 Calculate the voltage produced by a cell of a lead–acid battery

$Pb/Pb_{(s)}SO_4/H_2SO_4\ (5M)\ \|\ Pb_{(s)}SO_4/Pb_{(s)}O_2/Pb$

Solution

The given standard electrode potentials of the electrode reactions are

$$PbSO_4 + 2e \rightarrow Pb + SO_4^{2-};\ E^o_{Pb^{2+}/Pb} = 0.36\ V$$
$$PbO_2 + 4H^+ + SO_4^{2-} + 2e \rightarrow PbSO_4 + 2H_2O;\quad E^o_{PbO_2/PbSO_4} = 1.69\ V$$
$$E_{cell} = E^o_{PbO_2/PbSO_4} - E^o_{Pb^{2+}/Pb}$$
$$= 1.69 - (-0.36) = 2.05\ V$$

Cell performance:

The performance and efficiency of lead acid batteries changes during idle storage or due to repeated charge/discharge cycle or due to poor construction of the electrode materials.

(i) Electrical short circuit due to the crumble down of the active materials at electrode plates.

(ii) Parasitic corrosion of the grid electrode can also occur. The charging of the battery minimizes these problems.

(iii) Any variation of the specific gravity of sulphuric acid alters the cell performance and hence sp. gr. of 1.2 has to be maintained.
(iv) The accumulation of lead sulphate causes the electrode plates to collapse under stress.
(v) Over voltage leads to nonadhesion and shedding of active materials.

2. Nickel–cadmium rechargeable cell

This is another secondary storage alkaline battery. A Ni - Cd cell is a rechargeable battery. The cell can be represented as

Cd/CdO/KOH (6M) || NiOOH. $Ni(OH)_2$/Ni
Anode—Spongy Cd/CdO
Cathode—Nickel oxide hydroxide/Ni
Electrolyte—KOH (6M)

Rechargeable Ni–Cd batteries are composed of a +ve electrode plate which uses nickel oxy hydroxide as its main active material, a –ve plate which uses cadmium as its main active material, a separator made of a thin nonwoven fabric, an alkaline electrolyte and a metal case a sealing plate provided with a self-sealing safety valve. In the Ni–Cd cell as shown in Fig. 3.6, a cell cup made up of steel coated with nickel acts as the cathode and a cell cap in contact with the anode of the cell is at the centre and anodic.

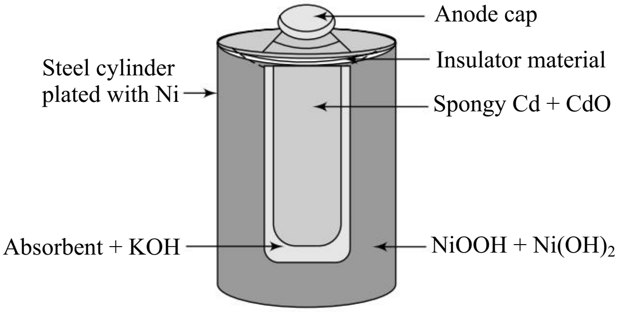


Figure 3.6 *An Ni–Cd cell*

- *Anode composition*—The composition of the active materials of the anode are spongy Cd with 78 percent $Cd(OH)_2$, 18 percent Fe, 1 percent Ni and 1 percent graphite which are pressed into a tablet form and wrapped in nickel wire gauge.
- *Cathode composition*—The outer cup contains a mixture of 80 percent NiOOH and $Ni(OH)^2$, 2 percent $Co(OH)_2$, 18 percent graphite and very minute quantities of barium compounds to increase the efficiency of the active materials and also the cycle life.

The two electrodes are separated by an absorbent material soaked with 6M KOH as an electrolyte. Insulated washer at the top separates the anode and the cathode compartments without any contact. The discharging and charging reactions of Ni–Cd cell are reversible.

For rechargeable Ni–Cd cell, the charge and discharge reactions are as follows:

At anode

When the cell is supplying current to an external circuit, electrons flow from this electrode, and the reaction proceeds to form cadmium hydroxide—the cadmium is *oxidised* to form cadmium hydroxide

$$Cd_{(s)} + 2OH^- \leftrightarrow Cd(OH)_2 + 2e$$

When the cell is being charged this reaction is reversed. Electrons are supplied to the cadmium electrode, and the hydroxide is reduced back to metallic cadmium.

At cathode

The positive electrode is a hydrated oxide of nickel. During discharge it is reduced to divalent nickel hydroxide.

$$2NiO(OH) + 2H_2O + 2e \leftrightarrow 2Ni(OH)_2 + 2OH^-$$

When the cell is supplying current to an external circuit, electrons flow into the cathode from the external circuit and the reaction proceeds to form nickel hydroxide.

The overall reaction is

$$Cd_{(s)} + 2NiOOH + 2H_2O \leftrightarrow Cd(OH)_2 + 2Ni(OH)_2$$

The voltage of the Ni–Cd cell is 1.40 V. It is very compact, lightweight, rechargeable battery with very low internal resistance, good cycle life, capacity and long shelf life. They are used in calculators, pacemakers, transmitters, hearing aids and a host of others.

Advantages:

It is very reliable, sturdy, long-life battery, which can be operated effectively at high discharge rates and a wide temperature range. It has good retention properties, and it can be stored for long periods of life in any condition without deterioration. Little maintenance is needed for this battery.

Applications:

Ni–Cd batteries are used in photography, phones, computers, transmitters, hearing aids, emergency lights and so on. Vented Ni–Cd batteries also find applications in railways, for marine duties, aircraft and traction batteries.

3.1.8 Modern Batteries

Many of the commercial batteries are not able to meet the performance requirements of many applications. Thus, a continual need exists for both the conventional battery technology with improved performance and advance battery technologies with characteristics such as high energy level, long life, low cost little or no maintenance and safety.

1. Zinc–air battery

Anode—granulated Zn powder
Cathode—Air/C
Electrolyte— KOH 6M

The production of electrochemical energy in Zn–air battery is due to the use of oxygen from the atmosphere. The diffused oxygen acts as a cathode reactant in the battery. The air cathode catalytically promotes the reaction of oxygen with an aqueous alkaline electrolyte and is not consumed or changed during the discharge. When an alkaline electrolyte is used in the Zn/air battery, it is necessary to increase only the amount of zinc present to increase the battery capacity.

The air cathode acts only as a reaction site and is not consumed. The reason for the increased energy density in the Zn–air cell is because of larger volume containing the active material. Since the air cathode has infinite life, the electrical capacity of the cell is determined only by the *anode capacity*, resulting in at least a doubling of energy density. Theoretically, the air cathode has infinite use life and its physical size and its electrochemical properties remain unchanged during cell discharge. A schematic representation of a typical Zn–air cell is shown in Fig. 3.7.

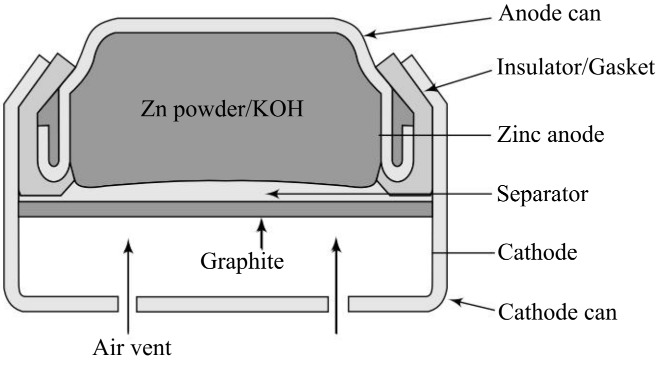


Figure 3.7 *A typical zinc–air cell*

A loose granulated powder of Zn mixed with the electrolyte (KOH) acts as the *zinc anode material* and in some cases, a gelling agent is used to mobilize the composite and ensure adequate electrolyte contact with zinc granules. The outer metal (button type) acts as the cathode of the battery and a plastic gasket insulates the anode active materials and the cathode as shown in Fig. 3.7.

The chemistry of the electrode reactions taking place in zinc–air battery is as follows:

At anode,

$$\mathrm{Zn} \rightarrow \mathrm{Zn}^{2+} + 2e \qquad E^{\mathrm{o}}=1.20\ \mathrm{V}$$

$$\mathrm{Zn}^{2+} + 2\mathrm{OH}^{-} \rightarrow \mathrm{Zn(OH)}_2$$

$$\mathrm{Zn(OH)}_2 \rightarrow \mathrm{ZnO} + \mathrm{H_2O}$$

At cathode,

$$\frac{1}{2}\mathrm{O}_2 + \mathrm{H_2O} + 2e \rightarrow 2\mathrm{OH}^{-} \qquad E^{\mathrm{o}} = 0.45\ \mathrm{V}$$

The overall reaction of Zn–air battery,

$$\mathrm{Zn} + \frac{1}{2}\mathrm{O}_2 \leftrightarrow \mathrm{ZnO} \qquad E^{\mathrm{o}} = 1.65\ \mathrm{V}$$

The output of the zinc–air battery is 1.65 V.

Advantages:

Zn–air battery technology offers the following advantages for many applications.

1. High energy density
2. Flat discharge voltage
3. Long shelf life
4. No ecological problems
5. Low cost
6. Capacity independent of load and temperature.

Applications:

Zn–air batteries have been most successfully employed as a power source for hearing aids. Other applications include electronic pagers, voice transmitters, portable battery chargers, various medical devices, and so on.

2. Nickel–metal hydride cell

(i) *Anode—A metal hydride, MH (*$\mathit{LaNi_5}$ *alloy grid)*
(ii) *Cathode—Nickel oxy hydroxide/Ni*
(iii) *Electrolyte—KOH*

A relatively new technology is adopted in the case of the chargeable sealed nickel–metal hydride battery with characteristics similar to those of the sealed Ni–Cd batteries. The Ni–MH battery uses *hydrogen* absorbed in a metal alloy for the active negative material whereas cadmium is used in the Ni–Cd battery and that makes the noticeable difference between the two.

A higher energy density can be achieved in the case of a metal hydride electrode than the cadmium electrode. Thus, a smaller amount of the negative electrode is used in the Ni–metal hydride. This allows for a larger volume for the +ve electrode, which results in a higher capacity or longer service life for the metal hydride battery. Moreover, as the metal hydride battery is free of Cd, it is considered more environmentally friendly than the Ni–Cd battery and may reduce the problems associated with the disposal of rechargeable nickel batteries.

Nickel–metal hydride batteries consist of a positive plate of a highly porous sintered or a felt nickel substrate impregnated with nickel hydroxide as its principal active material, a negative plate of a highly porous structure using a perforated $\mathrm{LaNi_5}$ alloy grid (a hydrogen-absorbing alloy). A synthetic nonwoven material separates the two electrodes, which serves as a medium for absorbing the electrolyte and a sealing plate provided with a self-resealing safety vent [Fig. 3.8(a)].

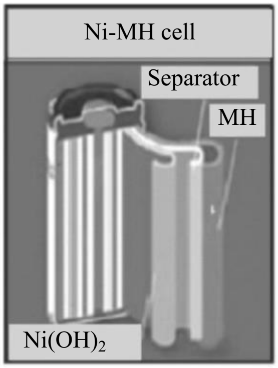


Figure 3.8(a) *Ni–MH cell.*

The chemistry of the electrode reactions of Ni-MH battery can be described as follows:

In the charged state of Ni–MH battery, nickel oxyhydroxide is the active material of the +ve electrode. This is same as the positive electrode in the Ni–Cd battery.

In the charged state of the Ni–MH battery, hydrogen is stored in a hydrogen-absorbing alloy as 'metal hydride' $LaNi_5$ (–ve active material). This metal alloy is capable of undergoing a reversible hydrogen absorbing-desorbing reaction as the battery is charged and discharged.

An aqueous solution of KOH is the major component of the electrolyte, with a minimum amount of the electrolyte absorbed by the separator and the electrodes. As can be seen from the overall reaction given below, the chief characteristics of the principle behind an Ni–MH battery is *that hydrogen moves from the +ve electrode to the –ve electrode during charge and reverse during the discharge, with the electrolyte taking no part in the reaction*; which means that there is no accompanying increase or decrease in the electrolyte.

The *discharge electrode reactions* of the Ni–MH battery are as follows:

At anode,

$$MH + OH^- \leftrightarrow M + H_2O + e \qquad E^o = 0.83\ V$$

At cathode.

The nickel oxyhydroxide is reduced to nickel hydroxide

$$NiOOH + H_2O + e \leftrightarrow Ni(OH)_2 + OH^- \quad E^o = 0.52\ V$$

The overall reaction on discharge is

$$MH + NiOOH \leftrightarrow M + Ni(OH)_2 \qquad E^o = 1.35\ V$$

The process is reversed during charge of the Ni–MH battery.

Advantages:

The following are the advantages of the Ni–MH battery:

1. High capacity
2. No maintenance required
3. Minimum environmental problem
4. Rapid recharging capability
5. Long cycle life
6. Long shelf life in state of charge.

Applications:

The nickel–metal hydride batteries are used in computers, cellular phones and other portable and consumer electronic applications where high specific energy is required.

3.1.9 Lithium Cells/Batteries

Lithium is a theoretically 'active material for negative electrode' of the electrochemical cells due to its least noble nature and low specific gravity.

(i) 'Primary cells' with metallic lithium electrodes and nonaqueous electrolytes were successfully introduced into the market. The outstanding features in comparison with conventional batteries with aqueous electrolytes are high voltage, high energy density (both volumetric and gravimetric energy densities are high), low self-discharging rate, and a wide range of operation.

(ii) Thus, 'secondary batteries' with metallic lithium negative electrodes have attracted much attention as a candidate for the battery with high energy density and much effort has been made in developing secondary lithium batteries. Many practical problems, however, have been encountered in development of rechargeable lithium batteries. Several among them are as follows:

1. Poor cycle life,
2. Need for long charging time, and
3. Poor safety characteristics.

Almost all these issues are due to the dendrite lithium formation during cycling. Dendrites are apt to penetrate into the separator and to cause an internal short circuit between +ve and −ve electrodes, which bring the problems including poor cycle performances and undesirable safety characteristics.

To overcome the weakness of metallic lithium −ve electrodes, some 'lithium storing' materials have been investigated as −ve electrode materials including Al, wood's alloy and carbonaceous materials and it has been confirmed that lithium/carbon alloy, 'lithium graphite intercalation' compounds being a typical one, is an excellent material for −ve electrode which inherits advantages of metallic lithium without the issue of lithium dendrite formation. Various investigations have been made on −ve and +ve electrodes, electrolytes and separators and others.

3.1.9.1 Primary Lithium Batteries

Lithium batteries use nonaqueous solvents for the electrolyte because of the reactivity of lithium in aqueous solutions.

Nonaqueous solvents such as acetonitrile, propylene carbonate (PC) and inorganic solvents such as thionyl chloride are typical. A complete solute is added to provide the necessary electrolyte conductivity.

Lithium batteries are classified into several categories, based on the type of electrolyte (or solvent) and cathode material used. They are

1. Soluble cathode cells
2. Solid cathode cells and
3. Solid electrolyte cells.

1. Soluble cathode cells

The cathodic materials used in these types of cells are liquid or gaseous such as SO_2, SO_2Cl_2, $SOCl_2$, that they dissolve in the electrolyte solvent. A reaction between the lithium anode and the cathode results in the formation of a protective thin film on the lithium anode.

Example:

Li/SO_2Cl_2 cell:

A lithium anode, a carbon cathode and the electrolyte/depolarizer of $LiAlCl_4$ in SO_2Cl_2 are used to construct lithium sulphuryl cell, which is similar to the $SOCl_2$ (thionyl chloride) cell. The discharge mechanism is
At anode,

$$2Li \rightarrow 2Li^+ + 2e$$

At cathode,

$$SO_2Cl_2 + 2e \rightarrow 2Cl^- + SO_2$$

The overall reaction

$$2Li + SO_2Cl_2 \rightarrow 2LiCl + SO_2$$

2. Solid cathode cells

(a) Lithium–manganese dioxide (Li–MnO_2) cell

The lithium–manganese dioxide is a primary lithium cell. It is available in many configurations (including coin, bobbin, prismatic) in multi-cell batteries and designs for low, moderate and high drain applications.

It has high cell voltage (3.0 V), energy density, good performance over a wide range of temperature, long shelf life and low cost.

The Li–MnO_2 battery uses lithium for the anode and a specially prepared heat treated form of manganese dioxide for the active cathode material and an electrolyte containing lithium salts in a mixed organic solvent (propylene carbonate and 1,2 dimethoxyethane).

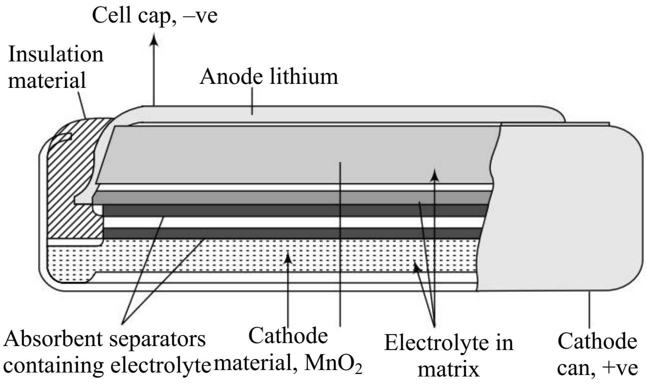


Figure 3.8(b) *$LiMnO_2$ cell.*

The chemistry of cell reactions are

At anode,

$$x.Li \rightarrow x.Li^+ + xe$$

At cathode,

$$Mn^{4+}O_2 + x.Li^+ + xe \rightarrow LiMn^{4+}_{1-x}Mn^{3+}_xO_2$$

Overall, $x.Li + Mn^{4+}O_2 \rightarrow LiMn^{4+}_{1-x}Mn^{3+}_xO_2$

The Mn^{4+} state is reduced to the Mn^{3+}state by the interstitially occupied lithium ions in the 'MnO_2 intercalation compound'. $LiMnO_2$ signifies the interstitial Li^+ ion in the host MnO_2 lattice [Fig. 3.8(b)]. It exhibits both electronic and ionic conductivity. The voltage of the cell is 3.5 V.

Characteristics:

High volumetric and gravimetric energy density, a wide operating temperature range (−20 to 55°C), performance at high discharge rates, relatively low cost.

Applications:

The lithium–manganese cell is used in a wide variety of applications such as long-term memory back up, safety and security devices, cameras, lighting equipment and many consumer electronic devices.

(b) **Lithium–vanadium pentoxide (Li–V_2O_5) battery**

Lithium–vanadium pentoxide battery is also a lithium primary battery. The Li–V_2O_5 battery is predominantly employed in 'reserve' type of battery configurations.

Lithium–vanadium pentoxide reserve battery consists of a lithium anode and a cathode that is composed of 90 percent V_2O_5 and 10 percent graphite on a weight basis. The electrolyte solution is 2M $LiAsF_6$ + 0.4M $LiBF_4$ in methyl formate (MF) because of its excellent stability during long-term storage and a microporous polypropylene film separator.

At anode,

$$Li \rightarrow Li^+ + e$$

At cathode,

$$Li^+ + V_2O_5 + e \rightarrow Li\ V^{5+}_{2-x}V^{4+}_{x}O_5$$

The lithium ions in the interstitial sites of the host V_2O_5 lattice reduce V^{5+} state to V^{4+} state. The voltage of the Li–V_2O_5 cell is 3.2 V.

This primary cell has high volumetric energy density, two-step discharge and used mainly in reserve batteries.

3. Solid electrolyte cells/batteries

Solid electrolyte lithium cells are known for their long storage life and are observed to have low discharge in the microampere range. They are used in applications, such as memory back up, cardiac pacemaker, and similar equipment, where current requirement is low but long life is most important.

Examples: Li/TiS_2, LiX/LiX-PEO, V_2O_5, Li/plasticized SPE/V_2O_5

Lithium Polymer electrolyte:

Thin film solid-state polymer electrolyte cells offer the possibility of an intrinsically safe battery design in combination with good high rate capability. In the construction of solid state batteries, polymer electrolyte materials are used. *The unique aspect of this battery is that the electrolyte is a solid flexible film composed of a polymer matrix [poly ethylene oxide (PEO)], and an ionic salt complexed into the matrix.* A wide variety of polymeric electrolytes has been used at an ambient normal temperature. The polymer electrolyte battery is based on thin film components that incorporate very large area electrolyte and electrode layers.

A typical cell is Li/LiCFSO, PEO/V_6O_{13} for which theoretical discharge reaction is

$$8Li + V_6O_{13} \rightarrow Li_8V_6O_{13}$$

3.9.2 Secondary (Rechargeable) Lithium Batteries

One of the basic concepts of investigation on rechargeable lithium battery is to develop batteries that have high energy density, power density, good cycle life and charge retention and to provide this high performance reliably and safely. In order to achieve the optimum balance for the construction of battery, it demands a compromise between the selection of cell components and the design of battery. The process is more complex for rechargeable batteries as the cell chemistry must be reversible and the reactions that occur during the charge affect all the characteristics and performance on subsequent cycling.

3.9.2.1 Classification of Rechargeable Lithium Batteries

A number of cathodic active materials (during discharge) such as intercalation compounds, soluble inorganic cathodes and polymeric material have been used lithium cells. Liquid aprotic and inorganic electrolytes are used in many cells. Solid polymer electrolytes are also popular as they may provide a safer design because of their lower reactivity with lithium. The secondary lithium batteries are classified into five categories.

1. Liquid organic electrolyte cells
2. Polymer electrolyte cells
3. Lithium ion cells
4. Inorganic electrolyte cells
5. Lithium alloy cells

In this chapter, the discussion of cells is confined to only the first three types of lithium batteries.

1. Liquid organic electrolyte cells

This type of solid cell uses intercalation compounds for cathode, a liquid organic electrolyte and a metallic lithium anode. *These cells use lithium metal for the negative electrode, a liquid aprotic for the electrolyte and transition metal compounds (oxides, sulphides and selenides) for the positive electrode.*

These transition metal compounds are insertion or intercalation compounds and possess a layered structure into which lithium ions can be inserted or from where they can be removed during discharge and charge, respectively.

$$x\text{Li} + \text{M}_y\text{B}_z \leftrightarrow \text{Li}_x\text{M}_y\text{B}_z$$

where M_yB_z is the transition metal compound (of the host lattice at the +ve electrode, $LiNiO_2$, MnO_2, V_2O_5).

The following materials have been used for the construction of these cells (Fig. 3.9):

(a) Anode—lithium metal.
(b) Cathode—($LiCoO_2$, $LiNiO_2$, MnO_2, V_2O_5).
(c) Electrolyte—$LiAsF_6$, PC/EC
(d) Separator—polypropylene.

Lithium ions are formed at the –ve electrode during the discharge, migrate through the electrolyte and are inserted into the crystal structure of the host lattice at the +ve electrode.

At anode,

$$x\text{Li} \rightarrow x\text{Li}^+ + xe$$

At cathode,

In the case of $LiCoO_2$ material at the cathode, the lithium insertion is

$$x\text{Li}^+ + x\text{e} + \text{LiCoO}_2 \rightarrow \text{Li}_x\text{CoO}_2$$

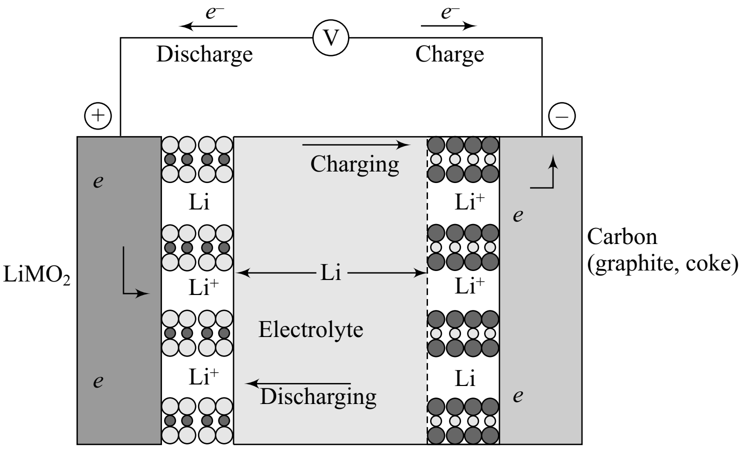


FIGURE 3.9 *A typical discharge/charge process of $LiCoO_2$.*

2. Polymer electrolyte cells

The cells are using and a lithium metal for the anode, polymer electrolyte and intercalation compounds for the cathode. The polymer electrolyte cell has high energy density, safer design, low electrolyte conductivity, poor high rate capability, poor low temperature performance, low self-discharge rate. Electronically conductive polymers are also used as cathode materials in rechargeable lithium batteries. The most popular are polyacetylene, polypyrrole, polyaniline and polythiophene, which are made conductive by doping with suitable anions.

An alternative to the liquid electrolytes is a solid polymer electrolyte (SPE) formed by incorporating lithium salts into polymer matrices and casting into thin films. These can be used as both the electrolyte and the separator. Initially, high molecular weight polymers such as $LiClO_4$, $LiN(CF_3SO_2)_2$ were used. Another class of polymer electrolyte is called gelled electrolyte-trapped $LiClO_4$ in polypropylene carbonate (PC) and ethylene carbonate (EC).The chargeable lithium cells which use a solid polymer electrolyte (SPE) are considered to have a safety advantage over the organic and inorganic liquid electrolytes because of their lower reactivity with lithium and the absence of the volatile organic solvent.

In most common form, as shown in Fig. 3.10, these cells use a lithium-conducting polymer membrane which acts both as the electrolyte and a separator, a thin lithium metal foil as the negative electrode material and transition metal oxide or chalcogenide such as V_2O_5, TiS_2, V_6O_{13}. Blended with carbon and the electrolyte and backed by a metal foil current collector as the +ve electrode.

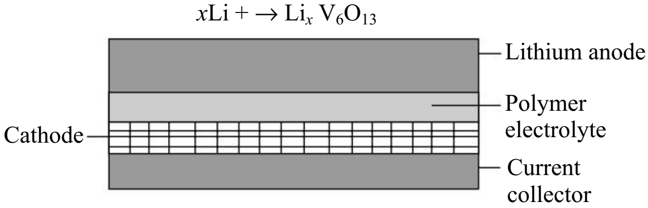


Figure 3.10 *Schematic cross-section of a solid polymer electrolyte cell.*

The basic structure can be represented as Li/PEO–LiX/M_yB_zPEo-LiX, C/M where M_yB_z is the intercalation, compound PEO-LiX, the polymer electrolyte and M the metal collector.

The cell reaction is similar to the liquid organic electrolyte cell-intercalation of lithium into the layered structure of the cathode during discharge and deintercalation of lithium ion from the charged cathode and deposition on the cathode during charge.

$$xLi + M_yB_z \rightarrow Li_xM_yB_z$$

Example, $Li/SPE/V_6O_{13}$cell,

$$xLi + \rightarrow Li_xV_6O_{13}$$

3. Lithium ion cells

Nowadays, successfully developed battery system is lithium ion battery (LIB) which has high volumetric and energy densities and is derived from lithium secondary batteries (LSBs) with metallic lithium anodes. Lithium secondary batteries have several drawbacks.

The cell reaction in LIB is merely the migration of lithium ions between +ve and –ve electrodes. No chemical changes were observed in the two electrodes or in the electrolytes. This results in chemical transformation of active electrode materials and electrolytes, and thus LIBs can overcome the weakness of LSBs. Graphite has layered structure that, as well known, can be intercalated or doped with lithium between the layers to form so-called graphite intercalation compound (GIC). It was confirmed that the electrochemical intercalation of lithium in an aprotic organic electrolyte containing complex lithium salts, such as $LiPF_6$, $LiBF_4$, $LiAsF_6$ and $LiClO_4$, reduces graphite to form Li-GIC and that Li-GIC can electrochemically oxidized by lithium deintercalation (undoping). This means that Li-GIC can be charged and discharged reversibly as rechargeable anodes for nonaqueous electrolyte cells. Lithium doping is 'charging' reaction and undoping reaction is discharging one. The principle of LiB is shown in Fig. 3.11. During charging of an LIB cell, lithium ions are extracted by electrochemical oxidation from $LiMnO_2$ cathode, and the extracted lithium ions are doped by electrochemical reduction into the carbon anode to form Li-GIC. Conversely, during discharging, lithium ions are extracted from the anode by electrochemical oxidation, and they are inserted into the cathode by electrochemical reduction.

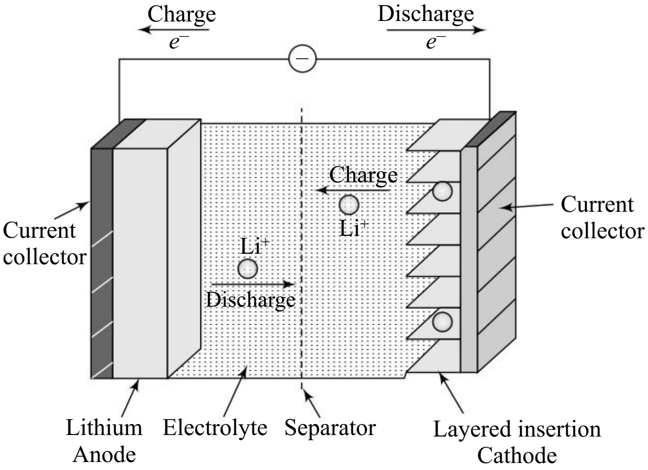


FIGURE 3.11 *A typical lithium ion cell working.*

The basic cell structure is then of the type

$$Li_xC_6/LiX \text{ in PC-PC}/Li_{1-x}MO_2$$

Where LiX is a lithium salt (such as $LiClO_4$), PC-EC is a mixed propylene carbonate-ethylene carbonate solvent and M is a transition metal.

Cell structure of lithium ion battery:

Copper and aluminium foils are used as the anode and cathode current collectors, respectively. $LiCoO_2$ powder was mixed with a polyvinylidene fluoride (PVDR) binder in an appropriate solvent to formulate paint for cathode active matrices and the paint was applied to both. The anodes were prepared in a similar way to the cathodes, namely, by mixing carbon powder and PVDF in a solvent to make a paste and then coating the paste on both surfaces of the copper foil (Fig. 3.12).

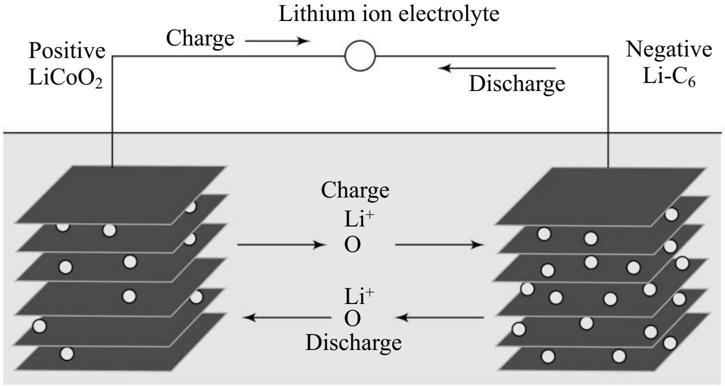


FIGURE 3.12 *Showing the typical structure of LIB.*

Microporous polythene film is used as a separator containing the electrolyte, lithium salts (LiX) in PC-Ec, which is a mixed propylene carbonate–ethylene carbonate solvent. The chemistry of LIB is described below:

The advantage of lithium ion cell is that it uses a lithiated carbon intercalation (LIC) material for the –ve electrode instead of metallic lithium. A lithiated transition metal intercalation compound is used for the +ve active material, and the electrolyte is aprotic organic (LiX in PC–EC solvent) solution.

The reactions at the electrodes and the overall cell reaction are

At anode,

$$\text{LiC} \leftrightarrow \text{C} + x\text{Li}^+ + \text{X}e$$

At cathode,

$$\text{LiMO}_2 + \text{XLi}^+ + \text{X}e \leftrightarrow \text{Li}_x\text{MO}_2$$

Overall reaction,

$$\text{LiC} + \text{LiMO}_2 \rightarrow \text{Li}_x\text{MO}_2 + \text{C}$$

Where Li_xMO_2 represents Li_xCoO_2, Li_xNiO_2 and Li_xMnO_2, the lithiated metal oxide compound. The lithium ions move back and forth between the +ve and –ve electrodes during the charge and discharge. The electrochemical process is the uptake of lithium ions at the –ve electrode during the charge and their release during the discharge, rather than lithium plating and stripping. The voltage of LIB is about 3.6–3.7V.

3.2 Fuel Cells

3.2.1 Introduction

Fuel cell technology offers the opportunity of creating environmentally friendly portable power supplies capable of producing enough energy to run devices and motor vehicles.

The electrochemical conversion of the free energy change of redox reactions at the electrodes into electrical energy is the working principle of any type of cell. The electrode reactions of primary batteries are irreversible and the cell produces EMF as long as the active materials are present in the cell, while the net cell reactions of secondary storage batteries are completely reversible and such batteries are rechargeable.

Contrast to these, a fuel cell, an electrochemical device, always operates with continuous replenishment of the fuel at the electrodes and consequently, it does not require any recharging.

A fuel cell is a device that converts the chemical energy of a fuel (hydrogen, natural gas, methanol, gasoline, etc.) and an oxidant (air or oxygen) into electricity.

A fuel cell also has two electrodes and an electrolyte. In the fuel cell device, fuel and oxidizing agents are continuously and separately fed into their respective electrodes, at which they undergo redox reactions generating electrical energy. A notable feature is that fuel cells produce electrical energy with continuous replenishment of the fuel at the electrode.

A fuel cell may be represented as *fuel/electrode/electrolyte/electrode/oxidant.*

Like any electrochemical cell,

At anode,

fuel undergoes oxidation, Fuel $\rightarrow$ oxidized product $+ne$

At cathode,

the oxidant gets reduced, Oxidant $+ne \rightarrow$ reduced product.

The electrical energy produced in a fuel cell corresponds to the free energy change of the electrode processes, as in the case any electrochemical cell.

$$-\Delta G = nEF$$

where $-\Delta G$ = the decrease of free energy of the fuel cell reaction,
n = the number of electrons involved,
E = EMF in volts, and
F = Faraday.

3.2.2 Merits of Fuel Cell

The fuel cells have merits over other kinds of electrochemical cells. They are as follows:

1. high efficiency of conversion of free energy of redox processes into electric energy.
2. silent operation
3. recharging of the fuel cell is not required
4. less attention is needed
5. harmless waste products and eco-friendly.

3.2.3 Classification of Fuel Cells

Fuel cells are classified on the basis of temperature of operation.

1. low temperature fuel cells (less than 100°C)
2. moderate temperature fuel cells (100–250°C) and
3. high temperature fuel cells (less than 500°C).

The electrode reactions of fuel cells are slow at low temperature and in the absence of certain specific catalysts. When current is drawn from the fuel cell, generally, the EMF decreases due to polarisation effect. In order to achieve better cell performance various kinds of fuel cells are designed with suitable electrodes and electrolytes to suit the desired needs.

3.2.4 Types of Fuel Cells

Fuel cells can be divided into *five major categories* named after the electrolyte used in each. The five types resulted from the knowledge that heat accelerates chemical reaction rates and thus the electrical current. The materials used as electrolytes have their best conductance only within certain temperature ranges and thus, other materials must be used in order to take advantage of the temperature increase.

1. *Alkaline fuel cells (AFC).*
2. *Phosphoric acid fuel cells (PAFC).*
3. *Molten carbonate fuel cells (MCFC).*
4. *Polymer electrolyte membrane fuel cells (PEMFC)*
5. *Solid oxide fuel cells (SOFC) and*
6. *Biochemical fuel cells (BCFC).*

Fuel cell technology offers the potential to excel today's motive power techniques in terms of environmental compatibility, consumer's profit, costs of maintenance and efficiency. A brief description of working of various types of fuel cells is mentioned below:

3.4.4.1 Alkaline Fuel Cell (AFC)

In alkaline fuel cells (AFCs) as shown in Fig. 3.13*, liquid electrolytes such as solutions of NaOH or KOH are used*. Pure oxygen or air (free from CO_2) is used as the oxidant. The fuels of the cells are *hydrogen or any hydrocarbons*. Low temperature AFCs operate at 60–80°C. At these temperatures, highly active catalysts (Pt family) are required. Silver and high surface Ni catalysts are also used. Cheaper catalysts normally require higher operating temperature. e.g. 200–250°C for the Ni catalyst. At these temperatures, either a high pressure or high concentration solutions must be used to prevent water loss.

The operating temperature of AFCs is about 70°C and their power output is 10–100 kW. They have been widely used for space and defence applications, where pure hydrogen is used. Their excessive cost and sensitivity to CO_2, have restricted their research and development, no matter their high efficiency and power density.

- Fuel: Hydrogen or any hydrocarbon
- Electrolyte: KOH
- Electrolyte state: immobilised liquid
- Electrodes: porous carbon or graphite
- Catalyst: platinum/rhodium; or Ni
- Charge carrier: OH^- ions
- Operating temp: 60–80°C
- Co-generating heat: low quality
- Fuel cell efficiency: 40–45 percent

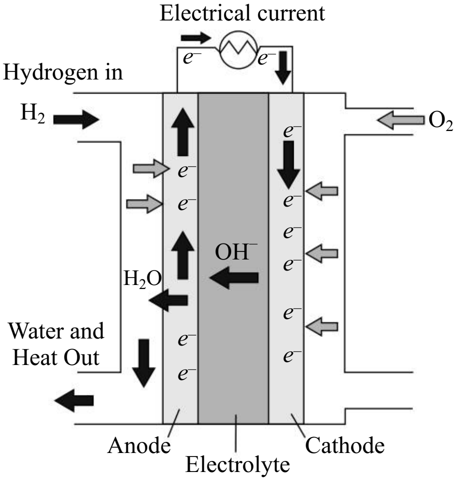


FIGURE 3.13 *A typical alkaline fuel cell.*

The chemistry of the electrode reactions of AFCs are as follows:

Anode reaction: $H_2 + 2OH^- \rightarrow 2H_2O + 2e$
Cathode reaction: $\frac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^-$
Overall reaction: $H_2 + \frac{1}{2}O_2 \rightarrow 2H_2O$

Applications:

AFC is widely used in the U.S. space program to produce electrical energy and water on-board spacecraft. Examples:

1. Hydrogen–oxygen alkaline fuel cell

A H_2–O_2 fuel cell device consists of two porous carbon electrodes as shown in Fig. 3.14. Anode is a porous carbon electrode coated with a nickel/platinum catalyst and the cathode of the porous carbon is impregnated with the Ag-catalyst to facilitate the desired electrode reaction. Hydrogen is flushed into the anodic compartment of the cell, where oxidation reaction occurs, while O_2, the oxidant, diffuses into the cathodic porous carbon containing silver catalyst.

Oxygen gets reduced to free OH-ions. The electrolyte is usually a warm solution of KOH, which is at the central compartment of the fuel cell.

The electrode reactions of the cell are

At anode,

$$2H_2 \rightarrow 4H^+ + 4e \text{ (oxidation)}$$

At cathode,

$$O_2 + 2H_2O + 4e \rightarrow 4OH^- \text{ (reduction)}$$

The net cell reaction is $2H_2(g) + O_2(g) \rightarrow 2H_2O(l)$

The H_2–O_2 fuel cell generates a voltage of 1.15 V, as long as the continuous supply of H_2 and O_2 is maintained.

2. Methyl alcohol–oxygen alkaline fuel cell

In this fuel cell, CH_3OH is used as a fuel and O_2 as an oxidant to generate electrical energy. Figure 3.15 shows a typical CH_3OH–O_2 fuel cell.

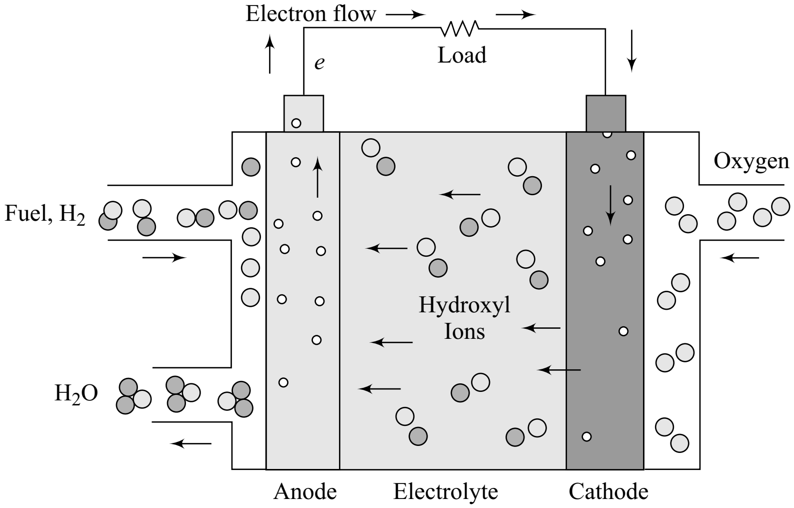


Figure 3.14 *The construction of a H_2–O_2 fuel cell.*

The methyl alcohol–oxygen fuel cell has two electrodes. The anode consists of a porous nickel electrode impregnated with a Pt/Pd catalyst. The porous nickel electrode coated with the silver catalyst constitutes a cathode of the cell. The electrolyte, KOH, is taken in between the two electrodes. CH_3OH and O_2 are sent continuously into their respective electrodes as shown in Fig. 3.15(a) and the electrical energy is produced with the continuous replenishment of the fuel, CH_3OH at the anode.

At anode,

$$CH_3OH + 6OH^- \rightarrow CO_2 + 5H_2O + 6e$$

At cathode,

$$3/2O_2 + 3H_2O + 6e \rightarrow 6OH^-$$

Overall reaction,

$$CH_3OH + 3/2O_2 \rightarrow CO_2 + 2H_2O$$

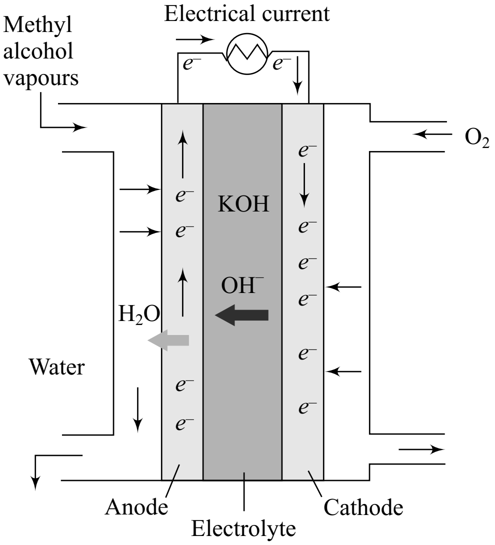


Figure 3.15(a) *A typical CH_3OH–oxygen fuel cell.*

3.4.4.2 Phosphoric Acid Fuel Cell (PAFC)

- *Fuel: Hydrocarbons.*
- *Electrolyte: phosphoric acid*
- *Electrolyte state: immobilized liquid*
- *Electrodes: porous carbon or graphite*
- *Catalysts: platinum/rhodium/gold/tantalum*
- *Charge carrier: H^+ ions*
- *Operating temp: 200°C*
- *Co-generating heat: Low quality*
- *Fuel cell efficiency: 40–45 percent*

The design of the phosphoric acid fuel cells is similar to AFCs [Fig. 3.15(b)] except that the phosphoric acid is used as the electrolyte in the case of hydrocarbon fuels.

The acid electrolyte cells are more tolerant to CO_2 and allow the use of *normal air and impure hydrogen.* But the corrosion problem restricts the choice of construction materials especially for the electrodes and catalysts. *The electrodes can be made out of gold, tantalum, titanium and carbon and only platinum group metals can be used as catalysts.* The phosphoric acid electrolyte is used in the cell so that only water is lost by evaporation. The electrolyte in the PAFC is a paper matrix saturated with phosphoric acid, transporting the hydrogen ions. The operating temperature is around 200°C.

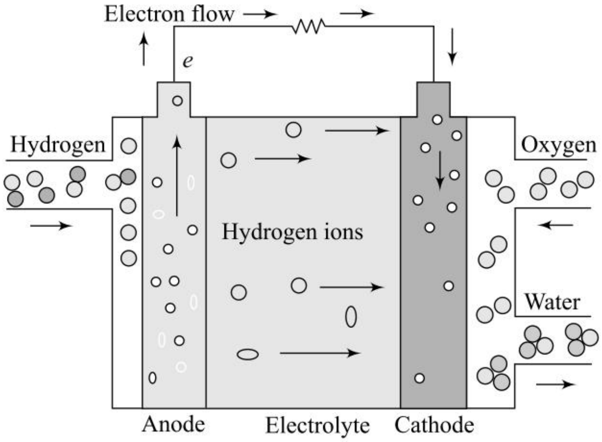


Figure 3.15(b) *A typical phosphoric acid fuel cell.*

PAFCs have an operating temperature of 200°C. The efficiency of this system is much lower than that of the other systems at 40 percent. It is the FC that has mostly been exploited, mainly due to its high grade heat, which can be used in small-scale CHP especially at military sites and UPS systems fuelled with hydrogen, natural gas, LPG and methane from waste water purification plants. The power output varies from 200 kW to 20 MW. The main disadvantage is that it has no self-starting capability, because at lower temperatures (40–50°C) freezing of concentrated phosphoric acid occurs. In order to reduce losses, the cathode catalyst and the reformer need to be improved.

The fuel cell reactions are

Anode reaction: $H_2 + 2OH^- \rightarrow 2H_2O + 2e$
Cathode reaction: $\frac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^-$
Overall reaction: $H_2 + \frac{1}{2}O_2 \rightarrow 2H_2O$

Applications:

PAFC is typically used for stationary power generation, but some PAFCs have been used to power large vehicles such as city buses.

3.4.4.3 Molten Carbonate Fuel Cell (MCFC)

It is known that the rate of electrode reactions and the electrical conductivity of the electrolyte increases appreciably at high temperatures. Majority of the fuel cells use fairly high concentration of aqueous electrolytes to minimize the polarization effects. At high temperature, loss of water by evaporation could be there for aqueous electrolytes. In order to overcome these difficulties, fused carbonates (or a eutectic mixture of lithium carbonate and sodium carbonate or potassium carbonate in $LiAlO_2$ matrix) are used as electrolytes (Fig. 3.16). The high temperature not only improves the electrical conductivity of the electrolyte but also increases the rate of electrode reactions.

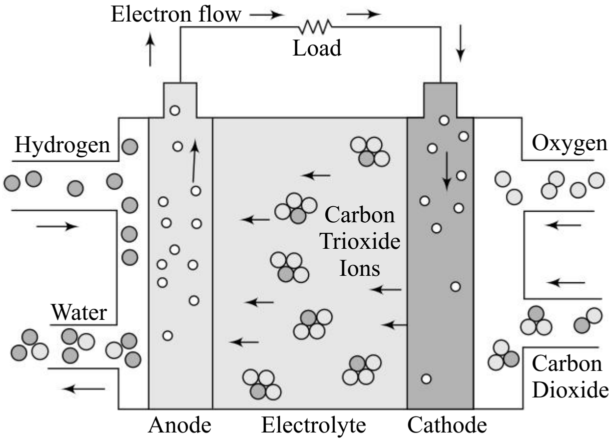


Figure 3.16 *A typical MCFC cell.*

- *Fuel: hydrogen*
- *Electrolyte: mixture of alkali carbonates*
- *Electrolyte state: immobilised liquid*
- *Electrodes: (i) Anode: porous Ni powder alloyed with Cr*
  - *(ii) Cathode: porous NiO doped with Li*
- *Catalysts: nickel oxide*
- *Charge carrier:* CO_3^{2-} *ions*
- *Operating temp: 650°C*
- *Co-generating heat: High*
- *Fuel cell efficiency: 50–60 percent*

They operate at 600°C and can use CO_2 as input on the cathode side but need hydrogen on the anode. The temperature is high enough to be used for additional power production through cogeneration of steam. The efficiency of these types of fuel cells has risen to 50 percent in a combined (electrical and steam) cycle. They can also be used in mega-watt size power plants because of their heat.

The electrolyte in this fuel cell is usually a combination of molten alkali (Na, K, Li) carbonates, the composition of the electrolyte varies, but typically consists of lithium carbonate and potassium carbonate. The electrolyte is suspended in a porous, insulating and chemically inert lithium aluminium oxide ($LiAlO_2$) ceramic matrix. The anode is a highly porous-sintered nickel powder, alloyed with chromium to prevent agglomeration and creep at operating temperatures. The cathode is a porous nickel oxide material doped with lithium. At the operating temperature of about 650°C, the salt mixture is liquid and a good ionic conductor.

They operate with NiO as a catalyst material (cathode) at a temperature of 600°C where CO_2 reacts with the O_2 and electrons to form CO_3^{2-} ions, which carry the ionic current through the electrolyte. At the anode, these ions are consumed in the oxidation of H_2 (fuel), which also forms water vapour and CO_2 to be transferred back to the cathode.

The anodic reaction of the MCFC is

$$H_2 + CO_3^{2-} \longrightarrow H_2O + CO_2 + 2e$$

The cathodic reaction is

$$\tfrac{1}{2}O_2 + CO_2 + 2e \longrightarrow CO_3^{2-}$$

Overall reaction of MCFC is

$$H_2 + \tfrac{1}{2}O_2 \longrightarrow H_2O$$

Applications:

MCFCs are currently being developed for natural gas and coal-based power plants for electrical utility, industrial and military applications.

3.4.4.4 Polymer Electrolyte Membrane Fuel Cell (PEMFC)

Steadily increasing requirements for portable electric power have stimulated the interest to develop more efficient and more energetic fuel cells. *A polymer membrane fuel cell (PEMFC) is one of the best candidates as a portable power source for commercial applications primarily because of its light weight, high energy, high power, nonemission and low temperature operation***.** The PEMFCs with 'Nafion' membrane as an electrolyte have received much attention. Nafion is a perfluorinated cation exchange polymer membrane. It has shown good conductivity and stability up to 100°C under the fuel cell operating environmental conditions.

- *Fuel: Hydrogen*
- *Electrolyte: Ion exchange polymer membrane (Nafion)*

- *Electrolyte state: Solid*
- *Electrodes: Carbon or metal based*
- *Catalysts: Platinum*
- *Charge carrier: H^+ ions*
- *Operating temp: 80°C*
- *Co-generating heat: None*
- *Fuel cell efficiency: 60 percent*

PEM fuel cells operate at around 80°C and a practical efficiency of 60 percent. Power output is in the range of 5–200 kW. They are ideal for transportation and portable power. Additional advantages are their high response, the small size and low cost.

The PEMFC is unusual in that its electrolyte consists of a layer of solid polymer which allows protons to be transmitted from one face to the other. Porous carbon is used as an anode and a cathode of the PEMFC as shown in Fig. 3.17.

It basically requires H_2 and O_2 as its inputs, though the oxidant may also be ambient air, and these gases must be humidified. It operates at low temperature because the limitations imposed by the thermal properties of the membrane itself. The operating temperatures are around 90°C and a Pt catalyst is used.

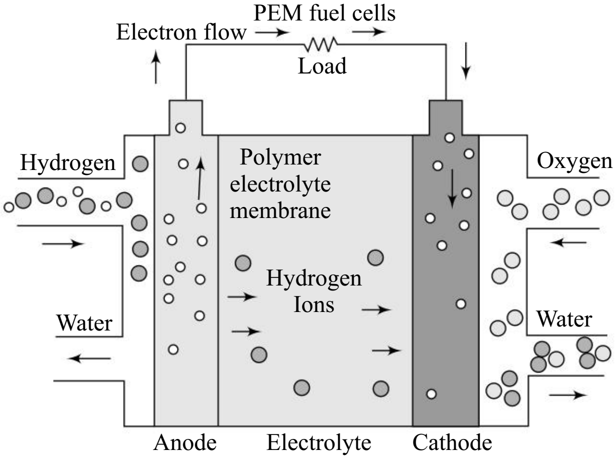


FIGURE 3.17 *A typical PEMFC.*

The PEMFC electrode reactions are

$$\text{Anodereaction: } H_2 \rightarrow 2H^+ + 2e$$
$$\text{Cathode reaction: } ½O_2 + 2H^+ + 2e \rightarrow H_2O$$

Overall reaction of PEMFC is

$$H_2 + ½O_2 \rightarrow H_2O$$

Most recently, more and more PEMFC stacks (multi-cells) were developed with a variety of types and functions. *The PEMFC stack has much higher operating voltage and stronger power and better fuel-energy efficiency*.

Applications:

PEM fuel cells are used primarily for transportation applications and some stationary applications. Due to their fast start-up time, low sensitivity to orientation, and favourable power-to-weight ratio, PEM fuel cells are particularly suitable for use in passenger vehicles, such as cars and buses.

3.4.4.5 Solid Oxide Fuel Cell

High temperature solid oxide fuel cells (SOFC) offer a clean, pollution-free technology to electrochemically generate electricity at high efficiencies. These fuel cells provide many advantages over traditional energy conversion systems including high efficiency, reliability, fuel adaptability and very low levels of NO_x and

SO_x emissions. The high oxygen–ion conductivity over a wide range of temperature and oxygen pressure in stabilized cubic zirconia has led to its use as a solid–oxide electrolyte in a variety of electrochemical applications. Zirconia-based sensors are widely used for high temperature solid–oxide fuel cells (SOFCs).

- *Fuel: hydrogen*
- *Electrolyte: yettria stabilized zirconia*
- *Electrolyte state: solid*
- *Electrodes: carbon or metal based*
- *Catalysts: ceramic*
- *Charge carrier: O^{2-} ions*
- *Operating temp: 1000°C*
- *Co-generating heat: high*
- *Fuel cell efficiency: 50–60 percent*

They operate at high temperatures (1000–1100°C) and a practical efficiency of 50–60 percent. They are not the most reactive because of the low conductivity of its ionic conducting electrolyte (yettria-stabilised zirconia). Because of the conductivity and heat, they have been used in large power plants, which can use the cogeneration of steam for additional power.

Operating principle:

SOFCs consist of two porous electrodes separated by a dense, oxygen-ion conducting electrolyte. The operating principle of such a cell is illustrated in Fig. 3.18. Current technology employs several ceramic materials for the active fuel cell components. The anode is typically constructed from an electronically conducting Ni/yettria-stabilized zirconia cermets (Ni/YSZ). The cathode is based on a mixed conducting perovskite, lanthanum manganate ($LaMnO_3$). Yettria-stabilized zirconia (YSZ) is used as the oxygen-conducting electrolyte. To generate a suitable voltage, fuel cells in the same stack are interconnected with a lanthanum manganate ($MnMnO_3$) or doped lanthanum chromate ($La_{0.8}Ca_{0.2}CrO_3$) joining the anodes and cathodes of adjacent units.

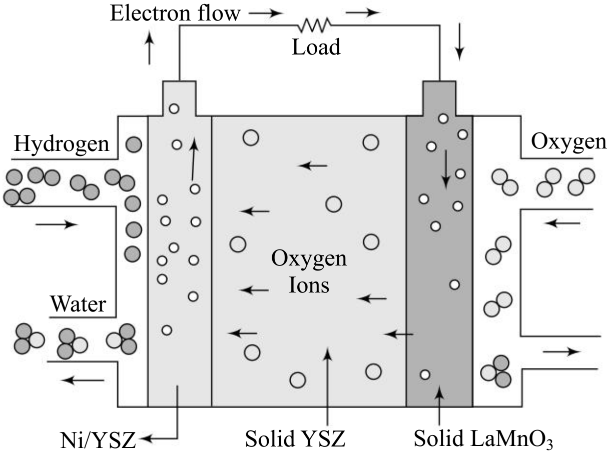


FIGURE 3.18 *Operating principle of a solid oxide fuel cell (SOFC).*

Oxygen supplied at the cathode (air electrode) reacts with incoming electrons from the external circuit to form oxygen ions, which migrate to the anode (fuel electrode) through the oxygen-ion conducting electrolyte. At the anode, oxygen ions combine with hydrogen in the fuel to form water, liberating electrons. Electrons flow from the anode through the external circuit to the cathode.

(a) Cathode

The air electrode operates in an oxidizing environment of air or oxygen at about 1000°C and *participates in the oxygen reduction reaction.*

$$\frac{1}{2}O_2 + 2e \rightarrow O^{2-}$$

that is, oxygen in the gas phase is reduced to oxide ions, consuming two electrons in the process.

(b) Anode

The fuel electrode must be stable in the reducing environment of the fuel. It should be electronically conducting, and must have sufficient porosity to allow the transport of the fuel away from the electrolyte/fuel electrode interface where *the fuel oxidation reaction takes place*, that is

$$O^{2-} + H_2 \rightarrow H_2O + 2e$$

The reducing conditions present on the fuel side of an SOFC permit the use of yettrium-stabilized zirconia (YSZ) as the fuel electrode.

(c) Electrolyte

Solid oxide fuel cells are based on the concept of an oxygen ion conducting electrolyte through which the oxide ions (O^{2-}) migrate from the air electrode (cathode) side to the fuel electrode (anode) side where they react with the fuel hydrogen to generate an electrical voltage. Fluorite structured oxide materials such as yettria stabilized zirconia (YSZ), rare earth doped ceria, and rare earth doped bismuth oxide have been widely used as electrolytes for fuel cells.

High temperature SOFCs provide a highly efficient, pollution-free power generation technology. Atmospheric and pressurized SOFC-based power systems are ideal distributed power generation systems— reliable, clean, environmentally friendly and fuel conserving.

Applications:

SOFCs are now being considered for large power plants and for industrial applications because of its electrolytic resistance to poisoning which allows internal reforming of many carbon compounds into hydrogen to create power.

Review Questions

Battery Technology

1. Distinguish between a cell and a battery.
2. What is the working principle of a battery?
3. What are the basic concepts of a battery?
4. Why a battery is known as commercial battery?
5. Mention the few characteristics of a battery.
6. Give an expression for EMF of a battery when the net cell reaction is

$$M_1 + M_2^{n+} \rightarrow M_1^{n+} + M_2$$

7. What are the different types of cells/batteries? Give an example in each case.
8. What is a primary battery? Mention few examples.
9. What is a secondary battery? Give few examples.
10. What is a reserve battery? Give an example.
11. Describe the working principle, construction and chemistry of $Zn–MnO_2$ cell.
12. What are the limitations of primary cells?
13. Mention the cell performance of primary cell. Give few uses of primary cells.
14. Describe the working principle, construction and chemistry of RM cell.
15. Describe the working principle, construction and chemistry of $Zn–Hg_2O$ cell.
16. What is the working principle of a secondary battery? Mention the types of secondary batteries available in the market.
17. Give the construction and the working of a lead acid storage battery.
18. Describe lead–acid battery. Give the electrode reactions during a discharge process.
19. Discuss the charging electrode reactions of a lead storage battery.
20. Give the cell performance of secondary lead–acid battery.
21. Describe the construction and the chemistry of working of a nickel–cadmium storage battery.

22. Describe a Zn/air cell. Mention its electrode reactions.
23. What are the advantages and applications of Zn/air battery?
24. Write a note on nickel-metal hydride battery? Why is it preferred to Ni–Cd battery?
25. Describe the working of an Ni–MH battery. Give the electrode reactions during its discharge.
26. Mention the advantages of Ni–MH battery over the conventional batteries.
27. What are lithium batteries? Mention few outstanding features in comparison with conventional batteries.
28. How are lithium batteries classified? Name the different categories of primary lithium batteries?
29. Write a note on the working of lithium-vanadium pentoxide reserve battery.
30. What are secondary lithium batteries? What are the advantages?
31. Write notes on the following primary lithium cells:
    (a) soluble cathode lithium cell
    (b) solid cathode lithium cell
    (c) solid electrolyte lithium cell
32. Mention the components of lithium–MnO_2 cell. Discuss the chemistry of working of this cell.
33. Write a note on liquid organic electrolyte secondary cell.
34. Write a note on polymer electrolyte secondary cell
35. Describe a lithium ion cell and write the electrode reactions.
36. Give the cell structure and working of a lithium ion cell with an example.
37. Describe the construction and working of zinc–air battery. [VTU June/July, 2015]
38. (a) What are batteries? How are they classified?
    (b) Describe the construction and working of Lead acid battery. Mention any two applications of Li–MnO_2 battery. [VTU June/July 2016]
39. Explain the construction and working of Zn–air cell. [VTU, CBSC, 2016]
40. Explain the construction, working and application of Ni-metal hydride battery. [VTU June/July, 2016]

Fuel cell

1. What is a fuel cell? How does it differ from other type of cells?
2. Mention few merits of a fuel cell.
3. What is the working principle of a fuel cell?
4. What are most commonly used fuel cells? Name each type.
5. What is an alkaline fuel (AFC) cell? Mention an example.
6. What is a phosphoric acid fuel cell?
7. What is a molten carbonate fuel (MCFC) cell? Why is it preferred to AFC? What are their electrode reactions?
8. Give a brief description of the design of a polymer electrolyte membrane (PEMFC) cell.
9. Describe the construction and working of a solid oxide fuel (SOFC) cell.
10. What is a biochemical cell?
11. Describe the working of H_2–O_2 fuel cell.
12. Describe the working of CH_3OH.
13. Describe the construction and working of Methanol-oxygen fuel cell. Mention any two applications. [VTU June/July, 2015]

4

Corrosion Science

Chapter Outline

Introduction to corrosion:

Introduction to corrosion. Types of corrosion—dry corrosion, wet corrosion. Electrochemical theory of corrosion—liberation of hydrogen, absorption of oxygen, neutral and alkaline medium. Forms of wet corrosion—differential metal (or galvanic) corrosion, concentration corrosion cell—differential aeration corrosion, metal ion concentration cell, pitting corrosion, water-line corrosion, stress cracking corrosion, intergranular corrosion, leaching corrosion, fatigue corrosion erosion corrosion. Galvanic series. Factors affecting corrosion—metallic factors, environmental factors, nature of the corrosion products. Galvanic series. Electrochemical series versus Galvanic series. Corrosion control—proper selection and design of materials. Protective coatings. Metal coating—anodic metal coating, cathodic coating. Techniques of metal coating—hot dipping, cladding, spraying, cementation, electroplating, inorganic coating. Anodization of aluminium, anodization in sulphuric acid. Phosphate coating, organic coating. Cathodic protection—sacrificial anodic protection, impressed current cathodic protection. Anodic protection. Corrosion inhibitors—-anodic inhibitors, cathodic inhibitors. Passivity of metals, polarization curves for passivating metals and the effect of pH and potential on corrosion and Pourbaix diagram of iron and applications.

4.1 Introduction to Corrosion

A wide variety of metals/alloys are being used in daily life, equipment, plant machineries, industrial structures, bridges, buildings, etc., and all these have a natural tendency to get tarnished or deteriorated by the environment to which they are exposed. The best example of the process of corrosion can be seen in the case of 'rusting of iron', in which the metal is deteriorated due to corrosive environment.

It is observed that few metals like Al, Cr, Ni, etc., may be tarnished due to corrosion or even get passivated gradually due to the surface layer oxidation of these metals in O_2/air environment. Therefore, it is evident, when metals are exposed to air, water or soil, they suffer chemical/electro-chemical attack under these conditions. Thus, the process of deterioration of such metals is known as corrosion.

Corrosion can be defined as the destruction or deterioration of metals by the surrounding environment through chemical or electrochemical changes.

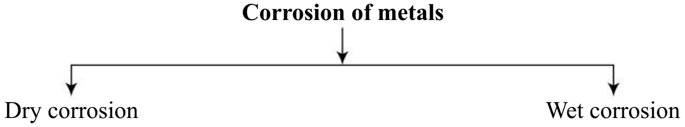
Why Do Metals Corrode?

Naturally occurring compounds of metals are in thermodynamically stable state (least energy state), while the extracted metals from these ores have high-energy state because of the taking up of large amount of energy during its metallurgy. The metals, thus, obtained when exposed to natural environment will have a tendency to go to their stable lower energy state whenever an opportunity arises.

Today, unprotected environment due to rapid industrialization has led to increased pollution of atmosphere and water. Such a created situation has marked effect on metals and alloys; and obviously, has offered a challenge to *Materials Engineers* to plan and design materials to protect them from corrosion. In view of the above, the understanding of the concepts of corrosion process of metals is, therefore, highly desirable.

4.2 Types of Corrosion

The process of corrosion of metals depends on the environment to which the metals are exposed. The corrosion of metals can be either (1) dry corrosion or (2) wet corrosion.



4.2.1 Dry Corrosion or Chemical Corrosion or Direct Chemical Attack

If a metal is exposed to dry environment, the metal gets corroded due to the direct attack of O_2 or corrosive gases like SO_2,Cl_2 or chemical action on metals through chemical reactions.

This is, generally, known as direct chemical corrosion or direct chemical attack. A tarnished surface of a metal is the best example of dry corrosion. Metals like Al, Zn and even Fe, etc., have a natural tendency to get oxidized and form a thin film of their oxides on the surface of the metal.

Dry corrosion may be due to

(i) Oxidation corrosion
(ii) Corrosion by gases
(iii) Liquid metal corrosion.

(i) Oxidation corrosion

Metals are attacked by dry oxygen (absence of moisture) at extreme conditions like low or high temperature leading to oxidation type of corrosion. The alkali and alkaline earth metals are more prone and rapidly oxidized at low temperatures; while metals like Fe, Al, etc. (except noble metals), are oxidized at high temperatures by the interaction with oxygen.

Examples,

(a)
$$\begin{array}{l} Mg \rightarrow Mg^{2+} + 2e \\ \frac{1}{2}O_2 + 2e \rightarrow O^{2-} \\ \hline Mg + \frac{1}{2}O_2 \rightarrow MgO \end{array}$$

(b)
$$\begin{array}{ll} 2Al \rightarrow 2Al^{3+} + 6e & \text{(oxidation)} \\ [\frac{1}{2}O_2 + 2e \rightarrow O^{2-}] \times 3 & \text{(reduction)} \\ \hline 2Al + 3/2O_2 \rightarrow Al_2O_3 & \end{array}$$

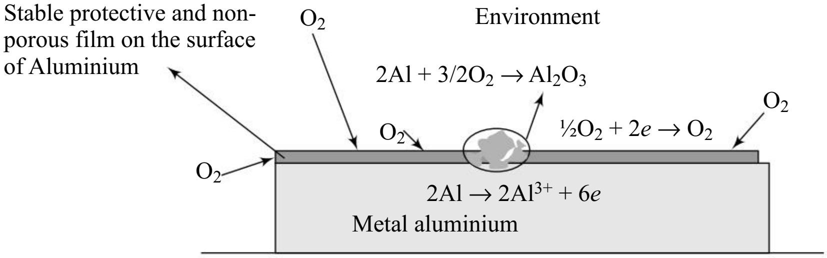


FIGURE 4.1 *Dry corrosion of aluminium metal by oxidative process.*

It is observed that thin film metal oxide layers are formed on the surface of the metals (Fig. 4.1). The nature of tarnished layer or the thin film layer formed on the metal surface decides the further action with the environment.

1. *Stable oxide film* formed on the surface of the metal is protective and shields the metal from further corrosion, since the oxide film is rigidly adhered to the surface and impervious in nature.
2. *Unstable oxide film* formed decomposes back into the metal and oxygen.

For noble metals like Pt, Au, etc., corrosion is not possible because the oxide layer formed is very unstable and decomposes back to metal and oxygen.

(ii) Corrosion by gases

In the absence of moisture of atmosphere, few gases like SO_2, Cl_2, F_2 attack the metal by dry corrosion and the degree of corrosion depends on the formation of thin film which may be *protective* and *non-protective* on the surface of the metal.

(a) Dry Cl_2 attacks silver metal and forms AgCl as a thin protective and non-porous layer on the metal. As a result of this protective layer on the metal surface, the intensity of corrosion decreases.

(b) However, it is observed in the case of Tin, the corrosion attack increases and the entire metal gets destroyed in due course of time. This is due to the fact that the thin layer ($SnCl_4$) formed on the tin surface is non-protective and porous.

(iii) Liquid metal corrosion

This type of corrosion occurs when a molten liquid is continuously passed on a solid metal surface or an alloy. This behaviour may be due to the following (a) dissolution of the molten liquid or (b) penetration of molten liquid into solid metal phase.

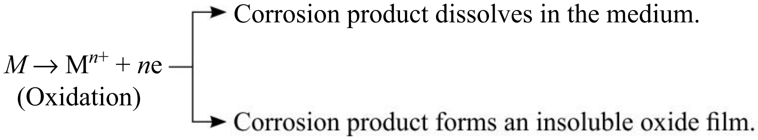
4.2.2 Wet Corrosion or Electrochemical Corrosion or Immersed Corrosion

Wet corrosion of metals is generally a common type of corrosion, where corrosion of metal takes place in an aqueous corrosive environment. This type of corrosion can be observed when

(a) a metal is in contact with an acid solution, and
(b) dissimilar metals are dipped partially in a corrosive environment.

It is observed that any impact of corrosive environment on a portion of a metal changes its electrode potential compared to the original value of electrode potential of the metal. This creates a *potential difference* within the metal system. It should be apparent that the potential difference created between the areas within a cell is a measure of the tendency for oxidation and reduction to take place—the more positive the cell potential created, the greater the tendency for the redox reaction to proceed spontaneously. A galvanic corrosion cell is created in the metal. Area of lower reduction potential created in the metal is known as the *anodic area* and that with higher reduction potential value is called the *cathodic area*. Therefore, process of corrosion takes place due to the formation of separate anodic and cathodic areas within the metal, and the corrosion current flows from anodic area to the cathodic area. Corrosion of the metal occurs always at the anodic area due to the oxidation process.

At the anodic area, *corrosion* starts due to the oxidation reaction.



corrosion product forms an insoluble oxide film.

The reduction process occurs at the cathodic area of the metal and, thus, *protects* it from corrosion.

At the cathodic area,

$$M^{n+} + ne \rightarrow M \qquad \text{(Reduction)}$$

4.3 Electrochemical Theory of Corrosion

Corrosion occurs when metals are exposed to air, water or soil and which appears to be an indispensable electrochemical process. A knowledge of the understanding of the corrosion process is, thus, quite essential. The problems related to corrosion of metals can be best explained on the basis of *electrochemical theory*.

According to the electrochemical theory of corrosion, when a metal (or dissimilar metals) is exposed to an acidic environment medium, the process of corrosion sets in by the formation of separate ‘anodic’ and ‘cathodic’ areas within the metal surface.

A driving force is necessary for electrons to flow between the anodes and the cathodes. This driving force is the difference in potential between the anodic and cathodic areas. This difference exists because each oxidation or reduction reaction has associated with it a potential determined by the tendency for the reaction to take place spontaneously. The potential is a measure of this tendency.

Corrosion always occurs at the ‘anodic area’ of the metal due to the oxidation process; and thus, electrons are liberated. The metal ions [M^{n+}] formed during destruction of metal either dissolves in the medium or forms a thin film of oxide on the metal surface

$$M \rightarrow M^{n+} + ne \qquad \text{(oxidation)}$$

On the other hand, ‘cathodic area’ of the metal is ‘protected’ as a result of the reduction reaction,

$$M^{n+} + ne \rightarrow M \qquad \text{(reduction)}$$

and the electrons set free at the anodic area flow through and are consumed at the cathodic area by the following processes:

(a) Liberation of H_2
(b) Absorption of oxygen

During corrosion, corrosion current flows from anodic to cathodic area of the metal.

4.3.1 Evolution of Hydrogen

(a) The evolution of H_2 type of corrosion (Fig. 4.2) is best illustrated by the action of very dilute acid solution on iron.

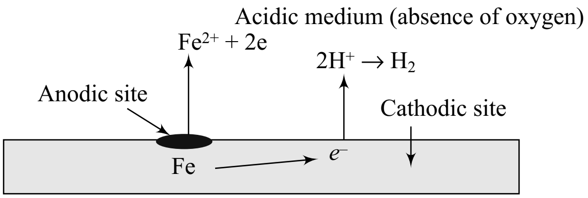


FIGURE 4.2 *Evolution of hydrogen type corrosion*

The process of dissolution of iron as Fe^{2+} ions takes place at the anodic area of the metal when exposed to an acid medium with the liberation of electrons due to oxidation process.

$$Fe \rightarrow Fe^{2+} + 2e \qquad \text{(oxidation)}$$

The liberated electrons flow from the anodic to the cathodic area of the same metal and give rise to the corrosion current. The electrons, thus, liberated are available for reduction of H^+ ions at the cathodic area. Thus, in acidic medium (absence of oxygen) hydrogen ions acquire electrons with the liberation of H_2 gas in cathodic reaction.

$$2H^+ + 2e \rightarrow H_2\uparrow \qquad \text{(reduction)}$$

4.3.2 Absorption of Oxygen

Rusting of iron: Rusting of iron takes place as follows. At the anodic area of the metal, iron dissolves as Fe^{2+} ions with the liberation of electrons.

$$Fe \rightarrow Fe^{2+} + 2e \qquad \text{(oxidation)}$$

The anodic area gets corroded. The liberated electrons move from the anodic area to the cathodic area where it is available for the reduction reactions with the absorption of oxygen and forms OH^- ions as shown in Fig. 4.3.

$$\frac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^- \qquad \text{(reduction)}$$

$Fe(OH)_2$ is precipitated on the surface of the metal due to diffusion of both Fe^{2+} and OH^- ions.

$$Fe^{2+} + 2OH^- \rightarrow Fe(OH)_2\downarrow$$

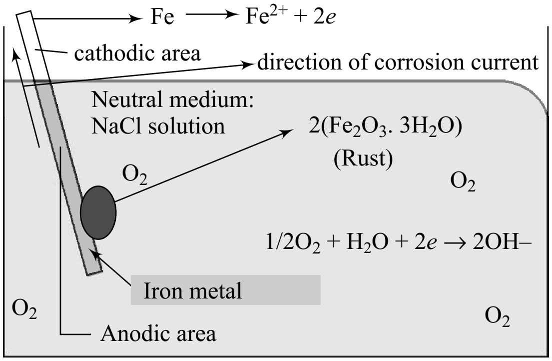


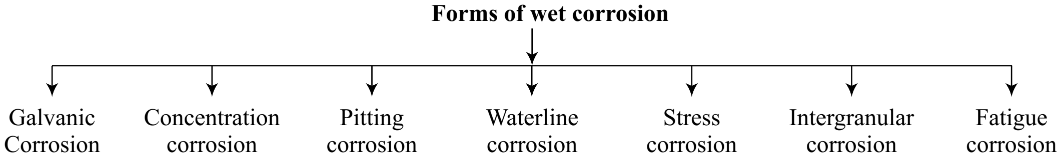
FIGURE 4.3 *Rusting of iron in neutral medium in the presence of oxygen.*

If the environment to which the metal exposed is rich in oxygen, $Fe(OH)_2$ formed gradually gets oxidized to $Fe(OH)_3$. The net corrosion reaction due to rusting of iron is, thus,

$$4Fe + 3O_2 + 6H_2O \longrightarrow \underset{\text{Rust}}{2(Fe_2O_3 3H_2O)} \downarrow$$

4.4 Different Forms of Wet Corrosion

Almost all corrosion problems and failures encountered in service can be associated with one or more of the basic forms of corrosion.



4.4.1 Galvanic Corrosion (or Differential Metal Corrosion)

When dissimilar metals (different potentials) like zinc and copper are electrically connected and are exposed to a neutral environment in the presence of oxygen, where the metal of lower reduction potential (or higher up) in the electrochemical series undergoes spontaneous oxidation and corrosion.

The mechanism of differential metal corrosion is illustrated in Fig. 4.4. In this, zinc is in electrical contact with copper and is immersed in a corrosive environment (soil). Zinc metal, being higher up in the electrochemical series acts as *anodic area* and gets corroded as shown in Fig. 4.4, whereas, the copper metal which is lower down in the series (+ve reduction potential) behaves as *cathodic area* of the dissimilar metals. The tendency of a metal to corrode in a galvanic cell is determined by its position in the 'galvanic series' of metals and alloys. As a result of this, corrosion current is pushed from the anodic area (Zn) to the cathodic area (Cu) by the driving force which is due to the potential difference between the two.

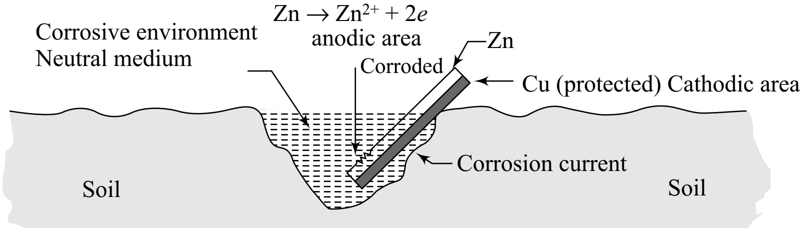


FIGURE 4.4 *Differential metal corrosion.*

The corrosion of zinc can be summed up as follows:

At the anodic area,

$$Zn \rightarrow Zn^{2+} + 2e \qquad \text{(oxidation and corrosion)}$$

At the cathodic area, we have

$$\tfrac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^- \qquad \text{(reduction and protection)}$$

Overall reaction is the formation of $Zn(OH)_2$.

In an acid medium, corrosion of Zinc occurs with the evolution of hydrogen and the formation of zinc salt,

$$Zn + 2HCl \rightarrow ZnCl_2 + H_2\uparrow;$$

while in the neutral medium, oxygen absorption occurs at the cathodic reaction producing OH^- ions which interact with zinc ions forming its hydroxide as stated above. In brief, in a galvanic corrosion cell of the type described above, Zn gets corroded while Cu is protected.

Factors of Galvanic Corrosion

The following factors determine thermodynamic and kinetic conditions of galvanic corrosion.

- Difference between the electrode potentials of the two metals. The greater the difference the higher the driving electric force of corrosion.
- Contact resistance at the boundary between the two metals. High contact resistance limits the electrons transfer through the boundary and decrease the corrosion rate.
- Electric resistance of electrolyte solution. Dilute solutions having high electric resistance provide low corrosion rate.
- Anode-to-cathode areas ratio. Large anode connected to a small cathode result in low corrosion rate.
- Presence of passive film.
- Electrolyte solution properties (pH, oxygen content, temperature and flow rate).

This type of corrosion can be observed in the following examples:

1. Steel pipe connected to copper
2. Zinc coating on mild steel
3. Tin coating on copper vessel
4. Lead–antimony solder around copper wires.

4.4.2 Concentration-Cell Corrosion

Concentration-cell corrosion occurs because of differences in the environment surrounding the metal. This form of corrosion is sometimes referred to as 'crevice corrosion', 'gasket corrosion' and 'deposit corrosion' because it commonly occurs in localized areas where small volumes of stagnant solution exist. Concentration-cell corrosion is the corrosion of metals in a metal-to-metal joint, corrosion at the edge of a joint even though the joined metals are identical or corrosion of a spot on the metal surface covered by a foreign material. Normal mechanical construction can create crevices at sharp corners, spot welds, lap joints, fasteners, flanged fittings, couplings, threaded joints and tube sheet supports. Metal ion concentration cells and oxygen concentration cells are the two general types of concentration cell corrosion. Areas on a surface in contact with an electrolyte having a high oxygen concentration generally will be cathodic relative to those areas where less oxygen is present (oxygen cell). Areas on a surface where the electrolyte contains an appreciable quantity of the metal's ions will be cathodic compared to locations where the metal ion concentration is lower (metal ion cell).

(a) Metal ion concentration cells

The solution may consist of water and ions of the metal which is in contact with water. A high concentration of the metal ions will normally exist under faying surfaces where the solution is stagnant, and a low concentration of metal ions will exist adjacent to the crevice which is created by the faying surface. An electrical potential will exist between the two points; the area of the metal in contact with the low concentration of metal ions will be anodic and corrode, and the area in contact with the high metal ion concentration will be cathodic and will not show signs of corrosion.

(b) Oxygen concentration cells

The solution in contact with the metal surface will normally contain dissolved oxygen. An oxygen cell can develop at any point where the oxygen in the air is not allowed to diffuse into the solution, thereby creating a difference in oxygen concentration between two points. Typical locations of oxygen concentration cells are under gaskets, wood, rubber and other materials in contact with the metal surface. Corrosion will occur at the area of low oxygen concentration (anode). Alloys are particularly susceptible to this type of crevice corrosion (Fig. 4.5).

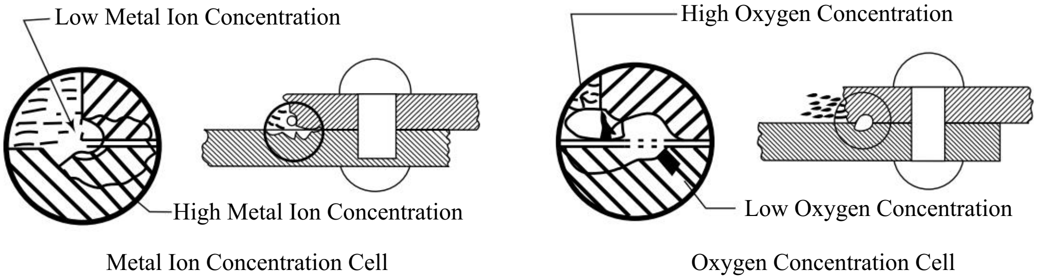


Figure 4.5 *Concentration cell corrosion: (a) metal ion concentration cell and (b) oxygen concentration cell.*

(a) Differential aeration (oxygen concentration cell) corrosion

Occasionally, we find a situation where a metal is exposed to different oxygen environment. The mechanism of corrosion of such metals can be explained by differential aeration environment.

Differential aeration corrosion occurs when one part of the metal is exposed to a different air/O_2 concentration from the rest of the part (Fig. 4.6). It can be noted that a portion of the metal surface exposed to less oxygen acts as 'anodic area' and gets corroded, while the more oxygenated part of the metal is protected because it behaves as cathodic area of corrosion cell.

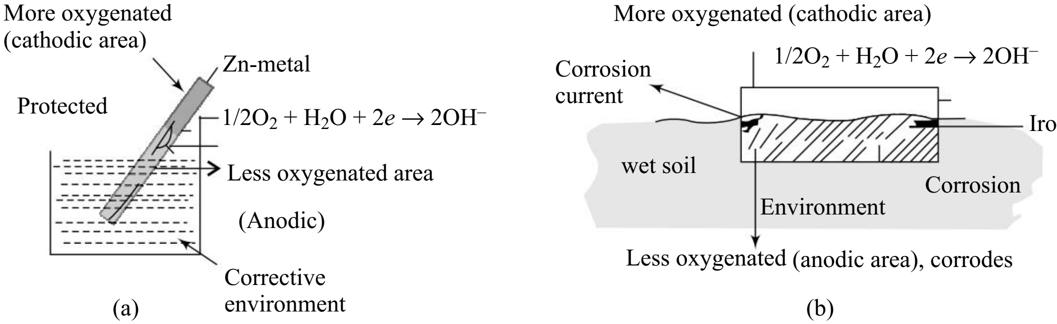


Figure 4.6 (a) and (b) *Examples of differential corrosion.*

According to the electrochemical theory of corrosion, the anodic area of the metal (M = Zn or Fe) gets corroded in a corrosive environment due to the oxidation process, liberating electrons

$$M \rightarrow M^{n+} + ne \qquad \text{(oxidation)}$$

e.g. $Zn \rightarrow Zn^{2+} + 2e$

or $Fe \rightarrow Fe^{2+} + 2e$

A potential difference arises due to the formation of anodic and cathodic areas of the metal surface, which causes the flow of corrosion current between the two differentially aerated areas of the metal.

The more oxygenated area of the metal (Zn or Fe) is protected and O_2 absorption occurs at the cathodic region. The reduction reaction is

$$\frac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^-$$

The overall corrosion products are as given

$$Zn^{2+} + 2OH^- \rightarrow Zn(OH)_2\downarrow$$

or

$$Fe^{2+} + 2OH^- \rightarrow Fe(OH)_2\downarrow$$

If the environment to which the iron metal exposed is rich in oxygen, $Fe(OH)_2$ formed gradually gets oxidized to $Fe(OH)_3$. The net corrosion reaction due to rusting of iron is, thus,

$$4Fe + 3O_2 + 6H_2O \rightarrow \underset{\text{Rust}}{2(Fe_2O_3.3H_2O)}$$

(b) Crevice corrosion

Consider a sheet of stainless steel that has been immersed in the ocean for some time, years perhaps. It has had a bolt with a washer on it to hold it in place. Corrosion has been noticed underneath the washer. The reason for the corrosion is that a *crevice* had been created under the washer (Fig. 4.7). Stagnate water, an electrolyte, is accumulated in the crevice. An electro-chemical, corrosion cell had been produced and resulted in corrosion. This cell is called a *differential aeration*, or *oxygen concentration* cell. It results because there is a difference in the composition of the electrolyte under the washer and that outside the washer. There is a higher oxygen concentration in the solution that surrounds the washer than that is under the washer. It may surprise you to know that corrosion takes place at the lower oxygen concentration whenever we have an oxygen concentration cell. You may have associated corrosion with a higher oxygen concentration which is a true general statement. In the case of a crevice, or a pit, there are two oxygen concentrations in the electrochemical cell that has formed. Thermodynamics is such that corrosion takes place at the lower oxygen concentration.

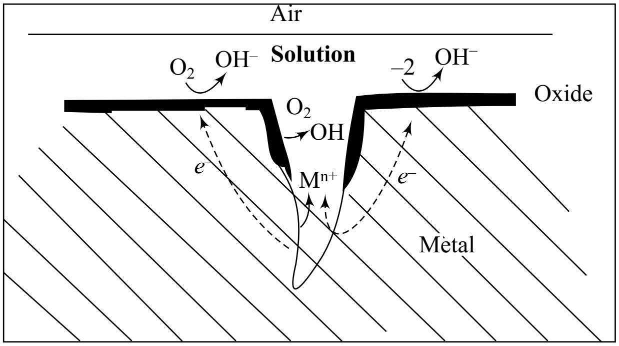


FIGURE 4.7 *Illustration of crevice corrosion*

4.4.3 Pitting Corrosion

Some materials are more subject to *pitting* than others. Consider the case of a stainless steel spoon lying on the beach near the ocean where it has been for several days. It has corroded. Little pits have formed. *Pitting corrosion* is a form of corrosion that is observed almost on a daily basis. This corrosion cell, that is, the electrochemical mechanism of the corrosion, is very similar to that of crevice corrosion. There is a stagnant solution at the bottom of the pit, that is, a lower oxygen concentration is outside of the pit than inside, resulting in another *oxygen concentration* cell. Corrosion takes place at the bottom of the pit. Generally, as the pit gets deeper, the solution at the bottom of the pit becomes more stagnant creating more driving force to promote corrosion (Fig. 4.8a and b). A pit is said to be *self-catalyzing*

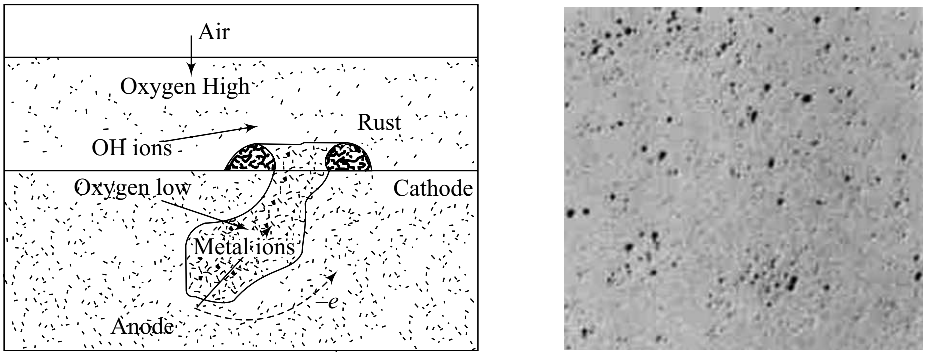


FIGURE 4.8 *(a) Pitting corrosion cell, and (b) photograph showing pitting corrosion on metal.*

This type of corrosion of metals is generally localized to very small area and accelerated corrosion takes place at the anodic region causing minute pits or pinholes on the surface of the metal. It is one of the most destructive types of corrosion that reduces the life of chemical equipment and other metal parts, which results when small extraneous or dust particles get adhered to a metal surface. Pitting type of corrosion can also be explained in the light of differential aeration corrosion.

A typical pitting corrosion cell is illustrated in Fig. 4.9. It can be noted that the impurity/dust-adhered region of the metal surface acts as the anodic region. This tiny portion of covered surface of the metal is poorly oxygenated compared to the large exposed area.

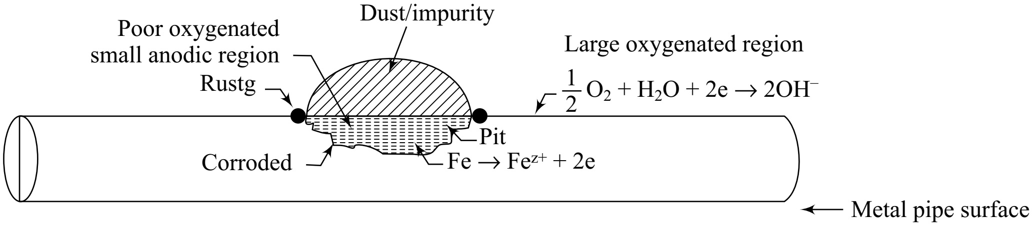


FIGURE 4.9 *Pitting corrosion*

Initially, pits or pinholes are formed on the surface of the metal beneath the dust. The corrosion rates of such minute pits or pinholes are much faster by releasing large number of electrons to the cathodic area of the metal for reduction reaction. This results in the enlargement of pits or pinholes.

Examples: Zn or Fe metal.

According to the electrochemical theory of corrosion, the anodic area of the metal (M = Zn or Fe) gets corroded in a corrosive environment due to the oxidation process, liberating electrons

$$M \rightarrow M^{n+} + ne \qquad \text{(oxidation)}$$

e.g. (i) $Zn \rightarrow Zn^{2+} + 2e$

(ii) $Fe \rightarrow Fe^{2+} + 2e$

A potential difference arises due to the formation of anodic and cathodic areas of the metal surface, which causes the flow of corrosion current between the two differentially aerated areas of the metal.

The more oxygenated area of the metal (Zn or Fe) is protected and O_2 absorption occurs at the cathodic region. The reduction reaction is

$$\frac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^-$$

If Zinc is the metal, in such a situation, the corrosion product is $Zn(OH)_2$.

$$Zn^{2+} + 2OH^- \rightarrow Zn(OH)_2\downarrow$$

If the material exposed material to the corrosive environment is iron, then the reaction is

$$Fe^{2+} + 2OH^- \rightarrow Fe(OH)_2\downarrow$$

When the iron metal is exposed to an environment rich in oxygen, $Fe(OH)_2$ formed gradually gets oxidized to $Fe(OH)_3$. The net corrosion reaction due to rusting of iron is, thus,

$$4Fe + 3O_2 + 6H_2O \rightarrow \underset{\text{Rust}}{2(Fe_2O_3.3H_2O)}$$

4.4.4 Water-line Corrosion

Water-line type of corrosion of metals can be illustrated in terms of the effects of different concentrations of oxygen on metal surface. This type of corrosion can be observed in water tanks, ocean going ships, etc., which are shown Fig. 4.10(a) and (b).

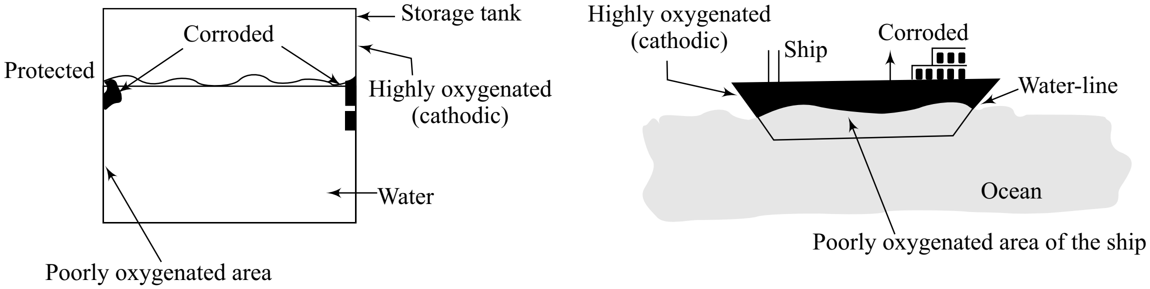


Figure 4.10 *(a) and (b) Illustration of water-line corrosion.*

Water used for domestic purpose is stored in storage metal tanks for long time. Corrosion occurs gradually in water storage iron tanks due to the effect of different oxygen concentrations inside and outside the water level as shown in Fig. 4.10a. In case of a water taken partially filled, the iron above the water line is exposed to more oxygen and acts as cathodic region and is unaffected by corrosion, while iron beneath the water line is poorly oxygenated. This region of the metal behaves as anodic area and slowly corrodes due to oxidation. The Fe^{2+} ions formed in the process dissolve in water.

$$Fe \rightarrow Fe^{2+} + 2e \qquad \text{(oxidation)}$$

The more oxygenated area of the metal (Fe) is protected and O_2 absorption occurs at the cathodic region. The reduction reaction is

$$\frac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^-$$

The overall corrosion products is as stated

$$Fe^{2+} + 2OH^- \rightarrow Fe(OH)_2\downarrow$$

If the environment to which the metal exposed is rich in oxygen, $Fe(OH)_2$ formed gradually gets oxidized to $Fe(OH)_3$. The net corrosion reaction due to rusting of iron is, thus,

$$4Fe + 3O_2 + 6H_2O \rightarrow \underset{\text{Rust}}{2(Fe_2O_3{\cdot}3H_2O)}$$

It is noticed that the effect of corrosion (or rusting) is maximum along a line beneath the level of water meniscus. Similar corrosion effects can be observed in ships and water-pumping pipes in water.

4.4.5 Stress Corrosion Cracking

Metals are widely used in reaction kettles, vessels, plants, machineries, fabricated structures, etc. These materials are under *tensile stress* due to any of the following mechanical operation such as cold working, welding, bending, pressing, riveting and hammering. When any portion of a metal is under stress, highly 'localized' corrosion takes place in a 'specific corrosive environment'. In stress corrosion, corrosive agents are also found to be 'specific and selective'.

Examples:

1. KOH/NaOH solution environment attacks stressed mild steel.
2. Stressed portion of brass corrodes in the presence of ammoniacal environment.

The stress corrosion of metals can also be explained on the basis of electrochemical theory. In a corrosive environment, the stressed portion of the metal acts as anodic area with respect to the unstressed (cathodic) area and undergoes corrosion.

A metal sheet shown in Fig. 4.11 has a stress at the bent region. The presence of such stress on metal results in strain, which causes 'localized area of higher electrode potential'. This strained localized zone, thus, becomes chemically reactive and sensitive to even mild corrosive environment; and the bent region is vulnerable for corrosion, while the unstressed (cathodic) region of the metal is protected from corrosion. The types of stress corrosion are: (i) season cracking, (ii) caustic embrittlement, and (iii) hydrogen embrittlement and stress-corrosion cracking.

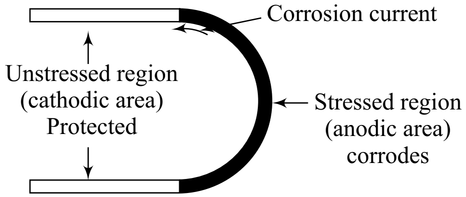


FIGURE 4.11 *Stress corrosion in a bent metal strip.*

1. Season cracking

Pure copper metal is less sensitive to stress type corrosion. However, the presence of few alloying impurities such as P, Zn, Al, etc., results in marked sensitivity for corrosion. It is observed that the stressed grain boundaries in brass gets corroded in mild ammonia solution and induces season cracking.

2. Caustic embrittlement

A type of stress corrosion known as *caustic embrittlement* occurs at the stress areas such as crevices or rivets, joints and bent portion of steam boilers and/or heat transfer equipment at high pressure in which alkaline water attacks the mild steel.

The water of the steam boiler is softened by the addition of small amounts of Na_2CO_3, which makes water alkaline due to the action of Na_2CO_3 at high pressure.

$$Na_2CO_3 + H_2O \rightarrow 2NaOH + CO_2\uparrow$$

The NaOH, thus formed moves into the crevices of rivets/cracks by capillary action. Due to the evaporation of water in boilers, the concentration of NaOH in these cracks and crevices increases. This corrosive environment attacks the mild steel, and as a result, sodium ferrate (Na_2FeO_2) is formed, which decomposes slightly away from the point of its formation.

$$2Na_2FeO_2 + 4H_2O \rightarrow 6NaOH + Fe_3O_4 + H_2\uparrow$$

The regenerated NaOH further enhances the corrosion attack and mild steel and Fe_3O_4 is deposited as a rust. This type of stressed corrosion is dangerous because it leads to boiler failure or even boiler surface bursting under pressure.

3. Hydrogen embrittlement corrosion

In general, '*embrittlement corrosion'* is corrosion that causes a ductile material to fail without localized yielding or shearing. More specifically, hydrogen embrittlement assumes several different forms with a general similarity. This damage takes place at the cathode, an area that we normally think is safe from corrosion—remember the saying that 'corrosion takes place at the anode', but it does not in this case. Hydrogen ions are reduced to hydrogen molecules at the cathode. Those atoms usually pair-up to become hydrogen molecules. These molecules harmlessly bubble off as hydrogen gas. However, some metals are very susceptible to letting hydrogen atoms permeate into the grains. This is done while the hydrogen exists as an atom, before it becomes a molecule. Once inside the metal the hydrogen atoms can do all sorts of mischief that results in *hydrogen damage* (Fig. 4.12).

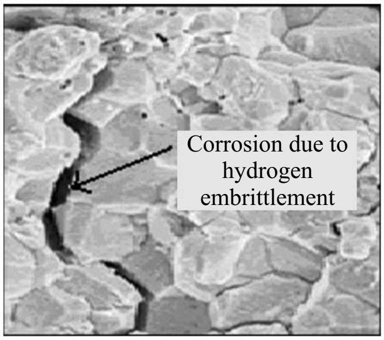


Figure 4.12 *Corrosion of ductile material due to hydrogen embrittlement*

4. Stress corrosion cracking

Some materials are subject to a corrosive attack into the grain boundaries, whenever that material is placed under tensile stress. Consider, for example, the landing gear of an airplane as it sits on the tarmac.

The landing gear is under stress and that stress is transferred to various parts of the metal parts. Parts that undergo tensile stress are subject to this problem. The result can be sudden and catastrophic cracking and failure of the metal.

Figure 4.13 *Stress corrosion cracking*

This form of *stress corrosion cracking* (SCC) (Fig. 4.13) involves constant or cyclic stress, acting in conjunction with a damaging chemical environment. The stress may be caused by internal or external loading.

(a) *Internal stress may be trapped* in a part of structure during manufacturing processes such as cold working or by unequal cooling from high temperatures. Most manufacturers follow up these processes with a stress relief operation. Even so, sometimes stress remains trapped. The stress may be externally introduced by riveting, welding, bolting, clamping, press fit, etc. If a slight mismatch occurs, or a fastener is over-torqued, internal stress will be present.

(b) *Internal stress is more important* than design stress, because stress corrosion is difficult to recognize before it has overcome the design safety factor. The level of stress varies from point to point within the metal. Stresses near the yield strength are generally necessary to promote stress corrosion cracking

(Fig. 4.14). However, failures may occur at lower stresses. Specific environments have been identified which cause stress corrosion cracking of certain alloys.

Possible causes and remedial mitigation for stress corrosion

1. Salt solutions and sea water cause stress corrosion cracking of high-strength, heat-treated steel and aluminium alloys.
2. Methyl alcohol-hydrochloric acid solutions will cause stress corrosion cracking of some titanium alloys.
3. Magnesium alloys may stress-corrode in moist air.
4. Stress corrosion may be reduced by
    - applying protective coatings,
    - stress relief heat treatments,
    - using corrosion inhibitors, or
    - controlling the environment.

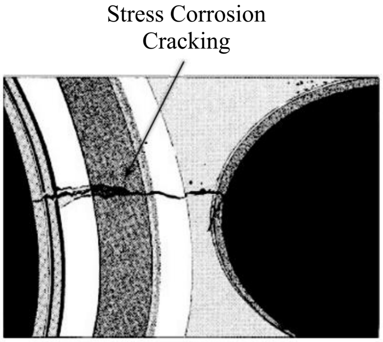


Figure 4.14 *Stress corrosion cracking.*

4.4.6 Intergranular Corrosion

Intergranular corrosion is an attack on the grain boundaries of a metal. A highly magnified cross-section of any commercial alloy shows the granular structure of the metal.

It consists of quantities of individual grains, and each of these tiny grains has a clearly defined boundary which chemically differs from the metal within the grain (Fig. 4.15). The grain boundary and the grain centre can react with each other as anode and cathode when in contact with an electrolyte.

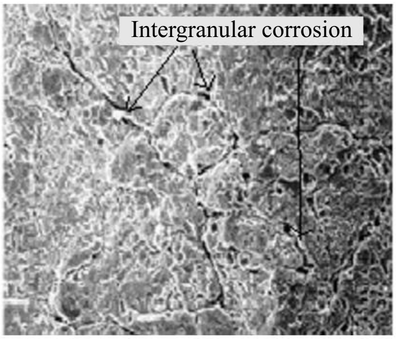


Figure 4.15 *Intergranular corrosion of grain boundaries.*

4.4.7 Fatigue Corrosion

Fatigue corrosion involves cyclic stress and a corrosive environment. Metals may withstand cyclic stress for an infinite number of cycles so long as the stress is below the endurance limit of the metal. Once the limit has been exceeded, the metal will eventually crack and fail from metal fatigue. However, when the part or structure undergoing cyclic stress is also exposed to a corrosive environment, the stress level for failure may be reduced many times. Thus, failure occurs at stress levels that can be dangerously low depending on the number of cycles assigned to the life-limited part.

(a) *Fatigue corrosion failure* occurs in two stages. During the first stage, the combined action of corrosion and cyclic stress damages the metal by pitting and crack formations to such a degree that fracture by cyclic stress will occur, even if the corrosive environment is completely removed.

(b) *The second stage* is essentially a fatigue stage in which failure proceeds by propagation of the crack (often from a corrosion pit or pits). It is controlled primarily by stress concentration effects and the physical properties of the metal. Fracture of a metal part due to fatigue corrosion generally occurs at a stress level far below the fatigue limit of an uncorroded part, even though the amount of corrosion is relatively small.

4.4.8 Erosion–Corrosion

Erosion–corrosion results from the disruption of protective passive films by erosive or abrasive processes. Once the protective or passive film is removed in an aqueous electrolyte, the electrochemical processes for the types of corrosion described above take place. *Erosion–corrosion, also known as flow-assisted corrosion, is the process of moving a corrosive liquid against a metal surface, leading to the accelerated loss of material.*

Erosion–corrosion refers to the repetitive formation (a corrosion process) and destruction (a mechanical process) of the metal's protective surface film.

This typically occurs in a moving liquid. Erosion may be impinging or sliding (pipe wall) when it occurs. An example is the erosion corrosion of copper water tubes in a hot, high velocity, soft water environment. Cavitation is a special form of erosion corrosion.

4.4.9 Leaching Corrosion

When one component of an alloy is removed selectively by corrosion because there exist differences in potential between different constituents of an alloy—a situation analogous to that of the galvanic corrosion of coupled dissimilar metals. This results in an increase in the concentration of the remaining components and selective leaching occurs. This type of corrosion is also called parting and, after the most common example, the selective leaching out of zinc from brass, dezincification.

4.5 Galvanic (or Corrosion) Series

Galvanic series relationships are useful as a guide for selecting metals to be joined, will help the selection of metals having minimum tendency to interact galvanically, or will indicate the need or degree of protection to be applied to lessen the expected potential interactions. In general, the further apart the materials are in the galvanic series, the higher the risk of galvanic corrosion, which should be prevented by design. Conversely, the farther the metal is, the greater the corrosion will be. However, the series does not provide any information on the rate of galvanic corrosion and thus serves as a basic qualitative guide only (Table 4.1).

A metal high up in the electrochemical series (low-reduction electrode potential) is more anodic and undergoes corrosion faster than the metal below it. However, it is observed in few cases, electrochemical series does not account for the corrosion of all metals and alloys, as expected.

Example, when aluminium and zinc are connected, zinc metal gets corroded while aluminium is protected; though, Zn is below Al in the electrochemical service. This reversal of the order in the above series is due to the formation of 'passivity' which results in the production of a thin film of Al_2O_3 on the surface of the Al, which protects it from corrosion.

A more sensible and informative series known as *galvanic series*, has been adopted by observing the corrosion behaviour of metals and alloys in sea water.

An arrangement of metals and alloys in order of their corrosion tendencies in a corrosive environment is referred to as galvanic series (Table 4.2).

	Magnesium
(more anodic)	Zinc
	Tin
	Steel
	Iron
	Stainless Steel (304 active)
	Stainless Steel (316 active)
	Nickel (active)
	Copper
	Lead
	Nickel (passive)
	Nickel (passive)
	Stainless Steel (304 passive)
	Monel
	Stainless Steel (316 passive)
	Titanium
	Silver
	Graphite
	Gold
	Platinum
(noble metals)	

Table 4.1

Galvanic series (nobler higher)

Electrochemical series	Galvanic series
1. Standard electrode potentials of metals are measured when they are in contact with their ions in solution at standard conditions (conditions of potentiometry).	1. Corrosion of metals and alloys are carried out in unpolluted sea water and their corresponding potentials are arranged as a series with more active metal (anodic) at the top of the series.
2. Based on the electrode potential value, the position of a metal is fixed in electrochemical series.	2. Metal may have a different position in galvanic series depending on the corrosion aspect, irrespective of its electrode potential
3. In electrochemical series, no position is given to alloys with their potential values.	3. Alloys are given position in galvanic series based on their corrosion behaviour.
4. Metals and nonmetals are included in the series	4. The galvanic series comprises both metals and alloys.
5. It predicts electrochemical behaviour in redox systems as well as displacement tendencies.	5. The relative corrosion tendencies of metals and alloys can be predicted.

Table 4.2

Electrochemical series versus Galvanic series

4.6 Factors Affecting Corrosion

Corrosion occurs in metals, and is an indispensable electrochemical process, when exposed to an environment. It can be noticed that the following factors affect the corrosion process in metals.

1. Metallic factors

The metallic factors include electrode potentials, hydrogen overvoltage and the presence of 'impurities' in metals.

2. Environmental factors

The environment to which the metal is exposed also affects considerably the corrosion.

They are as follows:

1. pH of the medium
2. Temperature
3. Oxygen environment
4. Humidity
5. Nature of the corrosion product
    (a) The nature of surface film formed
    (b) The solubility of the corrosion product.

A brief discussion of the factors that affect the rate of corrosion of metals is discussed here below.

4.6.1 Metallic Factors

(a) Anodic and cathodic area

When two metals are in contact and dipped in a corrosive environment, the corrosion of the 'anodic metal' is directly proportional to the ratio of anodic to the cathodic area of the metals. The process of corrosion is more rapid and the destruction of the metal is faster if the anodic area of the metal is small. The current density is high for smaller anodic area and the demand for electrons at larger cathodic area is met by severe and faster corrosion rate at the anodic area.

Example: The best example to be illustrated is the defective surface coating of tin on iron metal, as shown in Fig. 4.16.

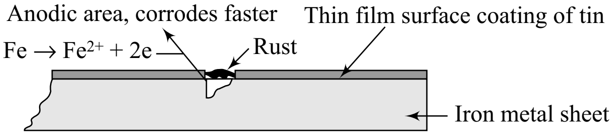


FIGURE 4.16 *Rusting of defective coated tin on iron.*

(b) Electrode potential values of metals

The rate and severity of corrosion of metals depends on the difference of electrode potentials or the positions of the metals in the galvanic series. When dissimilar metals are in contact, dipped in a corrosive medium, the corrosion at the anodic region occurs with a faster rate, because the anodic metal is higher up in the galvanic series. It can be remarked, therefore, that greater the electrode potential difference of the metals, faster is the rate of corrosion of anodic metal.

Corrosion of zinc metal is severe and faster than iron metal when it is in contact with copper in a medium. This can be attributed to the fact that the electrode potential difference between zinc and copper is greater than that of iron and copper.

(c) Overvoltage

The corrosion rate of a metal higher up in the galvanic series is reduced due to the phenomenon of 'overvoltage', which lowers the effective electrode potential of the metal to a small value. In other words, reduction of overvoltage of the corroding anodic metal increases the rate of corrosion.

(d) Physical state of metals

The grain size, stress, orientation of the constituent atoms in a particular physical state of a metal influences the corrosion. The smaller the grain size of the metal, the greater will be its corrosion.

The presence of stress even in the case of pure metal causes corrosion, since stress tends to be anodic area. A strained metal (anodic region) has higher electrode potential than the metal region without stress.

4.6.2 Environmental Factors

There is a noticeable effect of corrosion of metals when they are exposed to different environments.

(a) pH of the medium

The rate of corrosion of metal is much faster in an acid medium than an alkaline or a neutral condition. Corrosion rates almost always increase with decreasing pH (increasing acidity). This is a direct result of increasing the concentration of an aggressive ion (H+) and increasing the solubility of most potentially corrosive products. The effect of pH of a medium (and the potential) on a metal is discussed in Section 4.14.

(b) Humidity and temperature

A humid environment favours the corrosion of metals. Iron does not rust easily when exposed to dry air; however, it undergoes rusting when exposed to humid conditions. The metal gets corroded in humid atmosphere because atmospheric gases like CO_2, SO_2, etc., get dissolved in water and produce a medium that sets up an electrochemical cell in metal. The rate of corrosion of metal enhances with the rise in temperature.

(c) Conducting medium

If the medium to which the metal part is exposed is conductive, then it induces corrosion. If the electrical conductance of the soil or medium is high, the rate of corrosion of metal is severe and fast. *For example:* Rusting of iron part in underground structures.

4.6.3 Nature of the Corrosion Products

(a) The solubility of corrosion product

The corrosion rate of a metal proceeds with a faster rate, if the corrosion product formed is soluble in the medium to which it is immersed. If the corrosion product is insoluble, then it is likely to cover the entire metal surface as continuous thin film coating inhibiting the corrosion process.

(b) The nature of corrosion product

In an oxygen environment, if the tarnished metal gets a weak, thin, porous and nonadherent film of oxide, then the metal is susceptible for corrosion.

However, in few cases of metals, the oxide film so formed is nonporous, tough and adherent on the surface of the metal, which protects the metal from further corrosion.

4.7 Corrosion Control (or Prevention)

If, by any process or means, the ‘surface of a metal’ is protected, then the corrosion of the metal can be controlled from the environment.

The corrosion can be mitigated by any one of the following methods.

1. Proper selection and designing of metal equipment, fabrication, etc.
2. Protective coating
   (a) Metal coatings
   (b) Inorganic coating
   (c) Organic coating
3. Cathodic and anodic protection
4. Corrosion inhibitors

4.7.1 Proper Selection and Design of Materials

The use of acceptable engineering practices to minimize corrosion is fundamental to corrosion control. This is accomplished by engineering design. One of the most important factors in designing for corrosion control is to avoid crevices where deposits of water-soluble compounds and moisture can accumulate and are not accessible for maintenance. Any region where two surfaces are loosely joined, or come into proximity, also qualifies as a crevice site. Joining geometries also present various crevice corrosion problems. Examples include bolting, back-to-back angles, rough welds, weld spatter, sharp edges, corners, discontinuities and intermittent welding.

The following points may be noted for corrosion control:

1. Design an equipment or fabricate any metal structures by avoiding contact between two dissimilar metals in a corrosive environment and, thereby, differential metal corrosion can be controlled.
2. The rate of corrosion of metals can be controlled by maintaining larger ‘anodic area’ of the metal.
3. The metal for fabrication is selected in such a way that the dissimilar metals are as close as possible in the galvanic series so that the differential metal type corrosion can be minimized.

4. Isolation of dissimilar metals—Galvanic attack may be prevented by using an insulator to prevent contact (completion of electrical circuit) between dissimilar metals. The more noble metal can be insulated from the less noble metal through the use of plastic washers for fasteners. Ceramics or nonconductive insulating materials also may be used.
5. Design or fabricate equipment or metal parts in such manner that they have minimized sharp edges and corners, and also avoid as far as possible, crevices in joints, etc. This inevitably minimizes the corrosion of metals.
6. The metal parts, as far as possible, should be free from tensile stresses to avoid stress corrosion.
7. Select a pure metal to minimize corrosion.

4.7.2 Protective Coatings

Coatings and linings play a significant role in corrosion prevention and are used extensively for this purpose. Coatings also can perform other valuable functions such as provide easier clean ability, decontamination, add colour or a pleasing aesthetic appearance, increase light reflectance for a better working environment or for safety purposes, and provide a colourmarking or safety warning.

The surface of a metal can be protected from corrosion by any of these processes such as electrochemical treatment or inorganic or organic coating.

A protective coating protects the metal from corrosion by acting as a barrier between the metal and the environment.

Various types of protective coatings or metals are discussed below.

1. Metal coating

Except noble metals, all other metals have a natural tendency to get corroded when exposed to any environment and, therefore, it is highly desirable to adopt protective measures by preventing the metal surface from coming into contact with the environment. This can be achieved by metal coating the surface of the base metal and isolating it from environment.

Metallic coatings such as thermal sprayed zinc, aluminium or aluminium–zinc alloys, stainless steel and chromium can be used to protect against corrosion or provide increased wear or abrasion resistance. Zinc-rich coatings are widely used to provide galvanic corrosion protection to steel. Galvanizing (either strip or hot-dip) is another way of applying zinc to a steel surface for corrosion protection.

The process of covering the base metal with a layer of protective metal is known as metal coating.

Protective coating not only protects surface of a metal but also gives decorative value. Besides protective coating, it imports few physical and mechanical properties such as thermal insulating properties, electrical properties, wear resistance, oxidation, etc. The process of metal coating on to the surface of a base metal can be adopted by any one of the two techniques.

1. Anodic metal coating and
2. Cathodic metal coating.

4.7.3 Anodic Metal Coating

The metal coatings which are anodic to the base metal are known as *anodic metal coatings*. In order to protect a metal from corrosion, any metal which is higher up in the galvanic series such as Zn, Mg, Tin, etc., are coated on the surface of base metal. These active metals are anodic to the base metal. Anodic coating of Zinc on mild steel/iron is the best example. Iron is coated with more active metal zinc and, thus, zinc coating on the base metal iron acts as anodic area, while the iron is cathodic (see Fig. 4.17).

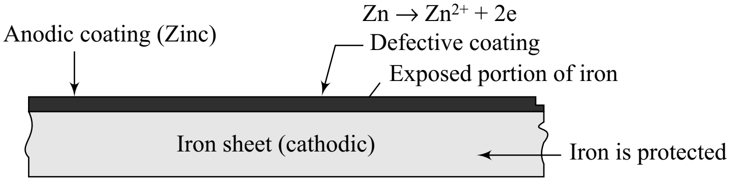


Figure 4.17 *Anodic coating of a base metal.*

It can be noticed that the anodic coating on the metal surface is defective or discontinuous (crack, etc.), the zinc coating being anodic corrodes ($Zn \rightarrow Zn^{2+} + 2e$) and the base metal iron remains cathodic and protected. The main advantage of this method of coating is that iron is completely protected till all the active material zinc coating is exhausted by corrosion.

Examples: Roof sheets, barbed wire, nuts and bolts.

4.7.4 Cathodic Coating

The surface of a metal can be protected from corrosion by coating a base metal with a more noble metal whose tendency for corrosion is lower than the base metal in the galvanic series.

The cathodic coating thus produced protects the base metal, because of its high corrosion resistance than the base metal. A thin film coating of metals like tin, copper, nickel, etc., are deposited on the base iron metal by cathodic coating techniques to prevent corrosion. If the cathodic coating on the base metal surface is continuous and not defective, the surface coating provides an effective protection to the base metal or otherwise, defective coating causes more corrosion damages when the base metal is made anodic.

Tin, chromium and nickel coating by electroplating technique on any base metal offers the best example of cathodic coating (see Fig. 4.18).

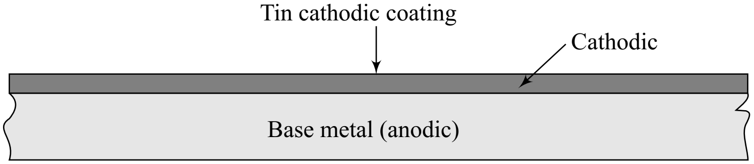


Figure 4.18 *Cathodic coating of a base metal.*

4.8 Techniques of Metal Coating

In order to protect a metal from corrosion, metal coating techniques have been adopted. They are as follows:

1. Hot dipping
2. Cladding
3. Spraying
4. Cementation
5. Electroplating

4.8.1 Hot Dipping

A thin film of uniform coating of metal (Zn, Sn, Pb, etc.) can be deposited on the metal surface by hot dipping technique. The process is carried out by immersing a clean metal in a bath of molten metal.

Examples: Zinc-coated iron (Galvanization), Tin-coated iron (Tinning)

(a) Galvanization

'The process of coating molten zinc on the base metal (iron) surface by hot dipping is known as galvanization'.

In the process of galvanization, the base metal iron/steel is coated with a thin film of zinc metal. Since, Zn is higher up in the corrosion series, the Zn coating acts as anodic with respect to the base metal (cathodic). With repeated exposure of galvanized metal to environment, only the coated active zinc metal gets corroded. Thus, the base metal iron/steel is protected. The galvanization process is carried out as follows (Fig. 4.19):

1. The metal is cleaned and degreased using organic solvents.
2. The metal is treated with dilute H_2SO_4 (pickling process) for 10 minutes at about 80°C to remove any rust or scales, if there is any.
3. The metal is further treated with flux materials, $ZnCl_2$ and NH_4Cl for best adhesion property.
4. Finally, the metal part is dipped in hot molten zinc at 430–470°C.
5. The excess Zn is removed from the surface of the coated metal by rolling, wiping or air blow techniques.

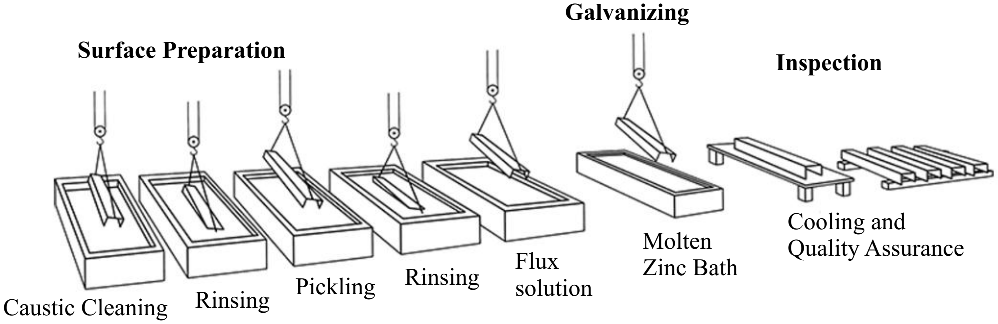


Figure 4.19 *Illustration of galvanization process.*

(b) Tinning

'The process of coating a thin film of molten tin on the base metal surface by hot dipping techniques is known as tinning'.

This is an example of cathodic metal coating. The process of tinning is carried out as follows (Fig. 4.20):

1. The clean surface of the base metal is treated with dil. H_2SO_4 for few minutes at 80°C to remove the rust or scales of the metal.
2. It is immersed in a flux, $ZnCl_2$ and NH_4Cl solution to facilitate the molten metal to adhere to the metal surface.
3. The metal is then dipped quickly in molten tin.
4. To avoid oxidation, the metal is treated with palm oil.

Tinning is widely used in food storage cans, tin-coated copper sheets for cooking utensils, etc. (Fig. 4.20)

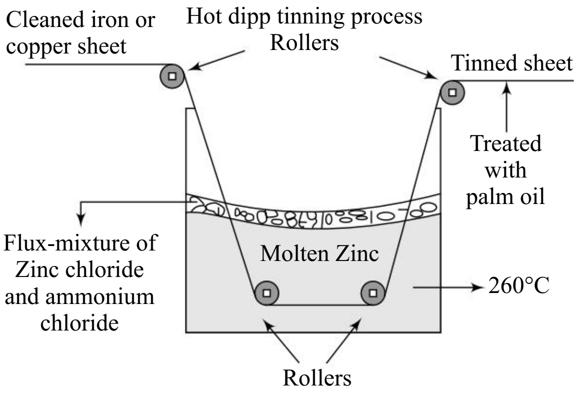


Figure 4.20 *Illustration of thin sheet tinning process.*

4.8.2 Cladding

The surface of the base metals can be bonded through a thin layer of corrosion-resistant metal (Fig. 4.21) like Ag, Ni, Pb, Cr, etc. This is done on the base metals like Al, Cu, etc.

Cladding is done by performing a sandwich of protecting layer and the base metal, which are passed through rollers and bonded under the action of heat and pressure.

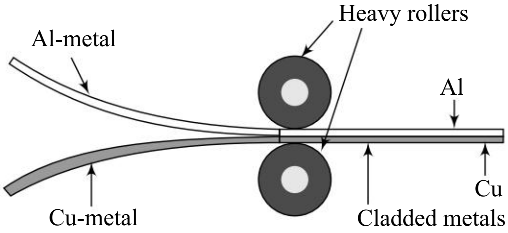


FIGURE 4.21 *Process of cladding of metals.*

4.8.3 Spraying

A good and continuous coating of a metal can be sprayed onto the surface of a base metal as corrosion resistant to environment.

Rough surfaces are essential for metal coating to adhere to the base metal. This process of spray coating is depicted in Fig. 4.22.

For huge structures, such as bridges, metal roofings, etc., spraying technique has been adopted.

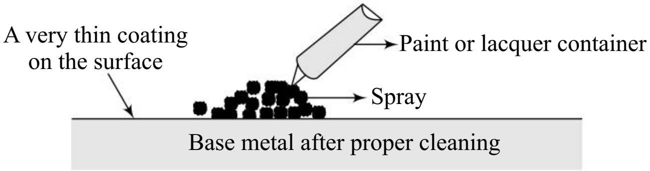


FIGURE 4.22 *Coating by spraying technique.*

4.8.4 Cementation or Diffusion Coating

This process of metal coating is suitable for coating small articles of uneven surfaces and different shapes.

In this technique, *a uniform surface coating is obtained by heating the base metal in a powder of the coating metal.* Diffusion of coated metal powders into the base metal takes place resulting in the formation of layers of alloy of varying composition and offers a protective coating to the base metal. Metal coating of nuts and bolts, thread metal parts, values, etc., are the examples.

4.8.5 Electroplating

'Electro-deposition' is a process by which the coating metal is deposited on the base metal by passing a direct current through an electrolyte solution containing the soluble salt of the coating metal.

Noble metals and the base metals are used for electroplating. When a base metal is electroplated with a noble metal, it evidently protects the base metal from corrosion.

Example:

(i) Electroplating of iron articles with Sn—(Tinning).
(ii) Electroplating of iron with Zn—(Galvanizing).

In electroplating, the object to be plated is made the cathode of the cell (Fig. 4.23). The electrolyte is a salt of the metal. The metal to be deposited is always made the anode, which dissolves to replenish the metal in the solution. In this way, Au, Ag, Cr, Ni, etc., can be electroplated on the surface of a base metal.

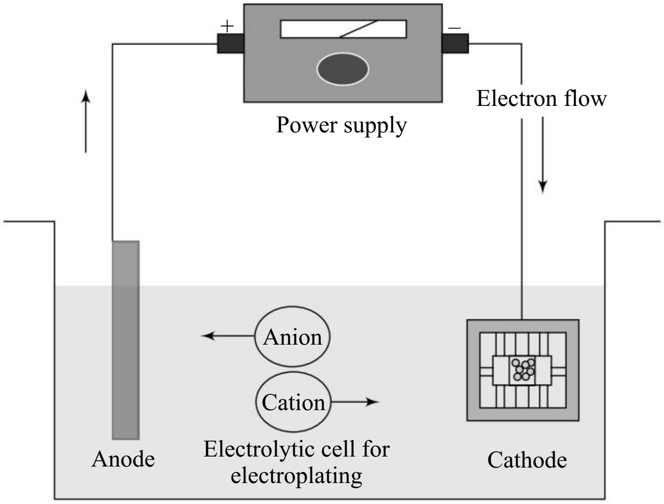


FIGURE 4.23 *Illustration electroplating of base metal by electroplating.*

4.9 Inorganic Coating

'*Inorganic coatings'* or chemical conversion coatings are carried out on the base metals in order to protect the metal from surrounding environment. A strong adherent coating is produced on the surface of the base metal by chemical or electrochemical techniques. In practice, there are two types of inorganic coating techniques and the process of coating is described below.

A metal due to 'passivation effect' exhibits much higher corrosive resistance than expected from its position in the galvanic series. Passivity of metal is due to the formation of a highly protective thin oxide film on the surface of the metal, which makes the metal more corrosion resistant. The metal like Al, Cr, Ni, Ti, etc., do show passivity due to the formation of their thin oxide film on the surfaces.

Anodization is a process in which a protective passive oxide film is artificially (chemically or electrochemically) produced on certain metals (Al, Zn, Mg, Cr, etc.) in oxidizing acid bath.

4.9.1 Anodization of Aluminium

This process is known as anodic oxidation of Aluminium. Al_2O_3 is formed on the surface of aluminium when Al is made the anode of the cell, in a suitable oxidizing electrolyte (H_2SO_4, chromic acid, oxalic acid bath) at temperatures of about 25–40°C and moderate current densities. A thin coating of oxide of aluminium gets deposited on the surface due to the process of anodic oxidation at the Al anode of the cell.

The sulphuric acid process enables film to produce high 'absorptive capacity' and 'resistance' to corrosion. Sulphuric acid is very suitable for anodizing almost all Al alloys. Sealing of oxide film is a must due to considerable 'porosity of the film' and which enhances the strength and corrosion resistance.

4.9.2 Anodizing in Sulphuric Acid

The solution compositions and operating conditions for anodizing aluminium and its alloys using d.c. is carried out as under:

Sulphuric acid : 190–200 g
Water : 1 L
Current density : 1–25 amp/cm^2
Voltage : 10–28 V
Temperatures : 15–23°C
Time : 30–50 minutes

Anodized articles are used in soapboxes, tiffin carriers, a host of household utensils, window frames, etc.

4.9.3 Phosphate Coating

Phosphate coating is a process in which a thin film of metal phosphate coating gets deposited on the metal surface due to the interaction of surface atoms of the base metal and certain metal phosphates in aqueous solution of phosphoric acid.

Phosphate coatings are generally used to coat iron, steel and zinc, and lesser extent to Al, Cd and tin. A chemical or electrochemical reaction between the phosphating solution and the base metal results in the formation of a surface film, consisting of zinc or manganese phosphates. The surface coating obtained as above on metal surface does offer complete resistance to corrosion; however, phosphating gives the surface a good point adhesion quality. Thus, phosphating techniques are adopted as an excellent base for paints, lacquers, oils and enamels. Phosphating of metals is usually carried out by immersion or spraying or brushing and electrophosphating.

The phosphating bath is a mixture of phosphorus acid and Zn/Mn phosphate solution. Accelerators such as copper salts, H_2O_2, nitrates, etc., are used to enhance the rate of deposition. The pH of the phosphating bath is to be maintained at 1.8–3.2. The electro-deposit of phosphate coating process is carried out at temperature of 35°C.

Phosphate coating is widely used in automobile industry, nuts and bolts, construction materials, etc.

4.10 Organic Coating

One of the methods adopted to prevent corrosion of metal is to have a thin coat of paint, lacquer or enamel on the exposed metal surface. Organic coatings are inert organic barriers introduced between the metal surface and the environment for protection from corrosion and also to impart decorative value. Protective coating of this type is preferred because of its chemical inertness to the environment and also possesses a good surface adhesion property.

Organic coatings are most commonly applied by brush, roller, or airless or conventionally atomized air-spray equipment. These coatings consist of an organic binder or film former with inhibitive, barrier and sacrificial pigments for corrosion protection. Because of the concerns regarding worker and environmental protection, many coating formulations are changing to the use of nontoxic, non-hazardous pigments and/or solvents, using water-based or high-solids binders.

(a) Paints

Paint is a mixture of one or more pigments in an organic medium consisting of oil. This mixture is thinned by a highly volatile solvent, known as thinner. When paint is applied to the surface of the metal to be protected the solvent evaporates while the drying oil slowly oxidizes forming a dry pigmented film on the surface.

(b) Enamels

It is an intimate dispersion of pigments in varnish. The drying of enamels on the surface of metal may take place either

1. at elevated temperature in the presence of O_2 or
2. enamels are applied at room temperature on iron and steel parts which adhere onto the surface of household utensils and or any other parts in industries.

4.11 Cathodic Protection

Cathodic protection is an electrical method of mitigating corrosion on metallic structures that are exposed to electrolytes such as soils and waters. *Corrosion control is achieved by forcing a defined quantity of direct current to flow from auxiliary anodes, through the electrolyte, and onto the metal structure to be protected.* Theoretically, corrosion of the structure is completely eliminated when the open-circuit potentials of the cathodic sites are polarized to the open-circuit potentials of the anodic sites. The entire protected structure becomes cathodic relative to the auxiliary anodes. Therefore, corrosion of the metal structure will cease when the applied cathodic current equals the corrosion current. There are two basic methods of corrosion control by cathodic protection. One involves the use of current that is produced when two electrochemically dissimilar metals or alloys (Table 4.1) are metallically connected and exposed to the electrolyte. This is commonly referred to as a sacrificial or galvanic cathodic protection system. The other method of cathodic protection involves the use of a direct current power source and auxiliary anodes, which is commonly referred to as an impressed current cathodic protection system.

The metal structures, equipment, ocean going ships, underground pipes, water tanks, etc., are made cathodic by attaching to it a more active 'anodic metal'. In this cathodic protection method, anodic material attached to the metal gets corroded while the base metal (cathodic) is protected, as long as the active material is present. The metals can be protected from corrosion by the following two methods.

1. Sacrificial anodic method,
2. Impressed current method.

4.11.1 Sacrificial Anodic Protection Method

Sacrificial anode cathodic protection systems provide cathodic current by galvanic corrosion. The current is generated by metallically connecting the structure to be protected to a metal/alloy that is electrochemically more active than the material to be protected. Both the structure and the anode must be in contact with the electrolyte. Current discharges from the expendable anode through the electrolyte and onto the structure to be protected. The anode corrodes in the process of providing protection to the structure. The basic components of a single, sacrificial anode cathodic protection installation are the structure to be protected, the anode and the means of connecting the structure to the anode. This is shown in Fig. 4.24.

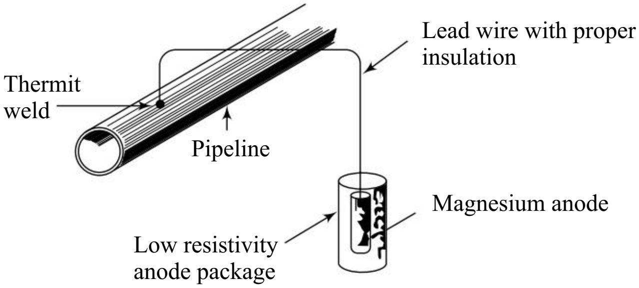


FIGURE 4.24 *Illustration of sacrificial anodic protection.*

In this sacrificial anodic protection, a more active anodic (higher up in the galvanic series) metal like Mg, Zn, etc., are attached to the base metal to be protected. As a result, corrosion gets 'localized' only at the more active anodic metal and the original metal part becomes cathodic and, thus, protected from corrosion. The more anodic active metals (Mg, Zn, etc.) used for this purpose generally known as 'sacrificial anode' and is replaced by a fresh anode when all the active metal is exhausted. This is illustrated in Fig. 4.24.

(a) ***Sacrificial anode cathodic protection systems have a number of advantages***:

- No external power is required
- No regulation is required
- Easy to install
- Minimum of cathodic interference problems
- Anodes can be readily added
- Minimum of maintenance required
- Uniform distribution of current
- Minimum right-of-way/easement costs
- Efficient use of protective current
- Installation can be inexpensive if installed at the time of construction

(b) ***Sacrificial anode systems also have disadvantages that limit their application***:

- Limited driving potential
- Lower/limited current output
- Poorly coated structures may require many anodes
- Can be ineffective in high-resistivity environments
- Installation can be expensive if installed after construction.

4.11.2 Impressed Current Cathodic Protection

Corrosion occurs in metal to the environment to which it is exposed. The corrosion current always flows from anodic area to the cathodic area of the metal.

In this impressed current method of cathodic protection, an impressed external current is applied in opposite direction to the metal that is corroding to nullify the corrosion current. By doing so, the metal subjected to corrosion is thus converted from anodic to cathodic and protects the metal from corrosion. The impressed current is obtained from a DC source with an inert anode graphite as shown in Fig. 4.25.

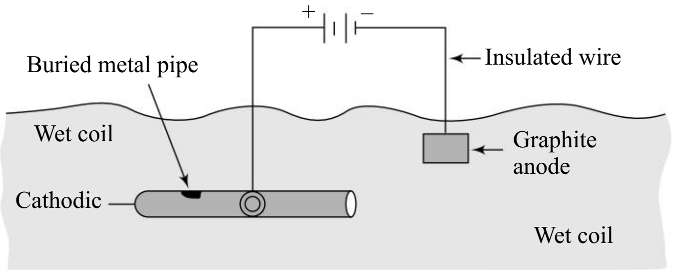


FIGURE 4.25 *Illustration of impressed current cathodic protection of the buried pipe.*

(i) ***Impressed-current-cathodic protection systems have a number of advantages:***

- Can be designed for a wide range of voltage and current

- High ampere-year output is available from single ground bed
- Large areas can be protected by single installation
- Variable voltage and current output
- Applicable in high-resistivity environments
- Effective in protecting uncoated and poorly coated structures.

(ii) ***Impressed-current systems also have disadvantages which limit their application***:

- Can cause cathodic interference problems
- Are subject to power failure and vandalism
- Have right-of-way restrictions
- Require periodic inspection and maintenance
- Require external power, resulting in monthly power costs
- Overprotection can cause coating damage.

4.11.3 Anodic Protection

'Passivity' is quite common for metal like Al, Cr, Ti, Ni, etc., and thus, in such cases cathodic protection technique cannot be adopted. In the anodic protection method, the metal part to be protected from corrosion is made 'more anodic'. An external impressed direct current is applied to the anodic part, which makes it more anodic. As a result of this application of external impressed anodic current, a thin oxide film is formed which protects the surface from further corrosion. A typical anodic protection setup of a tank is illustrated in Fig. 4.26.

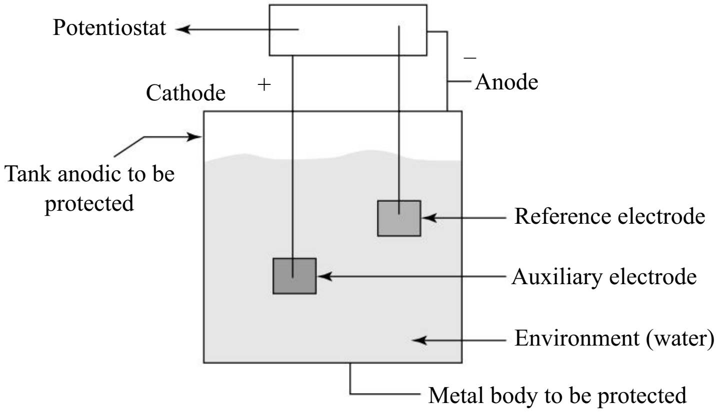
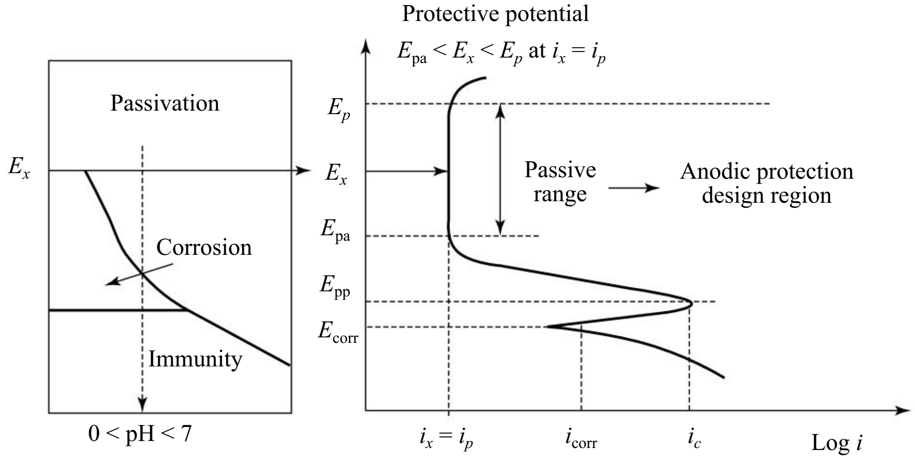


FIGURE 4.26 *Illustration of anodic protection of a metal body.*

The metal part (or tank) which is to be protected from corrosion is connected to a potentiostat, a device used to maintain a constant potential and also has a reference electrode attached to it (Fig. 4.26). An auxiliary electrode made up of corrosion resistant metal acts as cathode. The metal tank to be protected from corrosion acts as anode. The potentiostat is so adjusted, for a particular potential and current, the anodic metal tank gets oxidized and gets more passivated. The formation of a thin corrosion-resistant layer, thus, protects the metal from corrosion in the anodic protection method. This is illustrated in Fig. 4.27a and b showing schematic Pourbaix diagram and active–passive curve indicating the passive region for anodic protection.



(a) Pourbaix (b) Polarization

Figure 4.27 *(a) Schematic Pourbaix diagram and (b) active–passive curve indicating the passive region for anodic protection.*

4.12 Corrosion Inhibitors

Corrosion control of metals has been discussed in details in the previous section. It is observed that corrosion can also be controlled by adding 'inhibitors' to the medium. *Corrosion inhibitors can be defined as any substance which when added in requisite quantities to the medium reduces the rate of corrosion of metals*. Inhibitors are organic or inorganic compounds that are soluble in the medium and eventually form a protective layer either on the anodic or cathodic area of the metal.

4.12.1 Anodic Inhibitors

The metal phosphates, chromates, molybdates, etc., are oxygen rich compounds which when added to the medium interact with newly formed metal ions of the metal and forms a thin strongly adherent film of the metal phosphate on the surface of the anodic area, thus, protecting the anodic area from further corrosion. If insufficient quantity of inhibitors are present or the surface film is defective, then due to the formation of localized electrochemical cells, the corrosion rate may be more intense rather than protecting the metal.

4.12.2 Cathodic Inhibitors

If the reactions occurring at the cathodic area are controlled, then it is possible to inhibit the corrosion of metals. Corrosion takes place in metals depending on the environment to which it is exposed. The control of corrosion also can be achieved by the addition of cathodic inhibitors. These inhibitors are known for the slowing down of the reactions taking place at the cathodic area of the corrosion cell formed.

(i) In an acid medium, metal is corroded at the anodic area and the accompanying cathodic reaction is the liberation of hydrogen.

$$2H^+ + 2e \rightarrow H_2\uparrow$$

The reaction taking place at the cathodic area is controlled by the addition of cathodic inhibitors that reduces the diffusion of H^+ ions into the cathodic area or increases the hydrogen over-voltage by hydrogen evolution.

The addition of organic cathodic inhibitors like amines, substituted urea, thio urea, etc., reduces the diffusion of H^+ ions into cathodic area in acid medium. Another way is to deposit a thin film layer of arsenic or antimony on the surface of the metal as cathodic inhibitor, which induces hydrogen over voltage and as a result, corrosion reaction can be controlled at the anodic area.

(ii) In neutral solution, the dissolved oxygen of the medium gets reduced and forms OH^- ions at the cathodic area corrosion cell.

$$\frac{1}{2}O_2 + H_2O + 2e \rightarrow 2OH^-$$

The corrosion at the anodic region of the metal is controlled by either (i) eliminating oxygen from the medium or (ii) retarding diffusion of O_2 towards the cathodic area.

(iii) Depletion of oxygen content can be achieved by the addition of a reducing agent like sodium sulphite (Na_2SO_3) or the removal of oxygen by deaeration of the medium.

The addition of salts of Zinc, magnesium, nickel, etc. interact with OH^- ions produced at the cathodic area and forms insoluble metal hydroxides which forms a protective layer on the cathodic area of the metals and thus, protects the metal from corrosion.

4.13 Passivation of Metals

The oxide passivation film may form either on exposure to air or in electrolyte with presence of oxygen. The formation of passive film on the metal surface are determined by the Pourbaix diagram (Fig. 4.27) which depends on the electrode potential and pH of the medium. Low carbon steel does not corrode in concentrated nitric acid due to the protection effect of passive film; however, dilute nitric acid does not form stable passive film and, therefore, easily dissolves steel. Passive film is formed on the surface of aluminium, chromium, silicone and titanium in air, water and dilute acids.

Good corrosion and oxidation resistance of stainless steels is also a result of formation of chromium rich oxide passive layer. The passivation effect is achieved if the chromium content is not less than 10.5 percent. Other elements (nitrogen, titanium, nickel, molybdenum) enhance corrosion resistance of stainless steels.

Passivation is the formation of a thin film of oxidation products preventing further corrosion of the metal.

A damage of the passive film may cause intensive localized corrosion (pitting corrosion). The passive oxide layers are dissolved in electrolytes containing sulphates and chlorides. Phosphate and chromate ions stabilize passive films promoting repair of its defects.

4.13.1 Polarization Curve of Passivating Metal

One can notice that the *current density and corrosion rate* of certain metals and alloys increase as anodic polarization increases; while for others, the corrosion rate will drastically decrease to a stable value and become independent of the polarization applied. This observation is referred to as *passivating behaviour* and the environment of the metal is referred to as a *passivating system*. The behaviour metals and alloys are studied under various conditions and the information obtainable from polarization curves is important because they reveal the behaviour of materials for corrosion studies. The polarization curve technique is an excellent electrochemical tool that can be used to assess materials' resistance (passivation range) to corrosion; and a brief explanation of the curve is given below.

The solid line in Fig. 4.28 is a schematic potential versus current density ($E - I_g$) plot for a metal in a mild environment, such as an austenitic stainless steel in dilute sulphuric acid.

- AB represents *cathodic behaviour*.
- BG is the active zone. The metal is *not passivated* at its free-corrosion potential, B.
- AC and DC are *Tafel-type straight lines drawn for the reduction and oxidation reactions* of the normal metal dissolution ($M \rightarrow M^+ + e$).
- At *potentials more positive than B*, corrosion rate increases and reaches a maximum at the *passivation potential*, *G*, which is often given the symbol E_{pp}.
- The transition from active dissolution occurs as a solid species becomes more thermodynamically stable than the metal ion. A *protective film* begins to form and causes a sudden drop in corrosion current density in the region G to J.
- From J to P, the *passive zone*, the current density is maintained at a steady, low level, until
    - at *P*, *breakdown of the protective film begins*. It is here that the likelihood of pitting is greatest, and consequently
    - the potential E_c is often called the *critical pitting or breakdown potential*. It is a useful parameter in assessing pitting properties of materials. It should be noted that it is not an absolute parameter, and varies according to both metallurgical and environmental conditions.
- At potentials more positive than P, *the current density begins to rise* as more and more pits propagate.

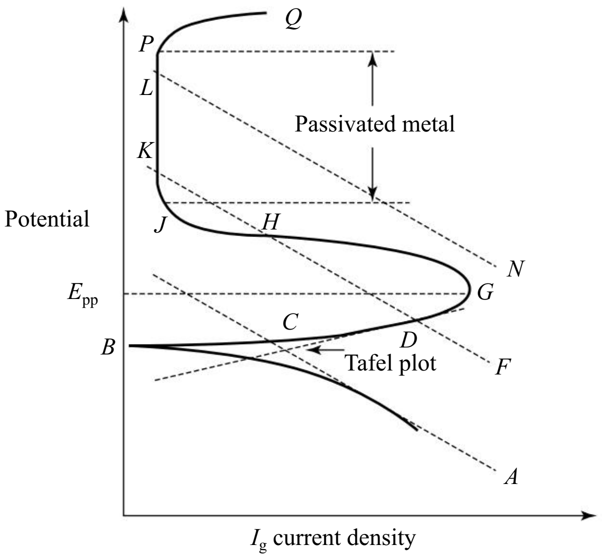


FIGURE 4.28 *The schematic potentiodynamic polarization behaviour of passivating metals.*

4.14 Effect of pH and Potential for Iron (Pourbaix Diagram) and the Polarization Curve of Iron

One can observe corrosion of metals and alloys in nature depending on the environment to which it is exposed. Corrosion reactions are electrochemical in nature. When a metal is subjected to corrosion, an electrochemical cell is set up within itself with the formation of anodic area and cathodic area. *Anodic reactions* involve oxidation: electrons appear on the right hand side of the equation. For example metallic iron can produce ferrous ions by the anodic reaction:

$$Fe \rightarrow Fe^{2+} + 2e^- \tag{4.1}$$

In a solution with *higher pH*, the anodic reaction produces a surface film of ferric oxide according to reaction (4.2).

$$2Fe + 3H_2O \rightarrow Fe_2O_3 + 6H^+ + 6e^- \tag{4.2}$$

Cathodic reactions involve electrochemical reduction: electrons appear on the left-hand side of the equation. In corrosion processes, the most common cathodic reaction is the electrochemical reduction of dissolved oxygen according to the equation:

$$O_2 + 2H_2O + 4e \rightarrow 4OH^- \qquad (4.3)$$

Hence *the reduction of oxygen at an electrode will cause a rise in pH due to the hydroxide ion production. If the medium is acidic, then the cathodic reaction may be the liberation of hydrogen at the cathode*

$$2H^+ + 2e \rightarrow H_2\uparrow \qquad (4.4)$$

The role of the electrode potential 'E' in defining the products of corrosion reactions can be readily seen in Fig. 4.29. This figure shows the corrosion products as a function of electrode potential and pH for iron at room temperature in the presence of water as solvent. *At negative potentials metallic iron itself is the stable form hence in this region no corrosion is possible, and this is referred to as the immunity condition*. At higher potentials and acidic pH values ferrous ions will form giving rise to active corrosion. Ferric ions are produced only at high potentials above 0.7 V.

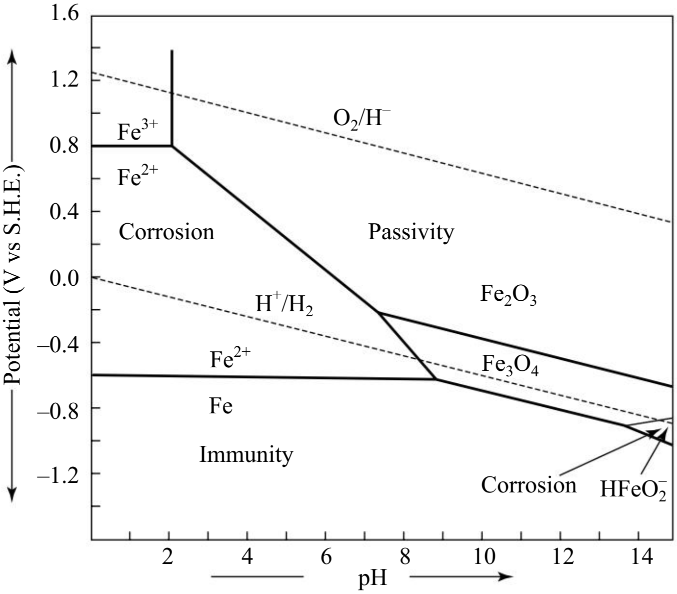


FIGURE 4.29 *Simplified potential-pH diagram for iron in a solution containing dissolved iron at 25°C.*

Figure 4.29 shows the stable forms of the iron element as a function of E and pH. If the pH lies on the alkaline side of neutral then *insoluble surface oxides* will form. The oxide Fe_3O_4, known as magnetite or black iron oxide, is produced at low electrode potentials. Low potentials are found in relatively stagnant conditions with a low oxygen partial pressure as in soil or inside boilers which have been treated to remove oxygen. The characteristic black surface of iron under these conditions is due to magnetite. At *more positive potentials*, the oxide formed is Fe_2O_3 and this is usually present as a thin adherent film. Since this oxide forms at the surface, its presence acts to block the surface reactions and hence corrosion rates are reduced. *This is called passivation and the oxide film on the surface is known as a passive layer.* The corrosion rate is very low in the passivation region of the diagram. Diagrams of the type shown in Fig. 4.29 are widely used in corrosion technology to predict the corrosion products which may be formed from a given metal under conditions specified by the axes of the figure.

In order to understand *the rate of the corrosion process*, *we must examine the electrochemical polarization curves* of the electrode reactions which take place on the metal surface. Figure 4.30 shows the polarization curve of iron in an acidic solution at room temperature. The rates of the electrode processes are controlled by the value of potential '*E*'. Thus, for a cathodic process in acidic solution producing hydrogen gas by the reduction of hydrogen ions, *the more negative the electrode potential the greater the surface concentration of electrons and the faster the reaction rate.*

$$2H^+ + 2e^- \rightarrow H_2(g)\uparrow \tag{4}$$

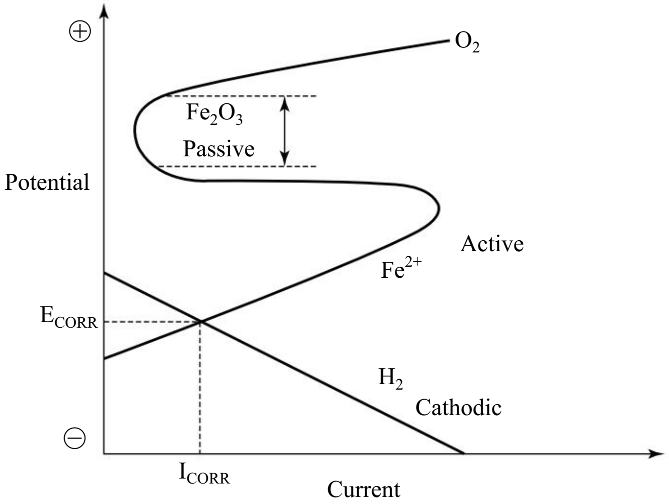


Figure 4.30 *Polarization curve of iron.*

Since the reaction rate is proportional to the flow of electrons (measured as a current *I*), the diagram shows the magnitude of *I* as a function of *E*.

The polarization curve of iron (Fig. 4.30) shows *how the potential, E, of the metal determines the electrochemical reaction rate and corrosion products*. Anodic reactions are accelerated by increasing potential in the positive sense. Ferrous ions are produced in the active state and this is the region in which corrosion will take place freely. At higher potentials, the reaction passes into the passivation region (as shown in Fig. 4.30) and passivation occurs. This is observed as a very small current flowing in this region. *The metal is protected by the passive film of ferric oxide on the surface.* It can be noticed that at very positive potentials the passive electrode surface will act as an anode to oxidize water to oxygen gas, but this does not occur in normal corroding systems. To find the corrosion rate under normal conditions it can be observed for the point on the diagram where the anodic and cathodic reactions intersect (Tafel plot). *At this point the rates of the anodic and cathodic reactions are equal* and the system is behaving as a closed circuit with all the electrons produced in the anode reaction being consumed in the cathodic reaction.

4.15 Application of Pourbaix Diagram for Corrosion Mitigation (or Prevention)

The types of electrochemical corrosion just described can be prevented or controlled by utilizing the current understanding of the principles underlying corrosion processes. This understanding has been the basis for the development of a number of corrosion-prevention measures.

Three corrosion-control measures are based on *electrochemical driving force* as shown in the Pourbaix diagram given in Fig. 4.31.

(i) Cathodic Protection

The corrosion control is achieved by cathodic protection by making the structure to be protected the cathode of a corrosion cell by bringing the potential of the structure into or near the immunity region of the Pourbaix diagram;

(ii) Anodic Protection

It makes the structure to be protected anode and an applied current brings the structure's potential into the passivity region of the Pourbaix diagram; and

(iii) Environment Modification

By controlling the pH of the structure to be protected into either the immunity or passivating regions.

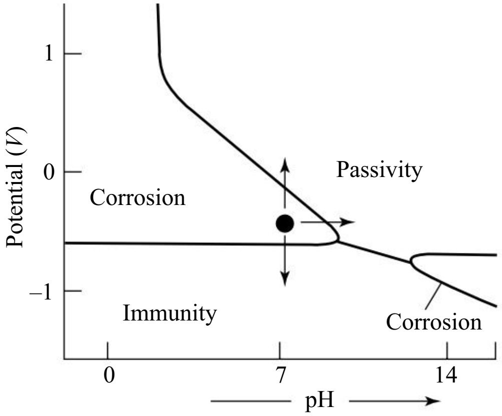


Figure 4.31 *Pourbaix diagram for iron showing how corrosion protection can be achieved by using measures that bring a corroding system into either the immunity or the passivity region.*

Review Questions

1. What do you mean by corrosion of a metal?
2. Mention the forms of corrosion of metals.
3. Describe the following types of corrosion:
   (i) dry corrosion and (ii) wet corrosion.
4. Discuss the electrochemical theory of corrosion. [VTU Mar, 2001, Jan, 2003]
5. What is metallic corrosion? Explain the electrochemical theory of corrosion of iron taking iron as an example. [VTU Feb, 2005]
6. Describe the corrosion of a metal with an example when the medium is
   (i) acidic, (ii) neutral, and (iii) alkaline.
7. Explain the electrochemical rusting of iron.
8. Explain the electrochemical mechanism of rusting of iron in humid atmosphere. Mention any four factors that affect the rate of corrosion. [VTU July, 2005]
9. Mention the types of corrosion observed in metals.
10. What is differential metal corrosion? Illustrate with an example. Illustrate the reactions involved in differential aeration corrosion with reference to the iron metal. [VTU August, 2000]
11. Explain the mechanism of differential aeration corrosion. Give two examples where differential aeration effect is seen. [VTU March, 2000]
12. What is differential aeration corrosion? Describe it with any example.
13. Explain pitting type corrosion in metals with an example.
14. Explain the type of corrosion occurring in the following cases.
    (i) Bolt and nut made from different metals are in contact with each other
    (ii) Deposition of small particles of dust on iron surface. [VTU Mar, 1999]
15. What is metallic corrosion? Pin holes on tin-coated iron are more prone to corrosion of iron than those on zinc-coated iron. Why? [VTU Mar, 1999]
16. What is the principle of waterline corrosion? Explain.
17. Describe stress corrosion with an example. Give reasons.
18. Explain what type of corrosion occurs when
    (i) screw and washer are made of different metals

(ii) presence of NaOH in mild steel boiler under stress. [VTU Feb, 2005]

19. Write a note on (i) Stress corrosion (ii) Waterline corrosion. [VTU Jan, 2003]
20. Describe season cracking with an example.
21. Describe caustic embrittlement.
22. Write a brief note on galvanic or corrosion series.
23. Mention the factors affecting the rate of corrosion.
24. Discuss the influence of following factors on the rate of metallic corrosion.
    (i) Nature of corrosion product. [VTU Jan, 2003]
    (ii) Polarization of the electrode surface. [VTU Aug, 2001]
    (iii) Electrode potential
    (iv) pH [VTU Jan, 2003]
    (v) Anodic and cathodic areas. [VTU Jan, 2004]
    (vi) Temperature [VTU Jan, 2003]
    (vii) Nature of metal [VTU July, 2004]
25. Describe in brief the factors affecting the corrosion of metal.
26. Describe how
    (i) anodic and cathodic area, and
    (ii) the electrode potential values affect the corrosion.
27. Explain the effects of the following factors on corrosion of metal:
    (i) pH, (ii) Humidity and (iii) The nature of the corrosion products.
28. What do you mean by 'corrosion control'?
29. How corrosion control can be achieved? Mention the various techniques adopted.
30. Write a brief note on proper selection and design of materials for corrosion control of metal.
31. What is protective coating? Mention the types of protective coatings.
32. Explain how corrosion can be minimized using the following techniques.
    (i) design and selection of materials, (ii) cathode protection [VTU Mar, 2001]
33. What do you mean by 'metal coating'?
34. Mention the two techniques of metal coating.
35. What is anodic metal coating? Explain the concept with an example.
36. Explain the anodic protection given to metals exposed to corrosion product? [VTU Aug, 1999]
37. What is cathodic metal coating? Discuss with an example.
38. Describe hot dipping of metal coating.
39. What is galvanization? Describe the galvanization process for iron/mild steel.
40. Discuss the following corrosion control methods: (i) Galvanizing and (ii) Tinning. [VTU Feb, 2002]
41. What is tinning process? Explain with neat diagram the tinning process.
42. What is electroplating of a metal? What are the merits of electroplating of a metal.
43. Describe the process of electroplating with a neat diagram.
44. What do you mean by 'inorganic coating'? Mention two types of inorganic coatings generally adopted.
45. What is anodization? How does it protect the metal from corrosion?
46. What is anodizing? Explain anodizing of aluminium. [VTU Mar, 2001, Aug 2001, Feb, 2002]
47. Describe the process of anodization of aluminium metal.
48. What is phosphating of a metal? Describe phosphating process of coating a metal.
49. Explain the following corrosion control methods: (a) Phosphating (b) Organic coatings.
50. What is organic coating?
51. What is the working principle of cathodic protection? Mention the methods of cathodic protection.
52. What is cathodic protection? Explain the sacrificial anode and impressed current techniques for prevention of corrosion? [VTU July, 2003, 2004]
53. Describe the sacrificial anodic method with an example.
54. Describe the impressed current cathodic protection method with an example.
55. What do you mean by 'anodic protection'. Describe anodic protection method to control corrosion.
56. What are corrosion inhibitors? Name few anodic and cathodic inhibitors.

57. What are anodic inhibitors? Explain how corrosion control can be achieved.
58. What are cathodic inhibitors? Describe cathodic inhibition to control the corrosion of metals.
59. Write a brief note on corrosion inhibitors. [VTU July, 2005]
60. (a) Explain the electrochemical mechanism of rusting of iron in humid atmosphere. Mention any four factors that affect the rate of corrosion.
    (b) Write notes on the following.
        (i) Cathodic Protection
        (ii) Corrosion inhibitors
    (c) Define the terms Decomposition potential and Overvoltage. [VTU Jan, 2005]
61. (a) Explain the electrochemical theory of corrosion with reference to iron.
    (b) Discuss the following corrosion control methods:
        (i) Cathodic protection
        (ii) Anodic protection system.
    (c) Explain why pinholes on tin-coated iron are more prone to corrosion than pinholes on zinc-coated iron. [VTU Jan, 2006]
62. (a) Explain the following types of corrosion:
        (i) Pitting corrosion.
        (ii) Waterline corrosion.
    (b) What is cathodic protection? Explain sacrificial anode and impressed current techniques for prevention of corrosion.
    (c) What are the types of protective coatings? How do protective coatings help in controlling corrosion? [VTU July, 2006]
63. (a) Explain stress corrosion with examples.
    (b) What are corrosion inhibitors? Explain how corrosion is controlled by using anodic and cathodic inhibitors?
    (c) Write a brief note on the effect of following factors on the rate of corrosion.
        (i) Nature of metal
        (ii) Hydrogen overvoltage
        (iii) Relative areas of anode and cathode. [VTU Jan, 2007]
64. (a) What is meant by wet corrosion? Explain the differential metal corrosion using galvanic series.
    (b) What are inhibitors? Explain the types of inhibitors employed to control corrosion with examples.
    (c) Write a short note on Phosphate coating. [VTU July, 2007]
65. (a) Explain the differential aeration corrosion with a suitable example
    (b) How does the following factors affect the rate of corrosion:
        (i) Nature of corrosion product
        (ii) Temperature (iii) pH.
    (c) Write a note on (i) Galvanizing (ii) Tinning. [VTU Jan, 2008]
66. (a) What is corrosion? Explain the electrochemical theory of rusting of iron in moist atmosphere.
    (b) Write a note on (i) Pitting corrosion (ii) Cathodic protection
    (c) Define (i) Decomposition potential and (ii) Overvoltage [VTU July, 2008]
67. What do you mean by passivation of metals?
68. Write a note on passivity of metals.
69. Describe Pourbaix diagram for iron.
70. Describe the polarization curves of a metal.
71. What is the effect of pH and potential for iron and discuss with its Pourbaix diagram.
72. Discuss the application of Pourbaix diagram for corrosion control or mitigation.
73. Describe the factors affecting the rate corrosion by nature of metal. [JNTU, 2014]
74. Explain the mechanism of oxidation corrosion. [JNTU, 2014]
75. Explain electrochemical corrosion by H_2 evolution and oxygen-absorption mechanism. [Pune University, Dec, 2014]
76. What is the principle of cathodic protection and explain it with any suitable method. [Pune University, Dec, 2014]
77. What is cathodic protection? Explain any one method in detail. [Pune University, May, 2016]
78. (a) What is Pilling–Bedworth ratio? Give four types of oxide films formed on the surface of metal with examples.
    (b) Explain galvanization with neat-labelled diagram to protect iron from corrosion,

(c) What is the principle of cathodic protection? Explain it with any one method.
[Pune University, 2014]

79. (a) Explain various factor affecting corrosion on the basis of nature of metal.
(b) Describe anodic protection of metal for corrosion.
(c) Compare: Galvanizing and Tinning.
[Pune University, April, 2013]

80. (a) Give the mechanism of electrochemical corrosion.
(b) Explain nature of metal factor affecting rate of corrosion.
(c) What are the types of metallic coatings? Which is preferred? Why?
[Pune University, Dec, 2013]

81. (a) Give the Pilling and Bedworth ratio and its significance. Give the oxidation reaction involved and state the type of film formed on the surface in the case of Mg, Cr and Mo.
(b) Explain the process of galvanizing with labelled diagram. Give the applications and limitations of this technique.
(c) State the different types of corrosion inhibitors with their examples. Discuss in brief their role in corrosion prevention.
[Pune University, Dec, 2015]

82. (a) What is dry corrosion? Explain the mechanism of oxidation corrosion with suitable figure and reactions.
(b) Explain how nature of metal affects the rate of corrosion.
(c) What are electroless coatings? Explain with suitable examples and Give its applications.
[Pune University, June, 2015]

83. (a) Explain the mechanism of dry corrosion. Discuss the oxidation corrosion in case of Mg, Cr and Mo.
(b) Give conditions under which the wet corrosion occurs. Explain the mechanism of wet corrosion by hydrogen evolution with example. [Pune University, Oct., 2012]

84. (a) What is stress corrosion? Explain stress corrosion in boilers due to alkali with chemical reactions.
(b) How does the following factor affect the rate of corrosion?
  (i) Nature of corrosion product,
  (ii) temperature, and
  (iii) pH
(c) What is cathodic protection? Explain sacrificial anodic method and impressed current method.
(d) Explain the electrochemical theory of corrosion by taking iron as example.
[VTU, June/July, 2015]

85. (a) Explain electrochemical theory with an example.
(b) Discuss the differential corrosion with an example. [VTU, June/July, 2016]

86. (a) Define the term corrosion. Explain pitting corrosion and waterline corrosion.
(b) Explain the following factors affecting rate of corrosion: (i) Nature of the metal and (ii) corrosion product.

87. (a) What is galvanization? Explain the process of galvanisation.
(b) What is cathodic protection? Explain sacrificial anode and impressed current method. [VTU, June/July, 2016]

5

Metal Finishing

Chapter Outline

Introduction to metal finishing. Technological importance of metal finishing. Methods of metal finishing. Manufacturing of electronic components—printed circuit board fabrication, electrolytic capacitor, aluminium electrolytic capacitors. Aluminium capacitor manufacturing process. Electrochemical techniques of forming, machining and etching—electroforming, electromachining, electrochemical polishing, electrochemical etching. Electrochemical etching of copper, electrolytic cell. Electrochemical deposition/plating process—electrolysis, metal deposition and the electrical relationships. Illustrated examples, current efficiency. Polarization, decomposition potential, overvoltage and the electro-deposit—application of potential for spontaneous and nonspontaneous irreversible process. Polarization. Decomposition potential–experimental determination of decomposition potential, significance of decomposition potential. Overvoltage, factors affecting overvoltage. Electroplating—mechanism of electroplating, characteristics of good and poor deposits, factors that influence the nature of deposit, experimental determination of throwing power—Haring-Blum cell. Electroplating process. Surface preparation of the substrate, steps involved in cleaning the surface for electroplating, electroplating of copper, applications of copper plating. Gold plating—basic principles of electroplating and electropolishing of gold, factors affecting the electroplating process, surface preparation, gold plating systems, applications. Electroplating of Nickel, applications of nickel plating. Electroplating of chromium, chemistry of electroplating of chromium. Electroless plating, advantages of electroless plating, process of electroless plating, comparison of electroplating and electroless plating. Electroless plating of copper, mechanism of electroless plating, electroless plating of Nickel, electroless plating of nickel on aluminium, preparation of printed circuit board (PCB) by electroless plating.

5.1 Introduction to Metal Finishing

Many elements known to us are metals. A large number of them are in common use. Starting from gold and silver, which are used in ornaments, we use aluminium, copper and their alloys and stainless steel in household articles and steel in the manufacture of machines and machinery parts. But none of these metals possess all properties such as (i) resistance to corrosion (ii) resistance to wear and tear (iii) resistance to impact, etc., required for their use in a variety of engineering applications. In order to satisfy the conditions of their applications, the surfaces of metals are suitably modified and any such surface modification is called metal finishing.

Metal finishing is defined as a surface phenomenon involving the deposition of a thin layer of a noble metal on a base metal or a polymer or conversion of a thin layer of the surface into the oxide (chromate or phosphate) of the metal.

5.2 Technological Importance of Metal Finishing

Many articles of daily use have undergone metal finishing. The primary object of metal finishing once was to give the object a decorative appearance, but there are a number of more important technological applications of metal finishing. These include the following:

- Provides a bright, shiny appearance.
- Corrosion resistant.
- High resilience resistance to wear to tear
- To impart thermal and electrical conductivity
- To increase abrasion and impact resistance
- To increase hardness of the surface
- To increase the solderability
- To produce reflectivity and appearance
- Chemical resistance
- Hardness.

Apart from these metal finishing also finds applications in the following areas:

1. Manufacture of electrical and electronic components like printed circuit boards
2. Electroforming
3. Electromachining
4. Electrochemical etching, polishing and engraving
5. Electroforming of objects
6. Electrotyping and
7. Material restorations.

5.3 Methods of Metal Finishing

Various metal finishing methods are available. The following are some of the most important methods.

1. Electroplating of metals and alloys
2. Electroless plating and immersion coating
3. Thermal spray coating
4. Vapour deposition technology of coating
5. Chemical vapour deposition (CVD) coating

6. Chemical conversion coatings involving the conversion of a thin layer of the metallic surface into its oxide, chromate, phosphate, etc. These are called anodizing, chromating and phosphating, respectively.

5.4 Manufacturing of Electronic Components

5.4.1 Printed Circuit Board Fabrication

The following steps have been adopted to manufacture PCB:

1. Components layout designing
2. PCB layout designing
3. Transferring the PCB layout design onto the PC board laminate
4. Developing or etching the PCB
5. Other operations like drilling, cutting, tinning, etc.

(i) Components Layout Designing

Component layout designing is nothing but planning the positions of different components constituting the circuit and then showing their interconnections as per the circuit diagram.

(ii) PCB Layout Designing

The *printed circuit board* layout designing is nothing but the mirror image of components layout. The components layout is drawn looking from the components side whereas the PCB layout is drawn from copper side.

(iii) Transferring PCB Layout onto PCB Laminate

Phenolic laminate boards are generally used for commercial applications.

The surface of the copper side of the laminate is first cleaned with petrol or acetone or alcoholic spirit to make it completely free from any contaminants.

Using a ballpoint pen and a carbon paper, the *printed circuit board* layout is transferred onto the copper side of the laminate The holes are drilled in proper place to insert the respective components of the circuitry. The circuitry track is then masked with paint or photo resist material or any corrosion-resistant stripes or else, the copper is covered with a photo-resist layer (photosensitive surface) and is selectively exposed to a UV light source. The photo-resist on the exposed areas is removed by a chemical agent and subsequently the exposed areas are chemically or electrochemically etched.

(iv) Etching

In Etching process, all the excess copper is removed leaving behind only the painted/photo-resist material/stripes of track pattern. To do etching, the masked portion of PCB board laminate is placed in a flat plastic tray with the copper side facing upwards. PCB should be fully immersed in ferric chloride solution. The solution is prepared by adding about 40–50 g of ferric chloride to every solution should be nicely stirred and a few drops of hydrochloric acid added if you want to speed up the etching process. The laminate is thoroughly washed in water after etching is complete. The paint/photo-resist material/stripe is then removed with a suitable solvent (alcohol or acetone).

(v) Drilling-Tinning

The next step after etching is drilling of holes. Holes are drilled with tiny drill bits made of tiny tungsten carbide. Drilling is made by automated drilling machines and the placements are controlled by tapes or drill file.

(vi) Tinning the track

Electrochemical process is to build copper in the holes and on the trace area. The oxidation of copper portions circuitry track can be prevented by either tinning by any technique or by giving it a coat of some insulating varnish.

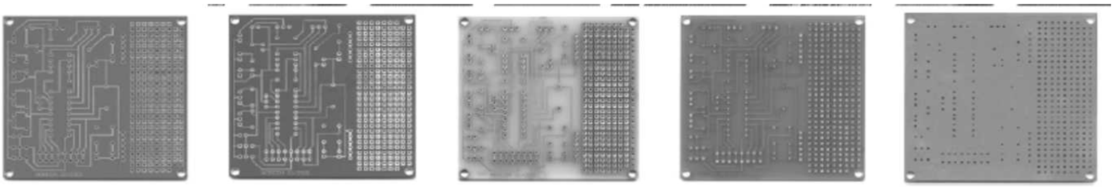
(vii) Solder mask

Apply solder mask area to entire board with the exception of solder pads.

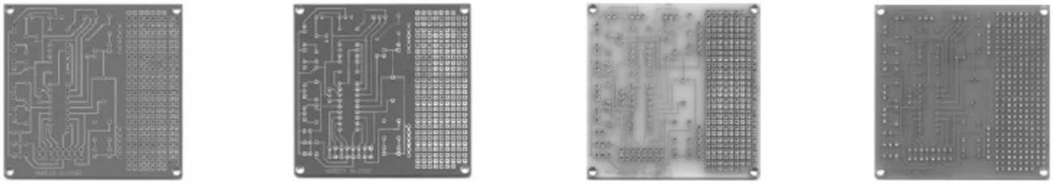
(viii) Solder coat

Apply solder to pads by immersing into tank of solder. Hot air knives level the solder when removed from the tank.

The entire process of manufacture of PCB is shown in Fig. 5.1(a)–(i).



(a) (Design) (b) (Copper laminate) (c) (Drilled holes) (d) (Electroless copper through holes) (e) (Apply developed film)



(f) (Pattern plate) (g) (Solder mask) (h) (Stripe & etch) (i) (Resist coat)

Figure 5.1 *(a)–(i) Process of PCB manufacture.*

5.4.2 Electrolytic Capacitor

(a) How capacitors work

When a voltage is applied across the electrodes consisting of a dielectric material layered between two metal electrodes, an electrical charge proportional to the voltage is stored in the capacitor; where, 'C' is the *capacitance* of the capacitor.

The capacitance 'C' of a capacitor can be calculated from the following parameters

(i) surface area of the electrodes (S)
(ii) the distance between the electrodes (t) and
(iii) the dielectric constant of the dielectric (ε).

$$C = \varepsilon_0 \cdot \varepsilon \cdot \frac{S}{t} = 8.855 \times 10^{-12} \times \frac{\varepsilon \cdot S}{t}$$

In the formula above, ε_0 represents the permittivity of free space (8.85×10^{-12} F/m).

It is possible to increase the capacitance by either increasing the dielectric constant/increasing the electrode surface area (*S*), or by decreasing the distance between the electrodes (*t*).

Example The dielectric constant of an aluminium oxide layer averages between 7 and 8. The capacitance of aluminium foil can be enhanced (100–120 times) by roughening the surface area of the foil.

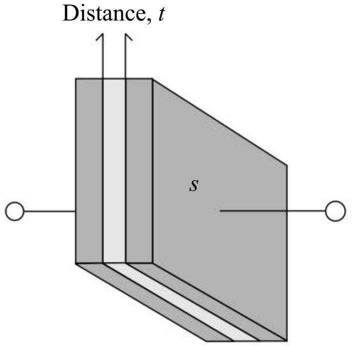


Figure 5.2 *Shows the basic concepts of how capacitors function.*

5.4.3 Aluminium Electrolytic Capacitors

The anode of the aluminium electrolytic capacitor consists of high purity aluminium foil with Al_2O_3 thin film dielectric on its surface. The construction of the capacitor is made using an electrolytic paper containing an electrolytic solution and an aluminium electrode foil for contacting the cathode. The thickness of the anode oxide thin film is the distance between the electrodes (*t*) in Fig. 5.2. The thickness of the anode oxide thin film in an aluminium electrolytic capacitor is selected by the required withstands voltage. By adjusting the distance '*t*', to minimum, large amounts of charge can be stored in the capacitor. The surface of the aluminium foil is roughened to increase the surface area (*S*) and the dielectric constant (ε) will also be large.

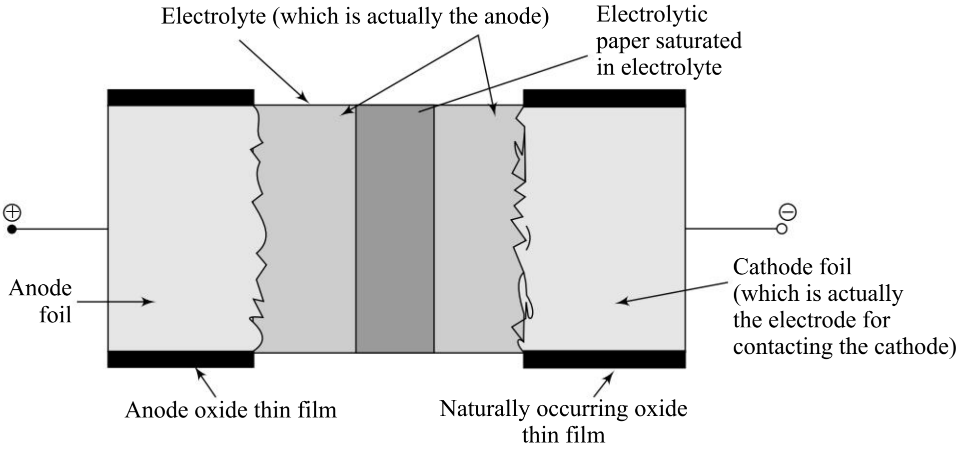


Figure 5.3 *A typical anode oxide thin film and the cathode foil of a capacitor.*

(a) Surface roughing (etching)

A high purity aluminium foil of thickness (50–100 μm thick) is used as anode of capacitor.

The cathode foil material uses an aluminium foil that is at least 99 percent pure and about 15–60 μm thick. *Because the capacitance is proportional to the surface area of the electrodes, the effective surface area is increased by roughening (etching) the surface of the aluminium foil before growing the dielectric film.* This process of surface roughening is referred to as 'etching'.

There are two typical etching processes.

(i) The first option submerges the aluminium foil in hydrochloric acid (physical etching)

(ii) A secondary option is electrolysis where the aluminium as the anode is placed in an aqueous hydrochloric acid solution (electrochemical etching).

The following parameters will play a major role on electrochemical etching:

(i) the waveform of the electrical current,
(ii) the composition of the solution, and
(iii) the temperature.

The etching method can be determined by the desired capacitor performance. Generally, it is possible to achieve etching multipliers (the ratio between the surface area of the smooth foil and the effective surface area of the etched foil) approximately between 3 and 120. The foil is then rinsed thoroughly with water.

(b) Forming (anode oxidation)

The anode oxidation or generally, known as 'forming' process which creates an electrically insulating oxide (to provide the withstand voltage) on the aluminium surface by performing anode oxidation in the electrolytic solution used for the growth. The produced chemical film is used as the anode thin film.

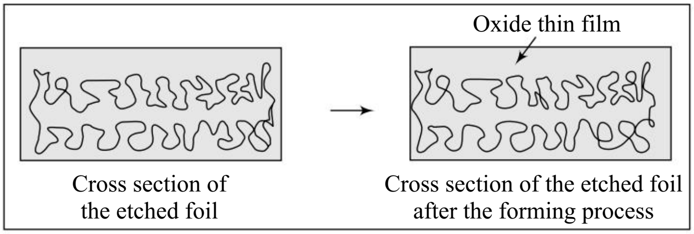


FIGURE 5.4(a) *Cross-section of Al foil before and after forming process.*

The anode oxidation is achieved, as shown in Fig. 5.4(b), by applying a voltage to the submerged foil found in the electrolytic solution used for growing the oxide film. Generally, the electrolytic solution is an aqueous solution such as ammonium, boric acid, ammonium phosphate or ammonium adipic for acid.

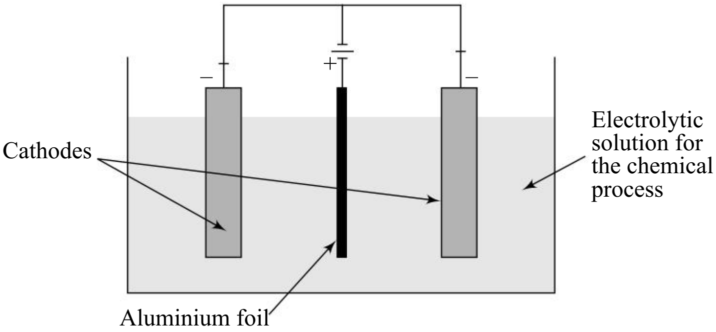


FIGURE 5.4(b) *Anodic oxidation of aluminium foil.*

During the anode oxidation (DC electrolysis), Al_2O_3 is produced by a reaction between the water and the aluminium's Al^{3+} ions. The thickness of the grown thin film is nearly proportional to the applied voltage with approximately 1.0–1.4 nm per volt.

(c) Electrolyte

Aluminium electrolytic capacitors are made by layering the electrolytic paper between the anode and cathode foils, and then coiling the result. The process of preparing an electrode facing the etched anode foil surface is extremely difficult. Therefore, the opposing electrode is created by filling the structure with an electrolyte. Due to this process, the electrolyte essentially functions as the cathode.

5.4.4 The Aluminium Electrolytic Capacitor Manufacturing Process

The manufacturing process of aluminium electrolytic capacitors are as follows:

1. Slit

Source rolls of anode foil, cathode foil and electrolyte paper are taken and specified width of each are cut as shown in Fig. 5.5a.

2. Lead attachment

By crimping or ultrasonic welding, the leads are attached to the cathode and anode foil electrodes (Fig. 5.5b).

3. Coiling

The electrolytic paper is inserted between the anode foil and the electrode foil in such a manner that it gives a coiled construction which separates the electrodes. When the structure is coiled in this manner it is known as an 'element' (Fig. 5.5c).

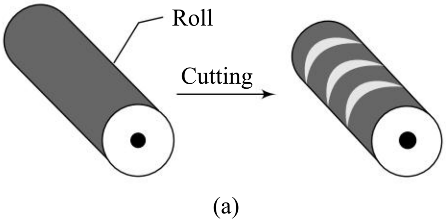


FIGURE 5.5(a) *Cutting the foil and the paper into specified length.*

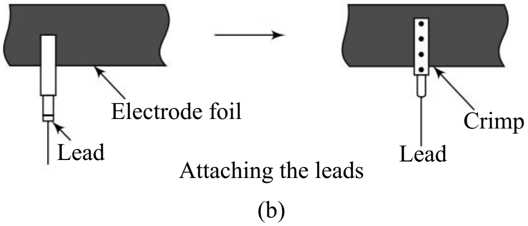


FIGURE 5.5(b) *Attaching the lead to the cathode and anode foil electrode.*

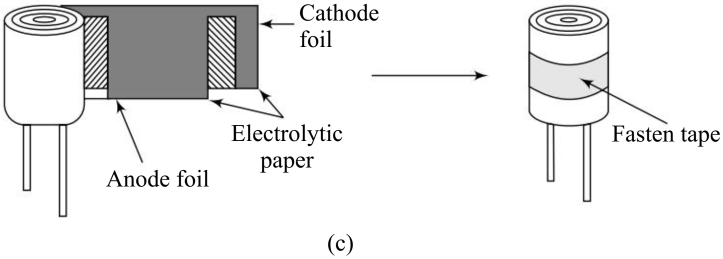


FIGURE 5.5(c) *The electrolytic paper is layered between the anode foil and electrode foil by coiling.*

4. Filling

Electrolyte is introduced to the element using vacuum technique.

5. Assembly

The rubber seal or plug is attached to the filled element. Then, the element is covered with a case (Fig. 5.5d).

6. Sleeve attachment

The case is slipped into a sleeve printed with the rated voltage, capacitance and polarity. After which, the sleeve is heat shrunk on to the case (Fig. 5.5e).

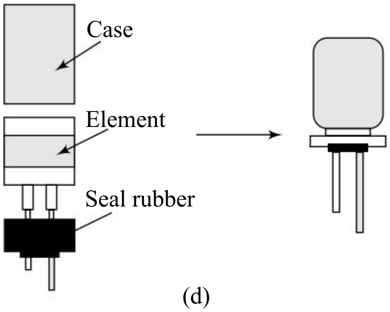


Figure 5.5(d) *Assembling of all the component parts into a capacitor.*

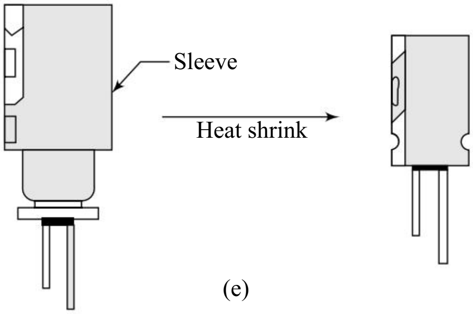


Figure 5.5(e) *Sleeve attachment to the capacitor.*

7. Ageing

The specified voltage is applied to grow an anode oxide thin film on the cut surface end and on any flaws in the formed oxide thin film.

5.5 Electrochemical Techniques of Forming, Machining and Etching

5.5.1 Electroforming

Electroforming is a highly specialized process of metal part fabrication using electrode position in a plating bath over a base form or mandrel which is subsequently removed.

A metal object of desired shape is synthesized by controlling the electrode position of metal passing through an electrolytic solution onto a metal. In the process, a metal skin is built up on a metal surface. Essentially, a metal part is fabricated from the plating itself.

The basic principle involved in the process is that metal particles are dissolved in an electroplating solution and electrochemically depositing them onto a conductive surface (Fig. 5.6).

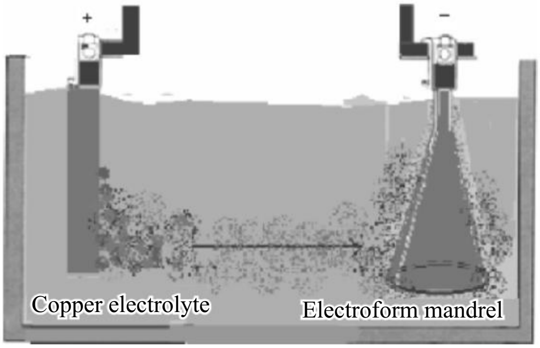


Figure 5.6 *An electroforming cell with a mandrel.*

Electroplating has the advantage over machining metal by the fact of being able to reproduce intricate patterns and also form complex shapes. A plated surface will have varying thickness; however, in tool manufacturing a thickness of 3–5 mm is generally used.

The electrodeposited material (skin) on the mandrel/mould is much thicker and can exist as a self-supporting structure if the original matrix is removed. The object being electroformed can be a permanent part of the end product or can be temporary (as in the case of wax), and removed later, leaving only the metal form, the 'electroform'.

The advantage of the electroforming process is that it faithfully reproduces the form or mandrel exactly, to within one micrometre, without the shrinkage and distortion associated with other metal forming techniques such as casting, stamping or drawing. And, since the mandrel is machined as an outside surface, close dimensional tolerances and high surface finishes can be held and maintained on complex interior configurations.

In the basic electroforming process, an electrolytic bath is used to deposit nickel or other electroplatable metal onto a conductive patterned surface, such as glass or stainless steel. Once the plated material has been built up to the desired thickness, the electroformed part is stripped off the master substrate. This process allows high-quality duplication of the master and therefore permits quality production—at low unit costs with high repeatability and excellent process control.

Electroforming is an additive process and herein lies electroform's superiority over etched metal and other forming technologies. The precision and resolution inherent in the photographically produced conductive patterned substrate, allows finer geometries to be produced to tighter tolerances while maintaining superior edge definition with a near optical finish.

(i) Comparison to other fabrication methods

1. Temperature and humidity does not appreciably affect electroforming
2. Electroformed parts have excellent light transmission when used in optical application (encoders, aperture plates, slits, etc.)
3. The advantage of the electroformed parts is that they have very low mass (weight and inertia)
4. The parts obtained from electroforming are electrically conductive and essentially unbreakable
5. Electroformed metal is extremely pure, with superior properties over wrought metal due to its refined crystal structure. Multiple layers of electroformed metal can be molecularly bonded together, or to different substrate materials to produce complex structures with 'grown-on' flanges and bosses.

(ii) Need for electroforming

Electroforming techniques are very useful for the production of the following

- Complex shaped objects can be manufactured
- the products manufactured are very precise and accurate
- Manufacture of large or small products
- Desired moulds can be electroformed
- Production of foils and mesh-products
- Metallic screen-printing on any part of the metal.

(iii) Applications

The applications of electroforming can be stretched to the following:

- Production of reflectors, Nose cones
- Manufacture of microcomponents
- Manufacture of complex waveguides

- Metal bellows
- Production Heat exchangers, microfilters
- Decorative ware

5.5.2 Electrochemical Machining (ECM)

Electrochemical machining (ECM) is an important method for removing metal by anodic dissolution in a conducting electrolyte. Metal removal is effected by a suitably shaped electrode (cathode tool).

Electrochemical machining (ECM) is a versatile method by means of which metal is dissolved anodically; and offers a number of advantages over other machining methods. ECM forming is carried out so that the shape of the tool electrode is transferred onto, or duplicated in the workpiece.

The basic process involved for electrochemical machining works is the electrolytic chemical reaction. ECM is a relatively fast method, with important advantages over more traditional machining methods (mechanical, laser, electrochemical discharge) since it can be applied to any electrically conductive material regardless of its hardness. Also there is no need to use a tool made of a harder material than the workpiece. Moreover, ECM is able to produce smooth, stress and crack-free surfaces, which is of major importance for workpieces which have to function in extreme environments (temperature, pressure, etc.).

From a methodology point of view, a difference can be made between:

- In the ECM imaging processes, the cathode tool shape moves with a certain feed rate (0.1–10 mm/min) towards the workpiece and its shape is reproduced in the workpiece
- Internal structures in a workpiece such as holes, cavities, tubes can be efficiently executed in ECM process
- Controlling the narrow electrolyte gap between cathode tool and workpiece is the major issue in ECM imaging process, since dissolved metal ions, gas and heat have to be removed from this gap.

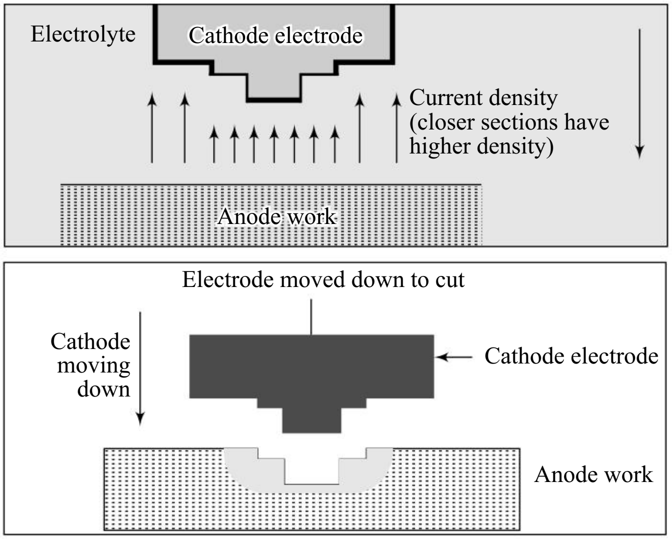


FIGURE 5.7(a) *Higher current density for a closer gap between the anode (workpiece) and the cathode for dissolution of metal ions at the anode.*

For internal ECM processes, a high forced electrolyte flow is still required for removal of metal ions, gas and heat, but in addition the shape of the produced structures depends strongly on the process conditions (electrolyte flow rate, DC/pulsed current or potential difference) and cell configuration (cathode tool shape, position and movement).

(i) Processes of electrochemical machining

A typical ECM has an electrode and *workpiece* (conductor) are placed in an electrolyte, and a potential voltage is applied. A direct current with high density and low voltage is passed between a workpiece (anode) and a preshaped tool (cathode). At the surface of the workpiece, metal is dissolved into metallic ions by the chemical reaction, and so the shape of the tool is copied into the workpiece.

On the anode (+ve) side the metal atoms ionize (lose electrons) breakfree of the workpiece, and travel through the electrolyte to the electrode (a cathode; has a −ve charge; a surplus of electrons).

- The work taking the electrode shape in the workpiece is due to the variation in the current density. The electrode is fed with a constant velocity, and the electrolyte is fed through the tool. The tool is designed to eliminate deposition of the ionized metal on the electrode
- For efficient working, an electrode gap typically of 0.1–0.2 mm is maintained
- The process is ideal for low machinability, or complicated shapes
- The process can be adopted for V = 8 to 20V, I = >1000A
- The tool wear and tear is very little.
- Forces are large with this method because of fluid pumping forces.

The electrolyte

A high velocity liquid solution forced to flow through the gap between the two electrodes.

It intensifies the mass/charge transfer through the layer near anode and to remove the sludge (dissolution products e.g. hydroxide of metal), heat and gas bubbles generated in the gap.

(ii) Characteristics of electrochemical machining

(a) Precise accuracy in shape and high rates of metal removal the process can be achieved at very high current densities at relative low voltage usually (8–30 V).
(b) A very narrow machining gap (of the order of 0.1 mm) is maintained by feeding the tool electrode in the direction of metal removal from the work surface, with feed rate from 0.1 to 20 mm/min.
(c) Dissolved material, gas, and heat are removed from the narrow machining gap by the flow of electrolyte pumped through the gap at a high velocity.

(iii) Materials used in electrochemical machining

(a) ECM is capable of machining any electrically-conductive material can be used in ECM.
(b) Removal rate in ECM is independent of the hardness, toughness and other properties of the material being machined, i.e., aluminium, titanium, steel, brass.

(iv) Advantages

The advantages of ECM technique are the following:

- All possible shapes and contours can be electromachined
- Current and metal removal rate can be controlled
- Tool once created can be used repeatedly without any damage
- No burrs created
- No residual stress introduced.

(v) Disadvantages

(a) Equipment and tools are generally expensive.
(b) Electrolytic solution is hazardous to environment as well as equipment.
(c) Machine parameters; Working voltage between the tool electrode (cathode) and workpiece (anode)
  - Machining feed rate
  - Inlet and outlet pressure of electrolyte (or flow rate)
  - Inlet temperature of electrolyte

(vi) Applications

(a) Drilling and sinking operations in the manufacture of die's, press and glass-making moulds, the manufacture of turbine and compressor blades for gas-turbine engine, the generation of passages, cavities, holes and slots in parts
(b) Electrochemical broaching as a method of making splines, gear sizing, reducing the wall thickness of shaped parts from high-temperature and titanium alloys, and preliminary generation of screw threads
(c) Drill complex-shaped hole or several such holes in a single run and also holes with an offset from the centreline

(vii) Characteristics of ECM

(a) Anode metal removal rate

If 'm_d' is the mass (kg) of metal dissolved in ECM process, $m_d = V \times d$

Where 'v' (m^3) is the corresponding volume and 'd' (kg/m^3) is the density of the anode metal.

By the use of Faraday's law. We have

The volumetric *removal rate* of anode metal (m^3/second) = $\dfrac{A \times I}{n \times F \times d}$

where

'A' (kg/mol) is the atomic weight of the anode metal,
'I' (ampere) is the *current flowing*,
'n' is the ionic charge of the anode metal, and
'F' Faraday constant = 96,487 coulombs/mol.

(b) Rates of machining

By use of Faraday's laws the rates at which metals can be electrochemically machined can be calculated.

$$m_d = \frac{A \times I \times t}{nF}$$

where

'm_d' (kg) is the mass of metal electrochemically machined by current,
'I' (ampere) passed for a time 't' (second).
The quantity '$A/n \cdot F$' is called the electrochemical equivalent of the anode metal as mentioned before.

5.5.3 Electrochemical Polishing

In electro polishing process, the object to be electro-polished is immersed in an electrolyte bath containing, typically, phosphoric and sulphuric acid and subjected to a direct electrical current. The object is maintained anodic, with the cathodic connection being made to a nearby metal conductor (see Fig. 5.8). *In electro-polishing, the metal is removed ion by ion from the surface of the metal object being polished.*

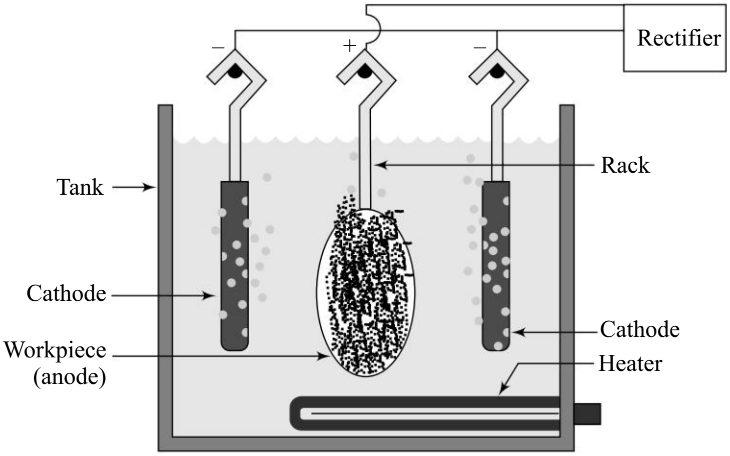
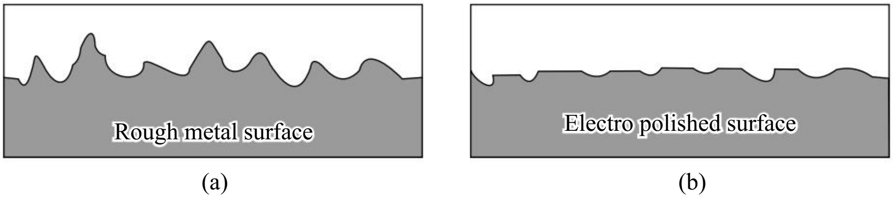


Figure 5.8 *An electropolishing cell.*

(a) The basics of the electropolishing process

Electropolishing is accomplished by creating an electrochemical cell in which the material to be polished is the *anode*. A *cathode* is formed to mirror the geometry of the work-surface and the two are submerged in a heated electrolyte bath (Fig. 5.8). When a DC current is applied, the electrical charge forces metal ions to be dissolved from the work-surface.

Electropolishing is an electrochemical process by which surface material is removed by anodic dissolution. Sometimes referred to as 'reverse plating', electropolishing actually removes surface material, beginning with the high points within the microscopic surface texture [Fig. 5.9(a)]. By removing these points, the electropolishing process will improve the surface finish, leaving a smoother and more reflective surface [Fig. 5.9(b)].



(a) (b)

Figure 5.9 *(a) Metal surface before (b) after polishing.*

The basic criterion for electropolishing is the difference in current density across the surface. The current density is observed to be greater at the high points and lesser at the low points within the surface profile. *The rate of the electropolishing reaction is directly proportional to the current density.* During electropolishing, due to the increased current density at the raised points forces the metal to dissolve faster at these points and thus tends to level the surface material. After the electropolishing treatment, the workpiece is passed through a series of steps to neutralize, rinse, clean and dry the surfaces.

Electropolishing delivers a smoother, more reflective surface that reduces product adhesion and improves surface cleanability. This process improves the near surface chemistry of the material, and promotes the formation of an improved corrosion-resistant surface layer.

(b) Benefits of electropolishing

(i) Improved corrosion resistance

All forms of corrosion begin on or near the surface. Unfortunately, all fabricating and handling practices invariably degrade surface conditions and surface properties. Electropolishing removes surface material and surface contaminants. Electropolishing dissolves free iron, inclusions, and embedded particles from the surface of the material steel.

Electropolished stainless steels are better suited to resist the onset of pitting corrosion, crevice corrosion, localized galvanic corrosion, stress corrosion cracking and microbiologically influenced corrosion.

(ii) Improved surface finish

Superior micro finishes can improve seals, lower friction, reduce real surface area, allow for easier sanitation, and improve heat and light reflection.

(iii) Reduced product adhesion and ease of cleaning

Electropolishing gives micro-finish product which provides poor adhesion and contamination build-up. Reduced adhesion can limit product build up and significantly lengthen duty cycles. When cleaning is necessary, cleaning operations can be completed in less time and with less effort.

(iv) Deburring

Electropolishing is the best suited process for deburring. During the electropolishing process, the current density is greater at high points and lesser at the low points within the surfaces profile. The rate of the electrochemical reaction is directly proportionate to the current density. The increased current density at the raised points forces the material to dissolve faster at these points and thus, tends to level the surface material. Electropolishing will simultaneously deburr and polish the surface.

(v) Appearance

A lustrous metal surface is obtained due to electropolishing. This is the striking benefit of electropolishing. The material is treated electrochemically, leaving a microscopically smooth surface that is highly lustrous.

5.5.4 Electrochemical Etching

The main process taking place in electrochemical etching of metallic specimen is redox reaction. All metals have a natural tendency to lose or gain electrons depending on their electrode potential values. The degree to which this reaction takes place may be recorded by measuring the electrochemical potential. Thus, micro structural elements of different electrochemical potential are attacked at different rates. This produces differential etching, resulting in micro structural contrast. Electrochemical etching may be considered as 'forced corrosion'. The differences in potential of the micro structural elements cause a sub-division into a network of very small anodic or cathodic regions.

A wide variety of etchants are available, including acid, bases, neutral solutions, mixtures of solutions, molten salts and gases. The stability of many etching solutions is limited; redox potentials change with time. Etching times range from several seconds to some hours. When no instructions are given, progress is judged by the appearance of the surface during etching. Usually, the surface will become less reflective as etching proceeds. Etching time and temperature are closely related; by increasing the temperature, the time can usually be decreased. Most etching is performed at room temperature.

(i) Conventional *chemical etching* is the oldest and most commonly applied technique for production of microstructural contrast. In this technique, the etchant reacts with the specimen surface without the use of an external current supply. Etching proceeds by selective dissolution according to the electrochemical characteristic of the component areas.

(ii) In electrolytic or *anodic etching*, an electrical potential is applied to the specimen by means of an external circuit. Typical setup consist, the specimen (anode) and its counter electrode (cathode) immersed in an electrolyte

(iii) *Potentiostatic etching* is an advanced form of electrolytic etching, which produces the ultimate etching contrast through highly controlled conditions.

On completion of any chemical or electrochemical etching process, the specimen should be rinsed in clean water to remove the chemicals and stop any reactions from proceeding further. After specimens are water rinsed, they should be rinsed in alcohol and dried in a stream of warm air. The use of alcohol speeds up the drying action and prevents the formation of water spots.

5.5.4.1 An Electrochemical Etching of Copper

The *pattern etching method* with adjustable etching rate is easily scalable and relatively low cost. One of the notable features of the process that the etching solution is reusable and minimizes the production of waste solution. The procedure for etching the copper layer is by *electrochemical method; and the process is referred to as* reverse of electroplating. The substrate being etched acts as the *anode* in the electrochemical cell. The cathode is also made of copper or any copper-coated board. The choice of plating bath composition dictates the pattern quality required for the anode and the potential reaction between the mask, e.g., photoresist, and the plating bath. The choice of the current density would therefore be a balance between acceptable etching times and the limitations of the underlying process itself.

- The polished PCB substrates were cut into squares. The substrates were then ultrasonically cleaned with a degreasing detergent and alcohol. Finally, they were air-dried and baked on a hotplate to remove the remaining moisture.
- Photoresist material was sprayed on the substrate until a relatively thick and smooth coating was obtained. A thick layer was found to be necessary for the mask to withstand prolonged etching times.
- The coated substrates were air-dried and subjected to a low temperature bake at 80–85°C
- The coated substrates were exposed in a flatbed ultraviolet (UV) exposure unit for a period of between 30 and 40 minutes.

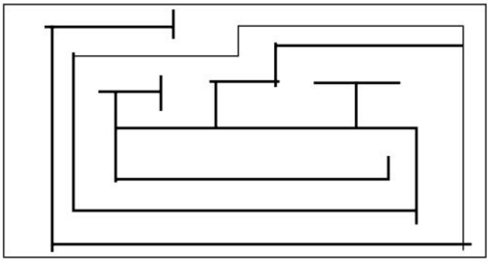


Figure 5.10 *Positive film pattern for PCB etching.*

After exposure, the substrates were immersed in a developer solution to produce the desired PCB pattern (Fig. 5.10).

(1) Etching of patterned substrate

(a) *Chemical etching* of the desired pattern is conducted in a heated 40 percent ferric chloride solution. Magnetic stirrer is used for complete agitation of the solution. Etching is carried out for about 2 hours or until the copper in the exposed surfaces was completely removed.

(b) *Electrochemical etching* was carried out in a typical copper electroplating setup shown in Fig. 5.11. A typical set-up consist of the specimen (anode) and

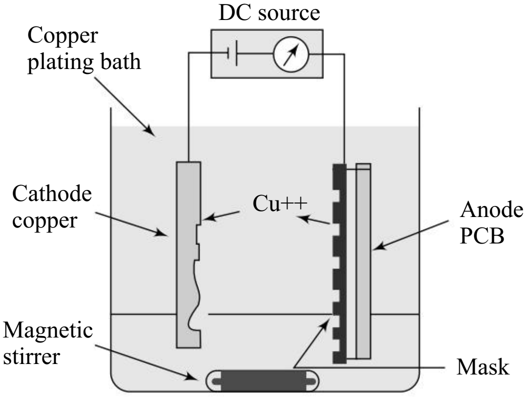


Figure 5.11 *An electrolytic anodic etching cell.*

its counter electrode (cathode) immersed in an electrolyte. In *electrolytic or anodic etching*, an electrical potential is applied to the specimen by means of an external circuit.

(c) '*Potentiostatic etching*' is an advanced form of electrolytic etching, which produces the ultimate etching contrast through highly controlled conditions. It consisted of a flat copper plate or copper-coated PCB that served as the *cathode* and the test substrate which served as the *anode*. The two electrodes were mounted parallel to each other at a distance. The surface area of the cathode was substantially larger than that of the exposed surface of the anode (test substrate). The entire fixture was then submerged in a copper plating bath.

Composition of the bath:

1.	$CuSO_4 \cdot 5H_2O$	200 g/l,
2.	H_2SO_4	5 g/l, and
3.	Rochelle salt	0.2 –1.0 g/l.

A magnetic stirring bar was used to provide agitation to the plating solution. DC plating current was provided by a power supply with a digital readout. The etching process remained in the constant current regime until the trenches became deep enough that it switched to the constant voltage regime close to the end of the process.

On completion of any chemical or electrochemical etching process, the specimen should be rinsed in clean water to remove the chemicals and stop any reactions from proceeding further. After specimens are water rinsed, they should be rinsed in alcohol and dried in a stream of warm air. The use of alcohol speeds up the drying action and prevents the formation of water spots.

5.6 Electrolytic Cell

5.6.1 Electrochemical Deposition/Plating Process

Electroplating is defined as a process in which a base metal is coated with a thin layer of a more noble metal by electrolytic deposition.

It is a process by which the surface properties of metals are modified by coating a thin layer of one metal on the top of a different metal. This is done to achieve the desired electrical and corrosion resistance, reduce wear and friction, improve heat tolerance and for decoration.

The theory of electroplating is based on the positively charged, conductivity and reactions of the plating metals and electrons. The process of electroplating is made of four important parts:

Cathode—the negative electrode in electrolysis, where negative electrons are created and positive ions are discharged, it is the object that is going to be plated

Anode—positive electrode in the electrolysis, where negative ions are discharged and positive ions are created, it is of the same material as the plating metal

Electrolyte—conducting medium where the flow of current is with the movement of matter. Most of the time, done in aqueous solutions such as acids, bases and salts

Direct current—is the electricity that passes from the anode to the cathode.

As result of passing direct current from anode through the electrolyte, the positive ions of coating metal gets reduced to metal at the cathode as metal coating. The metal coating bonds to the cathode and thus the electroplating process is complete. The theory of electroplating can be explained in a simple formula.

$$M \rightarrow M^{+n} + ne$$

In Fig. 5.12, M stands for the plating metal (the M is the metal considered to be plated and 'n' is the number moles of electrons involved in the reaction). The metal coating is due to the reduction reaction taking place at the cathode with the taking up of electrons.

$$M^{+n} + ne \rightarrow M$$

The physical embodiment of an electroplating process consists of four parts:

1. The external circuit, consisting of a source of direct current (dc), means of conveying this current to the plating tank, and associated instruments such as, ammeters voltmeters, and means of regulating the voltage and current at their appropriate values
2. The negative electrode or cathode, where the material to be plated, is called the *work*, along with means of positioning the work in the plating solution so that contact is made with the current source
3. The plating solution itself, almost always aqueous, called by platers the '*bath*';
4. The positive electrodes, the anodes, usually of the metal being plated but sometimes of a conducting material which serves merely to complete the circuit, called inert or *insoluble anodes*. The plating solution, of course, is contained in a tank, which must be of a material appropriate to the solution it contains: often plain mild steel for alkaline solutions, and of steel lined with resistant material for acid solutions. Such linings may be of rubber, various plastics or even glass or lead.

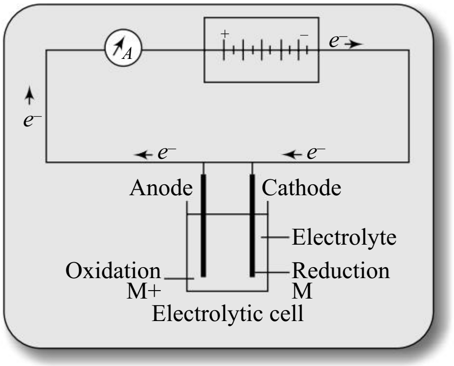


Figure 5.12 *A electrolytic cell.*

(1) Ingredients of a plating bath

The salient features of every plating bath contain ingredients that may have one or more of the following functions:

- to provide a source of the metal or metals being deposited
- to form complexes with ions of the depositing metal
- to provide conductivity
- to stabilize the solution, e.g., against hydrolysis
- to act as a buffer to stabilise the pH
- to modify or regulate the physical form of the deposit
- to aid in dissolving the anodes
- to modify other properties, either of the solution or of the deposit, peculiar to the specific case.

In plating process, there are two main purposes of forming complex ions of certain cations:

(i) To stabilize the cation. Some metal cations are not stable in the simple aquated form, e.g., gold. They are much more stable when complexed to some ligand. The presence of the ligand lowers the concentration of the free (aquated) ion.

(ii) To hold the aquated form at suitably low concentration allowing control of the evenness of plating. The cyanide ion, CN-, is a common ligand forming complex ions such as $Zn(CN)_4^{2-}$,$Cu(CN)_4^{2-}$-, $Ag(CN)^{2-}$ and $Au(CN)^{2-}$.

(2) The plating metals

Most electroplating coatings fall into one of the following six categories:

1. *Sacrificial coatings*—To protect base metals like iron from corrosion. Sacrificial denotes that the coatings 'sacrifice' themselves in the act of protecting the basis metal.

2. *Decorative protective coatings*, used primarily for adding attractive appearance to some protective qualities.
3. *Engineering coatings*—To impart specific properties to the surface, such as solderability, wear resistance, reflectivity, conductivity and many others.
4. Alloys—an almost unlimited number of alloys has been plated experimentally, since the possible combinations of the plateable metals, in various properties.

5.6.2 Electrolysis, Metal Deposition and the Electrical Relationships

Any processes or investigations related to electrolytic solutions, the basics of Faraday's laws are needed to understand and to apply for practical purposes.

They may be stated as follows:

1. *The amount of chemical change produced by an electrical current is proportional to the quantity of electricity that passes.*
2. *The amounts of different substances liberated by a given quantity of electricity are inversely proportional to their chemical equivalent weights.*

Consider an electrolytic cell as described in Fig. 5.13.

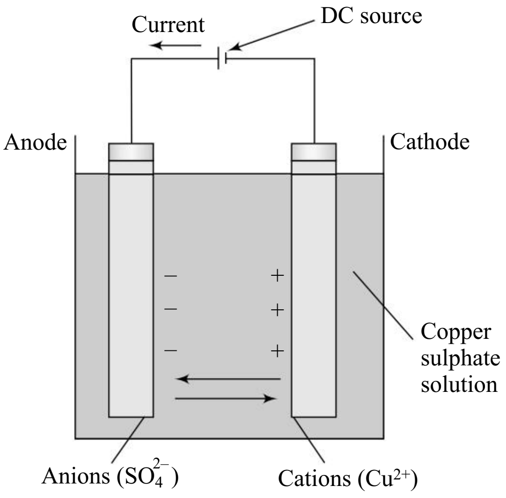


FIGURE 5.13 *Electrolysis of an electrolyte solution.*

If current of 'I' ampere is passed through an electrolyte solution of a metal for 't' seconds in an electrolytic cell, the metal ions get reduced to metal by involving 'n' mole electrons at the cathode, as given

$$M^{n+} + ne \rightarrow M$$

The total quantity of charge (Q) involved in the electrode process is the product of current x time ($Q = It$), which deposits 'W' g moles of a metal at the cathode.

If 'N_A' is the Avogadro's number and 'e', the electrical charge per electron (coulombs); then the total quantity of charge (Q) transported in whole electrode process is

$$Q = W \cdot n \cdot [N_A \cdot e] \quad \text{...(5.1)}$$

The product of $[N_A \cdot e]$ in Eq. (5.1) is the Faraday constant, F (= 96,484 Coulombs). Therefore, the *number of moles of metal* deposited at the cathode by charge 'Q' can be obtained as:

$$W = Q/(nF) \quad \text{...(5.2)}$$

The above Eq. (5.2) can be transformed into Eq. (5.3) to calculate the mass of the metal deposited at the cathode during the electrolysis process:

$$W = \frac{I \times t \times A}{n \times F} = \frac{(\text{Ampere} \times \text{Time} \times \text{Atomic weight})}{\text{Valence} \times \text{Faraday}} \quad \text{...(5.3)}$$

where

W = weight of plated metal in grams.
I = current in coulombs per second.
t = time in seconds.
A = atomic weight of the metal in grams per mole.
n = valence of the dissolved metal in solution in equivalents per mole.
F = Faraday's constant in coulombs per equivalent. F = 96,485.309 coulombs/equivalent.

The quantity '*A*/*nF*' is called the *electrochemical equivalent* of the metal.

The following illustration will describe how Faraday's Law can be used to find out what is the mass of the metal that can be deposited at the cathode of the electroplating cell.

5.6.3 An Illustrated Example of Nickel Plating

Consider nickel electroplating on a base metal.

The electrochemical reaction at the cathode will be

$$Ni^{2+} + 2e \rightarrow Ni \qquad \ldots(i)$$

Ni^{2+} ion takes up two electrons and gets reduced to metallic Ni which deposits on the cathode electrode.

(i) Calculate the weight of nickel deposit

Calculate the amount of nickel plated onto the surface of the base metal, if we pass a current of 1 ampere (1 coulomb/second) to the cathode for one hour (3600 seconds)?

Weight of plated Nickel metal can be obtained from Eq.(5.3)

$$W = \frac{[1 \times 3600 \times 58.69]}{(2 \times 96485.309)}$$
$$= 1.09 \text{ grams.}$$

(ii) Calculation of thickness of plating

(a) The *thickness of the metal* deposited at the cathode can be calculated using the following equation (ii):

$$T = \frac{(W \times 10{,}000)}{(D \times S)} \qquad \ldots(ii)$$
$$= \frac{(\text{weight of metal plated} \times 10{,}000)}{[\text{density} \times \text{surface area}]}$$

where

T = thickness in microns
D = density in grams per cubic centimetre
S = surface area of the plated part in square centimetres
10,000 is a multiplicative constant to convert centimetres to microns.

(b) If Eq. (5.3) and (iii) are combined we have the following equation for plated thickness, Equation (ii):

$$IT = \frac{(I \times t \times A \times 10{,}000)}{(n \times F \times D \times S)} \qquad \ldots(iii)$$
$$= \frac{(\text{ampere} \times \text{Time} \times \text{At. weight})}{[\text{valence} \times \text{Faraday} \times \text{density} \times \text{Surface area}]}$$

If we apply equation (ii) to our earlier example of nickel plating we find that the thickness of plated metal is:

$$W = \frac{(1 \times 3600 \times 58.69 \times 10000)}{(2 \times 96485.309 \times 8.90 \times 22.1)}$$
$$= 55.67\ \mu\text{m}$$

Here, we have assumed that the surface area of the coated part is 22.1 cm^2. Notice the importance of surface area in the equation. If all other parameters are kept the same, the plated thickness will increase as the surface area is decreased and vice versa.

(iii) Calculation of plating rate

Plating rate of a metal can be determined by eliminating time from equation (i) and rewriting it as:

$$R = \frac{(I \times A \times 600{,}000)}{(n \times F \times d \times S)} \qquad \ldots\text{(iv)}$$

where

R = plating rate in microns per minute.

600,000 is a multiplicative constant used to make R come out in units of microns per minute.

Applying equation (iv) to our nickel plating example we get:

$$R = \frac{(1 \times 58.69 \times 600000)}{(2 \times 96485.309 \times 8.90 \times 22.1)}$$
$$= 0.9278\ \mu\text{m/minute.}$$

This makes sense because if one multiplies 0.9278 μm/minute by 60 minutes (3600 seconds) one gets a thickness of 55.66 μm which matches the earlier calculation.

5.6.4 Current Efficiency

From the Faraday's law, it is evident that the amount of chemical change produced by an electrical current is proportional to the quantity of electricity that passes.

Often, however, we are interested in only one of the several chemical changes taking place, and any current used up in causing other changes is considered 'wasted'. In the usual electroplating situation, our interest focuses the quantity of metal deposited at the cathode or dissolved at the anode, and any hydrogen evolved at the cathode by the reaction

$$2H_2O + 2e \rightarrow H_2 + 2OH^-$$

oxygen at the anode by the reaction

$$2H_2O \rightarrow O_2 + 4H^+ (aq) + 4e$$

represents a waste of electricity and a reduction in the efficiency of the process. Thus, we speak of *percentage current efficiency as the ratio of the desired chemical change to the total chemical change multiplied by 100*:

$$\textit{Current efficiency}\ (\text{CE}) = 100 \times \frac{\text{Actual weight deposited}}{\text{Theoretical weight}}$$

where *CE is current efficiency* in percent, theoretical weight is to be expected from Faraday's laws if there were no side reactions. Cathode efficiency is current efficiency as applied to the cathode reaction, and anode efficiency is current efficiency as applied to the anode reaction.

5.7 Polarization, Decomposition Potential, Overvoltage and Electrodeposit

Electroplating of zinc, copper, nickel, chromium, silver, gold, etc., are commonly used in metal finishing. Electroplating involves the electrolysis of a suitable electrolyte from which metal ions are reduced and coated as metal on the object.

The process of Electroplating is made of four important parts:

- *Cathode*—the negative electrode in electrolysis, where negative electrons are created and positive ions are discharged, it is the object that is going to be plated
- *Anode*—positive electrode in the electrolysis, where negative ions are discharged and positive ions are created, it is of the same material as the plating metal
- *Electrolyte*—conducting medium, where the flow of current is with the movement of matter. Most of the time, done in aqueous solutions such as acids, bases and salts
- *Direct Current*—is the electricity that passes from the anode to the cathode.

As result of passing direct current from anode through the electrolyte, the positive ions of coating metal gets reduced to metal at the cathode as metal coating. The metal coating bonds to the cathode and thus the electroplating process is complete. The metal coating bonds to the cathode and thus the electroplating process is complete.

Electroplating involves electrolysis and hence three important electrochemical principles are associated with it for a successful deposition of the metal. They are as follows:

(i) Polarization,
(ii) Decomposition (or discharge) potential, and
(iii) Overvoltage.

5.7.1 Application of Potential for Spontaneous and Nonspontaneous and Irreversible Cell Processes

A galvanic cell produces electrical energy using the decrease in free energy involved in the spontaneous redox reaction. Under reversible conditions the free energy decrease is equal to the electrical energy by the equation $\Delta G = -nEF$; ΔG is free energy, n is the number of electrons involved in the cell reaction, F is Faraday constant (96,500 coulombs) and E is the EMF of the cell in volts. But if the conditions are irreversible, the electrical energy available will be less than the decrease in free energy.

For a cell to do any *useful work* (spontaneous) or for an *electrolysis (Nonspontaneous)* to occur, a significant current must flow. Whenever current flows, *three factors act to decrease the output voltage of a galvanic cell or to increase the applied voltage needed for electrolysis*.

These factors are called the

(i) Ohmic potential,
(ii) concentration overpotential (polarization) and
(iii) activation overpotential.

A nonspontaneous chemical reaction takes place only with the application of applied potential.

The potential applied to such a nonspontaneous reaction is given as

$$E_{\text{applied}} = E_{\text{back}} + IR$$

where

E_{back} = voltage required to cancel out the normal forward galvanic cell reaction.

$E_{\text{back}} = E_{\text{reversible(galvanic)}}$ + Overvoltage

$IR = IR_{drop}$. The work applied to force the nonspontaneous reaction to take place. R is the cell resistance. Overvoltage—It is the extra potential that must be applied beyond what predicted from the Nernst equation.

(i) Ohmic Potential

The voltage needed to force current to flow through the cell is called the ohmic potential and is given by Ohm's law:

$$E_{ohmic} = IR$$

where 'I' is the current and 'R' is the resistance of the cell.

- If the galvanic cell at equilibrium, then, $I = 0$, i.e., there is no ohmic potential.
- If a cell does an electrical work, the cell voltage decreases because part of the free energy released by the chemical reaction is needed to overcome the resistance of the cell itself.
- The voltage applied to an electrolytic cell must be great enough to provide the free energy for the chemical reaction and to overcome the cell resistance.
- It can be noticed that the voltage of a galvanic cell is decreased by IR, and the magnitude of the applied voltage in an electrolysis must be increased by IR in order for current to flow, provided other effects are absent in the cell.

Effect of ohmic potential:

The cell redox reactions, $Cd + 2AgCl \rightarrow Cd^{2+} + 2\,Ag + 2Cl^-$, are spontaneous for the cell $Cd/Cd^{2+}(0.167M)//Cl^-/AgCl/Ag$ and the EMF generated is 0.764 V.

(a) The cell resistance is 4.0 Ω and if a current of 35.0 mA is drawn from this cell, what will be the cell voltage?

(b) Suppose that the same cell is operated in reverse as an electrolysis. What voltage must be applied to reverse the reaction?

Solution

(a) In the absence of electron flow, the voltage (E_{eq}, means equilibrium voltage) is 0.764 V. With a current of 28 mA, the voltage will decrease to

$$E = E_{eq} - IR = 0.764 - (0.035\ \text{A})\ (4.0\ \Omega) = 0.624\ \text{V}.$$

(b) By convention, the voltage applied to an electrolysis in a cell is given negative sign. The voltage needed to reverse the spontaneous reaction will be

$$E = -E_{eq} - IR = -\ 0.764 - (0.035\ \text{A})\ (4.0\ \Omega) = 0.904\ \text{V}.$$

- It can be observed that the magnitude of any galvanic cell voltage is decreased by the ohmic potential.
- The magnitude of the voltage that must be applied to any electrolysis in a cell is increased by the ohmic potential.

5.7.2 Polarization

In a nonspontaneous electrolytic cell, external electrical energy is used to produce a chemical reaction. Electrolysis is a nonspontaneous process and hence involves an increase in free energy. Under reversible conditions, the free energy increase in electrolysis is equal to the decrease in free energy accompanying the same reversible reaction in a galvanic cell. But if the conditions are irreversible, *the EMF required for the electrolysis will be greater than the reversible EMF of the corresponding galvanic cell*. Such an electrolytic cell is said to be polarized and the phenomenon is called *polarization*.

Polarization is defined as a phenomenon involving variation of the electrode potential either due to the inadequate supply of the electroactive species from the bulk of the electrolyte to the electrode or due to one of the steps involved in the deposition of the ion being slow.

(1) Concentration polarization

The reduction potential E of an electrode in a cell is given but the Nernst equation as

$$E = E^\circ + \frac{2.303RT}{nF}\log M^{n+}$$

where E^o is the standard reduction potential of the electrode, R is the universal gas constant, T is the absolute temperature, F is the Faraday's constant, n is the number of electrons involved and M^{n+} is the concentration of metal ions.

In an electrolytic cell, during electrolysis metal ion gets reduced to metal.

$$M^{n+} + n\,e \rightarrow M$$

The potential of the electrode is found to decrease due to the decrease of the activity of the metal ions in the vicinity of the electrode. But equilibrium is re-established by diffusion of the ions from the bulk to the electrode. Diffusion is caused by the concentration gradient between the bulk and vicinity of the electrode. If the diffusion process is slow, the potential of the electrode changes from its equilibrium potential and the electrode gets polarized as shown in Fig. 5.14. This is called *concentration polarization*.

This type of polarization takes place at the electrode when the concentration of ions at the electrode surface is different from that in the bulk solution. This behaviour is observed when the rate of electrochemical reaction at the electrode surface is fast compared to the rate of diffusion of electro-active species from the solution bulk to the electrode surface.

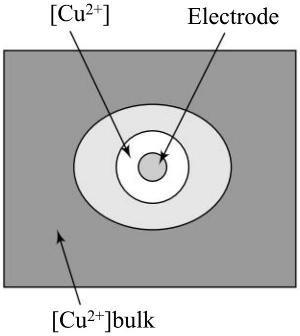


FIGURE 5.14(a) *An illustration of concentration polarization of Cd^{2+} ions around the electrode.*

Example on concentration polarization:

The oxidation reaction at the electrode is

$$\text{Cd} \rightarrow \text{Cd}^{2+} + 2e$$

The potential of the anode depends on $[\text{Cd}^{2+}]_{\text{electrode}}$, not $[\text{Cd}^{2+}]_{\text{bulk}}$, because $[\text{Cd}^{2+}]_{\text{electrode}}$ is the actual concentration at the electrode surface as shown in Fig. 5.14b.

The anode potential is given by the

$$E_{(\text{anode})} = E^\circ_{(\text{anode})} - (0.05916/2)\log[\text{Cd}^{2+}]_{\text{electrode.}}$$

If $[\text{Cd}^{2+}]_{\text{electrode}} = [\text{Cd}^{2+}]_{\text{bulk}}$, the potential of the anode will be that expected from the bulk

$[\text{Cd}^{2+}]$ concentration. If the current is flowing so fast that Cd^{2+} cannot escape from the region around the electrode as fast as it is made, $[\text{Cd}^{2+}]_{\text{electrode}}$ will be greater than $[\text{Cd}^{2+}]_{\text{bulk}}$.

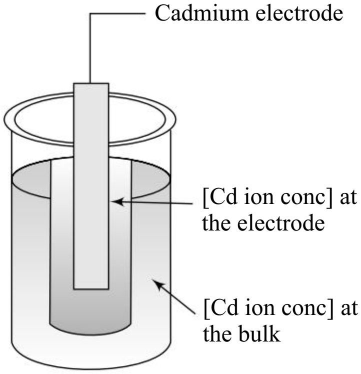


FIGURE 5.14(b) *Concentration of Cd^{2+} ions at the electrode surface and the bulk solution.*

When $[Cd^{2+}]_{electrode}$ does not equal $[Cd^{2+}]_{bulk}$, we say that concentration polarization exists. The anode will become more positive and the

$$\text{Cell voltage} = E_{(cathode)} - E_{(anode)},$$

will decrease.

The straight line shows the behaviour as expected. When ions are not transported to or from an electrode as rapidly as they are consumed or created, we say that concentration polarization exists, if only the ohmic potential (IR) affects the net cell voltage.

The deviation of the curve from the straight line at high currents is due to concentration polarization (Fig. 5.15).

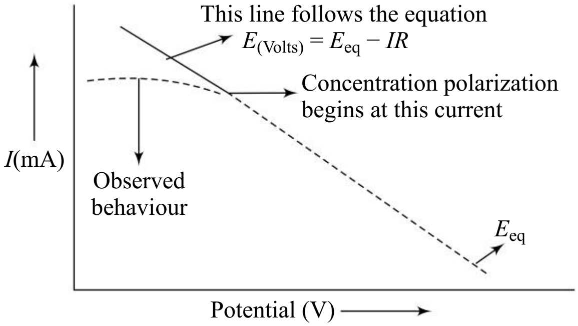


Figure 5.15 *Variation of current (I mA) with concentration polarization.*

- In a galvanic cell, concentration polarization decreases the voltage below the value expected in the absence of concentration polarization
- In electrolytic cells, the situation is reversed; reactant is depleted and product accumulates. Therefore, the concentration polarization requires us to apply a voltage of greater magnitude (more negative) than that expected in the absence of polarization
- Concentration polarization gets worse as $[M^{n+}]$ gets smaller.

Example: Concentration Polarization of copper:

$$[Cu^{2+}]_{bulk} = 10^{-2}\text{ M}$$

$$[Cu^{2+}]_{electrode} = 10^{-4}\text{ M}$$

Then,

$$E_{bulk} = 0.337 + \frac{0.0591}{2}\log(10^{-2}) = 0.278\text{ V}$$

$$E_{electrode} = 0.337 + \frac{0.0591}{2}\log(10^{-2}) = 0.219\text{ V}$$

Therefore, $\eta = E_{electrode} - E_{bulk} = -0.059$ V.

It can be noted that

The ions move towards or away from the electrode due to the following factors

- Diffusion,
- Convection,
- Electrostatic attraction or repulsion.

It can also be summarized, saying, that

(a) Raising the temperature increases the rate of diffusion and thereby decreases concentration polarization.
(b) Mechanical stirring is very effective in transporting species through the cell.
(c) Increasing ionic strength decreases the electrostatic forces between ions and the electrode.
(d) These factors can all be used to affect the degree of polarization.
(e) Also, the greater the electrode surface area, the more current can be passed without polarization.

Factors that affect concentration polarization

It is observed that the following factors affect the concentration polarization at the electrode:

(i) Increase of temperature
(ii) Increased agitation
(iii) Electrode surface area is increased: more reaction takes place
    (i) change ionic strength to increase or decrease attraction between electrode and reactive ion.

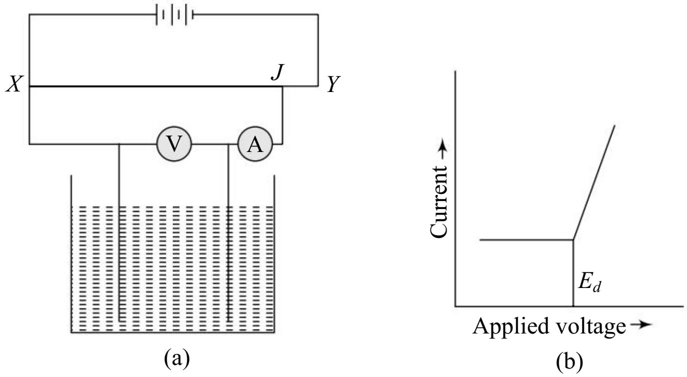
5.7.3 Decomposition (Discharge) Potential

Decomposition potential is defined as the minimum external potential required for the continuous electrolysis of the electrolyte.

5.7.3.1 Experimental Determination of Decomposition Potential

The cell consists of a glass vessel containing the electrolyte whose decomposition potential is to be determined. Two electrodes are kept in the electrolyte. The electrodes are connected to a voltmeter, *V*, an ammeter, *A*, and a movable Jockey *J* in series.

The Jockey moves over a uniform wire *XY*, which is connected to a source of current as shown in Fig. 5.16a.



(a) (b)

FIGURE 5.16 *(a) and (b) Polarization and overvoltage.*

The jockey is moved over the wire gradually increasing the applied voltage. For each applied voltage the corresponding current in amperes is measured. Up to a definite potential the current flow is very small and thereafter increases with increase in potential. A typical voltage *V*s current graph is shown in Fig. 5.16b. The intersection of the two straight-line portions of the graph gives the decomposition potential E_d.

5.7.3.2 Significance of Decomposition Potential

1. Decomposition potential refers to the minimum voltage required for the electrolysis of an electrolyte.
2. It is possible to predict the order of the discharge of ions from a solution containing several ions. For example from an electrolyte containing Cu^{+2}, Zn^{+2} and Cd^{+2} the order of deposition is copper, aluminium and then zinc. This is due to the fact that the $E°$ of Cu^{+2}, Cd^{+2} and Zn^{+2} are +0.34 V, –0.44 V, –0.76 V, respectively.

5.7.4 Overvoltage

An electrochemical cell is polarized, if its actual potential is different from that expected according to Nernst equation.

The extent of polarization is measured as overpotential.

$$\eta = E_{\text{applied}} - E_{\text{reversible (equil)}}$$

In some cathodic reactions, the reduction of the electroactive species may involve several steps and if any one of the steps is slow and irreversible it may result in polarization. This is known as overvoltage.

Polarization of a cell may take place at one or both the electrodes. Polarization may be caused by both the effects. But concentration polarization may be eliminated by stirring of the electrolyte and then polarization is mainly due to overvoltage

Overvoltage also can be defined as the excess voltage to be applied over and above the theoretical decomposition potential of an electrolyte, to start the electrolysis of the electrolyte.

Overvoltage, η = Experimental decomposition voltage – theoretical decomposition voltage

5.7.4.1 Factors Affecting Overvoltage

Overvoltage of an electrolyte is affected by the following factors.

- Nature and physical state of the electrodes
- Nature of the material deposited
- Current density
- Temperature
- Nature of electrolyte
- Hydrogen overvoltage.

The reversible standard EMF of a H_2/O_2 galvanic cell in which the net reaction is the formation of water is 1.23 V. The reverse process of H_2 and O_2 evolution by electrolysis of a dilute acid or base requires a voltage of 1.68 V. Thus, E_d experimental is 1.68 V and E_d theoretical is 1.23 V and overvoltage $\eta = 1.68 - 1.23 = 0.45$ V. Nearly same overvoltage is observed irrespective of the acid or base used indicating that the mechanism involved in the discharge of H_2 (and O_2) is the same.

Explanation

$\underset{\text{Bulk}}{H_3^+O} \longrightarrow \underset{\text{substrate}}{H_{ad}^+} + H_2O$ Bulk diffusion

$H_{ad}^+ + \bar{e} \rightarrow H_{ad}$ Reduction

$H_{ad} + H_{ad} \rightarrow H_{2ad}$ Molecularization

$H_{2ad} \rightarrow H_2\uparrow$ Desorption

The reduction of H^+ ions at the cathode and its release as H_2 gas involves the above 4 steps. Of these, desorption of H_2 gas is slow and results in an accumulation of H_2 gas at the electrode surface. This induces the formation of $H_{2\,ad}\,|H^+|$ Pt electrode, which produces a back EMF that opposes the applied EMF. Hence, unless an excess external EMF over the theoretical reversible discharge potential is applied no gas evolution takes place. Hydrogen overvoltage is a measure of the tendency of H_2 gas to get liberated on a metal surface. Higher the hydrogen overvoltage the more difficult is the evolution of hydrogen.

Significance of Hydrogen Overvoltage

Hydrogen overvoltage plays an important role in the electrode position of metals. The reduction of a metal during electrolysis is a single step process and involves no overpotential.

$$M^{n+} + n^e \rightarrow M\,[\eta = 0]$$

Thus, metal deposition at the cathode takes place nearly at the same reversible $M^{n+/}M$ electrode potential.

In an aqueous solution there is a competition between H^+ ions and M^{n+} ions to get reduced. If the potential required for metal deposition is less than that for hydrogen, the metal gets deposited and in the reverse case hydrogen gas is set free. Thus according $E°$ values, metals with a higher reduction potentials than hydrogen, will be preferentially deposited before H_2 is set free. In the case of metals with lesser $E°$ than hydrogen we expect the liberation of H_2 in preference to the metals. But in many such cases, the metal gets deposited rather than H_2 due to the hydrogen overvoltage.

5.8 Principle of Electroplating

Electroplating is defined as the process of depositing a layer of a metal by electrolysis, over the surface of a substrate, which may be another metal, polymer or ceramic.

Electroplating is carried out in an electrolyte cell (Fig. 5.17). The cell is filled with a suitable salt solution C of the metal being deposited. Anode A is the pure metal rod or plate and the cathode B is the object being plated. Upon electrolysis the anode metal M dissolves as M^{n+} ions and cathodic reduction of M^{n+} from the electrolyte and deposition of the metal on the substrate takes place.

$$\underset{\text{anode}}{M} \rightarrow \underset{\text{solution}}{M^{n+}} \; n\bar{e} \qquad \text{[anodic dissolution of the metal – oxidation]}$$

$$\underset{\text{solution}}{M^{n+}} + n\bar{e} \rightarrow \underset{\text{cathode}}{M} \qquad \text{[cathodic reduction of } M^{n+}\text{]}$$

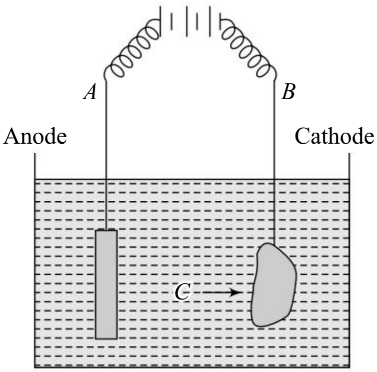


FIGURE 5.17 *Electroplating process.*

In an ideal situation, both the processes take place at the same rate and the concentration of M^{n+} ion in solution remains a constant. In some electroplating processes inert anode-like platinum is used and main anodic reaction is evolution of oxygen. In such cases M^{n+} concentration gets decreased during electrolysis and to maintain constancy of M^{n+} ion concentration the electrolyte is added from time to time.

5.8.1 Mechanism of Electroplating

Electroplating may be compared to crystal growth and it involves three steps:

- Diffusion of the metal ions from the bulk of the electrolyte to the cathodic surface
- Reduction of the metal ions on the substrate surface to form add atoms layer
- Growth of the deposit resulting in thickening of the layer into macroscopic deposit.

$$\underset{\text{bulk}}{M^{n+}} \rightarrow \underset{\text{substrate}}{M^{n+}} \qquad \text{[Bulk diffusion]}$$

$$M^{n+} + n\bar{e} \rightarrow \underset{\text{substrate}}{M_{\text{ad}}} \qquad \text{[Formation of atomic layer]}$$

$$\underset{\text{add atom}}{M} + \underset{\text{atoms}}{M} \rightarrow \underset{\text{substrate}}{M} \rightarrow \qquad \text{Thickening}$$

Formation of the add atom nuclei requires higher overvoltage. But once the nuclei are formed thickening takes place at lower potentials. The add atoms so formed move towards move favourable sites on the substrate surface. This is called surface diffusion. The most favourable site is called the kink site where the add atom can interact with three neighbours. Add atoms may also occupy edge sites with two neighbours. This is less favourable. Sometimes add atom may remain as an add atom with a single neighbour. Since this is least favourable it redissolves.

5.8.2 Characteristics of Good and Poor Deposits

The following Table 5.1 gives a comparative account of a good and poor deposit.

Good deposit	Poor deposit
1. Should have good adhesion and should be uniform and nonporous.	has poor adhesion, is nonuniform and porous.
2. Should be hard and ductile.	has burnt powdery structure and is brittle.
3. Should have a fine grained structure.	is coarsely crystalline.
4. Should be bright and lustrous.	is dull.
5. Is formed by definite spiral, block or layer arrangement.	is formed by step, ridge and poly crystalline block growth.

Table 5.1

Characteristics of good and poor deposits

5.9 Factors that Influence the Nature of Electrodeposits

A number of factors are found to influence the nature of electro deposits. The more important of these are (1) current density, (2) plating bath, (3) pH, (4) temperature, and (5) throwing power.

(1) Current Density

Current density is defined as *current per unit area*. Its unit is Am^{-2} or $mA\ Cm^{-2}$ of the electrode surface.

At low current density: Bulk diffusion and electron transfer are slow and hence the add atoms formed find time to move towards more favourable sites. This results in a well formed deposit.

At high current density: Electron transfer predominates and surface diffusion becomes slow. Number of nuclei formed increases and the add atoms may not move towards favourable sites. This results in a less ordered macroscopic deposit.

At higher current density (below limiting value): Bulk transfer predominates along with electron transfer. Surface diffusion is poor resulting in a rough, powdery, nonadhesive deposit.

At current densities above limiting value: Hydrogen evolution takes place resulting in a depletion of H^+ ion concentration in the vicinity of the cathode. As a result metal hydroxide or oxide gets precipitated. When these get included in the electrodeposit, the deposit appears dark, burnt and spongy.

(2) Plating bath

Plating bath contains an aqueous solution of the electrolyte used for plating, complexing agent, buffer to maintain pH and various organic additives to improve the quality of the deposit. The following points are important in the preparation of the electrolyte bath.

(a) Metal ion concentration

Metal ion concentration is normally kept high to reduce mass transfer. To improve the conductivity of the electrolyte, supporting electrolytes which do not participate in electrodeposition are also added. Sometimes such electrolytes also act as buffers.

(b) Complexing agent

Complexing agent is added to convert metal ions into complex ions so as to keep free metal ion concentration under check. This is done to ensure the formation of more adherent deposit. Complexing is also employed for the following purposes:

1. When the plating ion M^{n+} is known to react with the anode
2. To increase solubility of sparingly soluble electrolytes
3. To prevent passivation of the anode and make it dissolve easily and thus increase current efficiency
4. To increase the throwing power of the plating bath

Cyanides, hydroxides and sulphonates are commonly used as complexing agents. For example Gold, Silver, Copper and Cadmium are electroplated from cyanide baths.

(c) Organic additives

Certain organic additives have a remarkable influence over the nature of electrodeposits. They are classified as (i) brighteners, (ii) levellers, (iii) structural modifiers, and (iv) wetting agents according to the purpose for which they are added.

(i) Brighteners

Brighteners are substances that help in the formation of bright, light reflecting deposits. Only when the grain size of the deposit is smaller than the wavelength of height employed, the deposit becomes light reflecting. Thus, brighteners are used to produce microscopically fine deposits parallel to the substrate surface.

Aromatic sulphones, sulphomates and compounds containing groups like

$$–C \equiv N, –N = C = S, >C = 0$$

act as brighteners.

(ii) Levellers

Substrate surface may contain areas like dislocations where plating may take place faster than other areas. This leads to uneven deposits. When levellers are used they get adsorbed at such regions and prevent electron transfer. This ensures evenness of the deposit in which sodium alkyl sulphonates act as levellers.

(iii) Structural modifiers or stress relievers

All electrodeposits possess internal stress. If the stress level is high, the deposit may give way to impact and undergo cracking. Structural modifiers help in modifying the structure of the deposit and make it stress-free. Saccharin is a common stress reliever.

(iv) Wetting agents

During electrolysis sometimes evolution of hydrogen at the cathode may take place. The gas may get entrapped inside the metal deposit. The gas may try to escape when the plating is over and makes the deposit porous and brittle. This phenomenon is called hydrogen embrittlement.

Wetting agents release the hydrogen gas from the surface of the substrate and improve the adhesion of the deposit.

(3) pH

pH is defined as the negative logarithm to the base ten of H^+ ion concentration. In a neutral medium pH = 7, in acidic medium pH in less than 7 and in alkaline medium greater than 7.

If the pH of the electrolyte is low, H_2 gas evolution may take place causing the deposit to the brittle and burnt. At high pH deposits of metallic oxide or hydroxide may take place. Hence, an optimum pH between 4 and 8 is employed. To maintain the pH buffers are used.

(4) Temperature

At low temperatures electrode position may be uniform but slow. At high temperatures deposition may be fast but decomposition of organic additives, corrosion of equipment and hydrogen evolution may take place. Hence, an optimum temperature in the range of 308 – 333 K is used.

(5) Throwing power

Definition: Throwing power of a plating bath is defined as the ability of the bath to produce uniform and even deposit on the entire surface of the substrate.

Throwing power is important in the electroplating of irregular objects. Plating will be nonuniform if the object carries cavities or holes.

5.9.1 Experimental Determination of Throwing Power: Haring-Blum Cell

Haring-Blum cell is an electrolyte cell containing the electrolyte whose throwing power is to be determined (see Fig. 5.18). With reference to the figure, A is the anode and C_1 and C_2 are two cathodes kept at distances x_1 and x_2 ($x_1 > x_2$). Electroplating is carried out for sufficient time and the weights w_1 and w_2 of the metal deposited on the two electrodes C_1 and C_2 are measured. w_1 will be less than w_2 because $x_1 > x_2$. Then *percentage throwing power* is calculated using the formula

$$\text{Percentage throwing power} = \frac{100[X - Y]}{(X + Y - 2)}$$

where $$X = \frac{x_1}{x_2} \text{ and } Y = \frac{w_2}{w_1}$$

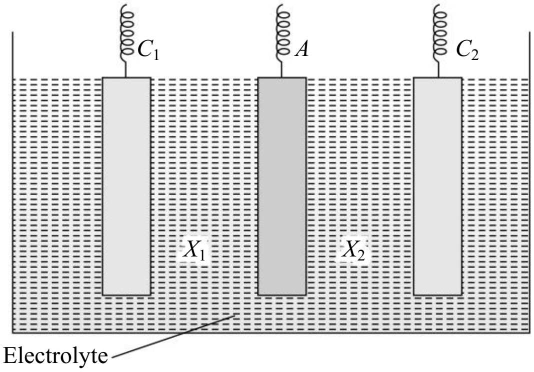


Figure 5.18 *Haring-Blum cell.*

Note: When $w_2 = w_1$, the value of $Y = 1$ and throwing power = 100. It means that when the throwing power is equal to 100, the plating bath produces equal quantities of deposits on the cathodes irrespective of their distances from the anode.

Significance of throwing power

Throwing power of a planting bath is the measure of the capacity of the bath to produce uniform plating on the entire surface of the object. Throwing power also gives us an insight into the following properties.

(a) Conductance of the electrolyte

The higher is the throwing power, the higher should be the conductance of the bath. This ensures uniform current distribution over the entire surface of irregular objects.

(b) Position of anode

From the measurement of throwing power it is possible to decide on the optimum position of the anode from cathodes. If needed additional anodes may be provided for better plating.

(c) Complexing agents

From throwing power experiments it is possible to choose correct complexing agent for obtaining a good deposit.

Example 5.1 Calculate the throwing power of a plating both in a Haring-Blum cell if the distances of the two cathodes from the anode are 6.0 cm and 5.0 cm and the quantities of metal deposited are 72 mg and 75 mg, respectively.

Solution

$x_1 = 6.0$ cm; $x_2 = 5.0$ cm; $x = \frac{c_1}{c_2} = \frac{6}{5} = 1.2$

$w_1 = 72$ mg and $w_2 = 75$ mg

$$\therefore \quad Y = \frac{w_2}{w_1} = \frac{75}{72} = 1.0416$$

$$\therefore \quad \text{Percentage T.P} = \frac{100(X-Y)}{(X+Y-2)} = \frac{100(1.2-1.0416)}{(1.2+1.0416-2)}$$

$$= \frac{100 \times 0.1584}{0.2416} = 65.56$$

Example 5.2 The throwing power of an electrolyte in a Haring-Blum cell is 75 percent. During an experiment 68 mg of the metal was deposited at the nearest cathode kept at a distance 4.8 cm from the cathode. At what distance must the cathode at a longer distance be kept if the metal deposited on it is 64 mg.

Solution

T.P = 75 $\quad c_1 = ?$ $\quad c_2 = 4.80$ cm

$w_1 = 64$ mg $\quad w_2 = 68$ mg

$$X = \frac{c_1}{c_2} = \frac{c_1}{4.8} \qquad Y = \frac{w_2}{w_1} = \frac{68}{64} = 1.0625$$

$$\text{T.P} = \frac{100(X-Y)}{(X+Y-2)}$$

$$75 = \frac{100(X-1.0625)}{(X+1.0625-2)}$$

$$75\,(X - 0.9375) = 100\,(X - 1.0625)$$

or $X = 1.4377$

$$\frac{c_1}{4.8} = 1.4377$$

or $c_1 = 1.4377 \times 4.80 = 6.90$ cm

5.10 Electroplating Process

Various types of plating tanks are in vogue depending upon the nature of the substrate to be plated, metal being deposited and the cost involved in the process. For large scale job jug mounting tanks are used. For batches of small objects barrel plating and individual mounting for large single objects are employed. The simplest of these is known as rectangular plating bath represented in Fig. 5.19.

Components of electroplating system

The different parts/components of an electroplating system are as follows.

(a) Plating tank [*A*]

The plating tank is made of wood or steel rectangular in shape. Steel tanks are thermally insulated with ceramic or polymeric materials. The capacity of the tanks is around 20–200 dm^3.

(b) Plating bath solution [*B*]

A suitable metal salt solution with high conductivity is used. Supporting nonparticipating electrolytes are added to increase conductivity and throwing power. Buffer to keep the pH and organic additives to improve the quality of plating are also present. The electrolyte should cover both the anode and the cathode.

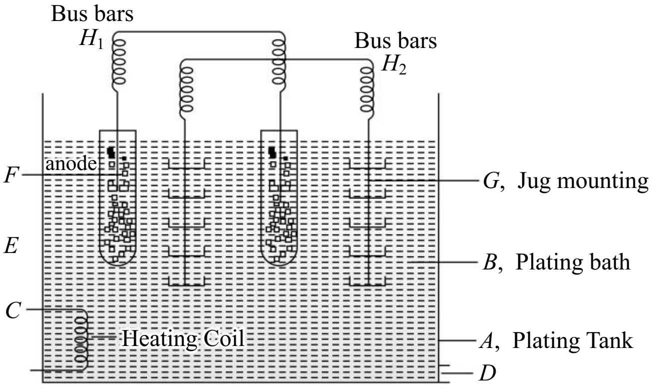


FIGURE 5.19 *Electrolytic process*

(c) Anode [*F*] and Cathode

Anode is the metal whose ion from the electrolyte is deposited on the cathode. Anode may be in the form of a rod, a plate or pellets. In some cases insoluble anodes may be used. The anode is usually enclosed inside an anode bag to retain the impurities in the anode material. This is done since the impurities may affect the quality of the deposit.

Cathode is the substrate with a cleaned surface mounted on suitable mountings like jug mounting *G*. The anodes and cathodes are suspended from separate bus bars H_1 and H_2 running above the tank. Electrodes are placed in such a way as to obtain uniform current distribution.

(d) Heating equipment [*C*]

Plating is usually conducted at an optimum temperature normally in the range of 308–333 K. Heating may be done using a heating coil-thermostat arrangement or by passing hot gases. Air is bubbled through inlet D to keep the electrolyte agitated.

(e) Filters

Smooth plating can be achieved only when the electrolyte is free from suspended materials. Hence, the electrolyte is purified by filtration using a pump and filter arrangement before it is kept in the tank.

(f) Electrical equipment

Plating is usually done at potential of 8–12 V and an operating current density of 1–200 mA/cm^2. There is provision to regulate the voltage and current. Initially, a high current density is applied for nucleation of start and then an optimum current density is employed.

5.10.1 Surface Preparation of the Substrate

Need for cleaning the surface

1. If the surface of the substrate contains any nonmetallic or nonconducting materials, electrodeposition of the metal at these points is prevented. Hence, the metal surface should be free from grease, oil, dirt and corrosion products.

2. Plating made on a rough surface will be less smooth than on a fine surface. Hence, electrodeposition is done on the substrate whose surface has a fine finish.

5.11 Steps Involved in Cleaning the Surface for Electroplating

Following steps are carried out on a substrate surface before it is sent for electroplating.

(i) Burn-off/Thermal cleaning

Burn-off or thermal cleaning systems use heat to remove grease, oil, paint or other organic compounds from the surfaces of parts.

(ii) Degreasing

Removal of oil, grease and other organic impunities from the surface of the substrate may be done by washing with organic solvents like 1:1:2 trichloro ethane, (or) dichloro methane.

Degreasing may also be done by alkali cleaning keeping the object cathodic.

The higher pH of the alkali helps in hydrolysis of the fat and the hydrogen liberated reduces unsaturated fatty acids.

(iii) Stripping/Coating removal

Stripping and coating removal services remove paint, coatings, or plated layers using a solvent, chemical or mechanical process.

(iv) Pickling

The object is immersed in 10 percent H_2SO_4 in order to remove excess alkali from alkali treatment, scales and rust present on the surface. This is called pickling. Oxide scales may be better removed by pickling in H_2SO_4 keeping the object anodic.

(v) Abrasive blasting

Blasting, sandblasting, or abrasive blast cleaning uses power spraying or a tumble chamber to remove all of a metal's visible rust, mill scale, paint and contaminants. It leaves the metal uniformly white or grey in appearance. Sandblasting provides a roughened surface with undercuts that improves the adherence of subsequently applied thermal spray or organic coatings. The substrate is then sand blasted and subjected to mechanical buffing to get a smooth bright surface.

(vi) Electropolishing/Electrolytic finishing

Polishing of the object may be carried out mechanically or electrochemically. In mechanical polishing silicon carbide grinding wheel is used. Electropolishing or electrolytic finishing is often called a reverse plating process. Electrochemical in nature, electropolishing uses a combination of rectified current and a blended chemical electrolyte bath to remove flaws from the surface of a metal part. The resulting surface is clean and bright.

(vii) Rinsing and drying

Substrate is then rinsed with distilled water, and dried before it is sent for electroplating.

5.12 Electroplating of Copper

Electroplating of copper is carried out in either a basic cyanide bath or an acidic sulphate bath.

Particulars	Cyanide bath	Sulphate bath
1. Bath composition	20–25 g CuCN 30 g NaCN, 20 g Na_2CO_3 per dm^3 solution pH = 12.5	200 g $CuSO_45H_2O$ 50 g H_2SO_4 per dm^3 pH = 1–2
2. Operating temperature	303–313 K	303–323 K
3. Current density	10–40 mA/cm^2	20–50 mA/cm^2
4. Additives	Phenol sulphonic acid, burnt	Dextrin, thiourea, gelatin, phenol sulphonic acid
5. Current efficiency	60–90 percent	90–100 percent
6. Anode	Pure copper	Pure copper
7. Cathode	Substrate	Substrate

Table 5.2

Basic and acidic bath for electroplating

In the cyanide bath (Table 5.2) cyanide concentration and carbonate concentrations are to be maintained. Low cyanide concentration produces greenish cathodic film while higher cyanide concentration leads to blistering deposit. Excess carbonate concentration gives roughness to deposit. The sulphate bath is economically simple, easy to operate and has high current efficiency.

5.12.1 Applications of Copper Plating

1. As an under coat for further plating
2. In the production of electrotype
3. In the production of printed circuit boards and other electronic components
4. In electro forming of objects
5. As a coating over steel cables to increase their electrical conductivity.

5.13 Gold Electroplating

Electroplating of gold is an electrochemical process by which gold is deposited on a substrate (at the cathode) by passing a current through the bath.

Let us consider the following aspects during gold plating.

- Basic principles
- Factors in electroplating to be considered
- Surface preparation in electroplating
- Gold plating–caratage, colour, types of electrolytes and deposits.

5.13.1 Basic Principles of Electroplating and Electropolishing Gold

Electroplating and electropolishing is carried out in an electrolytic cell, Fig. 5.20.

The electrolytic cell for gold plate comprises two electrodes that are electrically connected and immersed in a solution—called an electrolyte.

When DC current is passed into the cell, the metal ions obtained by the oxidation reaction gets dissolved in the solution, diffuses and is finally deposited on the negative electrode (the cathode), while the metal of the positive electrode (the anode) may be removed and dissolved in the electrolyte. Thus, metal passes from the anode into solution in the electrolyte and is then deposited on the cathode.

$\text{Au} \rightarrow \text{Au}^{3+} + 3e$ (Gold nuggets)

$\text{Au}^{3+} + 3e \rightarrow \text{Au}$

(at the negative electrode)

(from the electrolyte)

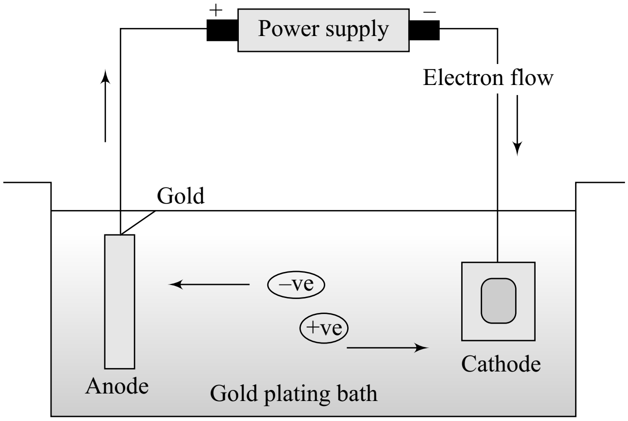


Figure 5.20 *An electrolytic cell for gold plating.*

Often, in electroplating, we use an inert anode, where metal is not dissolved away, and control the concentration of the depositing metal in the electrolyte solution by direct additions of the appropriate metal as a salt to the electrolyte.

These are the recognized classes of gold plating processes

1. Alkaline gold cyanide, for gold and gold alloy plating
2. Neutral gold cyanide, for high purity plating
3. Acid gold plating for bright hard gold and gold alloy plating
4. Non-cyanide, generally sulphite based for gold and gold alloy plating miscellaneous.

Dependent on the specific purpose and use, different kinds of electrolytes baths can be used, as given in Table 5.3.

Bath type:	**1**	**2**	**3**	**4**
Gold content, g/l	8–108	12–16	1–2	–
Bath temperature	60–70°C	50°C	70–75°C	50°C
pH	6–7	6	7.5–8	7
Plating rate, μm/min	0.1–0.6	0.5	0.6–24	0.1–0.2
Current density, A/dm^2	0.2–1.0	ca. 0.8	ca. 1–40	ca. 1.0
Additives	As/Ti/Pb	As	No As, Ti or Pb	–
Salts/acids	Citrate, phosphate, phosphoric acid	Citrate, phosphate	Phosphate, phosphoric acid	Phosphate
Deposit:				
Purity, %gold	99.9–99.99	99.9	99.9	99.9
Hardness, HV	70–90	250	70–100	100
Colour	Yellow	Deep Yellow	Yellow	Yellow
Appearance	Semi-matte	Bright	Semi-matte	Bright
Application	Electronics	Elec. Contacts, Decorative	Electronics	Decorative

Table 5.3

Different kinds of bath for electroplating of gold

5.13.2 Factors Affecting the Electroplating Process

A uniform gold content over all the surface and throughout can be achieved through the following:

- Control of electrolyte formulation and pH.
- Control of anode surface area and position.
- Control of electrical conditions.
- Control of temperature.

A good electrolyte will contain the metal (or metals) to be deposited in solution in a sufficient concentration. In cyanide-based gold baths, this will be in the form of gold potassium cyanide salt. It will also contain other additives to give good plating properties. These include, for example, additives to improve:

- For good uniform plating, the throwing power of the bath is to be controlled.
- Brighteners are added to assist the brightness of the deposit.
- The additives are added to relieve the stress of the deposits.
- The chemical stability of the electrolyte and may include buffering agents to control pH which is a measure of the acidity or alkalinity of the electrolyte.

Agitation of the electrolyte bath is desired during plating to maintain optimum plating conditions and uniformity of composition.

(i) The anode area and position are important to efficient electrodeposition and uniformity of deposit. There is a tendency for plating to be thicker on cathode areas closest to the anode and thinner in areas hidden from the anode. Correct positioning of the anodes (more than one may be used) and a large anode area (compared to cathode area) is desirable for good plating.

(ii) The electrical conditions during plating are also important for plating quality. In particular, the current density (the current divided by surface area of the piece) plays an important role, particularly in alloy plating where deposit composition is controlled by current density.

(iii) The temperature of the electrolyte can also play a role in getting good plating, particularly in alloy plating.

5.13.3 Surface Preparation

The condition of the surface to be plated is important for good quality electroplating and good adhesion of the deposit. Most plating defects arise from unclean surfaces prior to plating. The surface to be plated must be clean and free from grease, dirt, oxides and tarnish films, polishing compounds, etc. The surface to be plated (the 'substrate') can be prepared by normal polishing techniques and then cleaned in several ways

- Ultrasonic cleaning in detergent solution
- Degreasing in solvents, preferably in an ultrasonic bath
- Acid cleaning with pickling acids
- Steam cleaning under a high pressure jet of steam
- Electrolytic cleaning; this can also activate the surface.
- Chemical cleaning with reagents, often at high temperatures.

5.13.4 Gold Electroplating Systems

For plating pure gold and gold alloy deposits onto substrate material and onto base metals for decorative applications, different electroplating baths are employed. The electrolytes can be classified into cyanide and noncyanide-based and may contain small alloying additions to control colour and other properties. All cyanide-based electrolytes are based on the use of gold potassium cyanide salt, $KAu(CN)_2$, which contains about 68 percent gold. However, most electrolytes do not contain anything like this concentration of gold. Some electrolytes baths are acid, others neutral and others are alkaline, as shown in the classification in Table 5.4.

Electrolyte type	pH	Gold complex	Alloying metals
Alkaline	8–13	$KAu(CN)_2$	Cu, Cd, Ag, Zn
Neutral	6–8	$KAu(CN)_2$	Cu, Cd, Ag
Weakly acid	3–6	$KAu(CN)_2$	Co, Ni, In, Fe
Acid	0.5–2.5	$KAu(CN)_4$	Co, Ni, In, Sn
Cyanide-free, alkaline	8–10	$Na_3Au(SO_3)_2$	Co, Ni, In, Sn

Table 5.4

Electrolytes for gold alloy electroplating

This illustrates the high purity of the deposit and how the properties of the deposit are influenced by plating conditions and electrolyte composition.

5.13.5 Applications

- Its electrical conductivity stays uniform over long periods of time. It is ideally suited for electroplating applications.
- Gold plating offers good corrosion resistance, good solderability, and when alloyed with cobalt, it has very good wear resistance.
- Commonly used in electrical switch contacts, connector pins and barrels, and other applications where intermittent electrical contact occurs.
- Electroplating provides a much more permanent and durable finish.
- Gold plating is often used in electronics, to provide a corrosion-resistant electrically conductive layer on copper, typically in printed circuit boards.

5.14 Electroplating of Nickel

Electroplating of nickel is carried in a sulphate bath or sulphamate bath. The essential features of the two baths are given in Table 5.5.

Detail	Sulphate bath	Sulphamate bath
1. Plating bath composition	250 g $NiSO_4$ + 45 g $NiCl_2$ per dm^3	600 g Nickel sulphamate + 5 g $NiCl_2$ per dm^3
2. Buffer and pH	20 g boric acid buffer pH = 4.5	40 g boric acid buffer pH = 4.0
3. Temperature	313–323 K	323–333 K
4. Current density	20–50 mA/cm^2	50–400 mA/cm^2
5. Additives	Saccharin, Coumarin Sulphonic acid, etc.	Saccharin, sodium sulphonate
6. Current efficiency	95–100%	95–100%
7. Anode	High purity Ni pellets	High purity Ni pellets
8. Cathode	Substrate	Substrate

Table 5.5

Essential features of two baths

5.14.1 Applications of Nickel Plating

1. Corrosion resistance and decorative applications.
2. Undercoat for chromium plating of articles made of steel, brass, zinc, etc.
3. As an undercoat for gold and platinum coating.
4. As a high temperature oxidation resistant heavy plating on machinery parts.
5. In making black nickel-plated name boards military hardware, etc.

5.15 Electroplating of Chromium

Chromium plating is carried out as (i) decorative coating, and (ii) hard plating. The main aspects of the two processes are the same except that for hard chromium plating higher current density is applied. The special features of electroplating baths are given in Table 5.6.

Particulars	Cyanide bath	Sulphate bath
1. Bath composition	250 g chromic acid + 2.5 g conc. H_2SO_4 per dm^3	200 g chromic acid + 2.5 g conc. H_2SO_4 per dm^3
2. Current density	20–40 mA/cm^2	30–60 mA/cm^2
3. Temperature	313–328 K	313–328 K
4. Anode	Insoluble Pb – Sn alloy with PbO_2 coating	Insoluble Pb–Sn alloy with PbO_2 coating
5. Cathode	Substrate	Substrate

Table 5.6

Features of electroplating bath

5.15.1 Chemistry of Electroplating of Chromium

Chromium plating is carried out from a solution of chromic acid (H_2CrO_4) in which chromium is in an oxidation state VI. It gets first reduced to Cr (III) which forms a complex in the solution. Finally, Cr (III) in the complexed form gets reduced as metallic chromium on the substrate surface and forms the deposit.

In chromium plating, we do not use chromium anode due to the following reasons:

In the electrolyte only limited quantity of Cr (III) must be present. Cr (III) in excess produces a black porous deposit. To avoid this, chromium anode which may dissolve to form Cr (III) is not used. Inert anodes of Pb (or) Pb–Sn alloy or Pb–Sb alloy coated with PbO_2 are used. PbO_2 oxidises any free Cr (III) to Cr (VI). Chromium anode is that it may get passivated in the acid medium.

5.16 Electroless Plating

Electroless plating (also known as autocatalytic plating), is a plating process which involves deposition without any current applied. The process is a chemical reaction and is autocatalytic.

Electroless plating is defined as *the deposition of a metal from its salt solution on a catalytically active substrate surface using a chemical reducing agent without the use of electrical energy.*

Catalytically active

Metal ions + chemical reducing agent → metal deposit + oxidized product

The decrease in free energy involved in the redox reaction is responsible for plating. But the surface of the substrate must be prepared to be catalytically active. Activation of the surface of the substrate may be carried out by:

Etching the surface by acid treatment

Electroplating the surface with a thin layer of the same metal or another suitable metal followed by heat treatment.

Nonconducting surfaces like plastics, ceramics or PCB are activated by treating them alternatively with $SnCl_2$ and $PdCl_2$ so as to form a thin layer of palladium by reduction.

5.16.1 Advantages of Electroless Plating

1. Electrical power and electrical contacts are not required.
2. Insulators and semi-conductors can also be plated.
3. No elaborate arrangements are required.
4. Throwing power of the electroless plating bath is high and so objects of complex shapes can be plated.
5. The deposits are more compact and highly adherent.

5.16.2 Comparison of Electroplating and Electroless Plating

Following table contains the important features of electroplating and electroless plating.

Particulars	Electroplating	Electroless plating
1. Driving force	Electrical energy	Decrease in free energy of the redox reaction.
2. Cathodic reaction	$M^{n+} + n\bar{e} \rightarrow \mathrm{M}$	$M^{nt} + n\bar{e} \rightarrow M$
3. Anodic reaction	$M \rightarrow M^{n+} + n\bar{e}$	Reducing agent → Oxidised product + $n\bar{e}$
4. Site of cathodic reaction	Substrate surface	Substrate surface catalytically prepared
5. Site of anodic reaction	Separate anode	Reducing agent in the solution
6. Nature of deposit	Pure metal or alloy	Metal with reducing agent and oxidized products as impurities

Table 5.7

Characteristic features of electroplating and electroless plating

5.17 Electroless Plating of Copper

The composition of electrolyte bath for electroless plating comprises the following:

1. Salt: $CuSO_4\ 5H_2O$, 5g/dm^3
2. Reducing agent formaldehyde: 8g/dm^3
3. Buffer : NaOH 12g + Rochelle salt 14g per dm^3 , pH = 11

4. Complexing agent, EDTA : $5 g/dm^3$
5. Temperature: 298 K

The redox reaction takes place as

$$2HCHO + 4OH^- \longrightarrow HCOO^- + 2H_2O + H_2 + 2e$$

$$Cu^{2+} + 2e \rightarrow Cu$$

$$2HCHO + 4OH^- + Cu^{2+} \longrightarrow HCOO^- + 2H_2O + H_2 + Cu$$

Oxidizes

reduces

The metallic copper obtained as a reduced product in the bath deposits on the metal surface.

5.17.1 Mechanism of Electroless Copper Plating

The composition of the bath for the electroless plating of copper consists of the following:

1. Copper sulphate—source for copper.
2. Formaldehyde—reducing agent.
3. Caustic—basic medium.
4. Chelating agent

This governs the plating rate and has a marked influence on the properties of the deposit and the bath stability. The following chelating agents have been successfully used—(amines, gluconates, gluco-heptaonates, various EDTAs, and tartrates).

It is evident that the formaldehyde, a reducing agent, reduces cupric ions to metallic copper under a strongly basic medium and in the presence of a catalyst palladium.

$$2\ Cu^{+2} + HCHO^- + 30H^- \xrightarrow{Pd} \underset{\displaystyle \overset{\downarrow}{Cu + Cu^{+2}}}{2Cu^+} + HCOO^- + 2H_2O$$

However, the reduction proceeds through a cuprous state. An excess of cuprous oxide formation will cause the reduction reaction to proceed out of control. To inhibit the formation of cuprous oxide, air is bubbled slowly through the electroless copper solution and small complexing agents are added to the solution. The presence of these complexing agents turns the electroless copper bath into a chelated solution. Furthermore, in order for the reaction to progress, certain concentration levels of copper sulphate, formaldehyde, sodium hydroxide, and chelating agent, and the correct chemical ratios of these chemical components must be maintained. The electroless copper bath must be analysed at least every 4 hours to make sure optimum concentrations and ratios are maintained. Replenishment of copper sulphate, formaldehyde, sodium persulphate, and chelating agent must be provided as needed. Constant monitoring and maintenance are required to properly operate the electroless copper bath. Furthermore, in order for the reaction to progress, certain concentration levels of copper sulphate, formaldehyde, sodium hydroxide, and chelating agent, and the correct chemical ratios of these chemical components must be maintained.

5.18 Electroless Plating of Nickel

(i) ***Preparation of the substrate surface:*** The metallic surface is first degreased using organic solvents or by alkali treatment followed by pickling. Aluminium, iron, copper and brass need no preparation. Stainless steel is activated by dipping in 1:1 H_2SO_4. Magnesium alloys are given zinc or copper undercoat by electroplating. Nonmetallic objects are activated by dipping in $SnCl_2$ and then $PdCl_2$ to get a thin layer of Pd.

(ii) ***Plating****:* Plating bath solution contains $NiCl_2$ solution 20 g/dm^3, solution of hypophosphite [reducing agent] 20 g/dm^3, acetate buffer 10 g/dm^3 pH = 4 – 5 sodium succinate 15 g/dm^3 as complexing agent. The temperature is kept at 346 K. The chemical reactions involved are

$$\underset{\text{hypophosphite}}{H_3PO_2^-} + H_2O \longrightarrow \underset{\text{phosphite}}{H_3PO_3^-} + 2H^+ + 2e$$

$$Ni^{+2} + 2e^- \longrightarrow Ni$$

$$\overline{H_3PO_2^- + H_2O + Ni^{+2} \longrightarrow H_3PO_3^- + 2H^+ + Ni}$$

oxidises to

reduces to

Both oxidation and reduction take place on the same substrate surface. In the reaction H^+ ions are released and pH tends to decrease. This may affect the quality of plating.

Hence, the use of buffer is recommended.

5.19 Electroless Plating of Nickel on Aluminium

There is natural deterioration of the surface of Aluminium and its alloys due to greater affinity for oxygen and as result, it forms a thin layer film of aluminium oxide. It is difficult to plate onto an oxide film with good adhesion; though, the natural oxide film formed on aluminium helps to protect it in normal outdoor environments. Under this condition, good adhesion of nickel on the surface of aluminium is not possible. The removal of the oxide layer is desirable. Electroless plating of nickel (EN) is selected because of its ability to protect aluminium from hostile environments and to make a hard, wear-resistant surface that allows aluminium to be used for applications not otherwise suitable. EN deposits have a pleasing appearance and can be used for decorative as well as functional applications (Fig. 5.21).

(a) Problems of Plating on Aluminium

One of the major problems facing plating on aluminium is the presence of oxide layer on the surface due to its affinity towards oxygen. The other, aluminium being amphoteric in nature, i.e., aluminium is dissolved in both acidic and alkaline plating baths and unwanted ion-exchange processes are also likely to occur in the plating solution. All of these characteristics can be eliminated by using an intermediate pretreatment layer deposited by ion-exchange plating (zincate or stannate process).

The *electrode potential* of aluminium compared to some other metals shows that differences between aluminium and pure elements (as well as second phase elements) can be considerable. This shows that proper care should be taken when plating on aluminium to avoid alloying elements giving big electrode potential difference leading to unwanted electrochemical actions during the plating process.

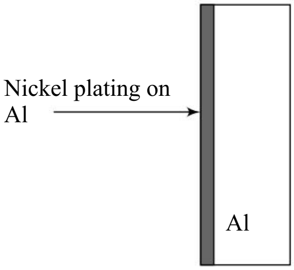


FIGURE 5.21 *Nickel-plated aluminium*

(b) Why Nickel Plate on Aluminium?

Parameters influenced by plating: Wear resistance

- Magnetic properties
- Electrical properties

- Solderability
- Weldability
- Surface hardness
- Temperature resistance
- Corrosion resistance
- Decorative use

(c) Pretreatment of Aluminium using Zincate solution

A precoat of zinc or tin is necessary prior to nickel coating on aluminium for good adhesion property. Using zincate solution, aluminium can be pretreated, and then only plated. Aluminium forms an oxide layer the minute it is exposed to air. This presents a problem when plating aluminium, as the oxide layer prevents the plate from sticking. By using the zincate process as a preplate dip, we chemically remove the oxide layer and at the same time, apply a layer of zinc. The zinc protects the aluminium until it is ready to be plated. As the part is lowered into the plating tank, the zinc is etched away by the nickel solution and plating proceeds onto a clinically clean surface.

Theory of the zincate solution immersion process

The basic reactions in the alkaline zincate solution are the dissolution of aluminium and the deposition of zinc. The basic reaction may be written as:

Anodic:

$$Al + 3\ OH^- \rightarrow Al(OH)_3 + 3e$$
$$Al(OH)_3 \rightarrow AlO_2^- + H_2O + H^+$$

Cathodic:

$$H^+ + e \rightarrow H \rightarrow 1/2\ H_2(g)$$

The zincate ion in strong alkaline solution is probably in the form $Zn(OH)_4{}^{2-}$.

The deposition of zinc may therefore be expressed as:

Cathodic: $Zn(OH)_4{}^{2-} \rightarrow Zn^{2+} + 4\ OH^-$

$$Zn^{2+} + 2e \rightarrow Zn\ (s)$$

The zincate solution is normally made up from zinc oxide and caustic soda and it has been shown that the ratio of these constituents has a profound effect on the adhesion of the subsequent deposit. The physical conditions, time of immersion and temperature, can also affect the degree of adhesion.

(d) Process of electroless nickel plating on aluminium

1. Thoroughly degrease with a suitable solvent such as acetone or lacquer thinner.
2. Remove any oxide, scale, corrosion or organic coatings by sanding with silicon carbide paper, or bead blasting.
3. Etch the workpiece in a tank of electrocleaner with the leads reversed. This will ensure that the corrosion is gone from the surface. Rinse the piece in clean tap water.
4. Make up the zincate solution by adding 1 litre of concentrate to 3 litres of distilled water. Heat the solution to about 25°C.
5. Submerge the piece in zincate for 15 seconds to 2 minutes. A uniform grey appearance is an indication of a properly zincated piece. If the piece reacts violently in the solution, then shorten the amount of time that you submerge it. Water rinse in clean tap water.

6. Plate the workpiece with Nickel solution with a nickel anode until a good covering is noticed. Water rinse with clean water.
7. Plate the workpiece with Nickel just long enough to have a full covering. Rinse in distilled water. The achievable properties that can be added to aluminium are mechanical, magnetic, electrical, thermal, corrosive and decorative. As an example it could be interesting to add wear-resistance (abrasive or adhesive wear), hardness and corrosion protection to aluminium.

5.20 Preparation of Printed Circuit Board (PCB) by Electroless Plating

The printed circuit boards are manufactured on laminates by electroless plating of copper. Metal-ion complexing agents and stabilizers are used in plating solutions. Complexing agents are essential to produce good plating. EDTA bath is used to plate printed-circuit boards by electroless copper plating. The composition for electroless plating is as given in Table 5.8.

Component	Concentration (g/l)
Copper sulphate	15
Formaldehyde	10
Ethylenediamine tetraacetic acid (EDTA)	20
Sodium potassium tartrate	10
Sodium hydroxide	15

Table 5.8

A typical electroless copper plating bath composition

Electroless plating process requires the presence of a catalyst. In electroless copper plating, palladium is generally used to initiate the reduction of copper ions to copper metal from ions in solution. Copper itself is also catalytic (although much less active than palladium).

Once palladium used up, the copper itself acts as the catalyst to sustain the process.

5.20.1 Process of Electroless Plating a PCB Circuitry with Holes

1. **Conditioning**—This ensures that all surfaces are clean and free from grease or metal contaminants—usually done by immersing the substrate in a chemical agent. Deburring of holes also done with rotating brushes.
2. **Pickling and etching**—Prepare the surface for deposition and removes any deformities left from previous stage. Heated sulphuric acid and hydrogen peroxide solutions are commonly used.
3. **Activating**—To create a chemically reducing surface. Two processes traditionally used:
   (a) the laminate is sensitized with stannous chloride solution to leave stannous (tin) ions on the board.
   (b) the board is seeded by dipping in acidic palladium chloride. Palladium ions reduced to a colloidal state during reaction with stannous ions.
4. **Pattern transfer**—The unselected areas of the circuitry paths (Fig. 5.22) are protected by photoresist and exposed to UV light.

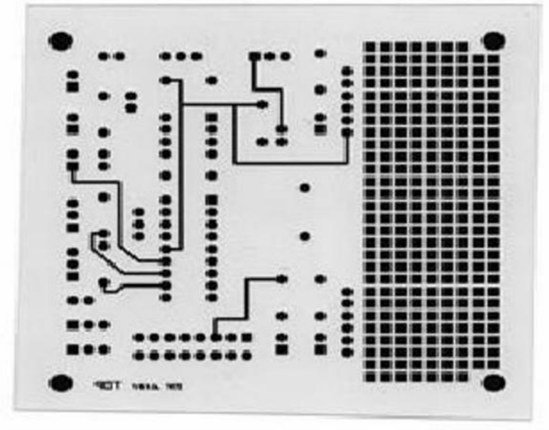


Figure 5.22 *Negative film for electroless plating.*

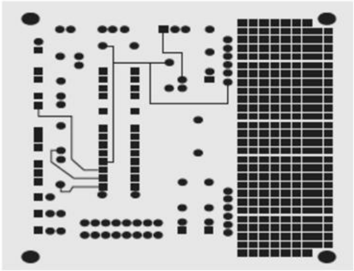


Figure 5.23 *PCB after electroless plating and thinning.*

The soft photoresist portion, except the circuitry pattern, is dissolved in a suitable solvent, cleaned and dried for electroless deposition.

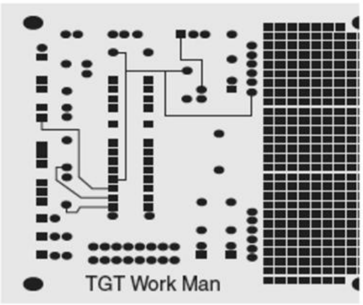


Figure 5.24

6. **Electrical contact—**Connection between the two sides of the laminate is made by drilling holes at required points on the board.
7. **Deposition/plating—**The laminate board is dipped into a bath containing copper ions. The chemically reducing surface of the board (colloidal palladium) causes a reaction and *copper is deposited onto the board along the circuitry pattern as well into the drilled holes making electrical contacts between the top and bottom sides of the PCB pattern* (Fig. 23). Thus, the circuit pattern or track gets printed on the board. Typical thickness is 1–5 μ. 25 μ is possible through for high build applications.

Thinning process—The entire circuit is subjected to tinning process to prevent oxidation of copper (Fig. 5.24).

Review Questions

Metal finishing

1. What is meant by metal finishing? Give a brief account of electroplating of Cu.
[VTU: March 2001]
2. Define the term metal finishing. What are its technological advantages?
[VTU: August 2000]
3. Explain the term electroplating.
4. Define polarization. Mention the factors that affect polarization.
[VTU: 2000 Aug, 2001 Feb]
5. What is decomposition potential? How is it determined? What is its significance?
[VTU: Aug 1999, VTU: Aug 2001]
6. Define overvoltage? Mention the factors that affect overvoltage. Explain the significance of overvoltage in electroplating.
7. Discuss the principle of electroplating.
8. Discuss the mechanism of electroplating.
9. Explain how the following factors affect the nature of electrodeposit
(i) CD (ii) pH (iii) organic additives and (iv) throwing power.
[VTU: Feb 2000, VTU: Aug 2001]

10. With suitable example explain the role of organic additives in electroplating.
11. Explain the term throwing power. How is it experimentally determined? Discuss the factors that affect throwing power.
12. Explain the characteristics of a good and poor deposit.
13. Explain the need for cleaning the metal surface before electroplating. [VTU: Aug 1999]
14. Discuss the steps involved in surface cleaning of the substrate to be electroplated.
15. What are the objectives of electroplating? Explain electroplating of Nickel. [VTU: Aug 1999, Aug 2000, Aug 2001]
16. Give a brief account of electroplating of (i) nickel, (ii) copper, and (iii) chromium.
17. Explain why chromium anodes are not used during chromium plating. [VTU: Aug 2000]
18. What is meant by electroless plating? Briefly explain the electroless plating of copper on PCB. [VTU: Aug 2000]
19. Give a brief account of the electroless plating of (i) copper and (ii) nickel. [VTU: March 1999]
20. Explain the differences between electroplating and electroless plating. [VTU: March 1999]
21. What are the advantages of electroless plating over electroplating? [VTU: August 2002]
22. Account for the following.
    - (i) Concentration polarisation can be minimised by constant stirring of the electrolyte.
    - (ii) Decomposition voltage of water is more than that of the potential for the formation of water.
    - (iii) Only D.C current is used in electroplating.
    - (iv) In the electroplating of Nickel, $NiCl_2$ is added along with $NiSO_4$.
    - (v) The electroless plating bath of Nickel.
23. (a) Define polarization, decomposition potential and overvoltage. Mention their significance with reference to electrode position
    - (b) How so the following affect the nature of electroplating?
      (i) Current density (ii) Temperature (iii) pH and (iv) Organic additives.
    - (c) What is electroless plating? Mention any two advantages. [VTU: July 2008]
24. (a) What is electroless plating? Write two advantages of electroless plating. Describe the process of electroless plating of nickel.
    - (b) Write briefly the process of electroplating of chromium
    - (c) Write a note on (i) Organic additives used in plating (ii) Throwing power of plating bath. [VTU: Jan 2008]
25. (a) What is electroplating? Explain how the following factors influence the nature of electro deposit: (i) Metal ion concentration. (ii) Wetting agents
    - (b) Discuss the electroplating of chromium
    - (c) Mention the technological importance of metal finishing
    - (d) Explain electroless plating of nickel and its applications
26. (a) Explain four factors influencing the nature of an electrodeposit.
    - (b) Define polarization, decomposition potential and overvoltage. Mention their applications with reference to electrodecomposition.
    - (c) Explain the electroless plating of Nickel. [VTU: July 2007]
27. (a) What is electroplating? Give the technological importance of metal finishing.
    - (b) Explain the following factors influencing the nature of deposit.
      (i) Complexing agents, (ii) Brighteners, (iii) Levellers (iv) Wetting agents.
    - (c) Discuss the electroless plating copper on PCB
    - (d) Write a note on overvoltage governing the metal finishing. [VTU: Jan 2007]
28. (a) What is decomposition potential? How is it determined? What is its significance in electroplating?
    - (b) Explain the principle involved in electroless plating. What are the advantages? Explain the electroless plating of copper.
    - (c) Explain the role of the following factors on the nature of the electrodeposit.
      - (i) Current density

(ii) Throwing power
(iii) pH [VTU: July 2006]

29. (a) Define electroplating? Explain process with a neat labelled diagram and its applications.
[Pune University, May, 2016]

30. What are electroless coatings? Explain with suitable example. Give its application.
[Pune University, June, 2015]

31. (a) Explain the following factors influencing rate of electrodeposit:
(i) Current density (ii) metal ion concentration (iii) throwing power
(b) Explain the process of electroplating of chromium for Engineering applications. Indicate the reasons for not employing Cr as anode. [VTU: June/July 2015]

32. (a) Explain the effect of the following plating variables on the nature of the electrodeposit:
(i) p^H of the electrolytic bath (ii) Temperature.
(b) Define the following terms:
(i) Polarisation, (ii) Decomposition potential, and (iii) Overvoltage
(c) What is elelctroless plating? Give any two differences between electroplating and electroless plating.
Explain the process of electroplating of Chromium. [VTU: June/July 2016]

6

Liquid Crystals

Chapter Outline

Introduction

Introduction, classification of liquid crystals-thermo-tropic liquid crystals, lyotropic liquid crystals. Chemical constitution and liquid crystalline behaviour, behaviour of thermotropic liquid crystals, behaviour of lyotropic liquid crystals. Liquid crystalline behaviour in PAA and MBBA Homologous series. Molecular ordering in liquid crystals-nematic liquid crystals, smectic liquid crystals, smectic–B, cholesteric liquid crystals, columnar liquid crystals. Identification of liquid crystal. Electro-optic properties of liquid crystals, reorientation of molecules in electric fields, switchable birefringence. Polymorphism in thermotropic liquid crystals. Applications of liquid crystals, Liquid crystal display. Explanation of liquid crystal display (LCD) work, LCD devices, applications of chiral liquid crystals in thermography.

6.1 Introduction

It is observed that many long-chain organic molecules when heated undergo one or more phase transitions in passing from the solid phase to the liquid phase, indicating the existence of one or more intermediate (meso) phases. The molecular ordering in these mesophases lies between that of a solid and an isotropic liquid. These mesophases are '*ordered fluid phases*'. It is this simultaneous possession of liquid-like and solid-like molecular order in a single phase with anisotropic molecules which gives rise to unique, fascinating and potential technologically relevant behaviour and exhibits a wide range of interesting properties.

The difference between the above three states can be attributed to the temperature of the substance. Temperature is a measure of randomness of the molecules and therefore the higher the temperature the less order exists and the increasing temperature will cause the transition from a solid to liquid state through the intermediate liquid crystal state.

The liquid crystalline mesophases between solid and isotropic liquid phases are shown in Fig. 6.1.

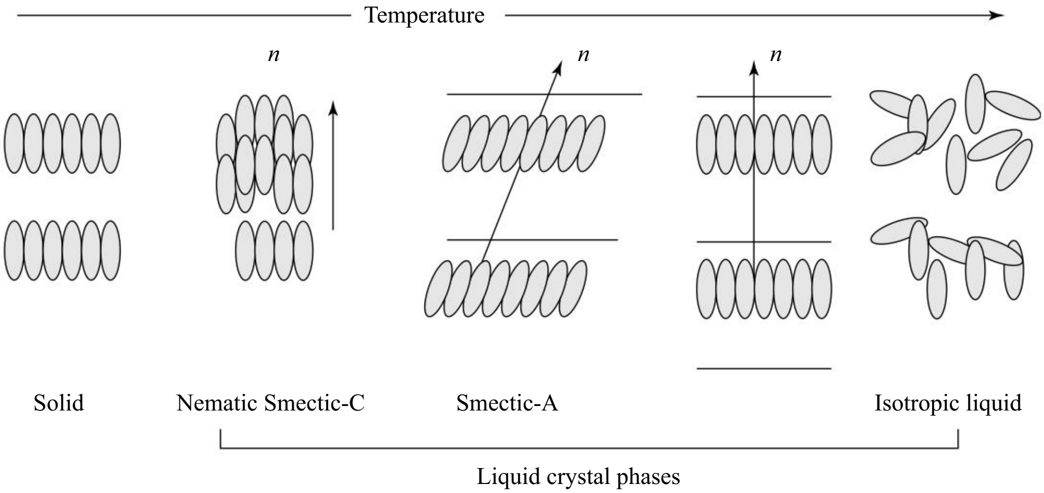


Figure 6.1 *Liquid crystalline mesophases between solid and isotropic liquid.*

Generally, when a solid is heated, the thermal agitation gradually overcomes the cohesive forces of attraction till its melting point is reached where the solid changes into the liquid form. Truly, in certain long-chain organic solids, the tendency towards an ordered arrangement is high so that the kinetic energy (at transition temperature) is sufficient to disrupt the binding between the ends of molecules, but is insufficient to overcome the strong lateral attraction between the long chains of molecules. Consequently, such long-chain organic solids do not melt directly to give isotropic liquid. Instead, they first transform sharply to 'turbid liquids' (mesophases) and then equally sharply to clear isotropic liquids at higher temperatures.

An ordered fluid mesophase of an organic long-chain molecule possessing both solid-like molecular order and liquid-like character is known as a '*liquid crystal*'.

In the solid state, there exists a rigid arrangement of molecules which stay in a fixed '*position*' and '*orientation*' with a small amount of variation from molecular vibration. In the isotropic liquid phase, the molecules have neither fixed '*position*' nor '*orientation*'. However, liquid crystals have '*orientational*' order of molecules with fluid-like properties and are anisotropic in its optical and electro-magnetic characteristics.

The regularity in the distance between the molecules is called *positional order* while the regularity in their orientation is called *orientational order.*

The liquid crystals are composed of moderate and long-sized organic molecules with highly exotic shapes. Because of elongated shape, under appropriate conditions, *the molecules can exhibit orientational order, such that all the axes line up in a particular direction. The average direction of the molecules is called the director, n.*

Liquid crystal materials have several common characteristics. Among these are *rod-like molecular structure*, *rigidness of the long axis* and *strong dipoles and/or easily polarizable.*

Liquid crystals exhibit 'anisotropic' physical (optical, mechanical, symmetrical, etc.) properties. The liquid crystals exhibit birefringence and give interference patterns in polarized light. On the contrary, true liquids are 'isotropic'.

The temperature at which an organic solid transforms into a liquid crystal phase is called the '*phase transition temperature*' and the higher temperature at which the true isotropic liquid is obtained is referred to as the '*melting temperature*'.

6.2 Classification of Liquid Crystals

The liquid crystals are classified into two types:

(a) *thermotropic* and
(b) *lyotropic* liquid crystals

6.2.1 Thermotropic Liquid Crystals

When long-chain organic solids are heated, they undergo sharp phase transitions at particular temperatures yielding liquid crystals.

The phase transitions involving these intermediate phases are most naturally affected by the changing temperature and hence, they are 'thermotropic liquid crystals'.

It is observed that every molecule participates on an equal basis in the long-range ordering of these liquid crystals.

Examples:

1. Ethyl-*p*-azoxycinnamate (Solid) $\xrightarrow{140^\circ\text{C}}$ Ethyl-*p*-azoxycinnamate (liquid crystal)
2. *p*-cholesteryl benzoate (Solid) $\xrightarrow{145^\circ\text{C}}$ *p*-cholesteryl benzoate (liquid crystal)

Thermotropic liquid crystals find extensive applications in electro-optic displays, temperature and pressure sensors, etc.

6.2.2 Lyotropic Liquid Crystals

When a high concentration of long-organic molecules of rod-like shape is dissolved in an isotropic solvent like water, 'lyotropic liquid' crystals are obtained.

The long-range ordering in these lyotropic liquid crystals are mainly due to the solvent–solute interaction. Like thermotropic liquid crystals, lyotropic liquid crystals are also 'fluids' with greater molecular ordering. They are quite different, however, in that they are necessarily systems of two or more components being composed of large organic molecules dissolved in a highly polar solvent. The interaction of the organic molecules with the polar solvent provides 'stability' effect within these ordered lyotropic liquid crystals. These are anisotropic liquid crystals.

Examples:

1. DNA—deoxyribonucleic acid
2. TMV—tobacco mosaic virus
3. Synthetic polypeptides

6.3 Chemical Constitution and Liquid Crystalline Behaviour

A brief salient feature of the molecular structure and liquid crystalline behaviour of organic solids are summarized in the text section.

6.3.1 The Behaviour of Thermotropic Liquid Crystals

The presence of common structural features in the majority of thermotropic liquid crystals makes certain generalizations regarding the type of molecules most likely to show liquid crystalline behaviour. *The two* *structural features essential for liquid crystals are (1) the 'long-chain' organic molecule and (2) the 'rigidity' of the molecule.*

It is evident that a vast majority of organic compounds do show liquid crystalline behaviour due to the presence of 'linkage group' and the 'terminal substituent groups' in the molecule as shown schematically below:

Substituent group–[*Aromatic group–linkage group–aromatic group*]$_n$–substituent group

where $n = 0, 1, 2, 3, \ldots$, etc.

Interestingly, it is observed that many of these molecules are nematic and smectic liquid crystals and if optically active, they may give rise to chiral nematic liquid crystals.

The above representation demands a requirement of 'long-chain molecule' in a liquid crystal which is satisfied, and the need of 'rigidity' is satisfied by restricting the linkage groups to those containing 'multiple bonds' (Fig. 6.2).

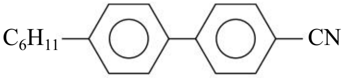


FIGURE 6.2 *p-pentyl-p-cyanobiphenyl (PCB).*

Examples:

1. In this compound, biphenyl is considered as a single aromatic group. Thus, $n = 0$, there is no linkage group and the substituents are $-C_5H_{11}$ and –CN (Fig. 6.3).

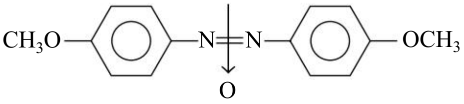


FIGURE 6.3 *p-Azoxyanisole (PAA).*

2. *p*-Azoxyanisole molecule illustrates the case for $n = 1$. If two phenyl groups are linked by an azoxy group and both substituents are $-OCH_3$. There is a double bond in PAA molecule between the nitrogen atoms of the 'linkage group' (Fig. 6.4).

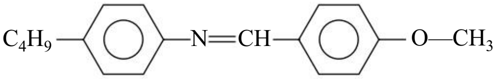


FIGURE 6.4 *4-Methoxybenzylidene, 4-butylaniline (MBBA).*

The MBBA molecule illustrates the case for $n = 1$ with substituents such as methoxy, $-C_4H_9$ and a double bond in the linkage group (Fig. 6.5).

3.

FIGURE 6.5 *2,6 di-(p-Methoxy benzylidine amino)-naphthalene.*

In the above example $n = 2$. The aromatic groups are phenyl, naphthyl and phenyl, respectively, with two –CH=N-linkage groups and two $-OCH_3$ substituents. Again, the linkage groups contain double bonds to provide the requisite rigidity.

Figure 6.6 shows the more common central linkages found in liquid crystals.

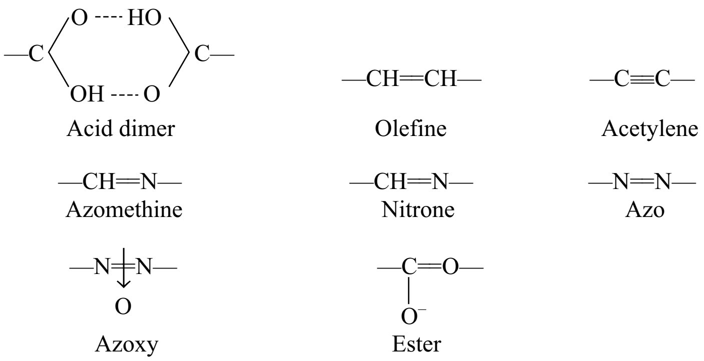
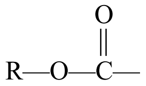
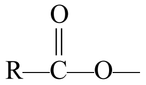


Figure 6.6 *Common central linkages of liquid crystal mesophases.*

The terminal substituent groups of liquid crystals vary widely in chemical nature. These are the few commonly found terminal substituent groups in liquid crystals.

$CH_3—CH_2—$	Alkyl, may be branched
R—O—	Alkyloxy
 $R—O—\overset{\overset{O}{\|}}{C}—$	Carbo alkoxy
 $R—\overset{\overset{O}{\|}}{C}—O—$	Alkyl carboxy
F, Cl, Br, I	Halogens
—CN	Cyano
$—NO_2$	Nitro
$(R)_2N—$	Amino

6.3.2 Behaviour of Lyotropic Liquid Crystals

It is observed that the lyotropic liquid crystals have two or more components. *Specifically, they are mixtures of 'amphiphilic' compounds and a polar solvent (water). Amphiphilic compounds possess two groups that differ in their solubility properties. One part of the molecule will be 'hydrophilic'—highly soluble in water, while the other portion will be 'lipophilic'—highly soluble in a hydrocarbon solvent.*

Typical hydrophilic groups are

$–OH, –COOH, –COONa, –SO_3K, –\ PO_4^{3-}$
$–O–(CH_2–CH_2–O)_n–H, –N(CH_2)_3\ Br, –CH_2–CH_2–NH_2$

Typical lipophilic or hydrophobic groups are

$–C_nH_{2n+1}$, $–C_6H_4$ or any other radical containing long chain of hydrocarbon, with or without aromatic rings included (Fig. 6.7).

Example: Sodium laurate—an aliphatic compound.

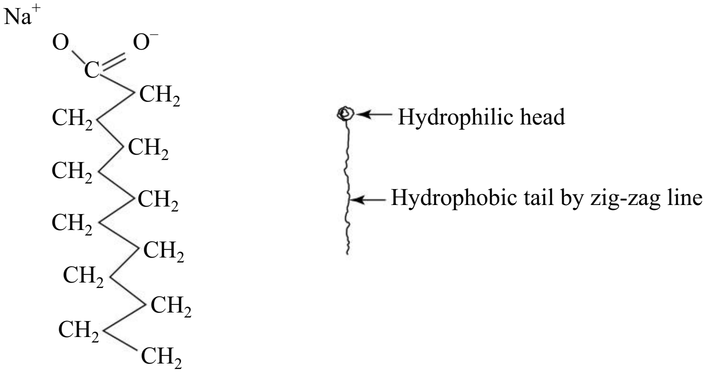
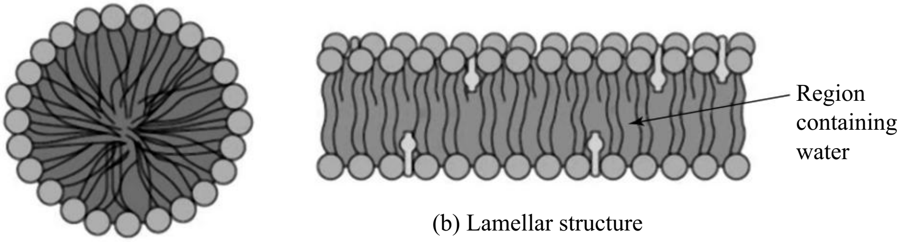


Figure 6.7 *Sodium laurate structure.*

The lyotropic liquid crystals give rise to two structural types as shown in Fig. 6.8.



(b) Lamellar structure

(a) Spherical structure

Figure 6.8 *Two types of lyotropic liquid crystals (a) micelle and (b) lamellar structure.*

6.4 Liquid Crystalline Behaviour in PAA and MBBA Liquid Crystal

Homologous series

The structure–property relationships of many groups of liquid crystal molecules have been studied by differing from one another by the number of methylene (–CH_2) groups only in the terminal substituents. The main work was to emphasize on the following aspects:

1. *Useful liquid crystals of technological importance having specific nematic temperature have been synthesized. They are widely used in liquid crystal display applications. Useful mixture combinations (eutectic) based on homologous series of liquid crystals have been used for device technology.*
2. *The basic studies of homologous series of liquid crystals are focused on a structure–property relationship when very small changes are made at each step in a series, emphasizing the effect of structural parameters and the size.*

In this topic, the discussion is restricted to liquid crystalline behaviour of *p*-azoxyanisole (*PAA*) and *p*-methoxybenzylidine *n*-butylaniline (*MBBA*).

PAA is a liquid crystal stable between 116 and 135°C. One can note that the substituent group is $-O-CH_3$ and two such groups are linked to benzene ring as shown in Fig. 6.9.

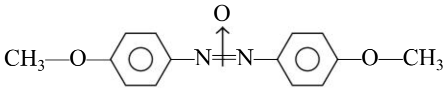


Figure 6.9 *Para-azoxyanisole (PAA).*

A number of homologous compounds are obtained by adding the additional $-CH_2$ group to its side chains. In a PAA homologous series, the substituent groups differ by $-CH_2$ and are $-CH_3$,$-C_2H_5$,$-C_3H_7$, C_4H_9, C_5H_{11}, $-C_6H_{13}$, etc.

The observed behaviour of PAA series is thus:

(i) *transition temperature decreases* with the increasing number of carbon atoms (C = 4, 6, 8, 10, 12, etc.) in side chains of PAA liquid crystal and

(ii) *nematic phases* are observed for PAA liquid crystals having 1–6 carbon atoms in the side chain of the liquid crystal, and the higher ones do show smectic and or both the phases.

1. It can be seen in *MBBA* liquid crystal, one alkyl ($-C_4H_9$) group is attached directly to one of the benzene ring while the $-CH_3$ group is linked to benzene through an oxygen atom. A number of homologous compounds are obtained by adding the $-CH_2$ group to the substituent groups of MBBA (Fig. 6.10).

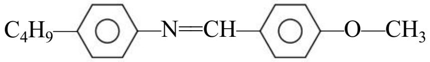


Figure 6.10 *p-methoxybenzylidiene (n-butylaniline), MBBA.*

The results of the studies of PAA and MBBA liquid crystal homologues series can be summarized as:

1. The liquid crystal to isotropic transition observed decreases with increasing chain length.
2. As carbon chain length increases for a homologous series, if more than one intermediate phase is observed for liquid crystals, then the smectic phase is thermally more stable than the nematic and cholesterol liquid crystals. It is often seen, while the lower homologues are purely nematic, the higher are purely smectic and the intermediate homologues exhibit both mesophases.
3. The transition temperatures for odd numbers of carbon atoms in a chain can be correlated by one smooth curve, while homologous with even numbers of carbon atoms fit another smooth curve.

It is observed that the transition temperature difference between two adjacent odd and even members of the series may be only few degrees.

6.5 Molecular Ordering in Nematic, Smectic and Columnar Type Liquid Crystals

The '*degree of molecular motion*' observed in various phases of liquid crystals differentiates the nematic, smectic and other phases and, therefore, it is reasonable to assert that intermolecular attractive forces play a major role in liquid crystal 'stability'.

It can be seen that 'long-range ordering' is found in liquid crystals as the organic molecules become less spherical in shape. Thus, one would anticipate that molecular associations capable of withstanding temperatures higher than the crystalline melting point, that is, liquid crystals would be more frequently observed for molecules of rod-like shape. The '*thermal stability*' of liquid crystals is mainly dependent on the following: 1. *long-chain molecules*, 2. *rigidity*, 3. *permanent dipole moment*, 4. *multiple bonds*, and 5. *aromatic groups*.

The liquid crystals are classified using a scheme based primarily upon their 'symmetry'. Organic long-chain liquid crystals are mainly classified as *nematic* (thread-like), *smectic* (soap-like), *cholesterol* and *columnar*-type.

6.5.1 Nematic Liquid Crystals

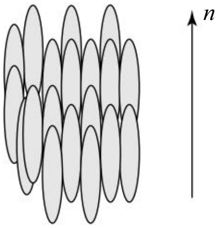


Figure 6.11 *Schematic representation of molecular order in nematic crystals*

The molecular order characteristic of nematic liquid crystals is shown schematically in Fig. 6.11.

The nematic liquid phase is characterized by molecules possessing only orientational but no positional long-range order.

The behaviour of nematic liquid crystals is closer to the true 'anisotropic' liquids. These liquid crystals exhibit normal liquid flow characteristics having low viscosity; flow readily, yet turbid and anisotropic.

The following salient features can be drawn from Fig. 6.11.

(a) There is a *long-range orientational order* in liquid crystals of this type, as the molecules align themselves parallel to each other to some principle axis *n*.

(b) The nematic crystal is a 'fluid' and there are no Bragg's peaks in the X-ray pattern.

The nematic liquid crystals appear to be '*thread-like*' when viewed through polarized light. These liquid crystals are also uniaxial like the smectic phases. When viewed in the direction of lines force of a magnetic field, the nematic crystals appear quite clear.

The phase transition temperatures of few nematic liquid crystals are given in Table 6.1.

Compound	Transition temperature, °C	Melting temperature, °C
p-Azoxyanisole	116	135
p-Azoxyphenetole	137	167
p-Methoxycinnamic acid	170	186
Anisaldizine	165	180
Dibenzalbenzidine	234	260

Table 6.1

Few thermotropic liquid crystals with phase transition temperature

6.5.2 Smectic Liquid Crystals

The smectic state is another distinct mesophase of liquid crystal substance. Molecules in this phase show a degree of *translational order* not present in the nematic. In the nematic state, the molecules maintain the general *orientational* order of nematics, but also tend to *align themselves in layers or planes.*

The optical observations indicated different macro-structures for these smectic phases and, therefore, there are A, B and C types of smectic liquid crystals. The smectic liquid crystals do not flow as normal liquids. A striking behaviour of smectic liquid crystal phase is that they flow in layers as if different planes or sheets are gliding over one another. Like nematic crystals, smectic liquid crystals are always uniaxial and are unaffected by the magnetic field.

(i) Smectic-Aliquidcrystal

Figure 6.12 shows the molecular ordering of smectic-A-type liquid crystals.

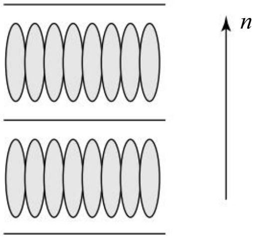


Figure 6.12 *Structure of smectic-A*

The following are the few characteristics of smectic-A liquid crystals:

1. *They do show layered structure*. The thickness of the layer is very close to the full length of the constituent molecules.
2. *There is no long-range molecular order* and each layer of smectic-A is a dimensional liquid.
3. *Smectic-A is optically active*.

(ii) **Smectic-C liquid crystals**

It is a class of smectic-type liquid crystal having the following characteristics:

1. *They form layered structures*,
2. *Each layer is still a two-dimensional liquid*, and
3. *The smectic-C phase is optically biaxial*.

The features as stated above can be explained by assuming that in this type the long *molecular axis is tilted* with respect to normal *n* axis of the layer as shown in Fig. 6.13.

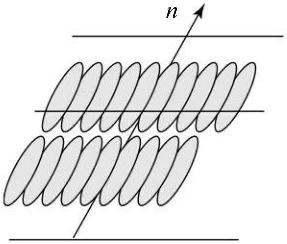


FIGURE 6.13 *Tilted molecular arrangement of smectic-C*

6.5.3 Smectic-B

The following are the characteristics of smectic-B liquid crystals:

1. *They form layered structures and these layers appear to have 'periodicity' and 'rigidity' of two-dimensional liquid.*
2. *Molecular order is found in each layer of the liquid crystal, as evident from XRD results.*
3. *The layers of smectic-B are not flexible.*

Thus, smectic-B appears as the most ordered of the three major phases A, B and C.

6.5.4 Cholesteric Liquid Crystals

The cholesteric (or chiral nematic) liquid crystal phase is typically composed of nematic mesogenic molecules containing a chiral centre which produces intermolecular forces that favour alignment between molecules at a slight angle to another. This leads to the formation of a structure which can be visualized as a stack of very thin2-D nematic-like layers with the director in each layer twisted with respect to those above and below. In this structure, the directors actually form in a continuous helical pattern as shown in Fig. 6.14.

It is observed that certain long-chain organic molecules are 'optically active' exhibiting nematic liquid crystal-like characteristics as well show strong colour effects in polarized light like the smectic phase.

'Liquid crystals having some nematic and some smectic characters are called cholesteric liquid crystals'.

In this cholesteric type, *molecular axis is aligned and the molecules are arranged in layers in which the orientation of axis shifts in a regular way in going from one to the next* as shown in Fig. 6.14.

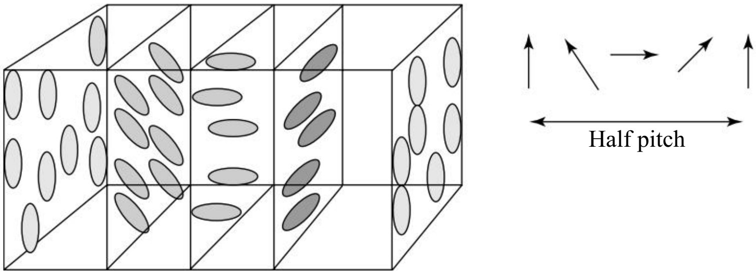


FIGURE 6.14 *Schematic representation of cholesteric molecular order.*

Chiral nematic liquid crystals

When the molecules that make up a nematic liquid crystal are chiral (i.e. they are not symmetric when reflected). Chiral refers to the unique ability to selectively reflect one component of circularly polarized light. These liquid crystals are characterized by their high 'optical rotation'. The term chiral nematic is used interchangeably with cholesteric. In this phase *the molecules prefer to lie next to each other in a slightly skewed orientation. This induces a helical director configuration in which the director rotates through the material* as shown in Fig. 6.15. An important characteristic of the cholesteric (chiral nematic) mesophase is the '*pitch*'.

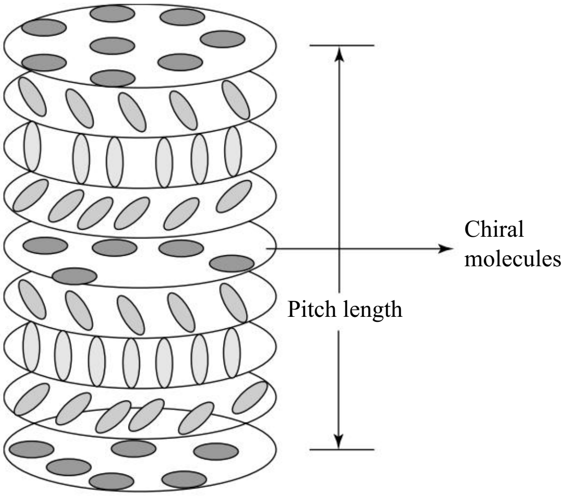


Figure 6.15 *A chiral cholesteric liquid crystal.*

The pitch, p, is defined as the distance it takes for the director to rotate one full turn in the helix as illustrated in Fig. 6.16.

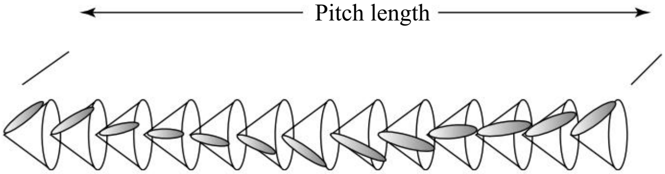


Figure 6.16 *Pitch length of a nematic chiral liquid crystal.*

A by-product of the helical structure of the chiral nematic phase is its ability to selectively reflect light of wavelengths equal to the pitch, so that a colour will be reflected when the pitch is equal to the corresponding wavelength of light in the visible spectrum. The pitch becomes shorter as the temperature is increased and vice versa. The pitch of the spiral decreases and the reflected colour depends sensitively on the temperature.

Example: Cholesteryl benzoate

Cholesteryl benzoate undergoes phase transition at 145°C to liquid crystal and melts to isotropic clear liquid at 178°C.

$$\text{Cholesteryl benzoate (solid)} \xleftrightarrow{145^\circ} \text{Cholesteryl benzoate (liquid crystal)} \xleftrightarrow{178^\circ} \text{Cholesteryl benzoate (isotropic liquid)}$$

6.5.5 Columnar Mesophases

Columnar liquid crystals are different from the previous types because they are shaped like *disks instead of long rods*. *This mesophase is characterized by stacked columns of molecules.* The columns are packed together to form a two-dimensional crystalline array as shown in Fig. 6.17. The arrangement of the molecules within the columns themselves leads to new mesophases.

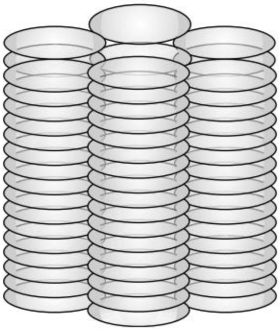


Figure 6.17 *Columnar liquid crystal phase.*

6.6 Identification of Liquid Crystals Using Optical Microscopy

The presence of a liquid crystalline phase is usually quite easy to identify but the identification of the phase type is often very difficult. '*Optical polarizing microscopy' is the most common method used to identify liquid crystal phases*. A small sample of liquid crystal is placed on a microscope slide with a cover slip. The slide is placed in a hot-stage of variable temperature which is placed under a microscope between crossed polarizers. When viewed between crossed polarizers, an isotropic liquid will appear black because polarized light will be extinguished by the second crossed polarizer. Liquid crystals have a *certain ordering* of their constituent molecules and are *birefringent.* Accordingly, plane polarized light is affected by the liquid crystal material and does not, in all cases, get extinguished by the second crossed polarizer; this generates a coloured texture or pattern. If a sample with an unknown liquid crystalline phase is mixed with a known and fully characterized liquid crystal then complete miscibility across the phase diagram indicates that the two phases are identical. Such miscibility studies are frequently employed in the identification of liquid crystals.

6.7 Electro-optic Properties of Liquid Crystals

The bulk order in liquid crystals has profound influence on the way light and electricity behaved in the material. *Liquid crystals are anisotropic not only in their shape, but also in all kinds of optical, electrical and magnetic parameters.* If one looks at the value of one parameter, it is different in the direction along the director of the molecules than the direction perpendicular to it. This gives rise to all kinds of interesting possibilities.

Two very interesting phenomena are the following: 1. *the orientation of the molecules in an electric field,* and 2. *birefringence of the molecules.*

6.7.1 Reorientation of the Molecules in Electric Fields

The dielectric constant of liquid crystal is anisotropic. If the dielectric constant is larger in the direction along the molecule than in the direction perpendicular to it, one speaks of +ve anisotropy. When molecules with such characteristic are brought in a sufficiently strong electric field, they tend to align themselves in the direction of field. This is illustrated in Fig. 6.18.

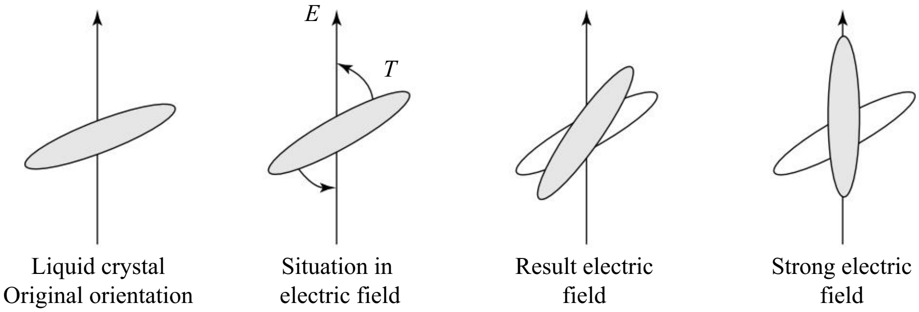


Figure 6.18 *Orientation of liquid crystals in electric field.*

Originally orientation is almost flat. When a field with direction E is applied there is a force T that tends to align the molecule parallel to the field. When the field is strong enough, the molecules will be almost parallel to the field. Liquid crystals with –ve anisotropy will have a tendency to orient themselves perpendicularly with the electric field. *The importance of this behaviour is obvious because it gives the possibility of a switchable medium by simply applying a voltage over a liquid crystal cell.*

6.7.2 Switchable Birefringence

Liquid crystals are found to be birefringent (having two indices of refraction) due to their anisotropic nature as shown in Fig. 6.19. A medium which has a refraction index depending on the direction is called a birefringent medium.

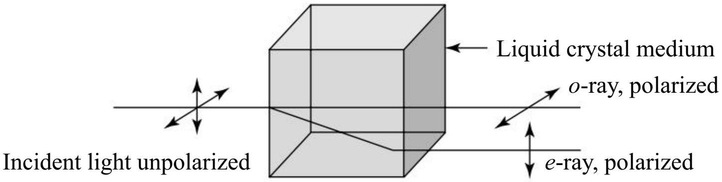


Figure 6.19 *Light travelling through a birefringent LC medium will take one of two paths depending on its polarization.*

Because the orientation of the molecules can be controlled using an electric field, this gives the possibility of a controllable birefringent medium. *Birefringence is very important for modifying and controlling the polarization of light propagating through the medium.* If the direction of the orientation varies in space, the orientation of the light (i.e. polarization) can follow this variation. *A well-known application of this phenomenon is the* liquid crystal display.

6.8 Polymorphism in Thermotropic Liquid Crystals

'Polymorphism'is common in long-chain organic molecules. It is observed that many of these solids pass through more than one mesophase between solid and isotropic liquids (Table 6.2). The temperature at which they transform to another phase is known as its 'phase transition temperature'. The stability of a particular type of liquid crystal depends on to the extent of destruction of molecular order by increasing temperature. Thus, the more ordered the liquid crystal phase, the closer in temperature it lies to the solid state.

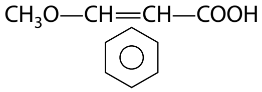
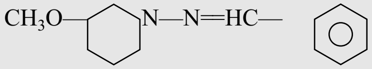
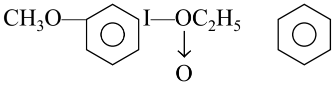
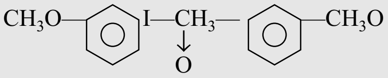
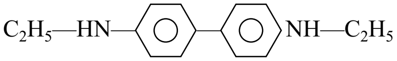
Compound structure	Transition temperature *t*°C	Melting temperature *t*°C
 methoxycinnamic acid	170	186
 Anisaldizine	165	185
 *p*-azoxyphenetole	137	167
 *p*-azoxyanisole	116	135
 di-ethyl benzidine	115	120

Table 6.2

Polymorphic liquid crystals

Example: The polymorphic forms of ethyl anisal *p*-amino cinnamate are depicted as follows:

Solid ⇔ Smectic-B ⇔ Smectic-A ⇔ nematic ⇔ isotropic liquid

In general, for a material having nematic and smectic tri-phases, the order of mesophase stability with increasing temperature will be

solid ⇔ smectic-B ⇔ smectic-C ⇔ Smectic-A ⇔ nematic ⇔ isotropic liquid

6.9 Applications of Liquid Crystals

Liquid crystal materials which have both crystalline and liquid properties are the key component in today's technology because they possess unique optical properties. Liquid crystal technology has had a major impact on many areas of science and technology, as well as device technology. Applications for this special kind of material are still being discovered and continue to provide effective solutions to many different problems.

6.9.1 Liquid Crystal Displays

Vertical and horizontal alignments of liquid crystal molecules have different optical properties. Accordingly, liquid crystals are ideal for display devices because

(i) *the mesophase is fluid and therefore the molecules are easily moved by the application of an electric field.*

(ii) *the phases are structured, and the alignment of the molecules in a thin film of mesophase can be controlled either by boundary conditions or by the application of a small electric field.*

The elongated liquid crystal molecules are surface aligned in one direction and give one optical property and when an electric field is applied, the fluid molecules reorient to give a different optical property—hence they use in displays.

A *liquid crystal display* (LCD) consists of an array of tiny segments (called pixels) that can be manipulated to present information. This basic idea is common to all displays. The liquid crystal devices (LCDs) are now used in a wide range of equipment and apparatus, such as watches, calculators, portable colour televisions, lap-top computer screens, car, ship and air craft instrumentation.

6.9.2 Explanation of LCD Display Work

LC molecules have rod-like structure and *rotate the direction of polarized light based on their alignment*. By applying an electric field, the alignment of the molecules can be controlled.

A twisted nematic liquid crystal display cell is made up of the following (Fig. 6.20).

- Two bonding glassplates, each with conductive transparent conductive coating (indium tin oxide, ITO) that act as an electrode;
- Spacers to control the cell gap precisely;
- Two crossed polarizers (the polarizer and the analyser) and
- A nematic liquid crystal material.

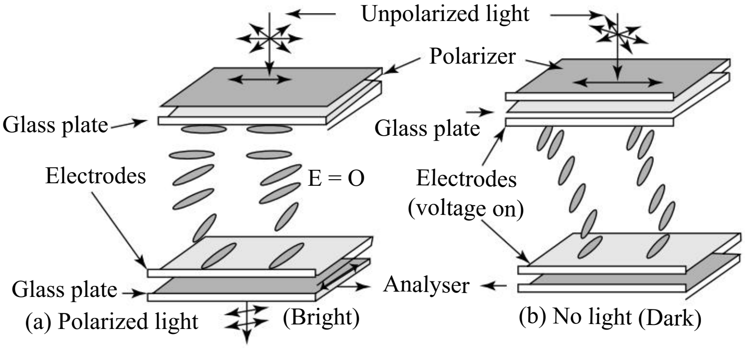


FIGURE 6.20 *Geometry of a working LCD cell.*

The polarizer and analyser, which are arranged parallel to the director orientation at their adjacent glass plates, are oriented at 90 degrees to each other.

In a display, the liquid crystal material is sealed between the two glass plates, each of which has a polarizer on its outer side. Light passing through one of the polarizers has its polarization rotated with the alignment direction of the liquid crystal. As the light reaches the second analyser, it can be passed or blocked, depending on the extent of rotation. The transparent conductor (usually, ITO) on the inner surface of the glass field is used to control the electric field on the cell, and thus the direction of liquid crystal molecules. The LC cell, as described, acts as a light switch.

When a voltage is applied to the electrodes, the liquid crystal molecules tend to align with the resulting electric field, E and the optical wave property of the cell is lost. The cell is *dark*, as it would be without the LC present (Fig. 20b). When the electric field is turned off, the molecules relax back to their twisted state and the cell becomes *transparent* again (Fig. 6.20a).

6.9.3 LCD Devices

The application of liquid crystal display technology can be found in the following devices: digital watches, calculators, flat TV displays, thermometers, which show the body temperature in colour and computers with fantastic colour graphics.

(i) Liquid crystal thermometers

Chiral (cholesteric) liquid crystals reflect light with a wavelength equal to the pitch. Because the pitch is dependent upon temperature, the colour reflected also is dependent on temperature. Liquid crystals make it possible to accurately gauge temperature just by looking at the colour of the thermometer.

(ii) Optical imaging

In optical imaging technology recording, a liquid crystal cell is placed between the two layers of photo-conductor. Light is applied to the photo-conductor, which increases the material's conductivity. This causes an electric field to develop in the liquid crystal corresponding to the intensity of the light. This electric pattern can be transmitted by an electrode, which enables the image to be recorded. This technology is still being developed and is one of the most promising areas of liquid crystal research.

(iii) Thermography

The cholesteric types of liquid crystals are employed for detecting tumours in the body by a method called thermography.

(iv) As polymer composite material

If a polymer is extruded in the liquid crystal phase, the anisotropic ordering of molecules confers an extremely high strength to the polymer. *Kelvar* is a good example of such high strength liquid crystal polymer and is used in bulletproof vests and car body panels.

6.10 Applications of Chiral Liquid Crystals in Thermography

Chiral nematic liquid crystals are of great technological importance *because of their ability to selectively reflect light of a wavelength equal to that of the pitch length*. So if the pitch length is of the order of the wavelength of coloured light then coloured light will be reflected. Additionally, the pitch length of the helix of a chiral nematic phase changes with temperature. At high temperature, the pitch length becomes wound up and short and hence the light reflected is *blue*, but at lower temperature the pitch unwinds and becomes long which causes the reflection of *red* light.

Liquid crystal (Chiral) thermography is used in a wide variety of applications. These applications include medical diagnosis, electronic cooling, gas turbine heat transfer, boiling heat transfer, fluid temperature measurements, detection of structural flaws, detection of hot spots in electronic circuits and aerodynamic testing for cooling of engine parts.

(a) Medical or clinical thermography

Clinical thermography is nonevasive, diagnostic imaging procedure involving the detection and recording of a patient's skin surface thermal patterns, using instruments which can provide visual and quantitative documentation of these temperature measurements.

Thermography is an imaging technology which provides information on the normal and abnormal functioning of the sensory and sympathetic nervous systems, vascular dysfunction, myofascial trauma and local inflammatory processes.

Liquid crystals can be used as a safe and effective means for the evaluation of complex pain states associated with arthritis, soft tissue injuries and back pain diseases. It facilitates identification of the damage to the nervous system.

For thermographic studies, liquid crystals having '*helical structure*' of molecules are used. The pitch of the helical molecules changes with temperature. It is noticed that the wavelength of the rays reflected from them are different at different temperatures. This is marked by changing colour of the reflected ray, since the local temperature of the tumour affected the part of the body is different from the healthier part, the rays of light from two parts reflected through such a crystal will have different wavelengths or different colours.

(b) Surface thermography in electronic industry

Thermography is the measurement of surface temperature. Measurement of the temperature distribution of an operating component junction and case temperature is necessary to complement and characterize the thermal and electrical performance of today's high power density electronic components. There is a need to identify 'thermal spots' that might lead to degraded electrical performance. Liquid crystal thermography is presently being used for high-resolution temperature measurement in applications ranging from submicron to larger electronic devices and complete circuit boards.

(c) Aerodynamic testing

In aerodynamic testing for cooling of engine parts, liquid crystal thermographic technique has been adopted.

Flow visualization has been extensively used in understanding the flow (fluid mechanics) phenomenon. An important requirement for flow visualization of aerospace applications is that the flow visualization technique should be able to operate at the speed and pressures of interest.

Liquid crystals in fact have been used successfully at high altitudes. The state of the boundary layer present on the aircraft dictates the heat transfer, being low where the boundary layer is laminar and high when this becomes turbulent. The regions of high heat transfer will be cool relative to those with low heat transfer regions and this is reflected in the colour display of the liquid crystals.

(d) Radiation sensing

Infrared thermography uses a specially designed infrared source to make images in LCD that show surface heat variations.

Thermal imaging as well as specialized liquid crystal thermal detectors has grown into an important technology that is applied directly by users on the ground or from flying aircraft. Using thermal sensors coupled with optical systems, one is able to see in the dark by detecting varying temperatures from different objects in the scene.

Review Questions

1. What is a liquid crystal?
2. How does a liquid crystal differ from that of a solid and an isotropic liquid?
3. What is phase transition? Mention any two examples of solid-liquid crystal phase transitions.
4. What are directors in liquid crystals? Illustrate with examples.
5. How are liquid crystals classified? Give an example in each case.
6. What do you mean by thermotropic liquid crystals? Illustrate with examples.
7. What are lyotropic liquid crystals? Give a brief illustration.
8. Illustrate the chemical constitution of liquid crystalline behaviour of thermotropic liquid crystals.
9. Name few common central linkages found in liquid crystals.
10. Name few common terminal substituent groups of liquid crystals.

11. Illustrate the chemical constitution of liquid crystalline behaviour of lyotropic liquid crystals.
12. Write a brief note on the liquid crystalline behaviour of PAA and MBBA homologous series.
13. How are liquid crystals classified on the basis of molecular ordering?
14. What are nematic type liquid crystals? Give their salient features.
15. What are smectic liquid crystals? How are they classified? Mention their salient features.
16. What are cholesteric liquid crystals? Mention a cholesteric liquid crystal.
17. What are columnar type liquid crystals?
18. Discuss the polymorphic behaviour of thermotropic liquid crystals with examples.
19. What are the applications of liquid crystals?
20. Enumerate the applications of chiral nematic liquid crystals in thermography.
21. What is liquid crystal thermography?
22. Explain the use of liquid crystals in thermal imaging.
23. Explain the application of liquid crystals in surface thermography.
24. Explain the use of liquid crystal in aerodynamic testing.
25. Explain radiation sensing by liquid crystals.
26. What are liquid crystals? Distinguish between smectic, nematic and cholesteric types with suitable examples. [VTU Jan, 2005]
27. Mention two characteristics of the mesophase state of matter. Define a director. Draw a relevant figure to represent the director in a crystalline solid and also in a liquid crystal. [VTU Jan, 2006]
28. What are liquid crystals? Explain the molecular ordering in the following liquid crystal phases:
    - Nematic crystal phase
    - Chiral nematic phase
    - Smectic phase [VTU Jan, 2008]
29. (a) Explain the following with examples:
    (i) Thermotropic liquid crystals and
    (ii) Lyotropic liquid crystal
    (b) What is homologous series? Explain the liquid crystalline behaviour of homologous of MBBA. [VTU Jan, 2007]
30. (a) With suitable examples, explain the liquid crystalline behaviour in the PAA homologous series.
    (b) Explain with examples, the liquid crystalline behaviour of compounds based on their chemical constitution.
    [VTU July, 2007]
31. (a) What are liquid crystals? Distinguish between thermotropic and lyotropic liquid crystals with examples.
    (b) Explain the working of liquid crystals in display systems. [VTU July, 2008]

7

High Polymers

Chapter Outline

Introduction

Introduction. Basic definitions. Polymer-monomer, polymerization, functionality, degree of polymerization, molar masses of polymers. Classification of polymers. Types of polymerization. Mechanism of free radical addition polymerization, condensation polymerization. Difference between addition and condensation polymerization. Methods of polymerization-bulk polymerization, solution polymerization, suspension polymerization, emulsion polymerization, mechanism of emulsion polymerization. Glass transition temperature, factors that affect the value of T_g, Structure and property relationship of polymers, crystallinity, impact and tensile strength of polymers, molar masses, tacticity, viscoelasticity, plastic deformation of polymers, chemical resistivity. Synthesis, properties & applications of commercial polymers. Polyethylene (polythene), polytetrafluoro ethylene—PTFE (teflon), polymethyl methacrylate: PMMA (Plexiglas), polyurethane (PU), polycarbonates (PC). Plastics—thermoplastic and thermosetting polymers, difference between thermoplastic and thermosetting polymers, phenol formaldehyde resins, urea-formaldehyde, novolacs, resoles. Elastomers—natural rubber, deficiencies of natural rubber, advantages of synthetic rubbers, vulcanization, synthetic rubbers. Adhesive, applications of adhesive, synthetic adhesives, epoxy resins (araldite), compounding of resins and plastics. Techniques of moulding of plastics—compression, injection, film blown and extrusion moulding. Polymer composites, materials used in polymer composites, the need to develop polymer composites: fibreglass, Kevlar, carbon fibre. Conducting polymers, conducting polyacetylene, conducting polypyrrole, conducting polyaniline, applications of polyaniline.

7.1 Introduction

Later part of twentieth century is called *plastic age*. Man's eagerness to compete with nature has led him to the *synthesis of polymers*. He has made a variety of polymers with wide ranging properties such as softness like silk and wool and strong like steel. We find use for plastics in all our activities of daily life, starting with brushes we use for cleaning our teeth, combs, mugs, buckets, ballpoint pens, chairs, carry bags, packing materials, rain coats, boxes, storage vessels, water tanks, clothes made of synthetic materials like polyesters are all polymers. Nature has also given us a number of high polymers such as proteins, carbohydrates, silk, wool, cotton, rubber, leather, etc.

The first modified (semi synthetic) polymer that came to the market was cellulose nitrate in 1860. Purely synthetic bakelite was introduced in 1910. From 1920 onwards a lot of developments in the science and technology of polymers have taken place. The plastic industry has grown very fast surpassing the growth of any other industry. In fact plastics are slowly replacing all our conventional materials. Scientists have made replacement for natural materials. In the recent years, research is being carried out on *conducting polymers* (*synthetic metals*), which is briefly discussed at the end of this chapter.

The only main drawback in the use of plastics is that even though many of them can be recycled most of them are not biodegradable. Plastic waste poses a major threat to environment.

7.2 Basic Definitions

7.2.1 Polymer

A polymer is a large molecule built by the repetition of small and simple chemical unit called monomer. The repeating chemical units are covalently linked to each other in a macromolecule.

The properties of a polymer are entirely different from those of the chemicals (or monomers) from which it is formed. The molar masses of polymers vary from 10^3 to 10^7.

7.2.2 Monomer

Monomer is defined as a simple molecule with two or more binding sites through which it forms covalent linkages with other monomer molecules to form the macromolecule.

Monomers are thus building blocks of polymers. All simple molecules cannot behave as monomers but only those with two or more bonding sites can act as monomers. Thus molecules like ammonia, water, ethanol, etc. are not monomers. Alkenes, vinyl chloride, adipic acid and glycol with two bonding sites act as monomers.

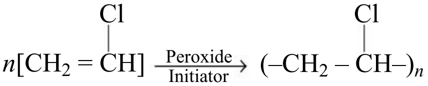
7.2.3 Polymerization

Polymerization is defined as the chemical reaction in which a monomer is converted to the polymer under specific conditions. Monomer alone cannot undergo polymerization, but requires the presence of a chemical called initiator.

$$\text{Monomer} + \text{Initiator} \xrightarrow[\text{pressure}]{\text{required temperature}} \text{Polymer}$$

e.g.

	Monomer	**Polymer**
1.	Vinyl chloride	Polyvinyl chloride
2.	Vinyl acetate	Polyvinyl acetate
3.	Ethylene	Polyethylene
4.	Styrene	Polystyrene



$$n[\mathrm{CH_2} = \overset{\mathrm{Cl}}{\overset{|}{\mathrm{C}}}\mathrm{H}] \xrightarrow[\text{Initiator}]{\text{Peroxide}} (-\mathrm{CH_2} - \overset{\mathrm{Cl}}{\overset{|}{\mathrm{C}}}\mathrm{H}-)_n$$

Vinyl chloride (monomer) PVC (polymer)

7.2.4 Functionality

Functionality of a monomer is defined as the number of bonding sites present in a molecule of the monomer.

Example: All double-bonded compounds $CH_2{=}CH_2$ (ethene), $CH_3{-}CH{=}CH_2$ (propene), $CH_2{=}CHCl$ (vinyl chloride) have two bonding sites, the double bond. Their functionality is, two and is bifunctional. Similarly, glycols $\begin{bmatrix} CH_2-OH \\ | \\ CH_2-OH \end{bmatrix}$, adipic acid, hexamethylene-diamine $[H_2N(CH_2)_6{-}NH_2]$, etc., are bifunctional. For phenol, its 2, 4, 6 positions act as active centres. Therefore, its functionality is 3, i.e. it is trifunctional.

7.2.5 Degree of Polymerization (DP)

Degree of polymerization is *defined as the number of repeating units present in a polymer molecule.*

$$n[CH_2{=}CH_2] \rightarrow (-CH_2{-}CH_2-)_n$$

In this, '*n*' is a whole number and it is called the degree of polymerization. In the case of addition polymerization, the molar mass of the polymer can be calculated using the formula:

Molar mass of addition polymer = Degree of polymerization × Molar mass of monomer.

Polymers produced in a reaction do not have the same degree of polymerization and show variation in molar mass. Hence, normally only average degree of polymerization and average molar mass ($\bar{M}_n$) are used.

$$\bar{M}_n = \underset{\left(\begin{smallmatrix}\text{average . degree . of} \\ \text{polymerisation}\end{smallmatrix}\right)}{\overline{\text{DP}}} \times \text{Molar mass of monomer}$$

7.2.6 Molar Masses of Polymers

Polymers normally contain a mixture of chains of different lengths (or different degree of polymerization). Thus, it is not possible to assign a definite molar mass to a polymer. It is usual to adopt statistical methods in assigning average molar masses of polymers. Two important averaging are adopted (i) number average molar mass, and (ii) weight average molar mass.

Number average molar mass

$\bar{M}_n$ is the ratio of the sum of molar masses of individual molecules to the total number of molecules in the mixture.

Thus if in a sample n_1 molecules have molar mass M_1, n_2 molecules with molar mass M_2, etc., till n_i with M_i then,

$$\bar{M}_n = \frac{n_1M_1 + n_2M_2 + \cdots n_iM_i}{n_1 + n_2 + \cdots + n_i} = \frac{\Sigma n_iM_i}{\Sigma n_i} \quad (1)$$

Weight average molar mass ($\bar{M}_W$)

In a mixture, if w is the mass of polymers with molar mass M_1, W_2 with molar mass M_2 up to W_i with molar mass M_i then,

$$\bar{M}_w = \frac{w_1M_1 + w_2M_2 + \cdots + w_iM_i}{w_1 + w_2 + \cdots + w_i} = \frac{\Sigma w_iM_i}{\Sigma w_i} \quad (2)$$

But the number of moles $n = \dfrac{w}{M}$ and $w = nM$

Therefore, replacing w_1 by n_1M_1, w_2 by $n_2M_2 \cdots w_i$ by $n_i\, M_i$ in Eq. (2) we have

$$\bar{M}_w = \frac{n_1M_1^2 + n_2M_2^2 + \cdots + n_iM_i^2}{n_1M_1 + n_2M_2 + \cdots + n_iM_i} = \frac{\Sigma n_iM_i^2}{\Sigma n_iM_i}$$

Usually $\bar{M}_w$ is greater than and therefore, $\dfrac{\bar{M}_w}{\bar{M}_n}$ is greater than or equal to one. This ratio is called distribution ratio.

If $\dfrac{\bar{M}_w}{\bar{M}_n} = 1$ the polymer is homogeneous and contains polymers of same chain length. The higher the deviation of the ratio from 1, the higher the degree of heterogeneity of the polymer.

7.3 Classification of Polymers

Several different methods of classification of polymers are possible. The most important methods are given below:

(i) *Natural polymers and synthetic polymers*
The classification depends on the origin of the polymer

Polymers

Natural polymers
Examples are
starch, cellulose, proteins, silk,
jute, wool, rubber

Synthetic polymers
(man-made polymers)
Examples are polythene, polypropene, polystyrene,
PVC, bakelite, nylon, polyesters, etc.

Some polymers are semi-synthetic polymers which are modified natural polymers like cellulose acetate, cellulose nitrate, etc.

(ii) According to the *nature of the chemical reaction* involved in the formation of the polymers, they are classified into (a) addition polymers, and (b) condensation polymers.

Polymers

Addition polymers

They are formed by the addition of olifinic monomers without the elimination of byproducts. (Molar mass of the homo polymer will be the product of the degree of polymerization and the molar mass of the monomer.
Examples: Polythene, polypropylene
Plexiglass PVC, etc.

Condensation polymers

They are formed by the intermolecular condensation of two different monomers with the elimination of byproducts like Ammonia water HCl etc. molar mass of the polymer will be the product of the degree of polymerization and the molar mass of the repeating unit.
Example: Bakelite, nylon polyester

(iii) According to the *structure and end use*, polymers are classified into four types.

(a) ***Elastomers*:** They undergo large elongation when stretched and retain their shape on releasing the force. The polymer chain is long and coiled and in between chains only weak forces of attraction exists.
Examples: Natural rubber, buna-s rubber, etc.

(b) ***Fibres***: Fibres are thin thread-like polymers. These do not undergo deformation. These contain long chain molecular strains held together through hydrogen bonding.
Examples: Wool, silk, cotton, nylon, terylene, polyesters, etc.

(c) ***Resins***: Resins are polymers of low molar mass in liquid or semi-solid form. They may undergo further polymerization to give plastics (higher molar mass).
Examples: Phenol-formaldehyde resin, urea-formaldehyde resin, epoxy resins, etc.

(d) ***Plastics:*** Plastics are polymers which can be moulded into desired articles by application of pressure and temperature.
Examples: PVC, plexiglass, etc.

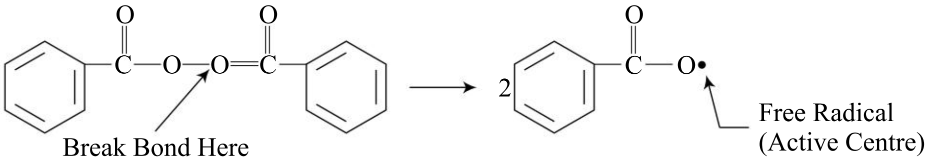
7.4 Types of Polymerization

Polymerization reaction can be carried out through one of the two ways:

- Addition polymerization
- Condensation polymerization

7.4.1 Mechanism of Free Radical Addition Polymerization

The most common type of addition polymerization is free radical polymerization. *A free radical is simply a molecule with an unpaired electron.* The tendency for this free radical to gain an additional electron in order to form a pair makes it highly reactive so that it breaks the bond on another molecule by stealing an electron, leading that molecule with an unpaired electron (which is another free radical). Free radicals are often created by the division of a molecule (an initiator) into two fragments along a single bond. The following diagram shows the formation of a radical from its *initiator*, in this case, e.g. benzoyl peroxide.



The stability of a radical refers to the molecule's tendency to react with other compounds. An unstable radical will readily combine with many different molecules. However a stable radical will not easily interact with other chemical substances. The stability of free radicals can vary widely depending on the properties of the molecule. The active centre is the location of the unpaired electron on the radical because this is where the reaction takes place. In free radical polymerization, the radical attacks one monomer, and the electron migrates to another part of the molecule. This newly formed radical attacks another monomer and the process is repeated. Thus the active centre moves down the chain as the polymerization occurs.

There are three significant reactions that take place in addition polymerization: initiation, propagation and termination. These separate steps are explained below in the polymerization of ethylene.

Polymerization of ethylene:

A polymer formed by direct addition of repeated monomers without the elimination of by-product is called addition polymer.

The addition polymerization follows the three major steps:

- Initiation
- Propagation
- Termination

(i) Free radical mechanism

Initiation:

An *initiator* (usually, peroxides, azo-compounds or peracids) in a reaction process breaks down to give free radicals by homolytic fission of a covalent bond. The initiators are thermally unstable compounds and their homolytic fission gives free radicals.

$$\text{R—R} \rightarrow 2\text{R}^\bullet$$

Initiator molecule free radicals

where R = peroxides, azo-compounds or peracids.

The endothermic energy required to obtain free radicals is supplied by the absorption of heat (thermal initiation) or by absorption of light (photo-chemical). The initiator free radical may combine with a monomer olefinic molecule to form monomer free radical. These two steps are called *initiation.*

Free radicals are very reactive. When a free radical gets close to a double bond, one of the bond/bonds of monomer is disrupted. One of the electrons in the double bond of monomer is attracted to the free radical (R^*). The double bond breaks, and a new single (σ) bond is formed and during the process the free radical site is shifted from the initiator to the monomer molecule as shown below:

Example: $R^* + CH_2–CH_2 \rightarrow R–CH_2–C^*H_2$

free radical ethylene monomer free radical

The reaction is exothermic with release of energy of 20.0 kcal/mole.

(ii) Propagation

The *monomer free radical* obtained as above attacks the double bond of a fresh monomer, ($CH_2–CH_2$) molecule to form a dimer free radical with a shift of free radical to the second monomer molecule. With another molecule of the monomer, the dimer free radicals add on to form a trimer free radical. The chain length goes on increasing and polymer free radical (a growing polymer) is formed. Each step involves growth of polymer by one unit. In each process, a monomer unit is added; there is an energy release of about 20 kcal/mole. This process is depicted as

$$R–CH_2–C^*H_2 + CH_2–CH_2 \rightarrow R—CH_2–CH_2–CH_2–C^*H_2$$

monomer free radical ethylene dimer free radical

$$\rightarrow R–CH_2–CH_2–CH_2–C^*H_2 + (n - 2)(CH_2–CH_2)$$

dimer free radical

$$\rightarrow R–(CH_2)_n–CH_2–C^*H_2$$

Polymer

where n = number of molecules/moles of monomer, ($CH_2–CH_2$) added to the chain during polymerization.

(iii) Termination

Chain length during the process of polymerization does not increase indefinitely. Depending on the conditions, chain propagation comes to an end at any stage when the polymer free radical is converted to a stable polymer molecule. Termination may take place in two different ways:

(i) By *coupling of one polymer free radical* with another polymer free radical or initiator free radical to form stable dead polymer

$$R–(CH_2)_{n1}–CH_2–C^*H_2 + R–(CH_2)_{n2}–CH_2–C^*H_2$$
$$\rightarrow R–(CH_2)_{n1}–CH_2—CH_2–(CH_2)_{n2}–CH_2–CH_2–R$$

Dead heavy polymer

(ii) By *disproportion of two polymer free radicals*, an H-atom attached to one of the free radicals is shifted to the other intermediate polymer; and as a result, two dead polymers, one unsaturated and other saturated are formed.

These steps are indicated below:

During the termination by disproportion, one H-atom from a growing polymer chain is shifted to the other growing polymer chain as a result two dead polymers, one saturated and one unsaturated, are formed.

$$R—(CH_2)_{n1}—CH_2—C^*H_2 + R—(CH_2)_{n2}—CH_2—C^*H_2$$
$$\rightarrow R—(CH_2)_{n1}—CH_2—CH_3 + R—(CH_2)_{n2}—CH{=}CH_2$$

Some examples of addition polymers are given in Table 7.1.

Name(s)	Formula	Monomer	Properties	Uses
Polyethylene low density (LDPE)	$–(CH_2–CH_2)_n–$	ethylene $CH_2{=}CH_2$	soft, waxy solid	film wrap, plastic bags
Polyethylene high density (HDPE)	$–(CH_2–CH_2)_n–$	ethylene $CH_2{=}CH_2$	rigid, translucent solid	electrical insulation bottles, toys
Polypropylene (PP) different grades	$–[CH_2–CH(CH_3)]_n–$	propylene $CH_2{=}CHCH_3$	atactic: soft, elastic solid isotactic: hard, strong solid	similar to LDPE carpet, upholstery
Poly(vinyl chloride) (PVC)	$–(CH_2–CHCl)_n–$	vinyl chloride $CH_2{=}CHCl$	strong rigid solid	pipes, siding, flooring
Poly(vinylidene chloride) (Saran A)	$–(CH_2–CCl_2)_n–$	vinylidene chloride $CH_2{=}CCl_2$	dense, high-melting solid	seat covers, films
Polystyrene (PS)	$–[CH_2–CH(C_6H_5)]_n–$	styrene $CH_2{=}CHC_6H_5$	hard, rigid, clear solid soluble in organic solvents	toys, cabinets packaging (foamed)
Polyacrylonitrile (PAN, Orlon, Acrilan)	$–(CH_2–CHCN)_n–$	acrylonitrile $CH_2{=}CHCN$	high-melting solid soluble in organic solvents	rugs, blankets clothing
Polytetrafluoroethylene (PTFE, Teflon)	$–(CF_2–CF_2)_n–$	tetrafluoroethylene $CF_2{=}CF_2$	resistant, smooth solid	non-stick surfaces electrical insulation
Poly(methyl methacrylate) (PMMA, Lucite, Plexiglas)	$–[CH_2–C(CH_3)CO_2CH_3]_n–$	methyl methacrylate $CH_2{=}C(CH_3)CO_2CH_3$	hard, transparent solid	lighting covers, signs skylights
Poly(vinyl acetate) (PVAc)	$–(CH_2–CHOCOCH_3)_n–$	vinyl acetate $CH_2{=}CHOCOCH_3$	soft, sticky solid	latex paints, adhesives
cis-Polyisoprene natural rubber	$–[CH_2–CH{=}C(CH_3)–CH_2]_n–$	isoprene $CH_2{=}CH–C(CH_3){=}CH_2$	soft, sticky solid	requires vulcanization for practical use
Polychloroprene (cis + trans)(Neoprene)	$–[CH_2–CH{=}CCl–CH_2]_n–$	chloroprene $CH_2{=}CH–CCl{=}CH_2$	tough, rubbery solid	synthetic rubber oil resistant

Table 7.1

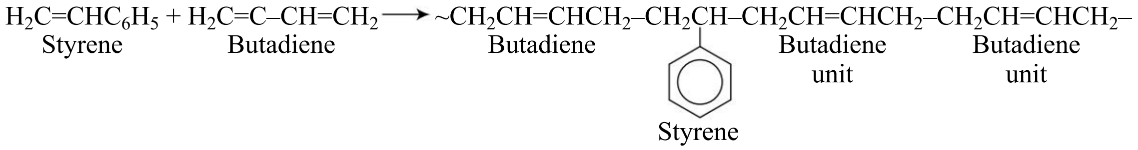
Some common addition polymers

7.4.2 Copolymerization

A specific type of addition polymerization process is known as copolymerization. Different types of monomers are involved in the process of polymerization to get the desired polymer product.

Polymer products obtained thus have unique properties of the monomers. Such unique properties can be enhanced or improved by selecting different types of monomers. This gives scope for the production of wide variety of polymer products of commercial importance.

Example: Styrene butadiene rubber



Copolymer: *A mixture of two polymers. Since a copolymer consists of at least two types of repeating units (not structural units).*

Copolymers can be classified based on how these units are arranged along the chain. These include

- *Random copolymer*: -A-A-B-B-A-A-A-B-A-A-B-B-A-B-A-A-A-B-A-A-A-B-B-
- *Alternate copolymer:* -A-B-A-B-A-B-A-B-A-B-, or $-(-A-B-)_n-$
- *Block copolymer:* -A-A-A-A-A-A-A-B-B-B-B-B-B-B-A-A-A-A-A-A-A-A-B-B-
- *Graft copolymer:* -A-A-A-A-A-A-A-A-A-A-A-A-A-A-A-A-A-A-A-A-B-B-B-B-B-B-B

7.4.3 Condensation Polymerization

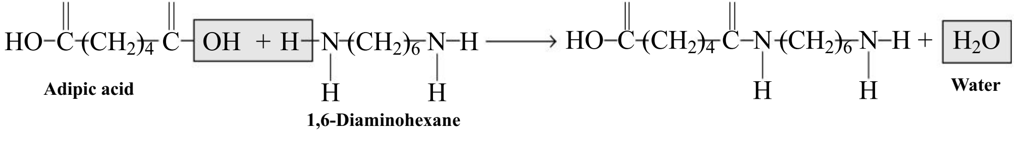
A polymer formed by the condensation of two or more monomers (*similar or different*) *with the elimination of molecules like water, hydrogen chloride, ammonia, alcohol, etc., is called condensation polymer.*

The monomers that are involved in condensation polymerization are not the same as those in addition polymerization. The monomers for condensation polymerization have two main characteristics:

1. Instead of double bonds, these monomers should have *functional groups* (like alcohol, amine or carboxylic acid groups).
2. Each monomer has at least *two reactive sites*, which usually means two functional groups.

Monomers involved in condensation polymerization have *functional groups* (*carboxylic amino and alcohol groups*). These functional groups combine to form amide and ester linkages. When this occurs, a water molecule is removed. Since water is removed, we call these reactions condensation reactions. When a condensation reaction involves polymerization, we call it condensation polymerization.

A carboxylic acid monomer and an amine monomer can join in an amide linkage.

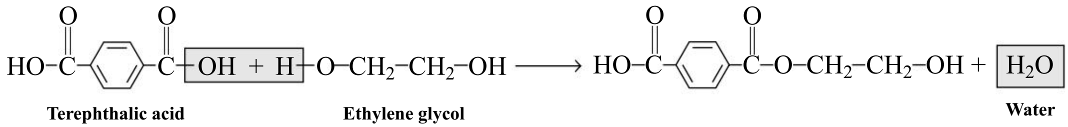


$$\underset{\textbf{Adipic acid}}{HO-\overset{O}{\overset{\|}{C}}-(CH_2)_4-\overset{O}{\overset{\|}{C}}-}\boxed{OH\ +\ H}\underset{\textbf{1,6-Diaminohexane}}{-\underset{H}{\underset{|}{N}}-(CH_2)_6-\underset{H}{\underset{|}{N}}-H} \longrightarrow HO-\overset{O}{\overset{\|}{C}}-(CH_2)_4-\overset{O}{\overset{\|}{C}}-\underset{H}{\underset{|}{N}}-(CH_2)_6-\underset{H}{\underset{|}{N}}-H + \underset{\textbf{Water}}{\boxed{H_2O}}$$

As before, a water molecule is removed, and an amide linkage is formed. Notice that an acid group remains on one end of the chain, which can react with another amine monomer. Similarly, an amine group remains on the other end of the chain, which can react with another acid monomer.

Thus, monomers can continue to join by amide linkages to form a long chain. Because of the type of bond that links the monomers, this polymer is called a polyamide. The polymer made from these two six-carbon monomers is known as nylon-6,6 (nylon products include hosiery, parachutes and ropes).

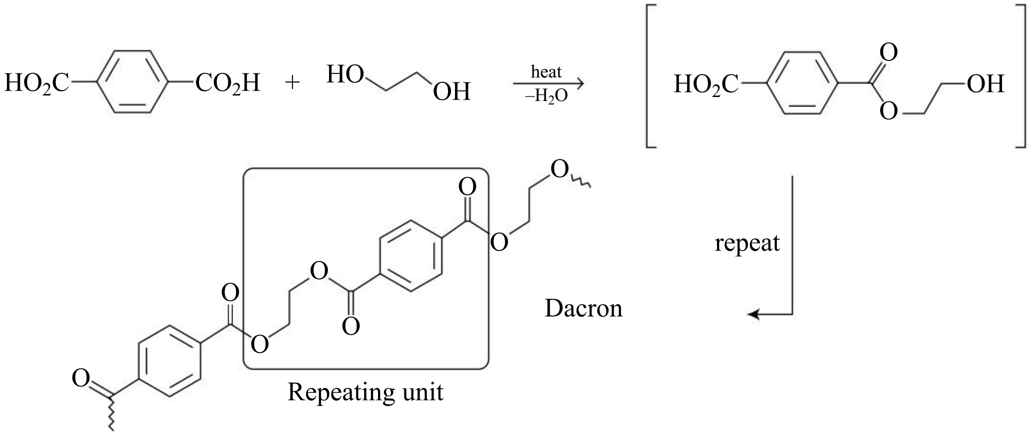
A carboxylic acid monomer and an alcohol monomer can join in an ester linkage.



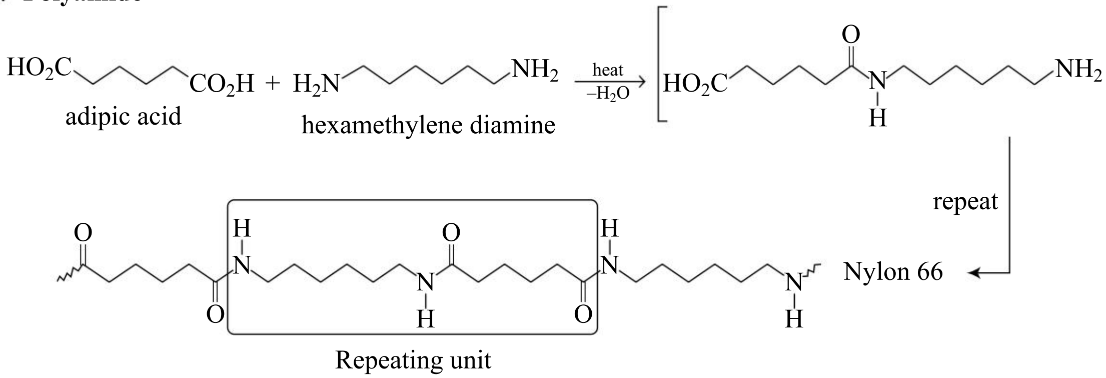
A water molecule is removed as the ester linkage is formed. Because the monomers above are all joined by ester linkages, the polymer chain is a polyester. This one is called PET, which stands for polyethylene terephthalate. (PET is used to make soft-drink bottles, magnetic tape and many other plastic products) As difunctional monomers join with amide and ester linkages, polyamides and polyesters are formed, respectively. We have seen the formation of the polyamide nylon-6,6 and the polyester PET.

Examples of condensation polymerization:

1. **Polyester**



2. **Polyamide**



7.4.3.1 Characteristics of Condensation Polymers

Condensation polymers form more slowly than addition polymers, often requiring heat, and they are generally lower in molecular weight. The terminal functional groups on a chain remain active, so that groups of shorter chains combine into longer chains in the later stages of polymerization. The presence of polar functional groups on the chains often enhances chain-chain attractions, particularly if these involve hydrogen bonding, and thereby crystallinity and tensile strength. The following examples of condensation polymers are illustrative. Note that for commercial synthesis, the carboxylic acid components may actually be employed in the form of derivatives such as simple esters. Also, the polymerization reactions for nylon-6 and spandex do not proceed by elimination of water or other small molecules. Nevertheless, the polymer clearly forms by a step-growth process.

Examples of condensation polymers are given in Table 7.2.

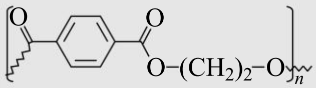
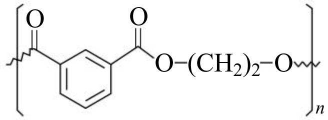
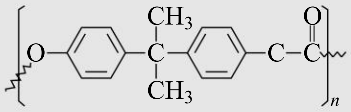
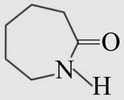
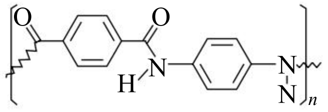
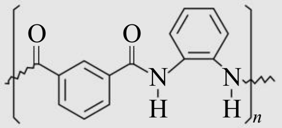
Formula	Type	Components	T_g(°C)	T_m(°C)
$\sim[CO(CH_2)_4CO\text{–}OCH_2CH_2O]_n\sim$	**Polyester**	$HO_2C\text{–}(CH_2)_4\text{–}CO_2H$ $HO\text{–}CH_2CH_2\text{–}OH$	<0	50
	**polyester** Dacron Mylar	para $HO_2C\text{–}C_6H_4\text{–}CO_2H$ $HO\text{–}CH_2CH_2\text{–}OH$	70	265
	**polyester**	meta $HO_2C\text{–}C_6H_4\text{–}CO_2H$ $HO\text{–}CH_2CH_2\text{–}OH$	50	240
	**polycarbonate** Lexan	$(HO\text{–}C_6H_4\text{–})_2C(CH_3)_2$ (Bisphenol A) $X_2C{=}O$ ($X{=}OCH_3$ or Cl)	150	267
$\sim[CO(CH_2)_4CO\text{-}NH(CH_2)_6NH]_n\sim$	**polyamide** Nylon-6, 6	$HO_2C\text{–}(CH_2)_4\text{–}CO_2H$ $H_2N\text{–}(CH_2)_6\text{–}NH_2$	45	265
$\sim[CO(CH_2)_5NH]_n\sim$	**polyamide** Nylon-6 Perlon		53	223
	**polyamide** Kevlar	para-$HO_2C\text{–}C_6H_4\text{–}CO_2H$ para-$H_2N\text{–}C_6H_4\text{–}NH_2$	–	500
	**polyamide** Nomex	meta-$HO_2C\text{–}C_6H_4\text{–}CO_2H$ meta-$H_2N\text{–}C_6H_4\text{–}NH_2$	273	390

Table 7.2

Condensation polymer type and their T_g data

The difference in T_g and T_m between the first polyester (completely aliphatic) and the two nylon polyamides (5th and 6th entries) shows the effect of intra-chain hydrogen bonding on crystallinity. The replacement of flexible alkylidene links with rigid benzene rings also stiffens the polymer chain, leading to increased crystalline character, as demonstrated for polyesters (entries 1, 2 and 3) and polyamides (entries 5–8). The high T_g and T_m values for the amorphous polymer Lexan are consistent with their brilliant transparency and glass-like rigidity. Kevlar and Nomex are extremely tough and resistant materials, which find use in bullet-proof vests and fire resistant clothing.

The differences between addition and condensation polymerization are given in Table 7.3.

Addition polymerization	Condensation polymerization
1. It demands the presence of multiple bonds in the monomer.	1. Both ends of the monomer should possess two reactive functional groups
2. It prefers the presence of multiple bonds in one or more monomers.	2. Minimum of two different monomers with functional groups
3. No elimination of by-product like water	3. It is always associated with the elimination H_2O, NH_3, etc.
4. A thermoplastic is obtained which is a homo chain polymer	4. A thermosetting or thermoplastic plastic (hetero chain) is obtained
5. It is observed that chain growth is at one active centre.	5. Chain growth is observed at two active centres

Table 7.3

Difference between addition and condensation polymerization

7.5 Methods of Polymerization

A number of techniques are used in the production of polymers from monomers. Four of them are important (a) bulk polymerization, (b) solution polymerization, (c) suspension polymerization, and (d) emulsion polymerization.

7.5.1 Bulk Polymerization

Monomer in the liquid state is mixed with a small quantity of initiator to form a homogeneous mixture. Initiation is carried out either thermally or photochemically. The mixture is constantly agitated and heated to the polymerization temperature. Once the reaction starts, heating is stopped since the reaction is exothermic and by careful control of the conditions temperature may be maintained steadily. The polymer formed is free from impurities and can be used without purification.

The polymer is obtained as a pure product

$$\underset{\text{(Liquid)}}{\text{Monomer}} + \text{Initiator} \rightarrow \underset{\text{(Pure Form)}}{\text{Polymer}}$$

The method is very simple and needs simple equipment. The percentage conversion is high. Product obtained is pure and has high optical clarity. But as polymerization proceeds, viscosity increases rapidly and agitation becomes difficult. It is difficult to control the degree of polymerization. Sometimes, uncontrolled chain reaction releasing huge quantity of heat takes place resulting in explosion. Hence bulk polymerization is not preferred industrially. However, the method is used for the manufacture of plexiglas, PVC, etc.

7.5.2 Solution Polymerization

Some of the disadvantages of bulk polymerization are eliminated in solution polymerization. In this method, the monomer and initiator are dissolved in a solvent to form a homogeneous mixture. The mixture is kept at the polymerization temperature and constantly agitated. After the reaction is over the polymer is used as such in the solution (or) the polymer is isolated by evaporating the solvent.

$$\underset{(\text{homogenous solution in solvent})}{\text{Monomer + Initiator}} \xrightarrow[P]{T} \text{Polymer}$$

In this method the increase in viscosity is negligible and so agitation and heat control are easy. But to get the pure polymer, evaporation of solvent is required. Method becomes costly since it uses costly solvents. Polymers of high molar mass cannot be formed as the solvent molecules may act as chain terminators. The method is used in the manufacture of polythene, PVC, etc.

7.5.3 Suspension Polymerization

The use of costly solvents for polymerization can be prevented by using water as the solvent. In this method polymer is produced in a heterogeneous medium. The water insoluble monomers are suspended in water as tiny droplets by continuous agitation. The droplets are prevented from coalescing by using small quantities of a water soluble polymer like polyvinyl alcohol. The Initiator solution only in the monomer droplets is added and the mixture is heated and agitated constantly. The method is like bulk polymerization except for the presence of water molecule. Polymerization takes place inside the droplet and the product formed being dissoluble in water produce spherical pearls or (beads) of the polymer. Hence, the method is also called pearl polymerization.

$$\underset{\text{Suspension in water}}{\text{Monomer}} + \underset{\text{Soluble in monomer}}{\text{Initiator}} \xrightarrow[T]{P} \text{Polymer}$$

Process is cheap since it involves only water instead of solvents. Viscosity increase is negligible. Agitation and thermal control are easy. Product isolation is easy since the product is insoluble in water. Product is also pure. But the method can be adopted only for water insoluble monomers and it is difficult to control polymer size, PVC, polyvinyl acetate, polystyrene etc., are manufactured in this method.

7.5.4 Emulsion Polymerization

This method is used to prepare polymers from water insoluble monomers. Emulsion is the colloidal dispersion of tiny droplets of a liquid in another liquid with which it is immiscible. To maintain the system stable small amounts of an emulsifier is required. Soaps and detergents are the normal emulsifiers and act as good surfactants. Both soap [RCOONa] and detergent [RSO_3Na] produce anions [RCOO– or RSO_3^-] in water, which contain a long chain non-polar R group at one end and polar COO^- or SO_3^- at the other end. Emulsifiers keep the insoluble liquid as uniform emulsified droplets. After emulsification, water-soluble initiator is added, and the system is kept agitated at the required temperature. The polymer formed is in the form of stable latex. It is isolated by coagulation by using an electrolyte or by freezing.

$$\text{Monomer insoluble in water + Water + Emulsifier + Initiator} \xrightarrow[\text{Constant-T}]{\text{agitation}} \text{Polymer}$$

The method is the most commonly used industrial method. PVC, polyvinyl acetate, PMMA, elastomers, etc., are manufactured by this method.

7.5.5 Mechanism of Emulsion Polymerization

Emulsifier contains a hydrophilic (water loving) polar end group (Head) and a hydrophobic (water hating) non-polar end group (tail). At very low concentration the soap (or) detergent dissolves completely. But at slightly higher concentration, the molecules form aggregates called *micelles* (Fig. 7.1).

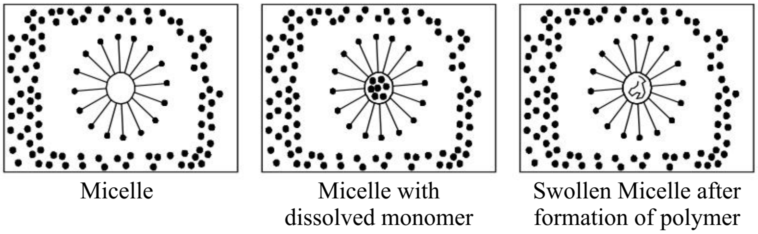


Figure 7.1 *Formation of emulsion.*

Each micelle contains nonpolar tail of emulsifier molecules inside and polar head upwards. The monomer molecules dissolve in the hydrocarbon centre of the micelles. The Initiator molecules diffuse into the micelles centre through the polar head. Reaction takes place at the centre of the micelles and polymer is formed and the micelles begin to swell. The monomer consumed inside the micelles is replenished by diffusion from aqueous phase. This continues till the size of the polymer is big enough to come out of the micelles.

The method is followed in many manufacturing units because

- rate of polymerization is high
- polymers with higher molar masses are formed
- thermal control and control over polymer molar mass is possible and
- no viscosity build up and hence agitation is easy.

But the polymer formed may contain emulsifiers and coagulant as impurities and needs further purification.

7.6 Glass Transition Temperature, T_g

Definition: *Glass transition temperature is defined as the temperature below which an amorphous polymer is brittle, hard and glassy and above the temperature it becomes flexible, soft and rubbery.*

$$\underset{\text{[Hard, brittle plastic]}}{\overset{T_g}{\text{Glassy state}}} \rightarrow \underset{\text{[soft, flexible]}}{\text{Rubbery state}}$$

Explanation: In the glassy state of the polymer, there is neither molecular motion nor segmental motion. When all chain motions are not possible the rigid solid results. On heating beyond T_g, segmental motion becomes possible but molecular mobility is disallowed. Hence, flexible and soft rubbery state is reached. On heating much above T_g both segmental and molecular motion become possible and the polymer flows like a viscous liquid. This temperature is usually called a flow temperature since any polymer is only a mixture and has no sharp melting point.

$$\underset{\text{[Frozen state]}}{\text{Glassy state}} \xrightarrow{\text{above } T_g} \underset{\text{[viscoelastic state]}}{\text{Rubbery state}} \xrightarrow{\text{above } T_t} \underset{\text{[viscofluid state]}}{\text{Viscous liquid}}$$

7.6.1 Factors that Affect the Value of T_g

A number of factors affect the value of T_g:

1. *Crystallinity:* In crystalline polymers the polymer chains are arranged in a regular parallel fashion. Each chain is bound to the other by strong forces like H-bonding. Hence, crystalline polymers have higher T_g than amorphous polymers.

2. *Molecular mass:* Generally T_g of a polymer increases with molar mass up to a particular value and beyond that there is no change.
3. The presence of side chain which hinders free rotation also increases T_g.
4. The presence of plasticizers' reduces T_g of a polymer (like impunities on melting point of pure solids).
5. Stereo regularity of polymers increases T_g. Thus T_g of an Isotactic polymer is greater than that of syndiotactic polymer whose T_g is greater than a tactic polymer.

Importance of T_g value

T_g value is not only a measure of flexibility of a polymer but also gives us idea of the thermal expansion, heat capacity, electrical and mechanical properties of the polymer. Thus the workability and usefulness of a polymer over a range of temperature can be obtained from its T_g.

7.7 Structure and Property Relationship of Polymers

We have 'tailor-made' polymers for any purpose with required properties. Polymers of wide ranging properties are produced and introduced into the market. The properties like crystallinity, tacticity, tensile strength, elasticity, resistance to chemicals, wear and tear, etc. depend mostly on the structure of polymers. The structure–property relationship is an important topic in polymer science.

7.7.1 Crystallinity of Polymer

A polymer is said to be crystalline if all the molecules are arranged in orderly manner with symmetrical orientation, with higher force of attraction between two chains.

Any polymer will contain a definite percentage of crystalline part and amorphous part. The degree of crystallinity depends on how best the polymer chains can be closely packed. Crystalline regions of a polymer are formed when the individual chains are linear (without branching), contain no bulky substituents, and are closely arranged parallel to each other. The strains of polymer may be held together by van der Waals force, hydrogen bonding or polar interactions. A polymer with high degree of crystallinity has high tensile strength, impact and wear resistances, high density and high fusion temperature. It has high T_g and melt viscosity.

Crystallinity of a polymer also depends on the stereo regular arrangement of their pendent groups. Polymers like HDPE, isotactic and syndiotactic polypropenes, PVC, etc., are highly crystalline. On the other hand, a tactic polypropylene, PVC, polystyrene which have their substituents in a random arrangement are less crystalline. Characteristics of crystalline polymers are as follows:

- High density
- Hard but brittle and strong
- Possess

7.7.2 Impact and Tensile Strength of Polymers and Molar Mass

Density, melt viscosity, impact and tensile strength are a few important mechanical properties of a polymer. These are highly influenced by the molar masses of polymers. Melt viscosity of polymers of low molar masses is low. With increase in molar mass melt viscosity increases and increases sharply at higher molar masses.

Tensile strength and impact strength also increase with molar mass of the polymer up to 2000 DP. Beyond that the increase is negligible. Actually, each polymer has a limiting molar mass at which its tensile strength and also workability are both good.

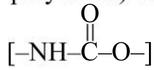
7.7.3 Plastic Deformation (Theology) of Polymers

Thermoplastics on heating to their T_g or on application of pressure become flexible and rubbery and undergo deformation. On further heating beyond their flow temperature they form viscous liquids and begin to flow. On cooling, the reverse change takes place and they return to their original state. This property is used in moulding of thermoplastics.

Thermosetting plastics do not undergo deformation. On heating first they soften and then set into a hard mass. On further heating no change is observed and on strong heating they undergo charring. This is because of the fact that in a well-set thermosetting plastics all the monomeric units are held together with strong co-valent bonds. These primary bonds cannot be easily broken by heat. Thus on strong heating charring occurs.

7.7.4 Chemical Resistivity

If a polymer is attacked by a reagent, it undergoes softening and swelling and loses strength. Chemical resistivity of the polymers depends on a number of structural factors like (i) the presence or the absence of residual unsaturation, (ii) the presence or the absence of polar groups, and (iii) molar mass, degree of crystallinity and extent of cross linking.

Polymers with non-polar groups (hydrocarbon polymer) undergo swelling and dissolution in non-polar solvents like benzene, toluene, CCl_4, etc. Polar polymers containing OH group or COOH group are soluble in polar solvents like water, alcohols etc., and polymers containing ester groups (e.g. polyesters) undergo hydrolysis with strong alkalis at high temperatures. Polyamides like nylon containing  $[-NH-\overset{O}{\overset{\|}{C}}-O-]$ group can be hydrolyzed using strong acid or alkali.

Polymers containing residual unsaturation (like natural rubber and some synthetic rubbers) easily undergo degradative oxidation in air in the presence of light or ozone. Thus, natural and some synthetic rubbers undergo degradative oxidation in air.

Polyalkenes, PVC, fluorocarbon, etc., are some polymers which have high degree of chemical resistance.

For a given polymer, chemical resistivity increases with the increase in the molar mass and the crystallinity. Linear polymers have lower resistivity than branched chain and cross-linked polymers.

7.7.5 Tacticity of Polymers

Polymer molecules have different arrangements of adjacent chiral centres within molecules which can be referred to as tacticity of polymers. They are classified on the basis of this (Fig. 7.2) and known as

- Isotactic polymers
- Atactic polymers
- Syndiotactic polymers

Example: Vinyl polymer

Vinyl polymers that have single substituent (e.g. propylene, the example below) or two unsymmetrical substituents (e.g. methyl methacrylate) have pseudoasymmetric carbon atoms in the backbone. Of course, there are many possible relative placements of the groups. The different stereoisomers may have very different physical properties. For example, atacticpolypropylene is a useless, gummy solid, while the isotactic version is a highly crystalline, and tough, but here are three common versions:

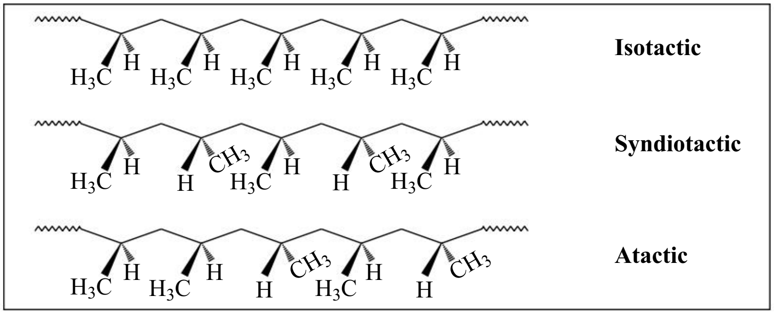


Figure 7.2 *Isotactic, syndiotactic and atactic structures.*

7.7.6 Viscoelasticity of Polymer

It is observed that few polymeric materials exhibit both *viscous* and *elastic* characteristics under certain conditions. This is referred to as viscoelasticity of polymers. Viscoelastic materials have elements of both of these properties and, as such, exhibit time dependent strain. Whereas elasticity is usually the result of bond stretching along crystallographic planes in an ordered solid, viscoelasticity is the result of the diffusion of atoms or molecules inside of an amorphous material.

The physical state of polymers depends on their chain length and molecular weight. They are either elastic solids or viscous liquids depending on their physical state. The change of the physical properties of polymers is mainly due to the effect of temperature on them. They transform to hard and brittle when cooled slowly or vice versa. The rate of cooling also affects the properties of polymers.

7.8 Plastics

Plastics are macromolecules, formed by polymerization and having the ability to be shaped by the application of reasonable amount of heat and pressure or some other form of force.

Depending upon their nature and properties, the polymers are classified as

(i) plastics, (ii) elastomers (or rubbers) and (iii) fibres.

If the polymer chains are very flexible, the intermolecular forces of attraction are low and the chains do not fit to a regular lattice structure easily, the material will tend to retract upon when external tension is released. This is the state typical for a rubber or an *elastomer*.

In the other extreme, if the polymer chains are inherently rigid, the intermolecular forces intense and the molecules fit readily into a crystal lattice, then the crystallinity once induced, will tend to be permanent. Such a material would be a typical *fibre.*

In the intermediate case, when the intermolecular forces of attractions are neither too high nor too low, the polymer is called *plastic*.

Generally speaking, *a plastic material should possess sufficient rigidity, dimensional stability and mechanical strength at room temperature to serve as a useful household article, gadget or structural part and still be of such a character that it may be moulded into shape by the application of reasonable temperature and pressure.*

7.8.1 Types of Plastics

There are mainly two types of Plastics:

- Thermoplastic plastics
- Thermosetting plastics

(1) *Thermoplastics are those polymers that are flexible, linear chain molecules, which once shaped or formed, can be softened by the application of heat and can be reshaped repeatedly, till it loses its property.*

Thermoplastics can be crystalline or amorphous (Fig. 7.3).

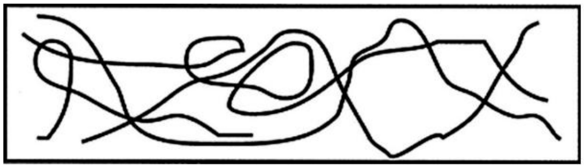


Figure 7.3 *A Thermoplastic polymer.*

Example: polyethylene, polypropylene, nylon, polycarbonate, etc.

Applications are: polyethylene buckets, polystyrene cups, nylon ropes, etc.

(2) *Thermosetting plastics are rigid three-dimensional network polymers, which once shaped or formed, cannot be softened by the application of heat* (Fig. 7.4). Excess heat will charge the material.

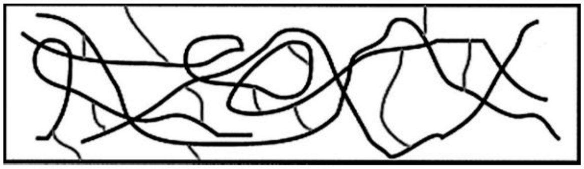


Figure 7.4 *A thermoset polymer.*

Example: Phenol formaldehyde, Urea Formaldehyde, Melamine Formaldehyde, Thermosetting Polyester, etc.

Applications are: Bakelite, Electrical switches, formica/sermica table tops, melamine cutlery, etc.

In thermoplastic polymers the links between chains are weak so the chains can move when heated (Fig. 7.5), becoming hard again as they cool and the weak links are reformed.

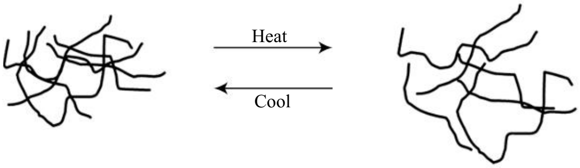


Figure 7.5 *Effect of heat on thermoplastic.*

7.9 Synthesis, Properties and Applications of Few Commercial Thermoplastic Polymers

First let us discuss the synthesis, properties and applications of some addition polymers.

7.9.1 Polyethylene (polythene)

It is the simplest homopolymer. polyethene is one of the simplest and most inexpensive polymers. It is a waxy, chemically inertplastic. It is a thermoplastic formed by the addition polymerization of ethylene. It has a structure $[–CH_2–CH_2–]_n$. There are two types of polyethylenes:

(i) Low-density polyethylene LDPE and (ii) high-density polyethylene HDPE.

(i) Low density polyethylene (LDPE)

Preparation: LDPE is manufactured by chain polymerization of ethylene gas at 350°C at a pressure of 1000–2500 atmospheres in presence of O_2 or any peroxide as catalyst.

$$n\,CH_2{=}CH_2 \xrightarrow[\substack{O_2 \text{ or peroxide catalyst} \\ \text{chain polymerisation}}]{\substack{350°C \\ 1000-2500\,\text{atm}}} [–CH_2–CH_2–]_n$$

(ii) High density polyethylene (HDPE)

Preparation: HDPE is manufactured by the coordination polymerization of ethylene in a hydrocarbon solvent using Ziegler–Nattacatalyst (Titanium tetra chloride and aluminium triethyl) at 50–70°C.

$$nCH_2{=}CH_2 \xrightarrow[\text{Catalyst}]{TiCl_4 + Al(C_2H_5)_3} [-CH_2-CH_2-]_n$$

A comparative study of the properties and applications of LDPE and HDPE is given in Table 7.4.

Property	LDPE	HDPE
1. Structure	Linear structure with extensive branching	Linear polymer with minimum branching
2. Degree of crystallinity	Low (40 percent)	High (90 percent)
3. Melting point	Low 110°C	Higher 140°C
4. Density	Low 0.912–0.935	High 0.96 –0.98
5. Tensile strength	Low	Low
6. Stiffness and hardness	Low	High
7. Chemical resistivity	Low swells and dissolves in hydrocarbon solvents	High does not swell or dissolve in solvents
8. Application	In films, sheets, etc.	In moulded articles

Table 7.4

Properties and applications of LDPE and HDPE

7.9.2 Polytetrafluoroethylene PTFE (Teflon)

Polytetrafluoroethylene, or teflon, is a linear polymer like polyethylene, the only difference being that instead of hydrogen atoms there are fluorine atoms.

Teflon (registered trade name of Du Pont company) is an addition polymer of tetrafluoroethylene $CF_2{=}CF_2$. It has a structural formula $(-CF_2-CF_2-)_n$ and is also a thermoplastic. Teflon can be synthesized by two methods:

1. The monomer tetrafluoroethylene is obtained by treating chloroform with HF to get difluorochloromethane, which on pyrolysis gives the monomer.

$$CHCl_3 + 2HF \rightarrow CHClF_2 + 2HCl$$
$$2CHClF_2 \rightarrow CF_2 = CF_2 + 2HCl$$

2. Emulsion polymerization of tetrafluoroethylene in presence of any peroxide initiator gives teflon.

$$n[CF_2 = CF_2] \xrightarrow[\text{initiator}]{\text{peroxide}} \underset{\text{PTFE [Teflon]}}{(-CF_2 - CF_2-)_n}$$

Properties:

Teflon is called a wonder plastic because of a number of technologically important properties.

Teflon is a linear polymer with high degree of crystallinity ~95 percent.

- Its flow temperature is high (327 C).
- Practically insoluble in all solvents.

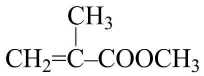
- It is chemically inert and not attacked by acids, alkalis, oxidising and reducing agents. Even concentrated sulphuric acid and fuming nitric acid do not have any action on Teflon.
- It is not wetted by oil or water.
- It has high resistance to impact and wear and tear.
- It is thermally stable and has a good electrical insulating property.

Applications:

1. Because of its excellent insulating capacity it is used for the insulation of generators, motors, transformer and other electrical appliances.
2. Oil and water do not wet Teflon and this property is used in coating non-stick cooking utensils, bakery trays and frying pans.
3. Its stability towards chemicals and high temperature is employed in making industrial gaskets, filters, etc. Articles coated with Teflon withstand atmospheric and chemical corrosion. Its lubricating capacity over a wide range of temperature is used in high and low temperature lubrication.

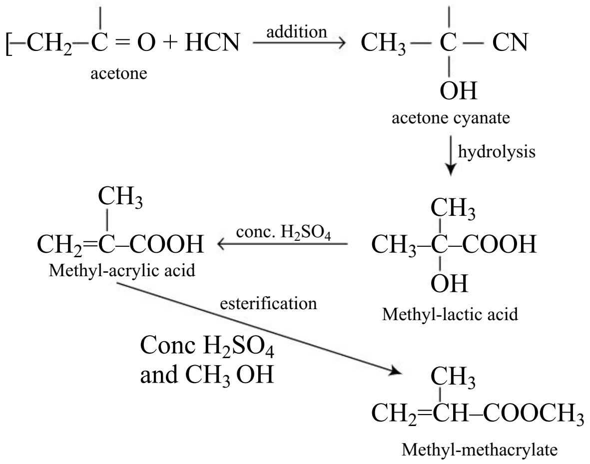
7.9.3 Poly (methyl-methacrylate) PMMA (Plexiglas)

It is an addition polymer formed from the monomer methyl-methacrylate

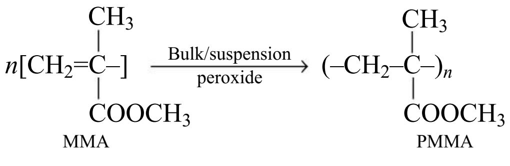


$$CH_2{=}\overset{\overset{\displaystyle CH_3}{|}}{C}{-}COOCH_3$$

The monomer methyl-methacrylate is obtained from acetone as follows.



Polymerization of MMA is carried out by bulk or suspension polymerization using peroxide as an initiator,



$$n[CH_2{=}\underset{\underset{\displaystyle COOCH_3}{|}}{\overset{\overset{\displaystyle CH_3}{|}}{C}}{-}] \xrightarrow[\text{peroxide}]{\text{Bulk/suspension}} (-CH_2{-}\underset{\underset{\displaystyle COOCH_3}{|}}{\overset{\overset{\displaystyle CH_3}{|}}{C}}{-})_n$$

Properties

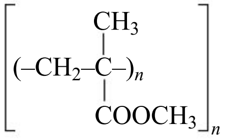
1. PMMA is a white transparent amorphous thermoplastic. It is amorphous due to the bulky pendent group ($-COOCH_3-$) present.

2. It has excellent optical clarity and not affected by sunlight or age.
3. It has excellent chemical resistivity but soluble in many organic solvents.

Applications:

1. Because of its optical clarity (flexible glassy nature) PMMA is used in making light fixtures, aircraft windows, plastic signboards, etc.
2. It is used for making artificial eyes and teeth.
3. It is used in making acrylic emulsion paints.
4. It is used in making transparent bottles, moulds, tubes, etc.

It has a structure



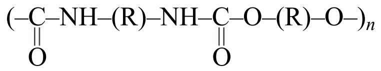
$$\left[(-CH_2-\overset{\displaystyle CH_3}{\underset{\displaystyle COOCH_3}{\overset{|}{\underset{|}{C}}}}-)_n\right]_n$$

(Polymethylmethacrylate)

Preparation over a wide range of temperature is used in high and low temperature lubrication.

7.9.4 Polyurethane (PU)

They are addition polymers with a wide variety of applications. They contain $[-NH-\overset{O}{\overset{|}{C}}-O-]$ group called methane group. They are linear polymers with the structure

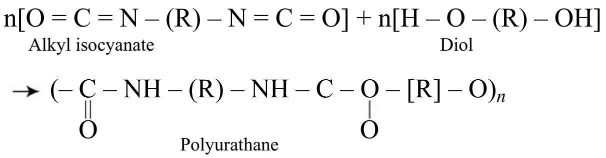


$$(-\underset{O}{\underset{\|}{C}}-NH-(R)-NH-\underset{O}{\underset{\|}{C}}-O-(R)-O-)_n$$

They resemble polyamides in their structure but the presence of one more 'O' in –NH–C–O group makes them more flexible than polyamides.

Preparation:

PU is made by addition of polymerization of alkyl di-isocyanate with a diol or triol. During addition the 'H' atoms of the OH group gets attached to the N atom to form the NH–C–O group as follows:



$$n\underset{\text{Alkyl isocyanate}}{[O=C=N-(R)-N=C=O]}+n\underset{\text{Diol}}{[H-O-(R)-OH]}$$

$$\rightarrow \underset{\text{Polyurathane}}{(-\underset{O}{\underset{\|}{C}}-NH-(R)-NH-C-\underset{O}{\underset{|}{O}}-[R]-O)_n}$$

Properties:

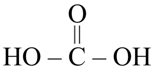
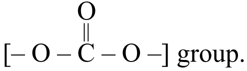
1. PU is spongy transparent linear thermoplastics.
2. They have low melting point and high degree of flexibility.
3. They are resistant to water, oil and corrosive chemical.

Applications:

- As PU is resistant to oil and corrosive chemical, it is used in tyre retreading.
- PU fibres are used in light weight water repellent garments like swim suits.
- As PU is abrasion resistant, it is used as floor coatings of gymnasium, dance floors, etc.
- Polyurethane foams [U foams] are one of the well-known materials used in cushioned furniture and car upholsteries.

Now let us discuss the synthesis, properties and applications of some condensation polymers.

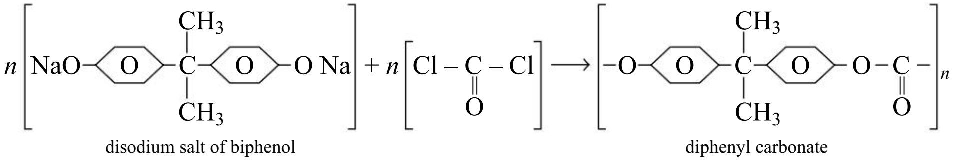
7.9.5 Polycarbonates (PC)

Polycarbonates are polyesters of unstable  $HO-\overset{O}{\overset{\|}{C}}-OH$ carbonic acid. They form a group of condensation polymers with wide range of application. They contain the  $[-O-\overset{O}{\overset{\|}{C}}-O-]$ group.

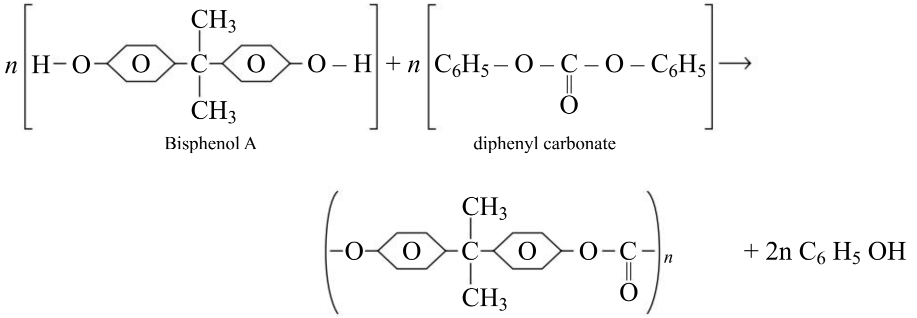
Preparation:

There are two methods of PC synthesis.

By condensation of di-sodium salt of bisphenol with phosgene with the elimination of NaCl.



By condensing bisphenol-A with di-phenyl carbonate involving elimination of phenol.



Properties:

1. PCs are white transparent amorphous solids.
2. They have high melting point.
3. They have high tensile strength and impact resistance.
4. They are soluble in a number of solvents.

Applications:

- They are used in making industrial safety glasses, because of their optical clarity.
- They are used in machinery hoses due to its resistance to abrasion and chemicals.

7.10 Thermoset Polymers

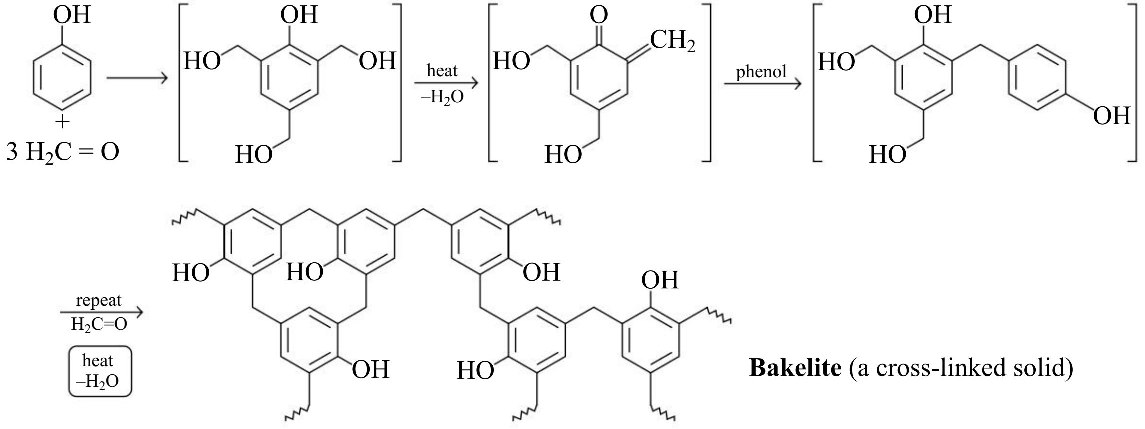
Most of the polymers described above are classified as thermoplastics. However, *another group of polymers, characterized by a high degree of cross-linking, resist deformation and solution, once their final morphology is achieved. Such polymers are usually prepared in moulds that yield the desired object. Because these polymers, once formed, cannot be reshaped by heating, they are called thermosets.*

7.10.1 Phenol-formaldehyde

One of the oldest known condensation polymers is phenol formaldehyde resin. A phenol ring has three binding sites (two ortho positions and one para position) while formaldehyde has two binding sites. The condensation is catalyzed by H^+ ions or OH^- ions. The nature of the product formed depends on the relative concentrations of phenol and formaldehyde and on whether H^+ or OH^- is used as the catalyst.

Initially, two intermediates called NOVOLACS and RESOLES which are linear polymers with low molar masses, fusible and soluble in organic solvents are formed. During moulding and curing, these intermediates undergo extensive branching producing highly cross-linked insoluble, hard, rigid product called bakelite.

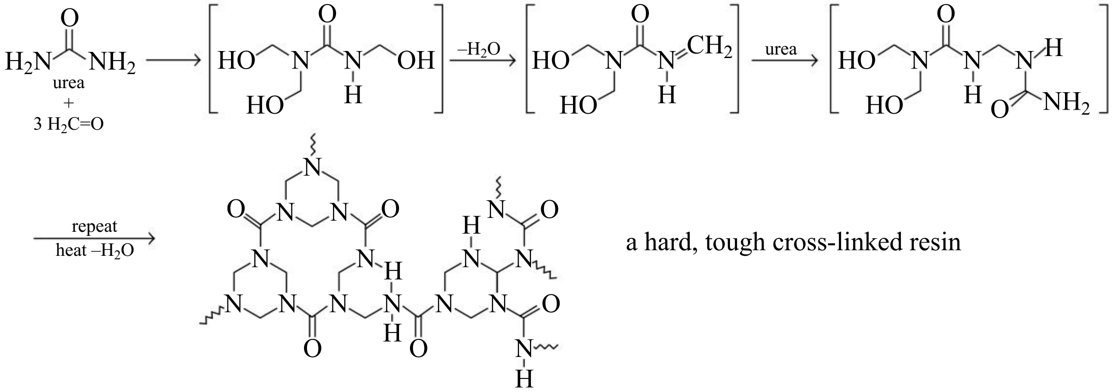
Phenol-Formaldehyde Resin



Bakelite (a cross-linked solid)

7.10.2 Urea Formaldehyde

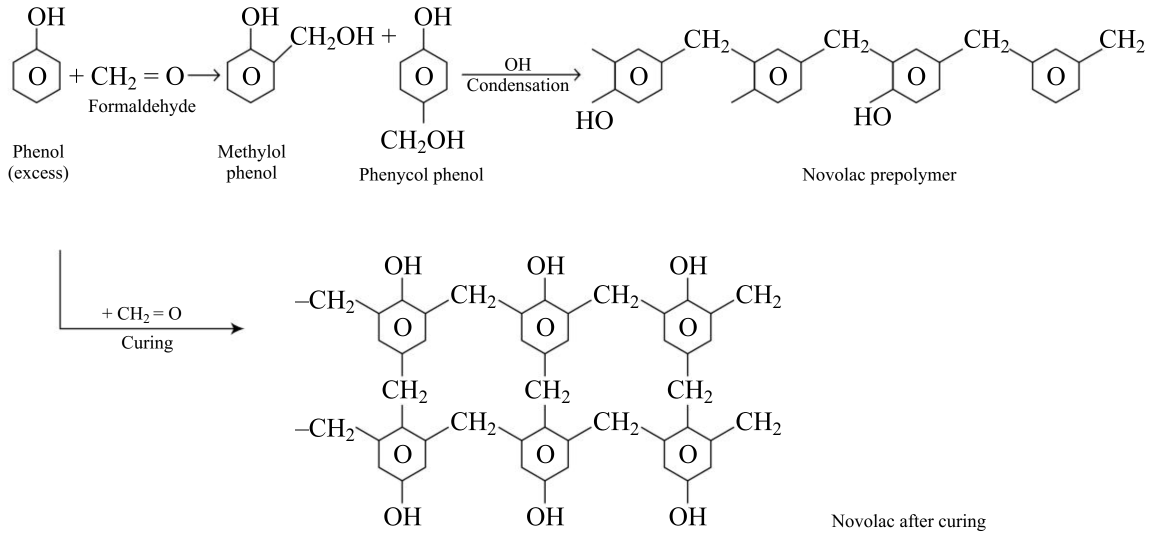
Urea-Formaldehyde Resin



a hard, tough cross-linked resin

7.10.3 Novolacs

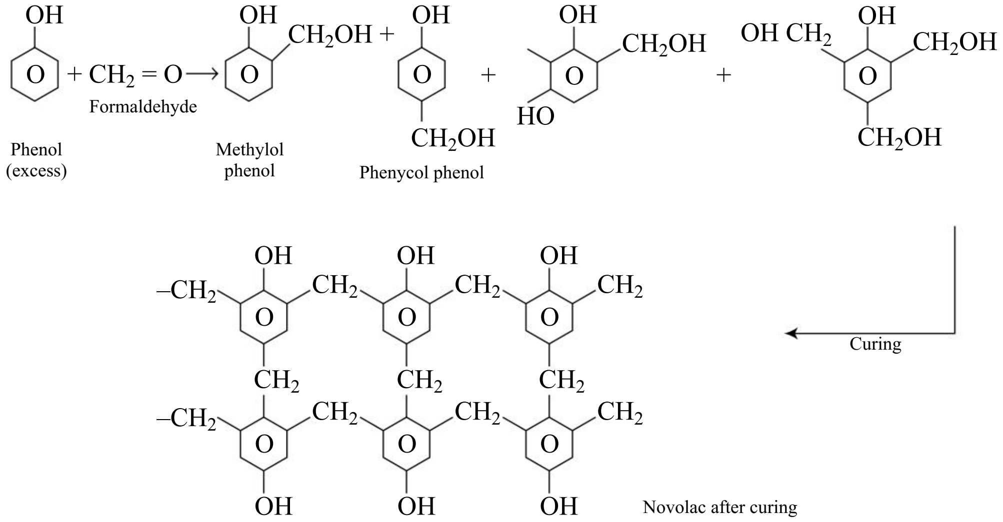
In the presence of an acid catalyst when the P to F ratio is greater than 1, novolac resin is formed.



First step in the reaction is the addition of phenol molecule to formaldehyde resulting in ortho- and para-methylol phenol. These undergo rapid condensation to form novolac. Novolac contains only five or six phenol rings and does not contain free CH_2OH group; therefore, no cross linking is possible. But when novolac is heated with formaldehyde cross linking takes place resulting in a thermosetting plastic.

7.10.4 Resoles

Under basic conditions and when the P to F ratio is less than 1, resoles formed as the pre-polymer. The first step in the reaction is the formation of mono-, di- and tri-methylol phenols. These undergo condensation through the CH_2OH group and form methylene bridged resoles. These resoles are soluble and fusible because they contain free OH groups. During curing, further reaction takes place producing cross linking resulting in a three-dimensional network.



Thus, both resole and novolac polymers almost have the same rigid three-dimensional network called *bakelite*.

Properties:

- Resole and novolac have adhesive properties.
- Both are soluble in organic solvents.
- After curing, they form hard, rigid insoluble cross-linked polymer called bakelite.

Applications:

- Pre-polymers are used in adhesive and in varnishes.
- Bakelite is used in making plugs, switches, handles of utensils, washing machine agitators, telephones, etc.

7.11 Distinction between Thermoplastics and Thermosettings

Differences between thermoplastic and thermosetting polymers are given in Table 7.5.

S.No.	Thermoplastics	Thermosetting
1.	These are the products of addition polymerizations	Products of condensation polymerization
2.	Long chain linear polymers	Have three-dimensional network structure
3.	Monomer used in the production is bi-functional	Used higher functional monomers
4.	Low molecular weight	High molecular weight
5.	They are soft, weak and brittle	They are hard, strong and brittle
6.	Soluble in organic solvents	They are insoluble
7.	Softened on heating readily	They do not soften on heating
8.	They can be softened, reshaped and recycled	Retain their shape and structure

Table 7.5

Distinction between thermoplastics and thermosettings

7.12 Techniques of Plastic Moulding

Generally, the following techniques have been used to obtain moulded plastic items in market. They are as follows:

- Compression moulding
- Injection moulding
- Blown film moulding
- Extrusion moulding

Brief procedures of moulding plastics are explained below:

7.12.1 Compression Moulding

The process for forming thermosets by applying heat and pressure (Fig. 7.6).

- A measured amount of thermoset powder, granules or pellets is fed into the mould cavity.
- Heat softens the material and pressure fills the cavity, then the material is cured.
- Heat actually causes the polymer to transform into a highly cross-linked and networked structure.
- Process is of limited use for thermosets because of the requirement of cooling time of the mould.
- Typical products include electrical insulators, pot handles and some automotive parts.

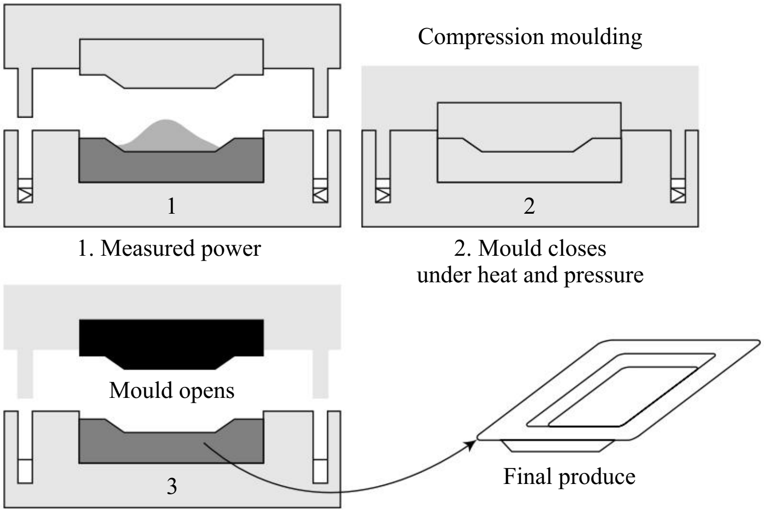


FIGURE 7.6 *A compression moulding unit.*

7.12.2 Injection Moulding

It is associated with processing thermoplastics (Fig. 7.7).

- However, with development of the reciprocating screw type equipment, thermosets can also be injection moulded.

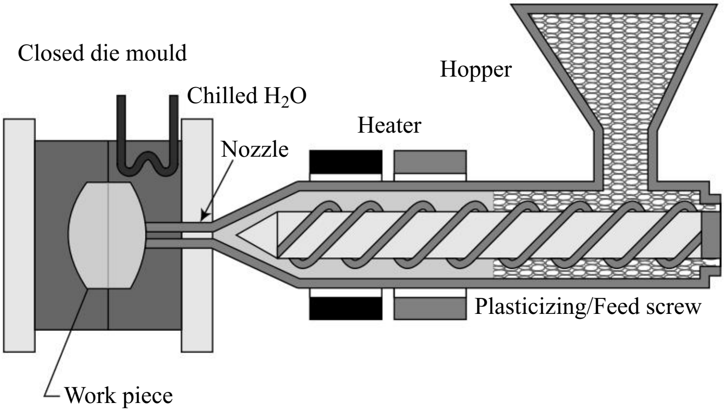


FIGURE 7.7 *Injection moulding unit.*

- The basic process includes plasticizing, injection, cooling and ejection.
- Granules are fed from a hopper into a screw that rotates to feed the material into a heated chamber to allow the material to change to a molten state.
- The material is then forced through a nozzle into the mould cavity.
- A cooling time is necessary to allow the polymer to become solid, and then is ejected from the mould by mechanical ejector pins associated with processing thermoplastics.

7.12.3 Blown Film Extrusion Moulding

Used to produce (Fig. 7.8) thin film hollow tubes.

- Somewhat of a combination of extrusion, blown moulding and calendering.
- As material is extruded, air is forced through the centre of a die, causing the material to expand to the diameter of the mould.
- Mould is open at the end, and the material is continuously taken up on rollers.
- During the take-up process, the walls on the tube may be seamed welded and perforated such as the case with garbage bags.

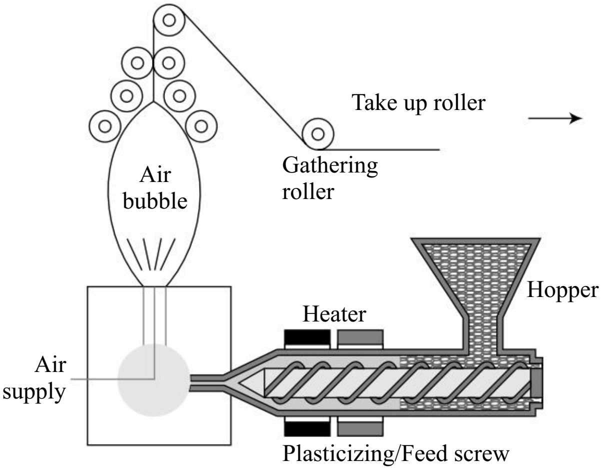


FIGURE 7.8 *Film blown extrusion unit.*

7.12.4 Extrusion Moulding

Continuous flow of molten material is forced through a die (Fig. 7.9).

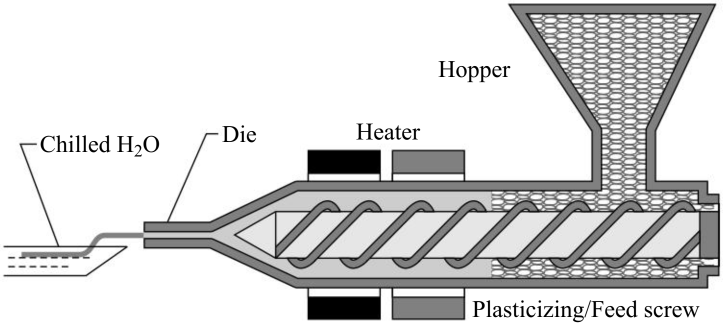


FIGURE 7.9 *An extrusion moulding unit.*

- Shape of the final product is determined by the shape of the die opening.
- Thermoplastic material is fed from a hopper, similar to the configuration of the screw system in injection moulding.
- The screw forces the material through a tapered opening in the die.
- Heat and friction causes plasticizing to occur, softens the material and forces it through the die opening.
- Material is cooled by either air or water.

Extrusion:

- Rate of cooling can be controlled and further forming is possible.
- Example, PVC pipe is extruded as electrical conduit. If allowed to be immersed in hot water, the conduit can be bent at 90 degree angles.
- Products that are extruded include tubing, rods, bars, mouldings, sheets and films.
- Extrusion is also used for coating wire and cable.

Thermoplastics can be made 'plastic' and malleable at high temperatures. Modern thermoplastic polymers melt anywhere between 65 and 200°C. In this state they can be moulded in a number of ways: they differ from thermoset plastics in that they can be returned to this plastic state by reheating. They are then fully recyclable.

7.13 Elastomers

Definition: Elastomer is defined as a long chain polymer which under stress undergoes elongation by several times and regains its original shape when the stress is fully released.

An elastomer is expected to have the following characteristics:

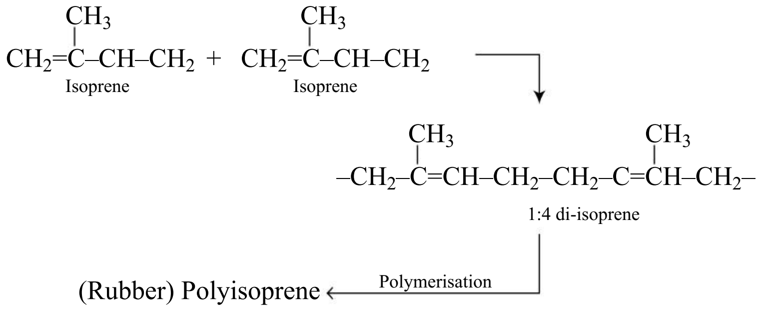
- It should have linear but highly coiled structure.
- Inter-chain cohesive forces should be negligible.
- Undergo elongation under stress.
- Regain the original size when the stress is fully released.

Examples of elastomers are natural rubber and all synthetic rubbers, some nylon materials, etc.

7.13.1 Natural Rubber

Natural rubber is obtained from rubber trees usually grown in tropical regions. The tree stores rubber as a milky white fluid called latex behind the bark of the tree. When a cut is made on the bark the liquid oozes out and collects in vessels kept for the purpose.

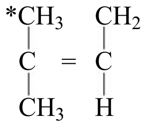
Natural rubber is an addition polymer formed from the monomer $CH_2 = \overset{CH_3}{\overset{|}{C}} - CH = CH_2$ called isoprene (2 methyl 1:3 butadiene). The average degree of polymerization of rubber is around 5000. Addition between molecules of isoprene takes place by 1:4 addition and one double bond shifts between 2 and 3 positions.



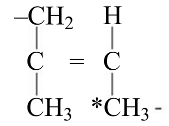
$$\underset{\text{Isoprene}}{CH_2{=}\overset{CH_3}{\overset{|}{C}}{-}CH{-}CH_2} + \underset{\text{Isoprene}}{CH_2{=}\overset{CH_3}{\overset{|}{C}}{-}CH{-}CH_2} \longrightarrow \underset{\text{1:4 di-isoprene}}{-CH_2{-}\overset{CH_3}{\overset{|}{C}}{=}CH{-}CH_2{-}CH_2{-}\overset{CH_3}{\overset{|}{C}}{=}CH{-}CH_2{-}} \xrightarrow{\text{Polymerisation}} \text{(Rubber) Polyisoprene}$$

Since each isoprene unit even now contains a C=C bond, polyisoprene can exist in two isomeric forms *cis* and *trans*.

Natural rubber is the *cis* isomer while the *trans* isomer is called *gutta-percha*.



CIS polyisoprene
Natural rubber



TRANS polyisoprene
Gutta-percha

7.13.2 Deficiencies of Natural Rubber

Natural rubber is an additional product of isoprene units and still contains a large number of double bonded carbon atoms.

Hence, it exhibits a large number of deficiencies.

- At low temperatures it is hard and brittle but as temperature rises it becomes soft and sticky.
- It gets easily air oxidized and produces bad smell even if kept as such for a few days.
- It is soluble in many organic solvents.
- It absorbs large quantities of water.
- Its chemical resistivity is low and is attacked by acids, alkalis, oxidizing and reducing agents.
- Its tensile strength, abrasion resistance, wear and tear resistance are low.

7.13.3 Advantages of Synthetic Rubbers

Synthetic rubber has slightly modified structures compared to natural rubber; they exhibit properties that are more conducive for their technical uses.

A comparative account of the properties of natural and synthetic rubbers is listed in Table 7.6.

Property	Natural Rubber	Synthetic Rubber
1. Tensile strength	Low	High
2. Chemical resistivity	Low: gets oxidized even in air.	High: not oxidized in air.
3. Action of heat	Cold condition it is hard and brittle, at higher temperatures soft and sticky.	Withstand effect of heat over a range of temperature.
4. With organic solvents	Swells and dissolves.	Does not swell and dissolve.
5. Ageing	Undergoes ageing quickly.	Resists ageing.
6. Elasticity	On increased stress undergoes permanent deformation.	Has high elasticity.

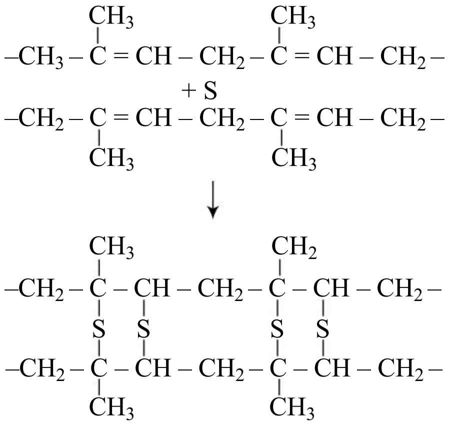
Table 7.6

Comparison of properties of material and synthetic rubbers

7.13.4 Vulcanization

Vulcanization is a process involving heating natural rubber with sulphur around 135°C for 1–4 hours. This is done in order to improve the properties of natural rubber. The S atoms form bridges between poly-neoprene chains involving a part of the double bonded carbon atoms.

The degree of vulcanization depends on the amount of S added. It increases the tensile strength, chemical resistance and wear and tear resistance. In tyre industry about 3–4 percent S and filler like carbon are added. If the percentage of S is increased beyond 30 percent full saturation of the bonds take place resulting in a rigid nonflexible rubber called ebonite.



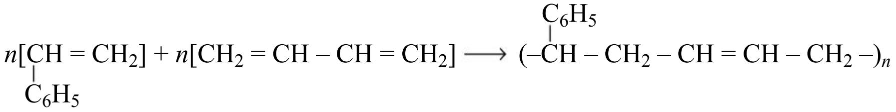
$$\begin{array}{c}
-CH_3-\overset{CH_3}{\overset{|}{C}}=CH-CH_2-\overset{CH_3}{\overset{|}{C}}=CH-CH_2-\\
+\,S\\
-CH_2-\underset{CH_3}{\underset{|}{C}}=CH-CH_2-\underset{CH_3}{\underset{|}{C}}=CH-CH_2-\\
\downarrow\\
-CH_2-\overset{CH_3}{\overset{|}{C}}-CH-CH_2-\overset{CH_2}{\overset{|}{C}}-CH-CH_2-\\
\quad\ \ |\quad\ \ |\qquad\qquad\quad |\quad\ \ |\\
\quad\ \ S\quad\ S\qquad\qquad\quad S\quad\ S\\
\quad\ \ |\quad\ \ |\qquad\qquad\quad |\quad\ \ |\\
-CH_2-\underset{CH_3}{\underset{|}{C}}-CH-CH_2-\underset{CH_3}{\underset{|}{C}}-CH-CH_2-
\end{array}$$

7.13.5 Synthetic Rubbers

Several synthetic rubbers have captured the natural rubber market. We can have tailor-made rubbers for any specific purpose. The following are some of the commonly used synthetic rubbers.

(1) Buna-S rubber

It is a copolymer formed by the emulsion polymerization of a mixture of 1:3 butadiene and styrene in presence of peroxide catalyst at 5°C. The rubber obtained is called *Buna-S* rubber or *SBR* (styrene butadiene rubber) or *cold rubber*. In Buna-S, *Bu* stands for *butadiene*, *Na* for *sodium* (used as a catalyst in earlier days), and *S* for *styrene*.



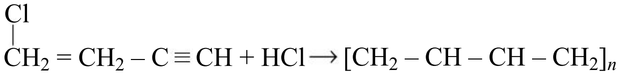
$$n[\underset{\displaystyle C_6H_5}{CH} = CH_2] + n[CH_2 = CH - CH = CH_2] \longrightarrow (-\overset{\displaystyle C_6H_5}{CH} - CH_2 - CH = CH - CH_2 -)_n$$

SBR contains unsaturation and is vulcanized with sulphur. It is used in tyre treads and shoe soles because of its high resistance to abrasion.

(2) Neoprene rubbers (chloroprene rubbers)

Neoprene rubber or chloroprene rubber is a polymer of 2-chloro 1:3 butadiene. The monomer is prepared from acetylene. Catalytic dimerization of acetylene gives vinyl acetylene which adds on a molecule of HCl to form 2-chloro 1:3 butadiene.

$$\underset{\text{Acetylene}}{CH \equiv CH + CH \equiv CH} \rightarrow \quad \underset{\text{Catalytic additions}}{CH_2 = CH - C \equiv CH}$$



$$\overset{\displaystyle Cl}{CH_2} = CH_2 - C \equiv CH + HCl \longrightarrow [CH_2 - CH - CH - CH_2]_n$$

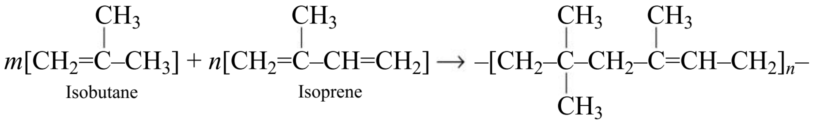
2, 1:3 chlorobutadiene

Emulsion polymerization of the monomer in the presence of peroxide catalyst gives chloroprene.

Neoprene is vulcanized using ZnO. The product so formed is used in making conveyor belts, hoses, etc.

(3) Butyl rubber

It is a copolymer formed from isobutylene with a small amount of isoprene (1–5 percent) at 90°C.



$$m[\underset{\text{Isobutane}}{CH_2{=}\overset{\displaystyle CH_3}{C}{-}CH_3}] + n[\underset{\text{Isoprene}}{CH_2{=}\overset{\displaystyle CH_3}{C}{-}CH{=}CH_2}] \rightarrow \underset{\text{Butyl Rubber}}{-[CH_2{-}\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{C}}{-}CH_2{-}\overset{\displaystyle CH_3}{C}{=}CH{-}CH_2]_n-}$$

Properties:

1. It is chemically inert.
2. It has permeability to air, about 10 percent of that of natural rubber.
3. It has good resistance to heat, abrasion and wear.
4. It is soluble in benzene.
5. It is not easily attacked by ozone.
6. It has good electrical insulating characteristics.

Applications:

- It is widely used in automotive inner tubes for the tyres.
- It is used for making conveyor belts.
- It is also used as insulating material in electrical appliances.

7.14 Adhesive

7.14.1 Natural Adhesive

Definition: An adhesive can be defined as a polymeric binding material which holds firmly two same or different materials by surface attraction.

Adhesives

Natural adhesives
Example Gum and Glue

Synthetic adhesives
phenol formaldehyde resin
urea formaldehyde resin, silicons, epoxides, etc.

Synthetic adhesives are far superior to natural adhesives in their adhesive capacity. The strength of adhesion depends on (i) surfaces being bound, (ii) solvent used, (iii) heat and (iv) light, etc. Adhesive for a particular purpose is chosen carefully. For example epoxy resins are used for binding metals, wood, glass, ceramics, etc., phenol formaldehyde is used for rubber.

(i) Advantages of adhesives over other methods of joining

- Adhesive can be applied to the surface of any material.
- The materials joined by adhesives are free from residual stresses.
- Adhesive binding is easy and quick.
- No heating is normally needed.
- The process of application of adhesives does not require special skill.
- Cost of bonding is low.
- Adhesive bonding does not require any after-finishing.
- Metals joined by adhesives can resist corrosion.
- The joints by adhesives are leakproof.
- Bonding by adhesives gives smoother surface.

(ii) Preparation of surface before application of adhesive

The surface of materials to be bound by adhesive must be properly cleaned and prepared for cohesive adhesion. Paper and wood surfaces are usually applied with adhesive without any preparation.

But metal surfaces are cleaned from grease, oil, and lubricants using organic solvents. Loose dirt and scales are removed by brushing or by rubbing with emery. Sand blasting is carried out to remove deposits on the surface.

(iii) Applications of adhesive

Uniform application of optimum quantity of adhesive over the prepared surface can be achieved by various methods.

1. By brushing the adhesive over the surfaces.
2. By spraying the adhesive if it is a liquid or in solution.
3. By roller coating.
4. By laying a film or sheet of dry adhesive.
5. By laying pressure sensitive tapes of adhesives.

If two chemically reactive substances are involved in the formation of the adhesives, one of them is applied to one of the bonding surface and then the second on the other surface and the two surfaces are compressed together.

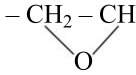
7.14.2 Synthetic Adhesives

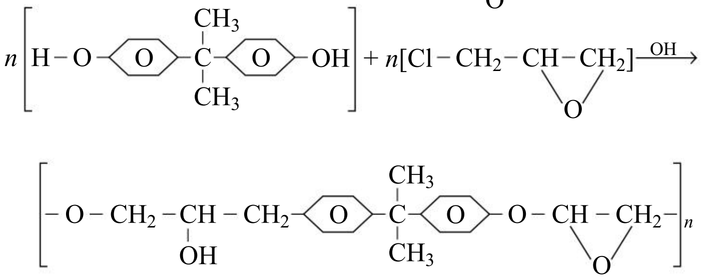
They are pre-polymers or resins of low molar mass. They are dissolved in suitable solvents and applied on the surfaces. As the solvent evaporates, curing takes place and they form thermosetting plastics that bind the surface.

Examples of such adhesives include PF, UF, MF, epoxy resins and silicones. The resin adhesives have high bonding strength, are water and corrosion resistant and are not affected by weather.

Epoxy resins (araldite):

Epoxy resins are condensation polymers produced by the condensation of excess of epichlorohydrin with bisphenol. They are characterized by the presence of epoxy $-CH_2-CH_2-$ group.





Properties:

1. Linear, low molecular weight pre-polymers.
2. Soluble in organic solvents.
3. During curing they form cross linking with di-ammines, di-carboxylic acids, etc.
4. After adhesion they are highly resistant to water, solvents, acids, alkalies and weather.

Uses:

1. As structural adhesive particularly for wood -wood, wood-laminates, etc., adhesion.
2. For lamination purposes.
3. To impart crease and shrink resistance to wrinkle-free clothes.

7.15 Compounding of Resins and Plastics Composites

The process of including various additives into plastics by mechanical means to get a homogeneous mixture is called compounding. These additives make the plastic colourful, improve the workability of plastics, change properties of plastics or produce new properties to the plastics. Depending on the purpose of which additives are used, they are classified into (a) fillers, (b) plasticizers, (c) stabilizers, (d) accelerators, (e) colourantsand (f) fire retarders.

(a) Fillers

Fillers are cheap inorganic or organic compounds incorporated into the polymers to either reduce the cost of the material or to improve certain properties of it. There are two types of fillers.

(i) *Particulate fillers* used mainly as dead weights to lower the cost of the plastic. They also help to increase hardness, improve tensile strength and retention of shape. Examples of such fillers are asbestos powder, clay, silica, talc, lime stone, mica and even wood floor.

(ii) *Fibrous fillers* used mainly to increase the tensile strength, impact and wear and tear resistance. They act like steel in reinforced concrete. Examples of such fillers are cotton thread, scraps of clothes nylon, polyester fibres, fibre glass, etc.

(b) Plasticizers

Some of the plastics are hard, brittle, glassy and have high T_g and melt viscosity. The moulding and workability of such plastics will be difficult. Plasticizers are high molecular weight organic compound that are freely miscible with the plastic and inert.

They reduce the T_g and melt viscosity of the plastic and make it workable. Examples of plasticizers are triphenylphosphates, phthalic esters, fatty acid esters, etc.

(c) Stabilizers

Many plastics undergo thermal or photochemical degradation during their processing or when they are put into use. Stabilizers protect the plastics from undergoing such degradation.

Elastomers like natural and synthetic rubber and addition polymers like PVC, polythene and some condensation polymers like polyesters undergo degradation by air and light during their usage. This is prevented by incorporating antioxidants like alkyl phenols, naphthols, phenyl salicylate, etc.

Organic phosphites like triphenylphosphite increase the stability of plastics towards heat during their processing.

(d) Colourants

Colourants are inorganic or organic pigments used to impart pleasing eye catching colours to the plastic. They do not have any other influence on the properties of the plastics.

Organic colourants give bright, less dense and transparent colours. Examples of organic colourants are as follows:

Carbon black: gives black colour anthraquinones
Yellow azodyes: yellow, orange, red phthalocyanins—green

Inorganic pigments produce less bright and opaque colours. For examples, barium sulphate and titanium dioxide—white, lead chromate and iron oxides—yellow, zinc chromate—green.

(e) Accelerators

They are substances that impart necessary rate of solidification at low temperatures: like, gypsum in cement.

7.16 Polymer Composites

Composite is any material that is made out of any two materials being put together with a defined interface, or one material stuck to, in or between one or more other materials.

Composites are combinations of two materials in which one of the materials, called the *reinforcing phase*, is in the form of fibres, sheets or particles, and are embedded in the other materials called the *matrix phase*. The reinforcing material and the matrix material can be metal, ceramic or polymer. Typically, reinforcing materials are strong with low densities while the matrix is usually a ductile or tough material.

If the composite is designed and fabricated correctly, it combines the strength of the reinforcement with the toughness of the matrix to achieve a combination of desirable properties not available in any single conventional material. The downside is that such composites are often more expensive than conventional materials. Examples of some current application of composites include the *diesel piston*, *brake-shoes and pads*, tyres and the *Beechcraft aircraft* in which 100 percent of the structural components are composites.

In other words, a composite is a physical mixture of two or more structurally different components, which combine to form a new class of material suitable for structural applications.

This combination of such materials gives composites with low density, high strength, abrasion, corrosion resistance, etc., which the individual materials do not have, and has led to the search and development of composite materials.

A polymer composite is a material made of two or more types of polymers, but the polymer composites are not chemically joined the way that copolymers are, instead, polymer composites have a defined interface between the polymers.

7.16.1 Materials Used in Polymer Composites

Modern composites are usually made of two components, a *fibre* and *matrix*.

The fibre is most often glass, carbon fibre, or polyethylene. The matrix is usually a thermoset like an epoxy resin, polydicyclopentadiene or a polyimide.

The fibre is embedded in the matrix in order to make the matrix stronger. Fibre reinforced composites have two things going from them. They are strong and light. They are often stronger than steel, but weigh much less. This means that composites can be used to make *automobiles lighter*, and thus much more *fuel efficient*. The materials generally used are as follows:

1. *Fibres*: glass fibres, carbon fibres and aramids.
2. *Resins*: (a) thermosetting: polyesters and epoxides.
   (b) Thermoplastic: polyolefins, polyimides, vinylic polymers, etc.
   polyacetals: polycarbonates and polyphenylenes.
3. *Additives*: For structural applications it is mandatory to achieve some degree of *flame retardancy.* Flame retardant additives are incorporated in the resin itself. Fillers and pigments are also used in the process.

7.16.2 Need to Develop Polymer Composites Based on their Properties

Today, *polymer composites* are needed which can stand up to the torture in various applications and they are preferred to many conventional metals and materials for the following advantageous properties.

- They are light in weight.
- They have high strength to weight ratio.
- They are much stronger and durable than conventional metals like steel, aluminium, etc.
- They are most suitable for aerospace applications due to the inherent characteristic properties.
- They have good corrosion resistance.
- They have high fatigue strength.
- They have high temperature resistance.

Applications of polymer composites:

Composites of phenolic resins and nylon are used in heat shields for spacecraft.

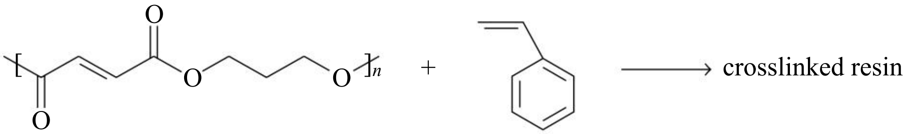
They are used in the aircraft and space industry.

They are suitable for automotive and railway applications.

They are used for civil construction works also.

7.16.2.1 Fibreglass

A common fibre reinforced composite is *fibreglass*. Its matrix is made by reacting a polyester with carbon-carbon double bonds in its backbone and styrene: a mix of the styrene and polyester over a mass of glass fibres.



The styrene and double bonds in the polyester react by free radical vinyl polymerization to form a crosslinked resin. The glass fibre are trapped inside, where they act as a reinforcement. The matrix adds *toughness* to the composite, while fibres have good tensile strength, they usually have awful compressional strength. The matrix gives *compressional strength* to the composite.

Fibre reinforced composites are so good at providing strength to materials that they are used in all kinds of things ranging from car engine components to aeroplane components to boat hulls.

7.16.2.2 Kevlar

Kevlar is an aramid (an aromatic polyamide). The chemical composition of Kevlar is poly paraphenyleneterephthalamide.

Aramids belong to the family of *nylons*. Common nylons, such as nylon-6,6, do not have very good structural properties, so the incorporation of para *aramids* improves the properties. The aramid ring gives Kevlar thermal stability, while the para structure gives it high strength and modulus.

Kevlar [1] *is a Polyamide , in which all the amide groups are separated by para-phenylene groups, that is, the amide groups attach to the phenylene rings opposite to each other, at carbons 1 and 4* as shown in Fig. 7.10.

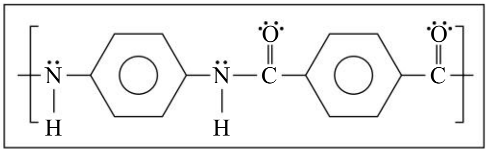


FIGURE 7.10 *Structure of a Kevlar material.*

(i) Advantages and disadvantages of Kevlar

Kevlar has the following merits:

- Lightweight and high strength.
- It has a slightly negative axial coefficient of thermal expansion and therefore, it is thermally stable.
- Kevlar is very resistant to impact and abrasion damage. It can be used as a protective layer on graphite laminates.
- Kevlar can also be mixed with graphite in hybrid fabrics to provide damage resistance, increased ultimate strains and to prevent catastrophic failure modes.

(ii) Demerits

The fibre themselves absorb moisture, so Kevlar composites are more sensitive to the environment than glass or graphite composites.

Although tensile and modulus are high, compressive properties are relatively poor.

(iii) Applications

The applications of Kevlar include lightweight boat hulls, aircraft fuselage, high performance race cars, bulletproof vest and puncture resistant tyres.

7.16.2.3 Carbon Fibre

Carbon fibres are a new breed of high strength composite materials. Carbon fibre has been described as a fibre containing at least 90 percent carbon obtained by the controlled pyrolysis of appropriate fibres. A large

---

[1] [a Product of Du Pont, USA]

variety of fibres called precursors are used to produce carbon fibres of different morphologies and different characteristics. The most prevalent precursors are *polyacrylonitrile* (PAN), cellulosic fibres (viscose rayon, cotton), petroleum or coal tar pitch and certain phenolic fibres.

Manufacture:

Carbon fibres are manufactured by the controlled pyrolysis of organic precursors in fibrous form. It is basically heat treatment of the precursor that removes oxygen, nitrogen and hydrogen to form carbon fibres. It is well established in carbon fibre literature that the mechanical properties of the carbon fibres are improved by increasing the crystallinity and orientation, and by reducing defects in the fibre. The best way to achieve this is to start with a highly oriented precursor and then maintain the initial high orientation during the process of stabilization and carbonization through tension.

(i) Carbon fibres from PAN

There are three stages in the conversion of PAN precursor into high performance carbon fibre.

Oxidative stabilization: The Poly acrylonitrile precursor is first stretched and simultaneously oxidized in a temperature of 200–300°C. This treatment converts thermoplastic PAN to a non-plastic cyclic or ladder compound.

Carbonization: After oxidation, the fibres are carbonized at about 1000°C without tension in an inert atmosphere (normally nitrogen) for few hours. During this process the non-carbon elements are removed as volatiles to give carbon fibres with a yield of about 50 percent of the mass of the original PAN.

Graphitization: Depending upon the type of fibre required, the fibres are treated at temperatures between 1500–3000°C which improves the ordering and orientation of the crystallites in the direction of the fibre axis.

(ii) Characteristics and applications of carbon fibres

In recent decades, carbon fibres have found wide application in commercial and civilian aircraft, recreational, industrial and transportation markets. Carbon fibres are used in composites with a lightweight matrix. Carbon fibre composites are ideally suited to applications where strength, stiffness, lower weight and outstanding fatigue characteristics are critical requirements. They also can be used in the occasion where high temperature, chemical inertness and high damping are important.

The production of highly effective fibrous carbon adsorbents with low diameter, excluding or minimizing external and intra-diffusional resistance to mass transfer, and therefore, exhibiting high adsorption rates is a challenging task. These carbon adsorbents can be converted into a wide variety of textile forms and nonwoven materials.

Cheaper and newer versions of carbon fibres are being produced in various chemical industries for work in extremely hostile environment, electromagnetic shielding and various other novel applications. The use of carbon fibres in non-woven is among new possible applications for high-temperature fire retardant insulations (e.g. Furnace materials).

7.17 Conducting Polymers

Introduction:

Conductive polymers have been the subjects of study for many decades as possible *synthetic metals*. Many of these polymers, especially those with a conjugated π-bond system, often yield higher conductivity once having undergone the doping process.

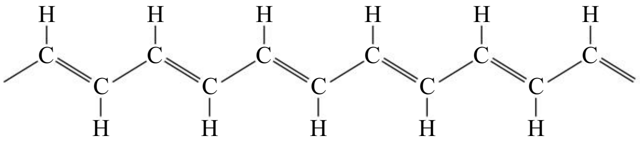
The electrical conductivities of the intrinsically conducting-polymer systems now range from those typical of insulators ($<10^{-10}$ S/cm (10^{-10} Ω^{-1} cm^{-1}) to those typical of semiconductors such as silicon ($\sim10^{-5}$ S/cm) to those greater than 10^{+4} S/cm (nearly that of a good metal such as copper, 5×10^{5} S/cm).

The most important aspect of conjugated polymers from an electrochemical perspective is their ability to act as electronic conductors such as the ‘conventional’ polymers (*e.g. polythiophene, polyaniline, polypyrrole and host of other series*) in charge storage devices such as batteries and supercapacitors, and also extended to new polymers with specialized conductivity properties such as low band gap and intrinsically conducting polymers. *The target is to obtain a material which could combine the process ability and environmental stability and weight advantages of a fully organic polymer with the useful electrical properties of a metal*.

Conjugated poly chain hydrocarbons (known as conducting polymers), are distinguished by alternating single and double bonds between carbon atoms on the polymer backbone.

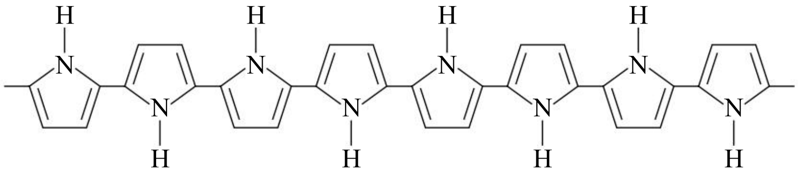
The conjugated polymer with the simplest chemical structure is *polyacetylene*.

(i) Polyacetylene

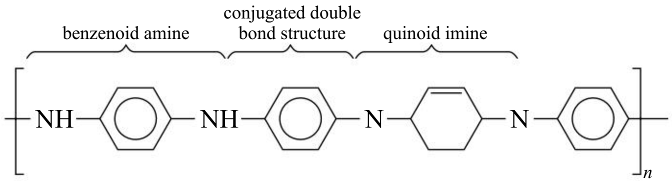


The other aromatic polymers *whose structures are*

(ii) Polypyrrole



(iii) Polyaniline



Conjugated polymers are organic semiconductors. They have a *band gap*

- They can emit light, the colour of which is tailorable through the chemical structure. They can generate a current upon absorbing light, and so can be used in photovoltaic devices.
- The conductivity of conjugated polymers can be varied.
- Like silicon, their conductivity depends on the doping level.
- The doping level depends on the oxidation state of the polymer, which can be electrochemically controlled.

Conductive polymers, which are almost always organic, may have extended *delocalized bonds* (often composed of linear or *aromatic* units) that create a band structure (Fig. 7.11a) similar to *silicon* (Fig. 7.11b), but with localized states (Fig. 7.11c). When *charge carriers* (from the addition or removal of electrons) are introduced into the conduction or valence bands the *electrical conductivity* increases dramatically.

Technically, almost all known conductive polymers are semiconductors due to the band structure, however, so-called zero *band gap* conductive polymers may behave like *synthetic metals*. The most notable difference between conductive polymers and inorganic semiconductors is the *mobility*, which until very recently was dramatically lower in conductive polymers than their inorganic counterparts, though recent advancements in self-assembly are closing that gap.

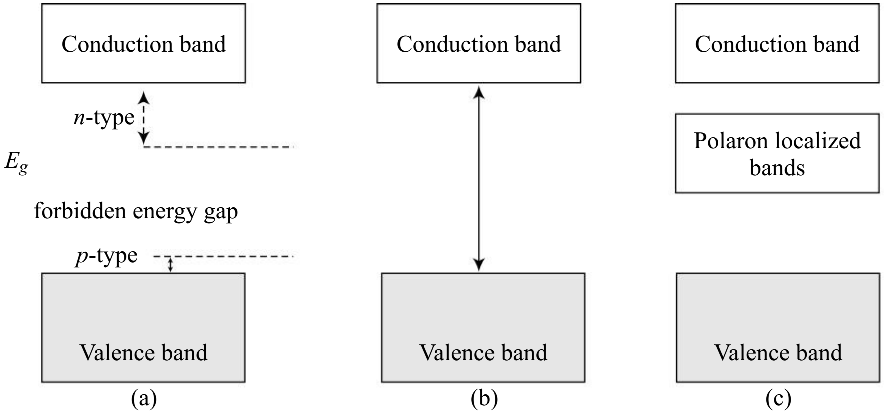


Figure 7.11 *Energy bands of semiconductors.*

Delocalization can be accomplished by forming a conjugated backbone of continuous overlapping orbitals, for example, alternating single and double carbon–carbon bonds, which leaves a continuous path of overlapping *p*-orbitals. In *poly acetylene*, but not in most other conductive polymers, this continuous string of orbitals creates degeneracy in the frontier molecular orbitals (the highest occupied and lowest unoccupied orbitals named HOMO and LUMO, respectively) which leads to the filled (electron containing) and unfilled bands (valence and conduction, respectively) that define a *semiconductor.*

However, conductive polymers generally exhibit very low conductivities. In fact, conduction in such relatively disordered materials is mostly a function of 'energy gaps' with *phonon assisted hopping, polaron-assisted tunnelling*, etc., between localized states and not band gaps as in crystalline semiconductors.

It is not until an electron is removed from the valence band (*p*-doping) or added to the conduction band (*n*-doping) does a conducting polymer become highly conductive. Doping (*p* or *n*) generates charge carriers which move in an electric field. Positive charges (holes) and negative charges (electrons) move to opposite electrodes. This movement of charge is actually responsible for electrical conductivity.

7.18 Conduction Mechanism of Polyacetylene

The polyunsaturated hydro carbons consist of hydrogen atoms and sp^2-hybridized carbon atoms that share two pairs in a double bond, which results in σ bonds (Fig. 7.12). The remaining 2*p* orbitals overlap above and below the nodal plane of the σ bond, orthogonal to the σ bond axis, give rise to π-bonds.

The semi-conducting poly acetylene $(CH_2)_n$ (Fig. 7.13) has a typical carbon-carbon backbone structure as given

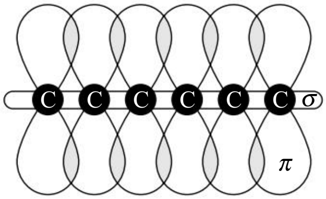


Figure 7.12 *σ and π bonds of a poly carbon-carbon backbone.*

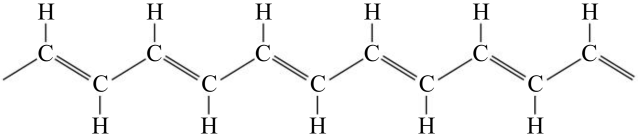


Figure 7.13 *Polyacetylene chain with conjugated double bonds.*

Electrons that are involved in π-bonds are desired in semiconducting polymers since they are delocalized at higher energies than localized σ bond electrons and can be moved more easily along the carbon chain in the vicinity of an applied electric field.

In summary, the localized electrons in the σ bonds form the backbone of the polymer chain and dominate the mechanical properties, *while the electrons in the π bonds are delocalized along the chain and responsible for the electrical and optical properties of a conjugated polymer.*

The σ bonds form completely filled low lying energy bands that have a *larger* energy gap than the π bond electrons (Fig. 7.14).

Before a current can flow along the molecule, *one or more electrons have to be removed* or inserted. If an electrical field is then applied, the electrons constituting the π-bonds can move rapidly along the molecule chain. The conductivity of the polymeric material, which consists of many polymer chains, will be limited by the fact that the electrons have to 'jump' from one molecule to the next. Hence, the chains have to be well packed in ordered rows.

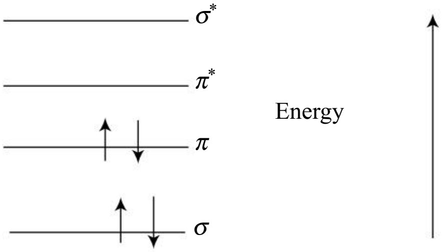


FIGURE 7.14 *Relative energies of MOs of bonding and antibonding of σ and π bonds of polyacetylene carbon-carbon backbone.*

(a) Doping of polyacetylene

Polyacetylene has alternating single and double bonds that give rise to mobile π-electrons that when doped (Fig. 7.15) become highly anisotropic metallic conductors.

There are two types of doping, oxidation or reduction. In the case of polyacetylene the reactions are written like this:

(i) Oxidation with halogen (p-doping):

$$[\mathrm{CH}]_n + 3x/2\mathrm{I}_2 \rightarrow [\mathrm{CH}]_n^{x+} + x\mathrm{I}_3^-$$

(ii) Reduction with alkali metal– (n-doping):

$$[\mathrm{CH}]_n + x\,\mathrm{Na} \rightarrow [\mathrm{CH}]_n^{x-} + x\mathrm{Na}^+$$

In the above reactions, oxidation, the iodine molecule attracts an electron from the polyacetylene chain and becomes I_3^-. The polyacetylene molecule, now positively charged, is termed as a radical cation, or *polaron* (Fig. 7.16). Polarons are created on doping but they may change into (p or n) solitons for polyacetylene.

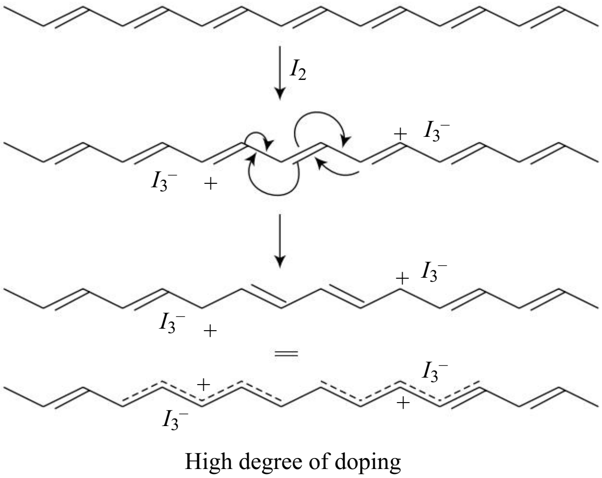


FIGURE 7.15 *Doping of polyacetylene with iodine.*

(a) Long chain of carbon atoms with attached hydrogen atoms is linked by alternating single and double bonds (neutral chain); (b) when the chain is doped with iodine, a hole is added (plus sign), and the negative iodine ion remains associated with the polymer. The distorted chain surrounding the hole is a positive polaron; (c) a topological kink, or soliton, develops if the alternating pattern of double–single bonds is reversed. The soliton shown here is positively charged and (d) formation of polaron, soliton energy levels and also solitons energy bands for conduction in doped polyacetylene.

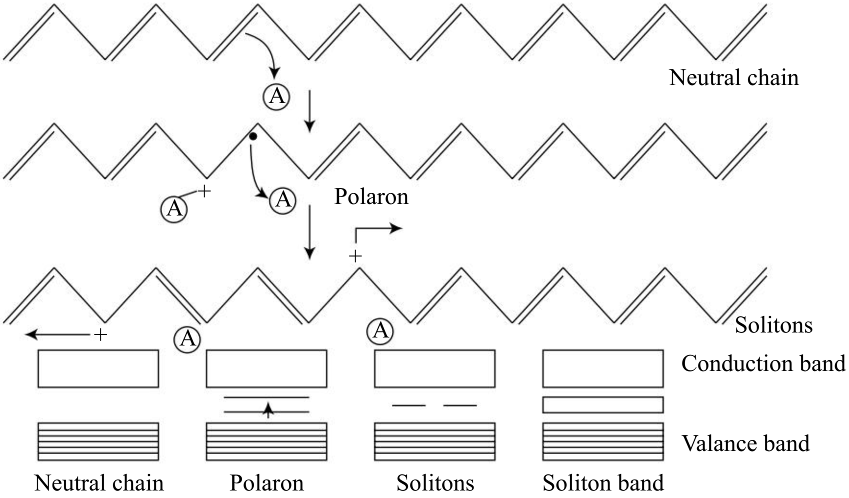


Figure 7.16 *Simple model of a π-conjugated polyacetylene.*

The *lone electron* of the double (π) bond, from which an electron was removed, can move easily. As a consequence, the double bond successively moves along the molecule. The *positive charge*, on the other hand, is fixed by electrostatic attraction to the iodide ion, which does not move so readily. If the polyacetylene chain is heavily oxidized, polarons condense pairwise into so-called *solitons*. These solitons are then responsible, in complicated ways, for the transport of charges along the polymer chains, as well as from chain-chain on a macroscopic scale. Mobile charged solitons which enable 'spinless' conductivity are unique to polyacetylene.

7.19 Mechanism of Conduction in Polypyrrole

The structure of polypyrrole is shown in Fig. 7.17. It behaves as a semiconducting polymer when doped with a requisite amount of iodine or bromine.

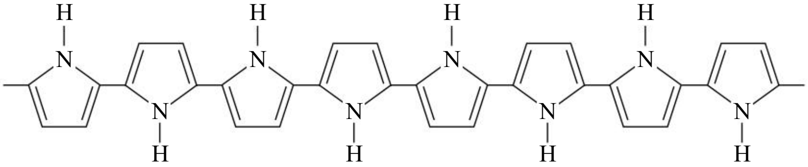


Figure 7.17 *Conjugated structure of polypyrrole.*

An electrically conductive polymer (ECP) is able to conduct electricity because of its *conjugated π-bond system, which is formed by the overlapping of carbon p orbitals and alternating carbon-carbon bond lengths*. The *conjugated double bonds* permit easy electron mobility throughout the molecule because the electrons are delocalized. Delocalization is the condition in which π-bonding electrons are spread over a number of atoms rather than localized between two atoms. This condition allows electrons to move more easily, thus making the *polymer electrically conductive*.

Electronically conducting polymers (ECPs) are extensively *conjugated* molecules, and it is believed that they possess a *spatially delocalized band-like electronic structure*. These bands stem from the splitting of interacting molecular orbitals of the constituent monomer units in a manner reminiscent of the band structure of solid-state semiconductors (Fig. 7.16).

It is generally agreed that the mechanism of conductivity in these polymers is based on the motion of *charged defects* within the conjugated framework. The charge carriers, either positive *p-type* or negative *n-type*, are the products of *oxidizing or reducing* the polymer, respectively. An oxidation doping (removal of electrons) can be done using iodine. The iodine attracts an electron from the polymer from one of the pi-bonds. Thus, the remaining electron can move along the chain.

Oxidation of the polymer initially generates a '*radical cation'* with both *spin and charge*. Borrowing from solid state physics terminology, this species is referred to as a *polaron* and comprises both the *hole site* and the *structural distortion* which accompanies it.

The *cation and radical* form *bound species*, since any increase in the distance between them would necessitate the creation of additional *higher energy quinoid units*. Theoretical treatments have demonstrated that two nearby *polarons* combine to form the *lower energy bipolaron*. One bipolaron is more stable than two polarons despite the coulombic repulsion of the two ions. Since the *defect* is simply a boundary between two moieties of equal energy: the infinite conjugation chain on either side, it can migrate in either direction without affecting the energy of the backbone, provided that there is no significant energy barrier to the process. It is this *charge carrier mobility* that leads to the high conductivity of these polymers. This charge transport mechanism is shown in Fig. 7.18.

It is generally agreed that the mechanism of conductivity in polypyrrole is based on the motion of *charged defects* within the conjugated framework. The charge carriers, either positive *p-type* or *n*egative *n-type*, are the products of *oxidizing or reducing* the polymer, respectively.

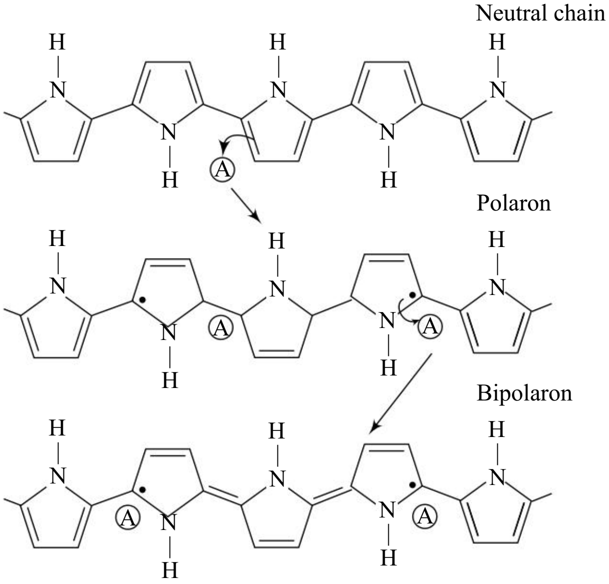


FIGURE 7.18 *Formation of charged defects and polarons in polypyrrole.*

Once doping has occurred in polypyrrole, the electrons in the pi-bonds are able to 'jump' around the polymer chain. As the electrons are moving along the molecule, an electric current occurs.

7.20 Conduction Mechanism of Polyaniline

Polyaniline is one of the oldest conductive polymers known. It was first prepared by the anodic oxidation of aniline in sulphuric acid. It has been known as an electrically conductive polymer [ECP] for the past 30 years. ECPs are able to conduct electricity because of their conjugated π-bond system, which is formed by the overlapping of carbon *p*-orbitals and alternating carbon-carbon bond lengths extending over large number of recurring monomer units. In polyaniline, nitrogen p_z orbitals and carbon rings are also the part of the conjugated system as shown in Fig. 7.19.

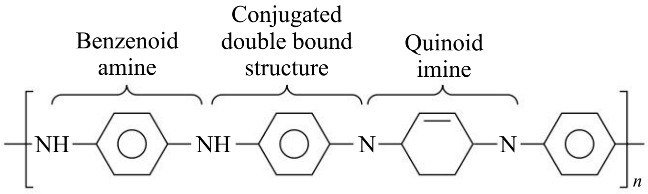


FIGURE 7.19 *Structural formula of undoped PANi, EB.*

The conjugated double bonds permit easy electron mobility throughout the molecule because the electrons are delocalized. Delocalization is the condition in which the π-bonding electrons are spread over a number of atoms rather than localized between two atoms. This condition allows electrons to move more easily, thus making the polymer electrically conductive.

In addition, it has a conjugated double bond structure, the benzenoid ring, between the quinoid imine and the benzenoid amine structures, which renders the polymer a candidate as an *electrically conductive polymer* (Fig. 7.19).

Polyaniline (PANi) exists in the three forms of oxidation states.

1. Leucoemeraldine (fully reduced or only benzenoid amine structure),
2. Emeraldine (neutral or partially reduced and partially oxidised) and
3. Pernigraniline (fully oxidised or only quinoid imine structure).

The emeraldine-based (EB) form of polyaniline was used because only doped EB polyaniline is conductive among the three oxidation states. The emeraldine-based form of polyaniline is also the most stable of the three states because leucoemeraldine is easily oxidized when exposed to air and pernigraniline is easily degraded.

(a) Preparation of polyaniline

Polyaniline (PANi) can be easily synthesized with a good yield of polymerizing aniline in protonic acid aqueous solution in the presence of an oxidant, or electrochemically. Polyanilines can be regarded as a conducting polymer under certain stimulating conditions like UV light, heat or addition of a suitable dopant to the polymer.

PANiis usually prepared by the redox polymerization of aniline using ammonium per disulphate $(NH_4)_2S_2O_8$ as the oxidant. Distilled aniline (0.2M) was dissolved in 300 mL of pre-cooled HCl (1.0 M) solution and maintained at 0.5° C. A calculated amount of $(NH_4)_2S_2O_8$ (0.5 M) dissolved in 200 mLof HCl (1M), pre-cooled to 0 –5°C, was added to the above solution. The dark green precipitate resulting from this reaction was washed with HCl (1M) until the green colour disappeared. This precipitate was further extracted with tetrahydrofuran. Polyaniline emeraldine base (EB), can be obtained by heating the salt with ammonia solution and vacuum dried.

(b) Doping of polyaniline

A polymer can be made conductive by *oxidation* (*p*-type doping) and or less frequently, *reduction* (*n*-type doping) of the polymer either by chemical or electrochemical means, generating the mobile charge carriers.

Doping is the process by which polymers that are insulators or semiconductors as synthesized are exposed to *charge transfer agents* (dopants such as para-hydroxybenzenesulphonic acid, PHBSA) in the gas or solution phase or through appropriate electrochemical oxidation or reduction. This process will increase the polymer's ability to conduct electricity because of the increased concentration of *charge carriers*.

Polyaniline in the undoped state is a poor semiconductor with conductivity of about 10^{-8} mho^{-1}/cm. However, once it is doped with PHBSA, its conductivity could increase by a factor of 10 ohm^{-1}/cm or more depending on the dopant and forms *polaron/bipolaron structure* (Fig. 7.20) that will increase conductivity due to increased delocalization. The '*bipolaron structure'* is the initial result of protonating polyaniline (emeraldine base) with PHBSA. The hydrogen ions from PHBSA will attach themselves to the quinoid nitrogen atoms. This is highly unstable because of the high energy, this structure possesses.

Thus, the C=N bonds of the quinoid imine structure will break, transforming the quinoid ring into 'more stable benzenoid ring, with lower energy', by aromatization, creating the bipolaron structure of doped polyaniline, as shown in Fig. 7.20.

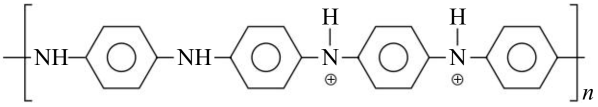


Figure 7.20 *Bipolaron structure of PANi, EB.*

The new benzenoid ring, although more stable than it was previously, still has high energy because of the repulsion force from the adjacent +ve charges. In order to stabilize this structure, the positive charge of one of the hydrogen ions will attract electrons from the neighbouring benzene ring, neutralizing the charge. This will create a new +vely charged nitrogen group with neutral nitrogen atom in between the two +ve ones. The increased distance between the two positive charges results in the polaron structure, which has a lower energy level than the bipolaron structure (Fig. 7.21).

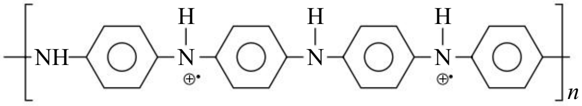


Figure 7.21 *Polaron structure of PANi, EB.*

Electronically conducting polymers are extensively conjugated molecules, and it is believed that they possess a *spatially delocalized band like structure*. The band structure of polyaniline is shown below as Fig. 7.22.

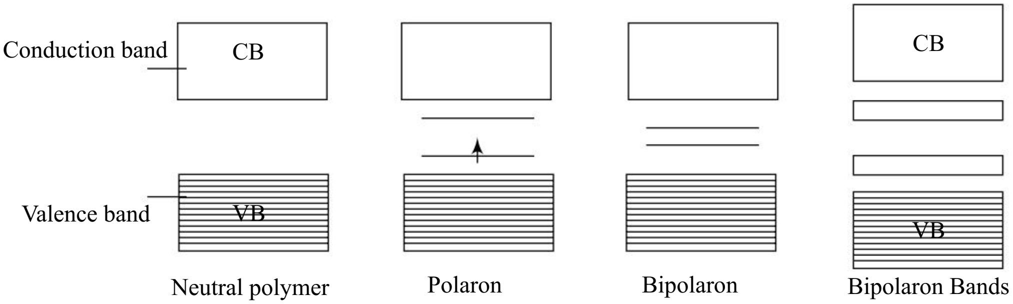


Figure 7.22 *Band structure of PANi, EB.*

The conductivity, σ, of a conducting polyaniline is related to the number of charge carriers 'n' and their mobility μ.

Therefore, $\sigma = ne\mu$.

7.21 Applications of Polyaniline

The extended π-bond system of conjugated polymers is highly susceptible to chemical and electrochemical oxidation or reduction. These alter the electrical and optical properties of the polymer, and by controlling this oxidation and reduction, it is possible precisely control these properties. There are two main groups of application for these polymers. First group utilizes their conductivity as its main property. The second group utilizes this electroactivity. They are shown in Table 7.7.

Group I	Group II
Electrostatic materials	Molecular electronics
Conducting adhesives	Electrical displays
Electromagnetic shielding	Chemical and biochemical sensors
Printed circuit boards electrolytes	Rechargeable batteries and solid
Artificial nerves	Drug release systems
Antistatic clothing	Optical computers
Thermal sensors	Ion exchange membranes
Piezoceramics	Electromechanical actuators
Active electronics	Smart structures
Aircraft structures	

Table 7.7

Applications of polymers

Review Questions

1. Define the following: (i) monomer, (ii) polymer, (iii) polymerization, (iv) degree of polymerization, (v) Functionality, (vi) glass transition temperature, (vii) addition polymerization and (viii) resins. [VTU: Aug. 1999]
2. Write a brief account of the classification of polymers.
3. Define number average and weight average molar mass. Distinguish between M_n and M_w. [VTU: Aug. 1999, Aug. 2000, Aug., 2001]
4. Discuss the mechanism of free radical addition polymerization with a suitable example.
5. Give mechanism of free radical polymerization of ethylene. [VTU: Mar. 2000; Aug. 1999]
6. Enumerate the properties of LDPE and HDPE and their uses.
7. What is condensation polymerization? Give two examples.
8. Mention the differences of addition and condensation polymerization.
9. Write the synthesis of a polyester and a polyamide.
10. What is copolymerization?
11. What is a copolymer? How are they classified?
12. Explain the following methods of polymerization (a) bulk polymerization (b) solution polymerization (c) suspension polymerization (d) emulsion polymerization. What are their advantages and disadvantages?
13. Explain the mechanism of emulsion polymerization.
14. Explain the term glass transition temperature, and various factors that affect T_g. [VTU: Aug. 2000]
15. Discuss the relationship between structure and following properties (i) crystallinity (ii) elasticity (iii) polymer strength. [VTU: March 2001, Aug. 2001]
16. Write a note on the following:
    (i) tacticity of polymers and (ii) viscoelasticity.
17. What is chemical resistivity of a polymer?
18. What are plastics? Explain the compounding of resins into plastics
19. Explain the terms resins and plastics.
20. Mention the types of plastics with examples.
21. What is a thermoplastic? Give an example.
22. What is a thermosetting plastic? Discuss.
23. Distinguish between thermoplastics and thermosetting plastics.
24. What are the techniques of plastic moulding? Discuss compression moulding of plastics.
25. Write a note on injection moulding process of plastics.
26. Explain Blown film moulding of plastics.
27. Write a note on extrusion moulding of plastics.

28. Explain the functions of different ingredients used in compounding of plastics.
29. How are the following polymers prepared (i) polyethylene (ii) polypropene (iii) polystyrene (iv) PVC (v) teflon (vi) plexiglas (vii) poly urethane (viii) polycarbonates (ix) resoles and novalocs (x) neoprene.
30. Describe the manufacture of phenol formaldehyde resins and its uses.
31. Describe the synthesis of urea formaldehyde.
32. What are elastomers? Write the structure of natural rubber and guttapercha? What are the deficiencies of natural rubber? [VTU: Aug. 1999]
33. Give the manufacture, properties and application of neoprene. [VTU: Aug. 1999]
34. How are the following polymers synthesized. Give their properties and uses. (i) neoprene and (ii) polystyrene.
35. Describe the manufacture of (i) polystyrene and (ii) neoprene. [VTU: Aug. 2001]
36. What are the deficiencies of natural rubbers? Explain the term vulcanization.
37. How are the following synthetic rubbers manufactured?
    (i) buna-S rubber (ii) neoprene
    (iii) butyl rubber. Discuss their uses.
38. What are adhesives? Write a note an epoxy resins.
39. What is a polymer composite material? Give two examples.
40. What is Kevlar? Give its structure.
41. What is carbon fibre? Mention an example.
42. What are conducting polymers? Illustrate with an example.
43. Explain the conduction mechanism of polyaniline
44. Mention the uses of conducting polymers.
45. (a) What are copolymers?
    (b) What are thermoplastics?
    [Annamalai University, Nov 2014]
46. (a) Discuss the mechanism of condensation polymerization in detail.
    (b) Explain any four of the important properties of polymers.
    [Annamalai University, Nov 2014]
47. (a) Describe the emulsion polymerization technique. Give two examples.
    (b) Write down the preparation, properties and uses of Nylon 6.6.
    [Annamalai University, Nov 2014]
48. (a) What is functionality of (i) Propylene (ii) phenol towards polymerization.
    (b) Define polydispersity index
    [Annamalai University, 2013]
49. (a) Explain the mechanism of polymerization through free radicals.
    (b) What are the different types of molecular mass of polymers? Explain each type with significance.
    [Annamalai University, 2013]
50. (a) What are the differences between thermoplastic and thermosetting polymers?
    (b) Explain the following:
    (i) emulsion polymerization (ii) solution polymerisation.
    [Annamalai University, 2013]
51. (a) Differentiate between thermoplastic and thermosetting polymers.
    (b) Define degree of polymerization
    [Annamalai University, May, 2016]
52. (a) Write the preparation, properties of
    (i) Nylon 6,6 and (ii) Epoxy resins
    (b) Distinguish between addition and condensation polymerization.
    [Annamalai University, May, 2016]
53. (a) Discuss the mechanism of addition polymerization.
    (b) Explain the number average and weight average molecular weight.
    [Annamalai University, May, 2016]
54. (a) What is meant by degree of polymerization?
    (b) Mention any two uses of epoxy resins.
    [Annamalai University, 2014]
55. (a) Discuss cationic polymerization mechanism in detail.
    (b) Distinguish thermoplastic and thermosetting plastics.
    [Annamalai University, Jan 2014]
56. (a) Discuss bulk polymerization technique. Mention any two polymers synthesized by this technique.
    (b) Explain the free radical polymerization.
    [Annamalai University, May 2014]

57. (a) Explain the anionic polarisation with mechanism.
    (b) Give difference between natural and synthetic rubber.
    [Gujarat tech University, May 2015]
58. (a) What are plastics? What are its constituents?
    (b) Write the differences between addition and condensation polymerization with examples.
    (c) Write the preparation, properties and engineering applications of Bakelite.
    [JNTU, June 2014]
59. (a) Give the preparation reaction, properties and uses of the following polymers.
    (i) LDPE and (ii) styrene-butadiene rubber
    (b) Distinguish thermoplastic and thermosetting plastic with examples.
    [Pune University, 2014]
60. (a) Give the preparation, reaction, properties and applications of the following:
    (i) Styrene-Butadiene rubber (ii) HDDE
    [Pune University, 2016]
61. (a) Explain bulk polymerization technique. Draw the figure and state its disadvantages.
    (b) Distinguish between thermo softening and thermosetting polymers with examples.
    [Pune University, 2016]
62. (a) Define addition polymerization. Explain free radical mechanism with example.
    (b) Distinguish between LDPE and HDPE. (any four points)
    (c) Give the preparation reaction, properties and uses of SBR.
    [Pune University, 2014]
63. (a) Give the purpose and examples of the following constituents during the compounding of plastics.
    (i) Fillers, (ii) lubricants and (iii) plasticizers
    (b) Explain solution polymerization technique? Draw the figure and state the disadvantage of this technique.
    [Pune University, Dec, 2015]
64. (a) What is vulcanization of rubber? Explain chemical reaction involved in vulcanization process.
    Compare natural rubber with vulcanized rubber.
    (b) Explain with suitable diagram bulk polarization technique to bring about addition polymerization.
    (c) Give the structure, properties and applications polyphenylene vinylene (PPV)
    [Pune university, 2015]
65. (a) Explain the free radical mechanism of polymerization taking vinyl chloride as a monomer.
    (b) Differentiate addition and condensation polymerization.
    (c) Give the synthesis of reaction of Teflon and polycarbonate.
    (d) Discuss the synthesis, properties and applications of epoxy resin.
    [VTU, June/July, 2015]
66. (a) Explain the free radical mechanism of addition polymerization of vinyl chloride.
    (b) Give the synthesis and applications of the following polymers:
    (i) PMMA (ii) polycarbonate
    (c) What are polymer composites? Explain, synthesis and properties of Kevlar.
    [VTU, CBSC, June/July, 2016]
67. (a) Explain the addition and condensation polymerization with suitable example each.
    (b) In a sample of a polymer 20 percent of molecules have molecular mass 15,000 g/mol, 35 percent of molecules have molecular mass 25,000 g/mol, 45 percent of molecules have molecular mass of 20,000 g/mol. Calculate the average and weight average molecular mass of a polymer.
    (c) Give the synthesis, properties and applications of polyurethane.
    (d) What are polymer composites? Explain the synthesis and properties of carbon fibre.
    [VTU June/July, 2016]
68. (a) Explain the free radical mechanism of addition polymerization of vinyl chloride.
    (b) What are adhesives? Explain the synthesis and properties of epoxy resin.
    (c) What is glass transition temperature? Explain any three factors affecting $T_{g.}$
    (d) What are conducting polymers? Explain the mechanism of conduction of polyaniline.
    [VTU June/July, 2016]

8

Water Technology

Chapter Outline

Introduction–sources of drinking water, impurities in water, contaminants in raw water. Water analysis. Total hardness of water, types of hardness, causes of hardness, units of hardness, removal of temporary hardness, disadvantage of hardness. Determination of hardness by EDTA method. Alkalinity. Determination of chloride, nitrate, sulphate, fluoride, total solids dissolved oxygen in water and its importance. Effect on the dissolved oxygen levels of water, sources of DO in water. Theory of Winkler's method to determine DO of water. Biological oxygen demand (BOD), introduction to BOD, determination of BOD. Chemical oxygen demand (COD), determination of COD. Sewage and its characteristics–sewage treatment, primary sewage treatment, secondary sewage treatment, tertiary sewage treatment. Potable water–purification of water for domestic purpose, water softening for domestic and industrial purpose. Cold Lime-soda process, hot lime process, Calgon process, zeolite process and ion exchange process. Boiler–boiler feed water. Water purification by other methods, flash evaporation of water, water purification by reverse osmosis, water filtration by electro-dialysis, what is electro-dialysis? electro-dialysis unit. Ion permeable membranes. Hazardous chemicals, characteristics of hazardous wastes.

8.1 Introduction

Water is one of the most basic and essential components for all lives. Yet, we find only 1 percent of the earth's water is available as fresh water. Water is indisputably the most essential resource the earth has to offer to the human race.

The distribution of water on earth is as follows:

Oceans:	**97.23 percent**
Ice caps and glaciers:	2.14 percent
groundwater:	0.61 percent
freshwater lakes:	0.01 percent
other:	0.01 percent

This negligible small amount of fresh water is not sufficient to meet the demands of people on earth, particularly in certain geographic areas of the world. Water is an essential resource for living systems, industrial processes, agricultural production and domestic use. Domestic water supplies often contain dissolved minerals, gases, bad smells and unsightly appearances that are objectionable and unhealthy to the consumers.

Water exists in nature in many forms such as clouds, rain, snow, ice and fog and we know that *all lives are dependent on water*; and chemically pure water does not exist for any appreciable length of time in nature (Fig. 8.1). During rainy seasons, water picks up small amount of gases, ions, dust and particulate matter from the atmosphere. Then, as it flows over or through the surface layers of the earth, it dissolves and carries with it some of almost everything it touches, including that which is dumped into it by man.

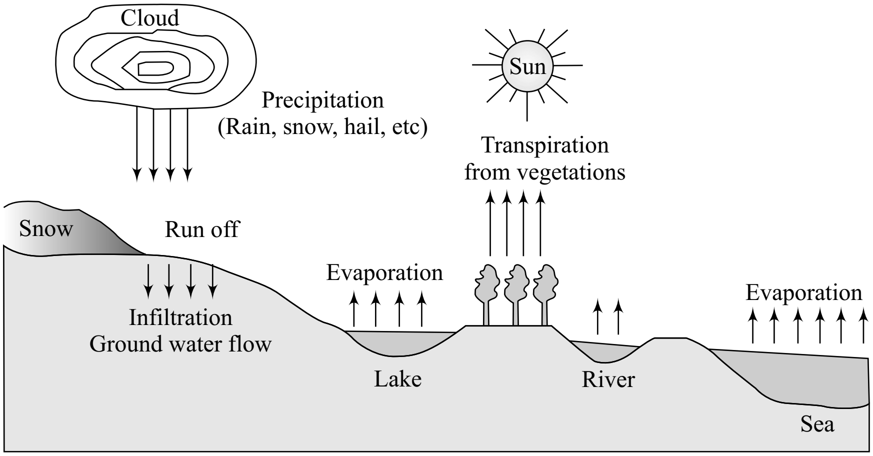


FIGURE 8.1 *Different sources of water.*

These contaminants may be classified as *biological, chemical* (both inorganic and organic)*, physical* and *radiological impurities*. They include industrial and commercial solvents, metal and acid salts, sediments, pesticides, herbicides, plant nutrients, radioactive materials, road salts, decaying animal and vegetable matter and living microorganisms, such as algae, bacteria and viruses. These impurities may give water a bad taste, colour, odour or cloudy appearance (turbidity) and cause hardness, corrosiveness, staining or frothing shown in Fig. 8.2. As a result, they may damage growing plants and transmit disease. Many of these impurities are removed or rendered harmless, however, in municipal drinking water treatment plants.

FIGURE 8.2 *A stream flowing with contaminants.*

8.2 Sources of Drinking Water

(i) Deep groundwater

The existence of deep groundwater table underneath is due to the rainfall for many decades or even hundred years ago. *Soil and rock* layers naturally filter the groundwater to a high degree of clarity before it is pumped to the treatment plant. Such water may emerge as springs or may be extracted from boreholes or wells. Deep groundwater is generally of very high bacteriological quality (i.e. a low concentration of pathogenic bacteria, pathogenic protozoa) but may be rich in dissolved solids, especially carbonates and sulphates of *calcium* and *magnesium*. There may be a requirement to reduce the iron or manganese content of this water to make it pleasant for drinking, cooking and laundry use. Disinfection is also required.

(ii) Shallow groundwaters

Water is also drawn from shallow groundwater such as bore wells or wells. The bacteriological quality can be variable depending on the nature of the catchments. A variety of soluble materials may be present including potentially toxic metals such as *zinc* and *copper*. Arsenic contamination of groundwater is a serious problem in some areas.

(iii) Upland lakes and reservoirs

Typically located in the headwater of river systems, upland reservoirs are usually sited above any human habitation and may be surrounded by a protective zone to restrict the opportunities for contamination. Bacteria and pathogen levels are usually low, but some bacteria, protozoa or algae will be present. Where uplands are forested or inorganic or organic matter can colour the water. Many upland sources have lower pH that requires adjustment.

(iv) Rivers, canals and low land reservoirs

Low land surface waters will have a significant bacterial load and may also contain algae, suspended solids and a variety of dissolved constituents.

Rain harvesting that collect water from the atmosphere can be used especially in areas with significant dry seasons and in areas experiencing fog even when there is little rain.

8.3 Impurities in Water

Water from an underground source generally has high total dissolved solids (TDS) and hardness level but a low organic content than the water samples obtained from other sources.

Seasonal variations in water quality are most apparent in surface waters. During the autumn and winter months, dead leaves and decaying plants release large quantities of organic matter into streams, lakes and reservoirs.

8.3.1 Contaminants in Raw Water

The water has a unique ability to dissolve any soluble matter that comes in contact with it. The major categories of impurities found in raw water include:

- Suspended particles, including colloids
- Dissolved inorganic salts
- Dissolved organic compounds
- Microorganisms
- Pyrogens
- Dissolved gases

(a) Suspended particles

Generally, water contains silt, pipe work debris and colloids as suspended particles. Colloidal particles (which can be organic or inorganic) present in water give rise to haze or turbidity.

(b) Turbidity

Turbidity is a measure of the relative clarity of water; the greater the turbidity the murkier is the water. Turbidity increases as a result of suspended solids in the water that reduces the transmission of light. Suspended solids are varied, ranging from clay and silt to industrial wastes and sewage.

With higher levels of turbidity, water loses its ability to support a diversity of aquatic organisms. Water becomes warmer as suspended particles absorb heat from the sunlight and cause oxygen levels to fall. Remember, warm water holds less oxygen than cool water.

The combination of warmer water, less light and oxygen depletion make it impossible for some form of aquatic life to survive. Suspended solid affects aquatic life in other ways as well. Suspended solid can clog fish gills, reduce growth rates, decrease resistance to disease and prevent egg and larval development.

(c) Dissolved inorganics

When water percolates through soil of the earth, there is every possibility of inorganic salts to get dissolved causing hardness of water. The bicarbonates of calcium and magnesium give rise to '*temporary hardness*', while the sulphates and chlorides cause '*permanent hardness*'.

The other inorganic impurities present in water include:

- Dissolved carbon dioxide/acidic carbonic acid
- Sodium salts
- Silicates leached from sandy river beds
- Ferrous and ferric iron compounds derived from minerals and rusty iron pipes

Chlorides from saline intrusion

- Aluminium from dosing chemicals and minerals
- Phosphates from detergents
- Nitrates from fertilizers

Environmental conditions such as drought, changing seasons, heavy rainfall, etc. can cause the concentration of dissolved salts in water to vary significantly. These dissolved salts (i.e. calcium, sodium, etc.) can directly affect plants health and over the time these render even the best soil useless.

(d) Water and Its pH

pH of water sample depends on the number of free hydrogen ions. A solution is more acidic when it contains more hydrogen ions. The level of acidity of the water is important for the plant and animal's life there. Most animals are adapted to living in neutral conditions. Changes in pH endanger the lives of the organisms in the water.

The *level of acidity* can be changed by human's actions. *Acid rain*, a result of air pollution and matter emitted from tailpipes and smokestacks, affects the pH. When these things combine with water in the atmosphere, they form sulphuric and nitric acids, then fall to the earth as acid rain, snow, hail and fog. This precipitation mixes with water already on the earth, in creeks, rivers, ponds and wetlands. Other pollutants carried by runoff from the land, also change the acidity of the water.

(e) Dissolved organics

Organic impurities in water arise from the decay of vegetable matter, mainly, organic acids, and from farming, paper making and domestic and industrial waste. These include detergents, fats, oils, solvents and residues from pesticides and herbicides. In addition, water-borne organics may include compounds leached from pipe work, tanks and purification media.

A water purification system can also be a source of impurities and so must be designed not only to remove contaminants from the feed water, but also to prevent additional recontamination from the system itself?

(f) Microorganisms

A wide variety of *microorganisms*, including *amoebae, bacteria, paramecia, rotifers*, *diatoms and algae* are found to be present on water surfaces. Since, most of the laboratory water comes from municipal water treatment plants, and is extensively treated to remove microorganisms, the chief microorganisms of concern for water purification systems are *bacteria*. A typical bacterial level for a potable laboratory water supply is *ten colony forming units* per millilitre (CFU/mL) or less. Bacteria are kept at these low levels by the use of residual levels of chlorine or other disinfectants. Once the disinfectants are removed during purification, bacteria have the chance to grow.

The challenges for an ultrapure water purification system are to

- remove the bacteria present in the feed water.
- prevent bacteria from entering the system and re-contaminating it.
- inhibit the growth of bacteria in the system.
- ensure that minimal bacteria are present in the product water

(g) Dissolved gases

Oxygen and carbon dioxide are the two gases that are most commonly found in natural waters. Carbon dioxide behaves as a weak anion and is removed by strongly basic anion exchange resins. Dissolved oxygen can also be removed by degassing or by anion exchange resins in the sulphite form, and the level of dissolved oxygen in the feed water can be monitored with oxygen specific electrodes.

(h) Total dissolved solids (TDS)

Total dissolved solids are the amount of particles dissolved in water. It may include all the suspended solids that may or may not pass through a filter. Dissolved solids may come from organic sources such as leaves, silt and industrial waste and sewage. Other sources come from runoff from the urban areas, road salts used on street and fertilizers and pesticides used in lawns and farms. Dissolved solids may come from inorganic materials such as rocks and air that may contain calcium bicarbonate, nitrogen iron, sulphur and other minerals. Rain water is almost pure with less than 10 mg/L of TDS.

A constant level of minerals in the water is necessary for aquatic life. Changes in the amounts of dissolved solids can be harmful because the density of total dissolved solids determines the flow of water in and out of an organism's cells. Many of these dissolved solids contain chemicals such as nitrogen, phosphorus and sulphur which are the building blocks of molecules for life. Concentration of dissolved solids that is too high or too low may limit the growth and may lead to the death of many aquatic organisms. High concentrations of total dissolved solids may reduce water clarity which contributes to a decrease in photosynthesis and to an increase in water temperature.

Levels of total dissolved solids:

- Drinking water may have a TDS reading of 25–250 mg/L. Drinking water should not exceed 500 mg/L TDS. Distilled water will have a TDS reading that will range from 0.5 to 1.5 mg/L.
- The amount of TDS ranges from 100 to 20,000 mg/L in rivers and may be higher in groundwater.
- Seawater may contain 3500 mg/L of TDS.
- Lakes and streams may have a TDS reading of 50–250 mg/L.

(i) The importance of temperature

The rates of biological and chemical processes depend on temperature. *Temperature affects the oxygen content of water* (oxygen levels become lower as temperature increases); the rate of photosynthesis by aquatic plants; the metabolic rates of aquatic organisms and the sensitivity of organisms to toxic wastes, parasites and diseases.

Oxygen content of water varies with weather, removal of shading stream bank vegetation, discharge of cooling water, urban storm water and groundwater inflows to the stream.

Thermal pollution is an increase in water temperature caused by adding relatively warm water to a body of water. Thermal pollution can come from storm water running off warmed urban surfaces (streets, sidewalks, parking slots) and industries that discharge.

(j) Salinity

Salinity is the concentration of dissolved salts in the water. Aquatic animals are adapted to live within certain salinity ranges. Sunfish, for example, common in ponds and freshwater streams, could not survive in salt water. Animals living in salty conditions tend to be more tolerant of a wider range of salinities. Many species of fish migrate from salt to fresh water in spawn. *Salinity* is affected by weather, especially at the surface of the water. During dry seasons, water evaporates, making water salty. When it rains, salty water is diluted by the added fresh water. Salinity is measured as a ratio of salts to water, and is expressed in parts per thousand (ppt), which means the number of units of salts per thousand units of water.

8.3.2 Water analysis

Determination of different constituents of raw water

A general chemical analysis of water will usually cover parameters such as:

(i) Hardness
(ii) Alkalinity
(iii) Chlorides
(iv) Nitrate
(v) Sulphate
(vi) Fluoride
(vii) Dissolved solids (TDS)
(viii) Dissolved gases

The following methods of analysis have been adopted for water analysis.

8.3.2.1 Total hardness of Water

Hardness measures the concentration of multivalent cations, particularly calcium and magnesium, present in a water sample. These ions precipitate easily and may form scale in hot water pipes, water heaters and appliances. They also react with soap to form difficult to remove scum. Hardness is not a water quality concern for natural water bodies. For most purposes, *hardness is defined as the sum of the calcium and magnesium concentrations; both expressed as calcium carbonate in mg/L (mg* $CaCO_3$*/L) or as equivalents/*m^3.

The hardness of water is of two types:

- Permanent, and
- Temporary

Based on the impact of hardness causing dissolved salts, the hardness of water can also be classified as follows:

(i) *Alkaline hardness:* The bicarbonates, carbonates and hydroxides of the dissolved salts of calcium and magnesium cause alkaline hardness.

(ii) *Nonalkaline hardness*: The nonalkaline (permanent) hardness of water is due to the effect of the sulphates, chlorides and nitrates of calcium and magnesium in water.

Permanent hardness is not destroyed by boiling and needs special treatment for removal of hardness. Depending on the need for the purpose, they are softened by a special technique.

(A) Cause for hardness

The cause for hardness of water is due to the presence of dissolved salts like chlorides, sulphates and bicarbonates of calcium and magnesium.

The presence of chlorides and sulphates of calcium and magnesium make water hard. This water is known as permanent hard water.

Hardness is that characteristic which prevents the lathering of soap. If such salts are present in water then water does not lather with soap solution. The reason is that these salts precipitate soap (sodium salt of stearic/palmitic acid) and thus prevent lathering.

$$CaCl_2 + 2Na\text{-}St \rightarrow Ca\text{-}St + 2NaCl$$

$$MgCl_2 + 2Na\text{-}St \rightarrow Mg\text{-}St + 2NaCl$$

$$MgSO_4 + 2Na\text{-}St \rightarrow MgSO_4 + Na_2SO_4$$

- *The presence of bicarbonates of calcium and magnesium gives temporary hard water* which readily get precipitated in boiling the water; and such temporary hard water can be soften easily.

Descriptions of hardness correspond roughly with ranges of mineral concentrations:

Soft: 0–20 mg/L as calcium
Moderately soft: 20–40 mg/L as calcium
Slightly hard: 40–60 mg/L as calcium
Moderately hard: 60–80 mg/L as calcium
Hard: 80:120 mg/L as calcium
Very Hard: >120 mg/L as calcium

(B) Unit of hardness

(i) Parts per million (ppm)*: It is defined as the parts of calcium carbonate equivalent hardness per 10^6 parts of water.*

That is 1 ppm = 1 part of calcium carbonate equivalent hardness in 10^6 parts of water.

(ii) Milligrams per litre (mg/L): *It is the number of milligrams of $CaCO_3$ equivalent hardness present per litre of water.*

Thus, 1 mg/L = 1 mg of $CaCO_3$ equivalent hardness in 1 L of water.
1 Litre = 1 kg = 1000 g = 1000×1000 mg = 10^6 mg
1 mg/L = 1 mg of $CaCO_3$ equivalent hardness in 10^6 mg of water or = 1 ppm.

(C) Disadvantages of hardness of water

The disadvantages of using hard water for domestic and industrial purposes are plenty.

(I) **For domestic use**

  (i) Washing—Presence of hardness causes wastage of soap, etc.
  (ii) Drinking—Excess calcium content in drinking water may be injurious to health.
  (iii) Cooking—It may affect the utensils if the water is hard. Effect the cooking process due to the presence of mineral contaminants.
  (iv) Bathing—Lack of foaming soap leads to uneasiness of bathing and wastage of soap.

(II) **For industrial use**

  (a) Paper—The presence of calcium and magnesium has impact on the properties and quality of paper and their products.
  (b) Textile—During dyeing process of cloths, the calcium and magnesium contaminants induce poor quality of shades.
  (c) Sugar—Causes hindrance to crystallization of sugar, if nitrates and sulphates of calcium and magnesium are present.
  (d) Boilers—Formation of scales corrodes the boilers. Wastages of fuel, etc.

(D) Determination of hardness of water by EDTA method

(i) Pipette out 50 mL of the given water sample into a 250 mL conical flask.
(ii) Add 3–5 mL of the (ammonia + ammonium chloride) buffer solution to elevate pH to approximately 10.0.
(iii) Add a tiny amount to dry eriochrome black (EBT) indicator. The solution turns wine red.
(iv) Titrate the sample of water *slowly* with 0.01 M EDTA until the last reddish tinge disappears from the solution.
(v) Add the last few drops at 3–5 second intervals to allow the end-point reaction to go to completion with the permanent blue colouration. Record the volume of EDTA used up at equivalence, V mL.

Example Calculate the hardness of the sample using: [Note: 1 L of water = 10^6 mg of water]

$$\text{The total hardness of water} = \frac{V \times B \times 1000 \text{ mg}}{\text{volume of water taken}} \ CaCO_3\text{/litre or ppm.}$$

where: V is the mL of the EDTA used in the titration 'B' is the mg $CaCO_3$ equivalent to 1 mL of EDTA solution $B \equiv$ 1 mL of 0.01 M EDTA solution $\equiv$ 1 mg $CaCO_3$

The total hardness of water can also be reported in terms of parts per million (ppm).

Worked Examples

1. The following chemicals are dissolved in a litre of water. Calculate the hardness of this sample of water.
$CaSO_4$ = 20.0 mg; $Mg(HCO_3)_2$ = 14.6 mg; $MgCl_2$ = 12.5 mg.
Calculation of $CaCO_3$ equivalent in the water sample:
Conversion factor = Molecular weight $CaCO_3$*/Mol weight of dissolved salt.*

	Conversion factor	$CaCO_3$ equivalent
1. $CaSO_4$	100/136	20.0 × 100/136 =14.70 mg/L
2. $Mg(HCO_3)_2$	100/146	14.6 × 100/146 = 10.0mg/L
3. $MgCl_2$	100/95	12.5 × 100/95 = 13.6 mg/L

Therefore, the total hardness due to dissolved salt = [14.70 + 10.0 + 13.6] = 38.3 mg/L
= 38.3 ppm

2. A water sample contains $Ca(HCO_3)_2$ = 22.4 mg/L; $Mg(HCO_3)_2$ = 19.2 mg/L; $CaSO_4$ = 23.5 mg/L. Calculate the temporary and permanent hardness.
Conversion factor = Mol weight $CaCO_3$*/Mol weight of dissolved salt.*

	Conversion factor	$CaCO_3$ equivalent
1. $Ca(HCO_3)_2$	100/162	22.4 × 100/162 = 13.83 mg/L
2. $Mg(HCO_3)_2$	100/142	19.2 × 100/142 = 13.5 mg/L
3. $CaSO_4$	100/136	23.5 × 100/136 =17.20 mg/L

Therefore, (i) Temporary hardness due to [$Ca(HCO_3)_2$ + Mg $(HCO_3)_2$ = 27.33 mg/L
(ii) Permanent hardness due to $CaSO_4$= 17.20 mg/L = 17.2 ppm.

3. A precipitate of 0.110 g of $CaC_2O_4.H_2O$ was obtained from 250 mL of water sample. Express the calcium content in the sample in ppm.
Gram molecular weight of $CaC_2O_4.H_2O$ is [40 + 24 + 64 + 18] = 146 g
If 146 g of [$CaC_2O_4.H_2O$] is dissolved in water and then it contains 40 g of Ca in solution.
Weight of calcium oxalate in 250 mL of water = 0.110 g
Therefore, weight of calcium oxalate in litre (L) = 0.110 × 4 = 0.440 g.

Now, 146 g of $CaC_2O_4.H_2O$ ≡ 40 g of Ca
0.440 g contains $0.44 \times 40/146 = 0.1206$ g/L of Ca.
Therefore, the calcium content in the sample = 120.6 mg/L or 120.6 ppm

4. The standard hard water contains 15 g of $CaCO_3$ per litre. 20 mL of this required 25 mL of EDTA solution. (i) 100 mL of sample of water required 18 mL of EDTA solution (ii) The same sample after boiling required 12 mL of EDTA solution. Calculate the temporary hardness of the given sample of water, in terms of ppm.

   A standard water contains 15 g of $CaCO_3$/L.
   1000 mL of standard hard water (SHW), therefore, contain 15,000 mg of $CaCO_3$.
   Therefore, 1 mL of standard water = 15,000/1000 = 15 mg of $CaCO_3$ equivalent.
   5 mL of EDTA solution = 20 mL of SHW contains [20 × 15] = 300 mg of $CaCO_3$ equivalent.
   Therefore, *1 mL of EDTA solution = 300/25 = 12 mg of* $CaCO_3$ *equivalent.*

   (i) **Calculation of Hardness of water**
   100 mL of water ≡ 18 mL of EDTA solution
   ≡ [18 × 12] = 216 mg of $CaCO_3$ equivalent.
   Therefore, 1000 mL of water ≡ 2160 mg of $CaCO_3$ equivalent.
   Hence, hardness of water = 2160 mg/L or ppm.

   (ii) **Calculation of permanent hardness of water**
   100 mL of water (after boiling) = 12 mL of EDTA solution.
   = [12 × 12] =144 mg of $CaCO_3$ equivalent.
   Therefore, 1000 mL of water contains 144 × 1000/100 =1440 mg of $CaCO_3$ equivalent.
   Hence, permanent hardness of water = 1440 ppm.

   (iii) **Calculation of temporary hardness of water**
   Temporary hardness of water = [Total – Permanent] hardness
   = [2160 – 1440] = 720 ppm.

5. 50 mL of a standard hard water containing 1 mg of pure $CaCO_3$ per 1 mL, consumed 20 mL of EDTA. 50 mL of a water sample consumed 25 mL of EDTA solution using EBT indicator. Calculate the total hardness of water sample.

   1 mL of SHW contains 1 mg of $CaCO_3$
   50 mL of standard hard water (SHW) contains 50 mg of $CaCO_3$
   20 mL of EDTA solution = 50 mL of SHW ≡ 50 mg of $CaCO_3$ equivalent.
   1 mL of EDTA solution ≡ 50/20 = 2.25 mg of $CaCO_3$ equivalent.
   Therefore, 25 mL of EDTA solution ≡ 2.25 × 25 = 56.25 mg of $CaCO_3$ equivalent.
   Hence, 50 mL of water contains 56.25 mg of $CaCO_3$ equivalent.
   Therefore, 1000 mL of water ≡ 1000 × 56.25/50 = 1,125 mg/L
   Hence, total hardness of water = 1125 ppm.

6. A water sample contains the following:

   $Mg(HCO_3)_2$: 95 mg/L
   $Ca(HCO_3)_2$:203 mg/L
   $CaSO_4$: 136 mg/L
   $CaCl_2$: 111 mg/L
   $MgCl_2$: 95 mg/L

   Calculate the temporary and permanent hardness of water.

Salt	mg/L	C. factor	$CaCO_3$ equivalent
$Mg(HCO_3)_2$	95	100/146	$95 \times 100/146 = 65.0$ mg/L
$Ca(HCO_3)_2$	203	100/162	$205 \times 100/162 = 125$ g/L
$CaSO_4$	136	100/136	$136 \times 100/136 = 100$ mg/L
$CaCl_2$	150	100/111	$150 \times 100/111 = 135$ mg/L
$MgCl_2$	95	100/95	$95 \times 100/95 = 100$ mg/L

$$CaCO_3 \text{ equivalent of salt} = \frac{\text{weight of salt (mg/L)} \times 100}{\text{Mol Weight of Salt}}$$

1. Temporary hardness of water = $[Mg(HCO_3)_2 + Ca(HCO_3)_2]$
   = [65.0 + 125]
   = 190.00 mg/L
2. Permanent hardness of water = $[CaSO_4 + CaCl_2 + MgCl_2]$
   = [100 + 135 + 100]
   = 335.0 mg/L (ppm)

7. An analysed sample of water containing only $Ca(HCO_3)_2$ is found to contain 150.00 mg/L of hardness. Calculate the amount of $Ca(HCO_3)_2$ dissolved in that water sample.

   (Ca = 40, H = 1, C = 12 and O = 16)

   M.F of $Ca(HCO_3)_2 = 40 + (1 + 12 + 48)_2 = 162$
   Therefore, 100 mg/L of $CaCO_3$ equivalent =162 mg/L of calcium bicarbonate

   $$150 \text{ mg/L of } CaCO_3 \text{ equivalent} = \frac{150 \times 162}{100}$$

   Therefore, amount of $Ca(HCO_3)_2$ = 243.0 mg/L (ppm)

8. A sample of hard water contains the following dissolved salts:
   $MgSO_4$ = 10.0 ppm
   $CaCl_2$ = 85.0 ppm
   $Ca(HCO_3)_2$ = 162.0 ppm
   $Mg(HCO_3)_2$ = 73.0 ppm
   $CaSO_4$ = 68.0 ppm.
   Find out the temporary and permanent hardness of water in ppm?

Salt	Quantity, ppm	C factor	$CaCO_3$ equivalent
$MgSO_4$	10.0	100/120	$10 \times 100/120 = 8.34$
$CaCl_2$	85.0	100/111	$85 \times 100/111 = 76.5$
$Ca(HCO_3)_2$	162.0	100/162	$162 \times 100/162 = 100$
$Mg(HCO_3)_2$	73.0	100/146	$73 \times 100/146 = 50$
$CaSO_4$	68.0	100/136	$68 \times 100/136 = 50$

Note: 1mg/L = 1 ppm.
Temporary hardness of water = $[Ca(HCO_3)_2 + Mg(HCO_3)_2]$
= {100 + 50} = 150 ppm
Permanent hardness = $[MgSO_4 + CaCl_2 + CaSO_4]$
= {8.34 + 76.5 + 50} = 134.84 ppm
Total hardness of water = 150 + 134.84 = 284.84 ppm

(E) Removal of temporary hardness on boiling hard water

Temporary hardness in water can be easily removed by boiling. On boiling, calcium/magnesium bicarbonate decomposes to give calcium/magnesium carbonate which is insoluble in water. Therefore, it precipitates out.

$$\mathrm{Ca(HCO_3)_2 \rightarrow CaCO_3\downarrow + H_2O + CO_2}$$

By the addition of slaked lime (Clark's process)

In Clark's process, slaked lime, $\mathrm{Ca(OH)_2}$ is added to temporary hard water. Insoluble calcium carbonate precipitates out and no longer produce hardness.

$$\mathrm{Ca(HCO_3)_2 + \underset{\text{(Slaked lime)}}{Ca(OH)_2} \rightarrow CaCO_3\downarrow + 2H_2O}$$

The methods used to remove permanent hardness given in the next section can also be employed to remove the temporary hardness. However, the above methods cannot be used to remove the permanent hardness

8.3.2.2 Alkalinity

Alkalinity measures the acid-neutralizing capacity of a water sample. It is an aggregate property of the water sample and can be interpreted in terms of specific substances only when a complete chemical composition of the sample is also performed.

The alkalinity of surface water is primarily due to the carbonate, bicarbonate and hydroxide content and is often interpreted in terms of the concentrations of these constituents. Alkalinity may also include contributions from borates, phosphates, silicates or other bases if they are present.

Alkalinity is used to determine the suitability of water for irrigation, industrial use, raw water characterization and water and wastewater monitoring. Alkalinity is also important as an indicator of water body's ability to resist pH change with the addition of acid from an accidental spill or acid precipitation.

Determination

1. Accurately measure 100 mL of your sample into a 250 mL conical flask. Insert a bar magnet and place sample on magnetic stir plate.
2. Record initial pH of the sample. If the pH is above 8.3, add several drops of phenolphthalein indicator.
3. Titrate the sample with 0.02 N $\mathrm{H_2SO_4}$ or HCl until the pH end-point (colourless) is reached. Note the total volume (V_1 mL) of acid needed to reach the end-point. This is the *phenolphthalein alkalinity, P*.

   1000 mL of 1 N $\mathrm{CaCO_3} \equiv 50$ g = 50,000 mg of $\mathrm{CaCO_3}$/L

   1 mL of 1 N $\mathrm{CaCO_3}$ = 50000/1000 = 50 mg of $\mathrm{CaCO_3}$

   100 mL of water (V) $\equiv V_1$ mL of N/50 $\mathrm{H_2SO_4}$

   Amount of $\mathrm{CaCO_3}$ equivalent in V_1 mL N/50 $\mathrm{H_2SO_4}$

   $$V_1 \times N/50\ \mathrm{H_2SO_4} \times 50 \text{ mg } \mathrm{CaCO_3}$$

   100 mL of water (V) = $V_1 \times N/50\ \mathrm{H_2SO_4} \times 50$ mg $\mathrm{CaCO_3}$ equivalent.

   $$\text{Therefore, 1000 mL (L) of water} = \frac{V_1 \times N/50\ \mathrm{H_2SO_4} \times 50 \times 1000 \text{ mg}}{\text{Volume of water } (V)}\ \mathrm{CaCO_3} \text{ equivalent.}$$

   In general,

   $$P = \frac{V_1 \times N \times 50{,}000 \text{ mg}}{\text{volume of water } (V)} \mathrm{CaCO_3/L}$$

   where V_1 is the volume in mL of the standard acid used N is the normality of the standard acid used [50 × 1000] = 50,000 is a conversion factor to change the normality into units of mg $\mathrm{CaCO_3}$/L.

4. If pH is below 8.3, add several drops of bromocresol green indicator. Titrate the water sample with 0.02 N H_2SO_4 or HCl until the pH 4.5 end-point (colour changes from blue to yellow) is reached. Note the volume (V_2) of acid needed to reach the endpoint. This gives the total alkalinity, T, of the sample.
5. The total alkalinity of the sample can be calculated using:

$$\text{Total alkalinity, } T = \frac{V \times N \times 50{,}000 \text{ mg}}{\text{volume of water } (V)} \text{ mg CaCO}_3\text{/L}$$

where $V = (V_1 + V_2)$, is the total volume in mL of the standard acid used. N is the normality of the standard acid used. 50,000 is a conversion factor to change the normality into units of mg $CaCO_3$/L. Calculation of alkalinity of water: (Table 8.1).
(Assuming all of the alkalinity is due to carbonate, bicarbonate or hydroxide)

Result of titration	OH^- alkalinity as $CaCO_3$	CO_3^- alkalinity as $CaCO_3$	HCO_3^- alkalinity as $CaCO_3$
$P = 0$	0	0	T
$P = ½T$	0	$2P$	0
$P > ½T$	$2P–T$	$2(T – P)$	0
$P < ½T$	$2T$	0	$(T – 2P)$
$P = T$	T	0	0

Table 8.1

Types of alkalinity in water

where P = phenolphthalein alkalinity T = total alkalinity

9. 100 mL of a water sample required 20 mL of N/50 H_2SO_4 for neutralization to phenolphthalein end-point. After this, methyl orange indicator was added to this and further acid required was again 20 mL. Calculate the alkalinity of water as $CaCO_3$ in ppm.
Volume of N/50 H_2SO_4 required to neutralize100 mL water to phenolphthalein end-point V_a = 20 mL.

$$P = \frac{V_1 \times N \times 50{,}000 \text{ mg}}{\text{volume of water } (V)} \text{ mg CaCO}_3\text{/L}$$

where V_1 is the volume in mL of the standard acid used N is the normality of the standard acid used [50 × 1000] = 50, 000 is a conversion factor to change the normality into units of mg $CaCO_3$/L.

$$\textit{Phenolphthalein Alkalinity, } P = \frac{20 \times 0.02 \times 50{,}000}{100} \text{ mg CaCO}_3\text{/L}$$
$$= 200 \text{ mg/L.}$$

Now 100 mL of water up to methyl orange end-point = 20 + 20 = 40 mL N/50 H_2SO_4

$$\textit{Total alkalinity, } T = \frac{[20 + 20] \times 0.02 \times 50{,}000}{100} \text{ mg CaCO}_3\text{/L}$$
$$= 400 \text{ mg/L.}$$

Now, P (= 200 mg/L) = ½ M [½ × 400], so the alkalinity is only due to CO_3^{2-}.

(a) $CO_3^{2-} + H^+ \rightarrow HCO_3^- \cdot P$

(b) $HCO_3^- + H^+ \rightarrow H_2O + CO_2]T$

Hence, total alkalinity (due to CO_3^{2-} only), T, is 400 mg/L

10. 100 mL of water sample required 20 mL of N/50 H_2SO_4 for neutralization to phenolphthalein end-point. After methyl orange addition to this, further acid required was 2.5 mL. Calculate the alkalinity of water in terms of $CaCO_3$ equivalent.

$$P = \frac{V_1 \times N \times 50{,}000}{\text{volume of water } (V)} \text{ mg CaCO}_3\text{/L}$$

where V_1, is the volume in mL of the standard acid used N is the normality of the standard acid used $[50 \times 1000] = 50{,}000$ is a conversion factor to change the normality into units of mg $CaCO_3$/L.

$$\text{Phenolphthalein alkalinity, } P = \frac{20 \times 0.02 \times 50{,}000}{100} \text{ mg CaCO}_3\text{/L}$$

$$= 200 \text{ mg/L.}$$

Now 100 mL of water up to methyl orange end-point in terms end methyl orange point.

$$= 20 \text{ mL} + 2.5 \text{ mL} = 22.5 \text{ mL of } N/50 \text{ H}_2\text{SO}_4$$

Total alkalinity of the sample can be calculated using:

$$\text{Total alkalinity, } T = \frac{V \times N \times 50{,}000 \text{ mg}}{\text{volume of water } (V)} \text{ mg CaCO}_3\text{/L}$$

where $V = (V_1 + V_2)$, is the total volume in mL of the standard acid used
N is the normality of the standard acid used
50,000 is a conversion factor to change the normality into units of mg $CaCO_3$

$$\text{Total alkalinity, } T = \frac{[20 + 2.5] \times N \times 50{,}000}{\text{volume of water } (V)} \text{ mg CaCO}_3\text{/L}$$

$$= \frac{22.5 \times 0.02 \times 50000}{100} = 225 \text{ mg/L.}$$

Now $P(= 200 \text{ mg/L}) > ½T\,[½ \times 225]$, so alkalinity is due to OH^- and CO_3^{-2}

(i) $OH^- + H^+ \rightarrow H_2O]P$

(ii) $CO_3^{-2} + H^+ \rightarrow HCO_3^-]T$

(iii) $HCO_3^- + H^+ \rightarrow HCO_3^{-i}$

Therefore, Alkalinity due to CO_3^{-2} ions $= 2[T - P] = 2[225 - 200] = 2 \times 25 = 50$ ppm.

And alkalinity due to OH^- ions $= [2P - T] = [2 \times 200 - 225] = 175$ ppm.

Therefore, Total alkalinity = (50 +175) = 225 ppm.

8.3.2.3 Chloride

Argentometric method

Take 100 cc of water in a conical flask and then it is neutralized with N/50H_2SO_4 using methyl orange as indicator. 2–3 drops of K_2CrO_4 indicator is added. The sample is titrated against N/50 silver nitrate solution till the colour changes from yellow to reddish brown.

The volume of $AgNO_3$ solution added is calculated as ppm of sodium chloride as $CaCO_3$ in the following manner:

$$Cl^- + AgNO_3 \rightarrow \underset{\text{(neutralized)}}{AgCl} + NO_3^-$$

$$\underset{\text{(few excess)}}{2Ag^+} + CrO_4^- \rightarrow \underset{\text{(reddish brown)}}{Ag_2CrO_4}$$

$$\text{Chloride (in ppm)} = \frac{V \times N \times 10^6}{1000 \times 50 \times 100}$$

where V is the no. of cc of N/50 $AgNO_3$ and N is the equivalent weight of $CaCO_3$.

8.3.2.4 Nitrate

Colorimetric method

Nitrates are the most oxidized forms of nitrogen and the end product of the aerobic decomposition of organic nitrogenous matter. The significant sources of nitrates are chemical fertilizers from cultivated lands, drainage from livestock feeds, as well as domestic and industrial sources. Natural waters in their unpolluted state contain only minute quantities of nitrates. The stimulation of plant growth by nitrates may result in eutrophication, especially due to algae. The subsequent death and decay of plants produces secondary pollution. Nitrates are most important for biological oxidation of nitrogenous organic matter. Certain nitrogen fixing bacteria and algae have the capacity to fix molecular nitrogen in nitrates. The main source of polluting nitrates is domestic sewage. Nitrates may find their way into groundwater through leaching from soil and at times by contamination.

Phenoldisulphonic method

Principle

Nitrates react with phenoldisulphonic acid and produce a nitrate derivative, which in alkaline solution develops yellow colour due to re-arrangement of its structure. The colour produced is directly proportional to the concentration of nitrates present in the sample.

Reagents

Phenol disulphonic acid: 25 g of phenol + 150 mL of concentrated sulphuric acid + 85 mL of sulphuric acid is further added + heated for about 90 minutes on a water bath and stored in dark bottles upon cooling.

Sodium hydroxide: About 50 g NaOH + 150 – 200 mL of water and cooled.

Conc. ammonium hydroxide

Nitrate solution: Water containing nitrate.

Stock nitrate solution: 721.8 mg AR potassium nitrate + distilled water and made up to 100 mL for stock solution with swirling.

Standard nitrate solution

Standard nitrate solution is prepared by evaporating 50 mL of the stock solution to dryness in the water bath. The obtained residue is dissolved in 2 mL of phenoldisulfonic acid and diluted to 500 mL, to give 1 mL= 10 µg. The solution of various strengths ranging from 0.0 (blank) to 1.0 mg/L at the intervals of 0.2 mg/L is prepared by diluting stock solution with distilled water.

Procedure

50 mL of the sample is pipetted into a porcelain dish and evaporated to dryness on a hot water bath. 2 mL of phenol disulphonic acid is added to dissolve the residue by constant stirring with a glass rod. Concentrated solution of sodium hydroxide or conc. ammonium hydroxide and distilled water is added with stirring to make it alkaline. This is filtered into a Nessler's tube and made up to 50 mL with distilled water. The absorbance is read at 410 nm using a spectrophotometer after the development of colour. The standard graph is plotted by taking concentration along the x-axis and the spectrophotometric readings (absorbance) along y-axis. The value of nitrate is found by comparing absorbance of sample with the standard curve and expressed in mg/L.

Calculation

$$\text{Nitrates (as mg/L)} = \frac{\text{Absorbance of sample} \times \text{Conc. of Std.} \times 1000}{\text{Absorbance of Std.} \times \text{Sample taken}}$$

The high concentration of nitrate in water is indicative of pollution.

8.3.2.5 Sulphate

Gravimetric method

Sulphate is a substance that occurs naturally in drinking water. Health concerns regarding sulphate in drinking water have been raised because of reports that diarrhoea may be associated with the ingestion of water containing high levels of sulphate. Of particular concern are groups within the general population that may be at greater risk from the laxative effects of sulphate when they experience an abrupt change from drinking water with low sulphate concentrations to drinking water with high sulphate concentrations. Sulphate in drinking water currently has a secondary maximum contaminant level of 250 milligrams per litre (mg/L), based on aesthetic effects (i.e. taste and odour).

The following reagents are prepared for the determination of dissolved sulphate in water

1. Barium Chloride solution

Dissolve 10 g of reagent grade barium dihydrate ($BaC1_2.2H_2O$) in water and dilute to 100 mL.

2. Hydrochloric acid

Concentrated (relative density 1.16 g/cm^3).

3. Bromine solution

A saturated solution of bromine in water.

Procedure

(i) Pipette out 200 mL of the filtered water into a 400 mL beaker. Add 3 mL of the bromine solution and 1 mL of the hydrochloric acid, mix and heat to boiling and boil until all the bromine has been driven off. Make up the volume of the solution to 200 mL by adding boiling water.

(ii) Add 10 mL of barium chloride solution drop by drop while stirring the contents of the beaker continuously. Boil the contents of the beaker for 2 minutes, then place the beaker on a water bath for at least 2 hours and stir the contents occasionally. Allow the precipitate to settle

$$SO_4^{2-} + BaCl_2 \rightarrow BaSO_4\downarrow + 2Cl^-$$

(iii) Filter off the precipitate, using Whatmans no. 42 and transfer the precipitate quantitatively onto the filter paper with the aid of a jet of warm water from a wash bottle. Wash the filter paper and precipitate with consecutive small volumes of warm water until the filtrate is free from chlorides.

(iv) Place the filter paper and precipitate in an ignited and tared platinum crucible, incinerate the paper slowly (without inflaming), then increase the temperature and maintain it at just under 600°C until all carbon has burnt off. Finally ignite at a temperature of 800–900°C for 30 minutes, cool the crucible in a desiccator to room temperature, and determine the mass of the residue (as $BaSO_4$).

(v) Carry out a blank determination by following the procedures described in (a)–(d) above using the same quantities of reagents but omitting the 25 g of test specimen, and correct the mass of the barium sulphate residue accordingly.

Calculation

Calculate, as follows, the water-soluble sulphates content of the fine aggregate (as relevant):

$$\text{Water-soluble sulphates content} = \frac{A \times 34.5 \times B}{5 \times 2}$$

where

A = mass of barium sulphate in 200 mL of the extract, as corrected, g
B = mass of the specimen taken for extraction, *g*

8.3.2.6 Fluoride

Colorimetric method

Fluorides have dual significance in water supplies. High concentration causes dental fluorosis and lower concentration (< 0.8 mg/L) causes dental caries. A fluoride concentration of approximately 1 mg/L in drinking water is recommended. They are frequently found in certain industrial processes resulting in fluoride rich waste waters. Significant sources of fluoride are found in coke, glass and ceramic, electronics, pesticide and fertiliser manufacturing, steel and aluminium processing and electroplating industries. It is calculated by SPADNS method.

Principle

The fluorides present in water reacts with zirconium SPADNS solution and forms a coloured 'lake'; which is greatly influenced by the acidity of the reaction mixture. Fluoride reacts with the dye 'lake', dissociating (bleaching) the dye into a colourless complex anion $(\mathrm{ZrF_6^{2-}})$. As the amount of fluoride increases, the colour produced becomes progressively higher or of different hue.

Reagents

Standard fluoride solution
Stock solution: Dissolve 221.0 mg of AR grade sodium fluoride in distilled water and make up to 1000 mL
Therefore, 1 mL = 100 µg of F^-
100 mL of the stock fluoride was diluted to 1000 mL to give 1 mL = 10 µg of fluoride
SPADNS Solution: Dissolve 958 mg of SPADNS in 500 mL of distilled water.
Zirconyl acid reagent: Dissolve 133 mg zirconyl chloride octahydrate ($ZrOCl_2.8H_2O$) in 25 mL of distilled water. Add 350 mL of conc. HCl and dilute to 500 mL with distilled water.
Zirconyl acid-SPADNS reagent: Mix equal volume of SPADNS and zirconyl acid reagent.

Procedure

A standard graph is prepared by using fluoride concentrations ranging from 0.005 to 0.150 mg/L at 570 nm. A reference solution is prepared by adding 4 mL of acid zirconyl-SPADNS reagent to 21 mL of distilled water. A known volume of filtered sample (21 mL) is taken in a test tube, 4 mL of acid zirconyl-SPADNS reagent is added to the sample along with a reference solution. The mixture is left for about 30 minutes for complete colour development and the optical density is read at 570 nm.

Calculation

$$\mathrm{F^-\,mg/L} = \frac{(\text{OD sample}) \times (\text{Conc. of the standard})}{(\text{OD standard}) \times (\text{sample taken})} \times (1000)$$

This method is applicable to the measurement of fluoride in drinking, surface and saline waters, domestic and industrial wastes. Concentration of fluoride from 0.1 up to 1000 mg/litre may be measured.

8.3.2.7 Total Solids

(a) The material residue left in the vessel after evaporation of the sample and its subsequent drying in an oven at a temperature of 103–105°C is referred to as the 'total solids'. Total solids include total suspended solids (TSS) and total dissolved solids (TDS).

Principle

50 mL of well-mixed sample is evaporated in a preweighed dish and dried to constant weight in an oven at 103–105°C. The increase in weight over that of the empty dish gives the total solids.

Procedure

50 mL of the well-mixed sample is measured into a preweighed dish and evaporated to dryness at 103°C in a steam bath. The evaporated sample is dried in an oven for about an hour at 103–105°C and cooled in a desiccators and recorded for constant weight.

Calculation

$$\text{Total solids (mg/L)} = \frac{(W_1 - W_2) \times (1000)}{\text{Sample volume (mL)}}$$

W_1 = Weight of dried residue + dish

W_2 = Weight of empty dish

(b) **Total suspended solids**

Suspended solids are the portions of solids that are retained on a filter of standard specified size (generally 2.0 μ) under specific conditions. Water with high-suspended solids is unsatisfactory for bathing, industrial and other purposes.

Principle

A well-mixed sample is filtered through a weighed standard glass fibre filter and the residue that is retained on the filter is dried to a constant weight at 103–105°C. The increase in the weight of the filter determines the total suspended solids.

Procedure

The known volume of vigorously shaken sample (50 mL) is filtered into a preweighed glass fibre filter disk fitted to suction pump, and washed successively with distilled water. The filter is carefully removed from the filtration apparatus and dried for an hour at 103–105°C in an oven, cooled in desiccators and weighed for constant weight.

Calculation:

$$\text{Total suspended solids} = \frac{(W_1 - W_2) \times (1000)}{\text{Sample volume (mL)}} \text{ (mg/L)}$$

W_1 = Weight of dried glass fibre filter + residue

W_2 = Weight of glass fibre filter disk before filtering

8.4 Dissolved Oxygen (DO) in Water and Its Importance

Oxygen is important to all lives. Dissolved oxygen is the form of oxygen in water that is freely available to aquatic plants and animals. Dissolved oxygen is vital to fish and other aquatic life and for the prevention of odours. Oxygen is transferred from the atmosphere into surface waters, as well as being produced by aquatic plants, algae and phytoplankton as a by-product of photosynthesis (Fig. 8.3). Once dissolved in water, oxygen diffuses throughout a

FIGURE 8.3 *A stream flowing with dissolved oxygen and pollutants.*

water body very slowly since distribution depends on the movement of aerated water by turbulence and currents, water flow and thermal upwelling.

Oxygen, in water, is measured as dissolved oxygen (D.O). One unit of *measure of dissolved oxygen* in water is parts per million (ppm) which is the *number of oxygen* (O_2) *molecules per million total molecules in a sample*. It is also defined as the number of moles of molecular oxygen (O_2) dissolved in a litre of water at a temperature, expressed as mg O_2/L (Fig. 8.4).

A high percentage of dissolved oxygen is conducive to supporting aquatic flora and fauna, such as algae, plants, fish, mollusks and invertebrates. A low percentage or depleted of dissolved oxygen indicates a negative impact on a body of water, which results in an abundance of worms and fly larvae. Aquatic life uses oxygen that is dissolved in the water and is in much smaller quantities than in the air. If more oxygen is consumed than is produced, dissolved oxygen levels decline and some sensitive animals may move away, weaken or die.

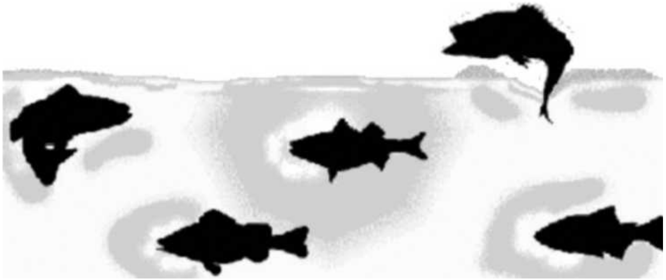


FIGURE 8.4 *Dissolved oxygen and aquatic life in water.*

One of the best indicators of the health of a body of water, such as a river, stream, lake or pond, ecosystem is the dissolved oxygen (D.O) parameter. Dissolved oxygen can range from 0 to 18 mg O_2/L. Most natural water systems require 5–6 mg O_2/L to support a diverse population. Oxygen enters the water either by direct absorption from the atmosphere or by plant photosynthesis. The oxygen is used by plants and animals for respiration and by the aerobic bacteria which consume oxygen during the process of decomposition. Algae growth increases when organic matter such as animal waste or improperly treated waste water enters a body of water causing the dissolved oxygen levels to decrease as the plant material dies off and is decomposed through the action of the aerobic bacteria.

8.4.1 Dissolved Oxygen Levels are affected by

1. Water temperature: when temperature decreases, more oxygen dissolves in water; when temperature increases, dissolved oxygen decreases.
2. Aquatic plant populations
3. Stream flow
4. Altitude/atmospheric pressure
5. Human activities
6. Water discharge: when water discharge (flow) increases, dissolved oxygen increases.
7. Organic waste (such as leaves, sewage, industrial waste, agricultural/fertilizer runoff and faecal matter) build up when it increases and dissolved oxygen decreases (Fig. 8.5).

FIGURE 8.5 *A completely polluted lake.*

Temperature is important for the ability of oxygen to dissolve and has different solubilities at different temperatures.

Degrees C	mg/L	Degrees C	mg/L
0	14.6	16	10.0
1	14.2	17	9.8
2	13.8	18	9.6
3	13.5	19	9.4
4	13.1	20	9.2
5	12.8	21	9.0
6	12.5	22	8.9
7	12.2	23	8.7
8	11.0	24	8.6
9	11.6	25	8.4
10	11.3	26	8.2
11	11.1	27	8.1
12	10.9	28	7.9
13	10.6	29	7.8
14	10.4	30	7.7
15	10.2	31	7.4

Weather, temperature and salinity affect amounts of dissolved oxygen (DO). Cold and fresh water holds more oxygen than warm or salty water. If DO levels are severely low, large quantities of fish may die (Fig. 8.6).

Low DO is also caused by fertilizer and manure runoff from streets, lawns and farms. Fertilizers and faecal matter encourage the growth of too much algae which uses up the oxygen quickly. Plants and animals die and are decayed by bacteria which also uses up a great deal of oxygen.

FIGURE 8.6 *Fishes die due to low level DO in water.*

The stream system gains oxygen from the atmosphere and from aquatic plants as a result of photosynthesis. Running water, because of its churning, dissolves more oxygen than still water. Respiration by aquatic animals, decomposition and various other chemical reactions consume oxygen. DO levels fluctuate seasonally and over a 24-hour period. They vary with water temperature and altitude. Cold water holds more oxygen than warm water (see table) and water holds less oxygen at higher altitudes. Thermal discharges, such as water used to cool machinery in a manufacturing plant or a power plant, raise the temperature of water and lower its oxygen content. Aquatic animals are most vulnerable to lowered DO levels in the early morning on hot summer days when stream flows are low, water temperatures are high and aquatic plants have not been producing oxygen since sunset.

8.4.2 Sources of DO in Water

It is dissolved by aeration as water moves over rocks and debris, aeration from wind and waves and photosynthesis of aquatic plants. When the dissolved oxygen levels decrease, it affects the number and types of aquatic macro-invertebrates which live in a water ecosystem. Species which cannot tolerate decrease in dissolved oxygen levels include mayfly nymphs, stonefly nymphs, caddisfly larvae and beetle larvae. As the dissolved oxygen levels decrease, these pollution-intolerant organisms are replaced by the pollution-tolerant worms and fly larvae.

Dissolved oxygen concentrations arise from the interaction between:

- oxygen produced by photosynthesis;
- oxygen consumed by aerobic respiration, nitrification and chemical oxidation within the water environment; and
- the exchange of oxygen with the atmosphere.

Natural processes (e.g. weather, tides and currents) and human pollution (*particularly organic matter*) can result in severe reductions in dissolved oxygen levels. Both anoxia (*no oxygen)* and hypoxia (*very low oxygen*) are harmful to most marine animals causing:

- animal kill,
- a decrease in the available habitat, and
- limiting animal movements.

Low dissolved oxygen levels can also result in reduction conditions within the sediments which may cause previously bound nutrients and toxicants to be released into the water column. Observed decrease in dissolved oxygen of individual water bodies is mainly related to increased organic matter load (e.g. from sewage treatment plants and industry, organic runoff or algal blooms) which leads to increased bacterial activity (decomposition by aerobic microorganisms). This increase in activity results in increased oxygen consumption and can deplete available oxygen. Low oxygen levels generally affect bottom waters first and most severely.

Dissolved oxygen levels change and vary according to the time of the day, weather and temperature. Large fluctuations in dissolved oxygen levels over a short period of time may be the result of an algal bloom. While the algae population is growing at a fast rate dissolved oxygen levels increase. Soon the algae begin to die and are decomposed by aerobic bacteria that use up the oxygen. As a greater number of algae die, the oxygen requirement of the aerobic decomposers increases, resulting in a sharp drop in dissolved oxygen levels. Following an algal bloom, oxygen levels can be so low that fish and other aquatic organisms suffocate and die.

8.4.3 Theory of Winkler's Method for DO Determination

The Winkler method (1888), otherwise known as the iodometric technique, is the most precise and reliable titrimetric procedure for DO analysis. The chemistry of this test is based on the addition of a manganese solution followed up by a strong alkali solution. The DO present rapidly forms hydroxide salts with the manganese. The colour of the precipitate formed is an initial indicator of how much DO is present. In the presence of iodide ions in an acidic solution, the oxidized manganese reverts back to the divalent state and the liberated iodine is equivalent to the original DO content. This iodine is then titrated with a standard thiosulphate solution.

The redox chemistry of Winkler's method is as follows:

Manganous sulphate reacts with the potassium hydroxide-potassium iodide to produce a white flocculent precipitate of manganous hydroxide:

$$MnSO_4 + 2KOH \rightarrow Mn(OH)_2 + K_2SO_4 \text{ (white)}$$

If there is any DO in the water, a second reaction between the $Mn(OH)_2$ and DO occurs immediately to form a brownish manganic oxide precipitate.

$$2Mn(OH)_2 + O_2 \rightarrow 2MnO(OH)_2 \text{ (brown)}$$

When the samples are ready to be titrated, requisite amount of H_2SO_4 (1:1 sulphuric acid) is added to each sample and the bottles are inverted several times in order to completely re-dissolve the brownish precipitate. Manganic sulphate ($Mn(SO_4)_2$) is yielded as the product of this reaction:

$$2MnO(OH)_2 + 2H_2SO_4 \rightarrow 2Mn(SO_4)_2 + 6H_2O$$

The $Mn(SO_4)_2$ immediately reacts with the potassium iodide (KI) added, liberating the number of moles of iodine exactly equivalent to the number of moles of oxygen present in the sample. The release of iodine (I_2) imparts a brown colouration to the water typical of iodine.

$$2Mn(SO_4)_2 + 4KI \rightarrow 2MnSO_4 + 2K_2SO_4 + 2I_2$$

Sodium thiosulphate ($Na_2S_2O_3$) then reacts with the released I_2 to give an accurate estimate of the DO in the original sample. The sample is then titrated to a pale, straw colour before the starch indicator is added. This will change the solution blue. The titration end-point is when the blue colour disappears.

Sources of error

As mentioned previously, the sources of error listed below are only found infrequently in sea water settings.

1. If nitrites are present in large quantities, high values for oxygen may be obtained because of the following chemical reaction:

$$2NO_2^- + 2I^- + 4H^+ \rightarrow 2NO + I_2 + 2H_2O$$

2. If hydrogen sulphide is present, the Winkler method for determining oxygen is not applicable. Hydrogen sulphide will react with the dissolved oxygen and with iodine.

For example, one possible reaction is:

$$2H_2S + O_2 \rightarrow 2H_2O + 2S\downarrow$$

The sulphur formed is almost colloidal in nature and such waters will be devoid of oxygen, except for the possibility of an interface zone where $O_2 + H_2S$ may be present. The hydrogen sulphide reacts with iodine and, if present in sufficient quantities, may be determined iodometrically.

Calculation of DO

1 mole of O_2 = 2 moles of I_2 = 4 moles of $Na_2S_2O_3$
1 mole of $Na_2S_2O_3 \equiv 1/4^{th}$ mole of $O_2 \equiv 8$ g of O_2
1 mL of 1 N $Na_2S_2O_3 \equiv 8$ mg O_2
'V' mL of 0.005 N $Na_2S_2O_3 \equiv 8 \times V \times 0.005N$
Therefore,

$$\text{DO of water sample} \equiv \frac{8 \times V \times 0.005N \times 1000}{25} \text{ at } t°\text{C}$$

8.5 Biological Oxygen Demand (BOD)

8.5.1 Introduction

The BOD is an important measure of water quality. *It is a measure of the amount of oxygen needed (in milligrams per litre or ppm) by bacteria and other microorganisms to oxidize the organic matter present in a water sample over a period of 5 days*. The BOD of drinking water should be less than 1. That of raw sewage may run to several hundreds. It is also called the 'biological' oxygen demand.

When organic matter decomposes, microorganisms (such as bacteria and fungi) feed upon this decaying material and eventually it becomes oxidized. Microorganisms such as bacteria are responsible for decomposing organic waste. When organic matter such as dead plants, leaves, grass clippings, manure, sewage or even food waste is present in water supply, the bacteria will begin the process of breaking down this waste. When this happens, much of the available dissolved oxygen is consumed by aerobic bacteria, robbing other aquatic organisms of the oxygen they need to live.

BOD directly affects the amount of dissolved oxygen in rivers and streams. The more rapidly oxygen is depleted in the stream, the greater the BOD. This means less oxygen is available. The detriment of high BOD is same as low dissolved oxygen; aquatic organisms become stressed, suffocate and die.

Sources of BOD include leaves and woody debris; dead plants and animals; animal manure; effluents from pulp and paper mills, wastewater treatment plants, feedlots and food-processing plants; failing septic systems and urban storm water runoff.

Biological oxygen demand (BOD) of a sewage is defined as the number of milligrams of dissolved oxygen required for the oxidation of biologically oxidizable impurities present in 1000 mL of waste water under aerobic conditions at 20°C over a period of 5 days.

Nitrates and *phosphates* in a body of water can contribute to high BOD levels. Nitrates and phosphates are plant nutrients and can cause plant life and algae to grow quickly.

When plants grow quickly, they also die quickly. This contributes to the organic waste in the water, which is then decomposed by bacteria. This results in a high BOD level. The *temperature* of the water can also contribute to high BOD levels. For example, warmer water usually will have a higher BOD level than colder water. As water temperature increases, the rate of photosynthesis by algae and other plant life in the water also increases. When this happens, plants grow faster and also die faster. When the plants die, they fall to the bottom where they are decomposed by bacteria. The bacteria require oxygen for this process so the BOD is high at this location. Therefore, increased water temperatures will speed up bacterial decomposition and result in higher BOD levels.

8.5.2 Determination of BOD

The 5-day biochemical oxygen demand (BOD_5) assay is a standardized assessment of the amount of oxidizable, or respirable, organic matter in water. *It is the amount of oxygen (in* mg/L*) required by bacteria to oxidize the organic molecules aerobically.* Oxygen consumed by inorganic compounds is also measured by this test which is why it is referred to as biochemical oxygen rather than just biological oxygen demand. It is used as an index of the amount of organic pollution of the water and is routinely employed to measure the efficiency of wastewater treatment plants in removing organic matter from wastewater or in assessing the effect of effluents on the tropic status of natural waters.

Any wastes, such as municipal sewage, abattoir discharge and some industrial wastes with high BOD, must be treated to remove or lower the BOD before release into the receiving water. This removal of BOD by the wastewater treatment plant is inevitably biological but may be either aerobic or anaerobic.

1. Overview

A sample of the water of interest is placed in a BOD bottle and its dissolved oxygen (DO) concentration is measured. The bottle is sealed and incubated in the dark for 5 days at 20°C. At the end of the incubation time the DO is measured again. The drop in DO over the incubation period is the amount of oxygen used by bacteria to oxidize the organic material in the sample. BOD is high in organically polluted waters and low in pristine water. It is very high in wastewater, some industrial effluents and slaughterhouse wastes.

2. Dilution

Waters with high concentrations of organic material will not contain sufficient dissolved oxygen for complete oxidation. In this case an undiluted sample will become anaerobic sometime during the five-day incubation period and no useful information will be gained from the procedure. It is necessary to dilute such samples so that there will be sufficient oxygen present to oxidize all the organic matter, with a little oxygen left over. The amount of dilution required by any given sample must be determined by trial and error method or past experience. It is customary to prepare multiple dilutions, usually three, of each sample to assure finding one that is usable.

Although the best dilutions to use must be determined by trial and error, the following guidelines can be used to approximate the correct dilution.

Sample	Dilution percent
Strong industrial wastes	< 1%
Raw and settled wastewater	1–5%
Biologically treated effluent	5–25%
Polluted receiving waters	25–100%

3. Dilution water

Bacterial growth requires inorganic nutrients as well as organic compounds and these must be added to the dilution water to ensure their presence. The dilution water is also buffered to ensure a pH suitable for bacterial growth.

4. Seeding

After dilution, the sample must be seeded with a bacterial source to ensure the presence of bacteria capable of degrading the organic material in the sample. Non-chlorinated treatment plant secondary effluent or raw sewage may be used for this but it is preferable to use commercial preparations of lyophilized bacteria such as polyseed®.

5. Calculation

Calculate the five-day BOD using the simplified expression

$$\text{BOD} = \frac{D_1 - D_2}{P}$$

where D_1 = initial DO, D_2 = final DO and P = decimal volumetric fraction of sample used, e.g. a 75 mL sample diluted to 300 mL would be a 25 percent solution and $P = 0.25$.

8.5.3 BOD by Winkler's Method

Principle

Winkler's method is based on the fact that in alkaline medium, DO oxidizes Mn^{2+} to Mn^{4+} which in acidic medium oxidizes I^- to free iodine. The amount of iodine released which can be titrated with a standard solution of sodium thiosulphate, is thus equivalent to the DO originally present.

The following reactions take place:

Manganous sulphate reacts with the potassium hydroxide-potassium iodide (alkaline-iodide) to produce a white flocculent precipitate of manganous hydroxide:

$$MnSO_4 + 2KOH \rightarrow Mn(OH)_2 + K_2SO_4 \text{ (white)}$$

DO in the water reacts with Mn(OH)$_2$ immediately to form a brownish manganic oxide flocculate

$$2Mn(OH)_2 + O_2 \rightarrow 2MnO(OH)_2 \text{ (brown)}$$

If the precipitate is white there is no DO.

Manganic oxide reacts with added H_2SO_4 to give manganic sulphate ($Mn(SO_4)_2$) as the product of this reaction:

$$2MnO(OH)_2 + H_2SO_4 \rightarrow 2Mn(SO_4)_2 + 6H_2O$$

$Mn(SO_4)_2$ immediately reacts with the potassium iodide (KI) (added initially as part of alkali iodide), liberating the iodine exactly equivalent to the number of moles of oxygen present in the sample. The release of iodine (I_2) imparts a brown colouration to the water typical of iodine.

$$2Mn(SO_4)_2 + 4KI \rightarrow 2MnSO_4 + 2K_2SO_4 + 2I_2$$

The liberated iodine is titrated with thiosulphate and the reaction is:

$$2S_2O_3^{2-} + I_2 \rightarrow S_4O_6^{2-} + 2I^-$$

From the above stoichiometric equations, we can find that:

$$\textit{1 mole of } O_2 = \textit{2 moles of } MnO(OH)_2 = \textit{2 moles of } I_2 \equiv 4S_2O_3^{2}$$

Therefore, after determining the number of moles of iodine produced we can determine the number of moles of oxygen molecules present in the water sample. Dissolved oxygen concentrations are generally expressed in mg O_2/L.

From the titre values BOD is calculated.

Procedure

A known volume of the sewage water is diluted to a known volume with fresh water. Equal quantities of the diluted water are taken in two BOD bottles. In the first bottle, BOD is determined immediately as follows:

To a known volume of diluted water, 5 mL of $MnSO_4$ solution and 5 mL of alkaline KI are added and shaken for about 15 minutes till MnO_2H_2O gets precipitated. Then precipitate is dissolved in 2 mL of 1:1 H_2SO_4 and the liberated I_2 is titrated against standard sodium thiosulphate. The titre value is called blank value and it indicates the total dissolved oxygen available at the start of the experiment.

The water in the second bottle is incubated under aerobic conditions for 5 days and BOD is determined as described above. The litre value indicates dissolved oxygen present after 5 days.

(I) When dilution water is not seeded:

$$BOD_5, \text{ mg/L} = \frac{D_1 - D_2}{P}$$

where

D_1 = DO of diluted sample immediately after preparation, mg/L

D_2 = DO of diluted sample after 5 days incubation at 20°C, mg/L

P = decimal volumetric fraction of sample used

(II) Alternative calculation:

Volume of sewage water taken = v_1 mL

Volume of water after dilution = V mL

Volume of diluted water taken for titration = V_2 mL

Volume of this required for blank = A mL

Volume of thiosulphate required after 5 days = B mL

Dissolved oxygen available at the beginning = A mL of $Na_2S_2O_3$

Dissolved oxygen available after 5 days = 'B' mL of $Na_2S_2O_3$ solution.
Dissolved oxygen used up for biological oxidation in 5 days = (A–B) mL of $Na_2S_2O_3$
Normality of $Na_2S_2O_3 = N$
1000 mL of I N $Na_2S_2O_3$ = 8 g of oxygen
= 8000 mg of oxygen
1 mL of I N $Na_2S_2O_3$ = 8 mg of dissolved oxygen
(A–B) mL × 'x' N $Na_2S_2O_3 = (A - B) \times N \times 8$ mg oxygen
V_2 mL of diluted wastewater requires $(A - B) \times N \times 8$ mg of oxygen

'V' mL of diluted wastewater requires $= \dfrac{(A-B)\times N\times 8\times V}{v_2}$ mg of oxygen

BOD of v_1 mL of wastewater $= \dfrac{(A-B)\times N\times 8\times V}{v_2}$

Therefore, BOD of 1000 mL of waste water $= \dfrac{(A-B)\times N\times 8\times v_2}{V}\times\dfrac{1000}{v_1}$ mg of oxygen/L.

where A = Blank titre value B = titre value after 5 days v_1 is the volume of waste water taken v_2 is the volume after dilution and V is the volume taken for titration.

11. 25 mL of waste was diluted to 500 mL and equal volumes are filled in two BOD bottles. In the blank titration 100 mL of diluted wastewater when titrated immediately required 6.1 mL of 0.02 N $Na_2S_2O_3$. 100 mL of the incubated sample after 5 days required 3.6 mL of same $Na_2S_2O_3$. Calculate BOD of the wastewater.

Solution:
Method I:
Volume of waste water = 25 mL
Volume of after dilution = 500 mL
Volume of dilute water used for titration = 100 mL
Blank titre value = 6.1 mL 02 N $Na_2S_2O_3$
Titre value after 5 days = 3.6 mL 02N $Na_2S_2O_3$
Dissolved oxygen at the beginning of the experiment is equivalent to 6.1 mL 0.02 N $Na_2S_2O_3$.
Dissolved oxygen after 5 days is equivalent to 3.6 mL 0.02N $Na_2S_2O_3$
Dissolved oxygen used up by wastewater in 5 days is,
= (6.1 – 3.6)
= 2.5 mL 02 N $Na_2S_2O_3$
1 mL1 N $Na_2S_2O_3$ = 8 mg oxygen
2.5 mL 02 N $Na_2S_2O_3 = 2.5 \times 0.02 \times 8 = 0.4$ mg oxygen
100 mL of diluted water = 0.4 mg oxygen
25 mL of original waste water = 500 mL dilute wastewater

$$= \frac{0.4 \times 500}{600}$$

= 2 mg oxygen

1000 mL of original waste water $= \dfrac{2 \times 1000}{25}$

= 80 mg oxygen

BOD of waste water = 80 mg oxygen/L

Method-II:

BOD of 1000 mL of wastewater = $\dfrac{(A - B) \times N \times 8 \times v_2}{V} \times \dfrac{1000}{v_1}$ mg of oxygen/L.

Given:

$A = 6.1$ mL, $B = 3.6$ mL, $v_1 = 25$ mL of wastewater taken; $V_2 = 500$ mL, the volume after dilution and $V = 100$ mL, the volume taken for titration and $N = 0.02$ N

$$= \frac{[6.1 - 3.6] \times 0.02 \times 8 \times 500 \times 1000}{100 \times 25} = 80 \text{ mg/L}$$

12. Calculate the BOD value of a sewage sample containing 9.2 mg/L of organic matter with the formula $C_6H_{12}O_6$.

    Gram Molecular weight of $C_6H_{12}O_6 = [6 \times 12 + 12 \times 1 + 6 \times 32] = 180$ g

    $$\underset{180\text{ g}}{C_6H_{12}O_6} + 6O_2 \rightarrow \underset{192\text{ g}}{6CO_2} + 6H_2O$$

    Therefore,

    180 mg/L of $C_6H_{12}O_6$ needs 192 mg/L of oxygen for complete oxidation under experimental conditions in sewage water.

    If sewage water contains 9.2 mg/L, then it needs $\dfrac{9.2 \times 192 \text{ mg/L}}{180}$ of oxygen to oxidize

    Hence, BOD of the sewage water = 9.813 mg/l.

13. An effluent water sample emerging out from a chemical factory is found to contain only organic compound of molecular formula $H_2C_2O_4$. If the wastewater sample contains 60 mg/L of the above compound, what would be the amount of oxygen required to oxidize by microorganisms under experimental conditions of BOD estimation. [C = 12, H = 1 and O = 16]

    Molecular weight of $H_2C_2O_4 = [1 \times 2 + 12 \times 2 + 16 \times 4] = 90$

    $$\underset{2 \times 90\text{ g}}{2H_2C_2O_4^{-}} + O_2 \rightarrow \underset{32\text{ g}}{4CO_2} + 2H_2O$$

    Therefore,

    180 mg/L of $H_2C_2O_4$ in waste water requires 32 mg of oxygen to oxidize into carbon dioxide and water. In a BOD experiment, if wastewater contains 150 mg/L of $H_2C_2O_4$, then the amount of oxygen needed to oxidize the organic impurity = $\dfrac{60 \times 32}{180}$ = 10.67 mg/L

8.6 Chemical Oxygen Demand

In environmental chemistry, the *chemical oxygen demand* (COD) test is commonly used to indirectly measure the amount of organic compound in water. Most applications of COD determine the amount of organic pollutants found in surface water (e.g. lakes and rivers), making COD a useful measure of water quality. It is expressed in milligrams per litre (mg/L) which indicates the mass of oxygen consumed per litre of solution. Older references may express the units as parts per million (ppm).

COD of wastewater is the number of milligrams of oxygen required to oxidize the impurities present in 1000 mL of wastewater using strong oxidizing agents like acidified $K_2Cr_2O_7$.

COD represents the total amount of oxygen required to oxidize all oxidizable impunities in a sample of sewage wastes. COD is always greater than BOD since in COD measurement both biodegradable and non-biodegradable load are completely oxidized. The difference in COD and BOD is equivalent to the quantity of biologically resistant organic matter.

8.6.1 Determination of COD

Principle

A known volume of the wastewater sample is refluxed with a known excess of $K_2Cr_2O_7$ solution in H_2SO_4 medium containing $HgSO_4$ (catalyst) and Ag_2SO_4 [which retains halides] for about 1½ hours for the oxidation to be completed. A part of the $K_2Cr_2O_7$ is used up for the oxidation of impurities. Remaining $K_2Cr_2O_7$ determined by titration with standard FAS solution using ferroin as an indicator. The end-point is the change of colour from blue green to reddish brown.

A blank is performed by titrating known volume of the acidified $K_2Cr_2O_7$ with the same FAS using the same indicator.

$$\text{COD of water sample} = \frac{(A - B) \times M \times 8000 \text{ mL}}{\text{volume of the sample}} \text{ mg /L}$$

where

A = Blank titre value of $K_2Cr_2O_7$ versus FAS and

B = volume of FAS consumed for unreacted $K_2Cr_2O_7$ of the solution.

M = molarity of FAS. solution.

Procedure

25 mL of wastewater is pipetted out into a round bottomed flask. 10 mL of $K_2Cr_2O_7$ is pipetted out into the same flask along with one test tube full of 1:1 H_2SO_4 containing $HgSO_4$ and Ag_2SO_4. The flask is filled with a reflux water condenser and the mixture is refluxed for 2 hours. The contents are cooled and transferred to a conical flask. Five drops of ferroin indicator is added to it and titrated against FAS taken in the burette till the colour changes from blue green to reddish brown. Same volume of $K_2Cr_2O_7$ is pipetted out, mixed with sulphuric acid and ferroin and titrated against same FAS to get blank titre value.

Calculation

Method-I

$$\text{Chemical oxygen demand of water} = \frac{(A - B) \times M \times 8000}{V} \text{mg/L}$$

where

'A' = FAS (mL) used for blank.

'B' = FAS (mL) used for sample.

'M' = Molarity of FAS

'V' = Volume of sample (mL).

Method-II

Volume of wastewater = V mL

Solution of standard FAS used in sample titration = B mL

Volume of standard FAS in blank titration = A mL

FAS equivalent of $K_2Cr_2O_7$ consumed by V mL wastewater = $(A - B)$ mL

Normality of FAS (N) = Molarity of F.A.S (M)

V mL of wastewater = used up $K_2Cr_2O_7$ = (Y-X) mLN normal FAS

1000 mL 1 M FAS = 8 g of oxygen = 8000 mg oxygen

1 mL 1 M FAS = 8 mg of oxygen

Therefore, $(A - B)$ mL normal FAS = $(A - B) \times M \times 8$ mg oxygen

V mL of wastewater requires $(A - B) \times M \times 8$ mg oxygen

1000 mL of water (COD) requires $\frac{(A - B) \times M \times 8000 \text{ mg/L}}{V}$

where V is the volume of waste, 'A' is the blank titre value 'B' is the titre value with the sample and M is the molarity of FAS.

14. 25 mL of wastewater was mixed with 25 mL of $K_2Cr_2O_7$, acidified and refluxed. The unreacted $K_2Cr_2O_7$ acidified required 8.2 mL of FAS. In a blank titration 25 mL of $K_2Cr_2O_7$ acidified required 16.4 mL of same 0.2 N FAS. Calculate COD of waste water.

Solution

Method-I

$$\text{COD of waste water} = \frac{(A - B) \times M \times 8000 \text{ mg/L}}{V}$$

$$= \frac{[18.4 - 16.4] \times 0.2 \times 8000}{25}$$

$$= 524.8 \text{ mg/L}$$

Method-II

25 mL waste + 25 mL $K_2Cr_2O_7$ = 8.2 mL 0.2 N FAS

25 $K_2Cr_2O_7$ alone = 16.4 mL 0.2 FAS

25 mL of waste water = used up $K_2Cr_2O_7$ = (18.2 – 16.4) mL 0.2 N FAS

= 8.2 mL 0.2 FAS

1 mL 1 N FAS = 8 mg of oxygen

8.2 mL 0.2 N FAS = 8 × 8.2 × 0.2 mg oxygen

= 13.12 mg oxygen

25 mL of waste water requires 13.12 mg oxygen

1000 mL of waste water requires $\frac{13.12 \times 1000}{25}$ = 524.8 mg oxygen

Therefore, COD of the waste water = 524.8 mg oxygen.

15. Calculate COD of effluent sample when 25 cm^3 of the effluent required 8.3 cm^3 0.001 M $K_2Cr_2O_7$ for oxidation [given molar mass of $K_2Cr_2O_7$ = 294].

Solution

Method-I

$$\text{COD of waste water} = \frac{(A - B) \times M \times 8000 \text{ mg/L}}{V}$$

$$6N\ K_2Cr_2O_7 \equiv 1\ M\ K_2Cr_2O_7$$

$$\text{COD of wastewater} = \frac{(8.3) \times 6 \times 0.001N \times 8000 \text{ mg/L}}{25}$$

$$= 15.976 \text{ mg/L}$$

Method-I

1000 1 mL 1 N $K_2Cr_2O_7$ = 8 g of oxygen

1 mL 1 N $K_2Cr_2O_7$ = 8 mg oxygen

Molar mass of $K_2Cr_2O_7$ = 294 and Eq. mass = 49

1 mole = 6 equivalence 1 M = 6 N

1 mL 1 mol = 6 mL 1 N $K_2Cr_2O_7$

Therefore, 1 mL 1 molar (6 N) $K_2Cr_2O_7 = 6 \times 8 = 48$ mg oxygen

25 cm^3 wastewater = 8.3 cm^3 0.001 M $K_2Cr_2O_7$

$= 48 \times 8.3 \times 0.001$ mg oxygen

25 cm^3 wastewater = 0.3994 mg oxygen.

$$\text{Therefore, } 1000 \text{ cm}^3 \text{ wastewater} = \frac{0.03994}{25} \times 1000$$

= 15.976 mg oxygen

COD of the sample = 15.976 mg oxygen

8.7 Sewage and Its Characteristics

Sewage is the liquid waste from toilets, baths, kitchens, etc. that is disposed of via sewers. In many areas, sewage also includes some liquid waste from industry and commerce. Any fresh sewage is usually green or yellowish green. With aging, it becomes dark brown. Fresh sewage is odourless but when it becomes stale it produces obnoxious smell due to the release of gases like H_2S, Phosphine, etc. Sewage water is normally turbid and has a temperature slightly higher than ordinary water. It contains dissolved and suspended impurities and also colloidal dispersion. Sewage contains both organic and inorganic impurities. The pH of fresh sewage is higher than 7 (alkaline) but after a few days it becomes less than 7 (acidic). Sewage also contains pathogenic (disease producing) bacteria and nonpathogenic bacteria. There are two types of nonpathogenic bacteria.

Much sewage also includes some surface water from roofs or hard-standing areas. Municipal wastewater therefore includes residential, commercial and industrial liquid waste discharges, and may include storm water runoff.

(i) *Aerobic bacteria,* which live and grow on free oxygen and dissolved oxygen. These bacteria act as catalysts for the oxidation of fresh sewage. The products of oxidation are nitrites, nitrates, sulphates, etc. These do not produce any offensive smell.

(ii) *Anaerobic bacteria* live and grow in the absence of free oxygen. When the dissolved oxygen content in water decreases below a particular level anaerobic bacteria try to decompose the organic compounds in water and produces methane, H_2S, phosphine, etc. responsible for the offensive smell. The water is said to be stale and the process of oxidation is anaerobic oxidation.

8.8 Sewage Treatment

In essence, sewage treatment has two functions

- to kill pathogens (whatever they are) and
- to eliminate harmful chemicals from the water
- Chemicals that are toxic and the bulk of organic chemicals that cause the depletion of life-supporting oxygen in the receiving waters of a stream, river or lake.
- Domestic sewage contains colour, bad odour, organic and inorganic impurities, pathogenic bacteria, etc. If this sewage is discharged into rivers, lakes and sea, the water in them get polluted and causes several harmful effects.
- It colours and produces bad odour affecting quality of water.
- It causes reduction in dissolved oxygen. This endangers the life of fish and other aquatic life.
- Pathogenic bacteria leads to water-bound diseases.

Domestic sewage therefore needs proper treatment before it is run into natural water bodies. Treatment of sewage involves removal of organic impurities, suspended and floating materials, inorganic salts, pathogenic bacteria, etc.

Sewage treatment is carried out in three stages as follows:

(a) Primary treatment
(b) Secondary treatment and
(c) Tertiary treatment, as depicted in Fig. 8.7

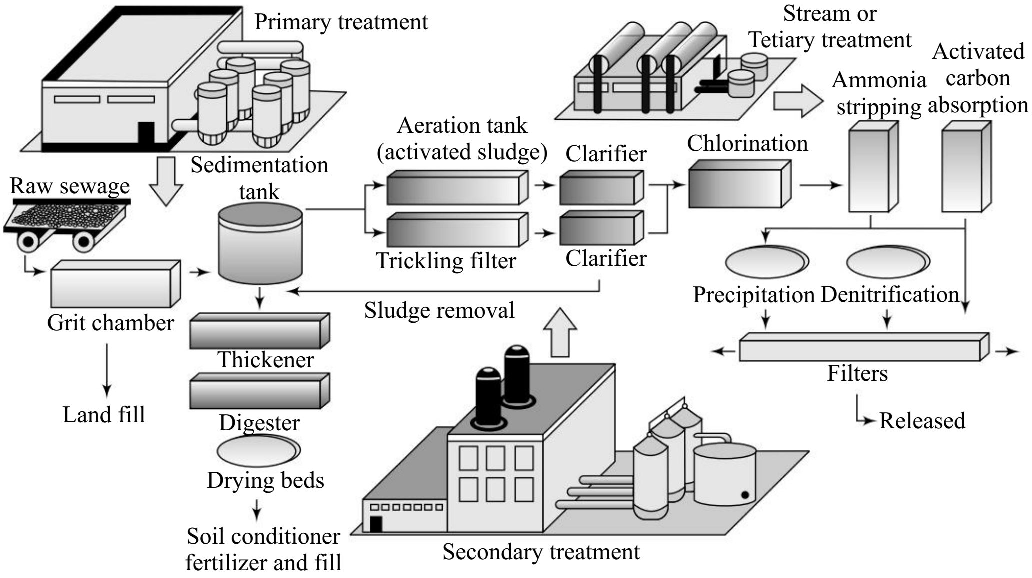


Figure 8.7 *Schematic flow diagram indicating the sewage treatments.*

8.8.1 Primary Treatment

Primary treatment is used to remove *suspended and floating solids* from wastewater by physical and chemical methods. It involves the following steps.

(i) *Screening*: Using bar screens and mesh screens floating, suspended and coarse particles are removed by passing sewage water through it.

(ii) *Slit and grit removal*: Sand, powdered glass, etc. called grit are removed by slowly passing sewage water through grit chambers. Heavier sand and broken glass settles down by gravity.

(iii) *Removal of oil and grease*: Sewage water is kept in a skimming tank and compressed air is blown through it. Oil and grease form froth and float on the surface. It is skimmed off.

(iv) *Sedimentation process*: In this process, fine suspended particles which do not settle down by gravity are coagulated by the addition of coagulating agents like alum, $FeSO_4$, etc. The sedimented particles are filtered off.

8.8.2 Secondary Treatment (Biological Treatment)

Activated sludge method:

Wastewater after sedimentation is mixed with required quantity of activated sludge (containing microorganisms and aerobic bacteria) in a aeration tank as shown in Fig. 8.8. The mixture is aerated by blowing air through

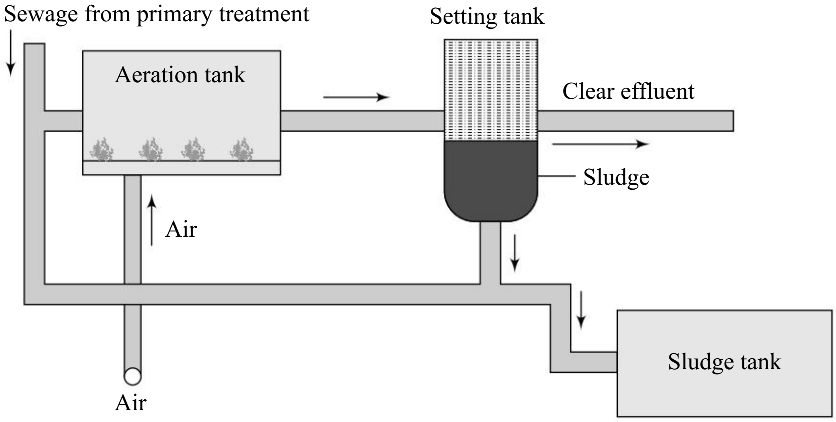


FIGURE 8.8 *Activated sludge process.*

it for several hours. Because of the aerobic conditions the organic matter in the sewage is fully oxidized. The purified water + sludge is sent to a tank where sludge settles down. A part of the sludge is used for purification of fresh batch of sewage while the rest is pumped to sludge disposal tank. Purified water is pumped out and collected separately.

8.8.3 Tertiary Treatment

Tertiary treatment involves further purification. It involves the following steps of removing dissolved salts, colloidal particles, gases, etc.

(i) *Removal of phosphate* is done by adding lime. A precipitate of calcium phosphate is formed at a pH of 10. Ammonium salts are converted to ammonia.

$$3Ca(OH)_2 + 2PO_4^{-3} \rightarrow Ca_3(PO_4)_2 + 6OH^-$$

$$NH_4^+ + OH^- \rightarrow NH_3 + H_2O$$

(ii) *Coagulation and sedimentation*: Disposed colloidal form of fine particles is coagulated in a sedimentation tank by the addition of alum or $FeSO_4$. The highly charged Al^{3+} and Fe^{+2} ions neutralize the charge on the colloidal particles. $Al(OH)_3$ and $Fe(OH)_2$ formed as flocculant precipitates bring down the coagulating colloids.

(iii) *Filtration*: The water is passed through conventional filter beds filled with gravels, coarse sand and fine sand to remove last traces of suspended matter.

(iv) *Degasification*: Dissolved gases are moved by trickling the water through tall tower filled with a number of perforated hot plates. Large surface area and heat liberates the dissolved gases like NH_3, CO_2, H_2S, etc.

(v) *Disinfection*: Harmful pathogenic bacteria are removed either by passing Cl_2 gas or ozone or UV light through the water.

8.9 Potable Water

Drinking water is water that is intended to be drunk by humans. *Water of sufficient quality (free from contaminants) to serve as drinking water or water that is safe for human consumption is called potable water whether* *it is used as such or not.* Although many fresh water sources are drinkable by humans, they may be a disease bearing or cause long-term health problems if they do not meet certain water quality guidelines. Unpolluted water is not harmful for human beings and is called safe water or potable water.

8.9.1 Purification of Municipal Water

Water purification is the removal of contaminants from raw to produce drinking water that is pure enough for human consumption. Substances that are removed during the process include bacteria, algae, viruses, fungi, minerals and man-made chemical pollutants. Many contaminants can be dangerous—but depending on the quality standards, others are removed to improve the water's smell, taste and appearance. A small amount of disinfectant is usually intentionally left in the water at the end of the treatment process to reduce the risk of re-contamination in the distribution system.

There are three principal stages in water purification:

(i) **Primary treatment**: Collecting and screening include pumping from rivers and initial storage.
(ii) **Secondary treatment**: Removal of fine solids and the majority of contaminants using filters, coagulation, flocculation and membranes.
(iii) **Tertiary treatment**: Polishing, pH adjustment, carbon treatment to remove taste and smells, disinfection and temporary storage to allow the disinfecting agent to work.

(i) Primary treatment

(a) *Pumping and containment*: The majority of water must be pumped from its source or directed into pipes or holding tanks. To avoid adding contaminants to the water, this physical infrastructure must be made from appropriate materials and constructed so that accidental contamination does not occur.
(b) *Screening*: The first step in purifying surface water is to remove large debris such as sticks, leaves, trash and other large particles which may interfere with subsequent purification steps. Most deep groundwater does not need screening before other purification steps.
(c) *Storage*: Water from rivers may also be stored in bank side reservoirs for periods between a few days and many months to allow natural biological purification to take place. This is especially important if treatment is done by slow sand filters. Storage reservoirs also provide a buffer against short periods of drought or to allow water supply to be maintained during transitory pollution incidents in the source river.
(d) *Preconditioning*: Waters rich in hardness salts are treated with soda-ash to precipitate out calcium carbonate utilizing the common ion effect.
(e) *Prechlorination*: In many plants the incoming water was chlorinated to minimize the growth of fouling organisms on the pipe-work and tanks. Because of the potential adverse quality effects (see Chlorine below), this has largely been discontinued.

(ii) Secondary treatment

There are a wide range of techniques that can be used to remove the fine solids, microorganisms and some dissolved inorganic and organic materials. The choice of method will depend on the quality of the water being treated, the cost of the treatment process and the quality standards expected of the processed water.

(a) *pH adjustment*: If the water is acidic, lime or soda ash is added to raise the pH. Lime is the more common of the two additives because it is cheaper, but it also adds to the resulting water hardness. Making the water slightly alkaline ensures that coagulation and flocculation processes work effectively and also helps to minimize the risk of lead being dissolved from lead pipes and lead solder in pipe fittings.
(b) *Coagulation and flocculation*: Together, coagulation and flocculation are purification methods which work by using chemicals that effectively 'glue' small suspended particles together so that they settle

out of the water or stick to sand or other granules in a granular media filter. Many of the suspended water particles have a negative electric charge. The charge keeps particles suspended because they repel similar particles.

Coagulation works by eliminating the natural electrical charge of the suspended particles so they attract and stick to each other. The joining of the particles so that they will form larger settling kind particles is called flocculation. The larger formed particles are called *floc*. The coagulation chemicals are added in a tank (often called a rapid mix tank or flash mixer) which typically has rotating paddles. In most treatment plants, the mixture remains in the tank for 10–30 seconds to ensure full mixing. The amount of coagulant that is added to the water varies widely due to the different source water quality. One of the more common coagulants used is aluminium sulphate, sometimes called *filter alum*. Aluminium sulphate reacts with water to form flocs of aluminium hydroxide. Coagulation with aluminium compounds may leave a residue of aluminium in the finished water. This is normally about 0.1–0.15 mg/L. It has been established that aluminium can be toxic to humans at high concentrations.

(c) *Flocculation:* In flocculation, coagulants are used but the resultant floc is settled out rather than filtered through sand filters. The chosen coagulant and the raw water are slowly mixed in a large tank called flocculation basin. Unlike a rapid mix tank, the flocculation paddles turn very slowly to minimize turbulence. The principle involved is to allow as many particles to contact other particles as possible generating large and robust floc particles. Generally, the retention time of a flocculation basin is at least 30 minutes with speeds between 0.5 feet and 1.5 feet per minute (15–45 cm/minute). Flow rates less than 0.5 ft/minute cause undesirable floc settlement within the basin.

(d) *Sedimentation*: Water exiting the flocculation basin enters the sedimentation basin, also called a *clarifier* or *settling basin*. It is a large tank with slow flow, allowing floc to settle to the bottom. The sedimentation basin is best located close to the flocculation basin so the transit between does not permit settlement or floc break-up. Sedimentation basins can be in the shape of a rectangle, where water flows from end to end, or circular where flow is from the centre outward. Sedimentation basin outflow is typically over a *weir* so only a thin top layer-furthest from the sediment exist. The amount of floc that settles out of the water is dependent on the time the water spends in the basin and the depth of the basin. As particles settle to the bottom of the basin a layer of sludge is formed on the floor of the tank. This layer of sludge must be removed and treated. The amount of sludge that is generated is significant, often 3–5 percent of the total volume of water that is treated. The cost of treating and disposing of the sludge can be a significant part of the operating cost of a water treatment plant. The tank may be equipped with mechanical cleaning devices that continually clean the bottom of the tank or the tank can be taken out of service when the bottom needs to be cleaned. An increasingly popular method of floc removal is by dissolved air flotation.

(e) *Filtration*: After separating most of the floc, the water is filtered as the final step to remove remaining suspended particles and unsettled floc. The most common type of filter is a *rapid sand filter*. Water moves vertically through sand that often has a layer of activated carbon or anthracite coal above the sand. The top layer removes organic compounds including taste and odour. The space between sand particles is larger than the smallest suspended particles, so simple filtration is not enough. Most particles pass through surface layers but are trapped in pore spaces or adhere to sand particles. Effective filtration extends into the depth of the filter. This property of the filter is the key to its operation: if the top layer of sand blocked all particles the filter would quickly clog. To clean the filter, water is passed quickly upward through the filter, opposite the normal direction (called *backflushing* or *backwashing*) to remove embedded particles. Prior to this, compressed air may be blown up through the bottom of the filter to break up the compacted filter media to aid the backwashing process which is known as *air scouring*. This contaminated water can be disposed of, along with the sludge from the sedimentation basin, or it can be recycled by mixing with the raw water entering the plant. *Slow sand filtration* may be used where there is sufficient land and space. These rely on biological treatment processes for their action rather than physical filtration. Slow sand filters are carefully constructed using graded layers of sand with the coarsest at the base and the finest at the top. Drains at the base convey treated water away for disinfection. Filtration depends on the development of a thin biological layer on the surface of the filter. An effective slow sand filter may remain in service for many weeks or even months if the pretreatment is well designed and produces an excellent quality of water which physical methods of treatment rarely achieve.

Ultrafiltration membranes are a relatively new development which uses polymer film with chemically formed microscopic pores which can be used in place of granular media to filter water effectively without coagulants. The type of membrane media determines how much pressure is needed to drive the water through and what sizes of microorganisms can be filtered out.

(iii) Tertiary treatment

Disinfection is normally the last step in purifying drinking water. Water is disinfected to destroy any pathogens that passed through the filters. Possible pathogens include viruses, bacteria and protozoas. In most developed countries, public water supplies are required to maintain a residual disinfecting agent throughout the distribution system, in which water may remain for days before reaching the consumer. Following the introduction of any chemical disinfecting agent, the water is usually held in temporary storage often called a *contact tank* or *clear well* to allow the disinfecting action to complete.

(a) *Chlorination:* The most common disinfection method is some form of chlorine or its compounds such as chloramine or chlorine dioxide. Chlorine is a strong oxidant that kills many microorganisms. Because chlorine is a toxic gas, there is a danger associated with its use. This problem is avoided by the use of sodium hypochlorite which is a relatively inexpensive solid that releases free chlorine when dissolved in water. A major drawback to using chlorine gas or sodium hypochlorite is that they react with organic compounds in the water to form potentially harmful levels of the chemical by-products trihalomethanes (THMs) and haloacetic acids, both of which are carcinogens. The formation of THMs and haloacetic acids is minimized by effective removal of as many organics from the water as possible before disinfection.

(b) *Ozonation: Ozone* is a relatively unstable molecule of oxygen which readily gives up one atom of oxygen providing a powerful oxidizing agent which is toxic to most waterborne organisms. It is a very strong, broad spectrum disinfectant that is widely used in Europe. It is an effective method to inactivate harmful protozoans that form cysts. It also works well against almost all other pathogens. Ozone is made by passing oxygen through ultraviolet light or an electrical discharge. To use ozone as a disinfectant, it must be created on site and added to the water by bubble contact. Some of the advantages of ozone include the production of relatively fewer dangerous by-products (in comparison to chlorination) and the lack of taste and odour produced by ozonation. Although fewer by-products are formed by ozonation, it has been discovered that the use of ozone produces a small amount of the suspected carcinogen bromate. Another one of the main disadvantages of ozone is that it leaves no disinfectant residual in the water.

(c) *UV radiation*: It is very effective at inactivating cysts, as long as the water has a low level of colour so the UV can pass through without being absorbed. The main drawback to the use of UV radiation is that, like ozone treatment, it leaves no residual disinfectant in the water. Because neither ozone nor UV radiation leaves a residual disinfectant in the water, it is sometimes necessary to add a residual disinfectant after they are used. This is often done through the addition of chloramines, discussed above as a primary disinfectant. When used in this manner, chloramines provide an effective residual disinfectant with very little of the negative aspects of chlorination.

8.10 Hard Water Treatment for Domestic and Industrial Purpose

The consumers of water have different perceptions about pure water. They are as follows:

- *Home owners* are primarily concerned with *domestic water* problems related to colour, odour, taste and safety to family health, as well as the cost of soap, detergents, 'softening,' or other treatments required for improving the water quality.
- *Chemists and engineers* working for industry are concerned with the purity of water as it relates to the scale deposition and pipe corrosion.
- *Farmers* are interested in the effects of irrigation waters on the chemical, physical and osmotic properties of soils, particularly as they influence crop production; hence, they are concerned with the water's total mineral content, proportion of sodium or content of ions 'toxic' to plant growth.

Therefore, demand for *high quality water* is growing around the world.

Water is drawn from different sources such as lakes, ponds, rivers, bore wells, etc. and is found to be contaminated with inorganic and other contaminants. Thus, water drawn from any such sources is unfit for domestic and industrial uses. Water in such cases is subjected to softening process. Generally, water quality is assessed in terms of hardness of water.

8.10.1 Softening of Water

The process of removal of inorganic contaminants like calcium and magnesium from hard water by any technique is referred to as softening of water.

Hard water cannot be used for domestic as well as for industrial purposes. It has to be subjected to softening process. By adopting softening process of water, the presence of dissolved salts causing hardness is minimized and the water can be liberally used for domestic and industrial purposes.

There are various softening processes. They are mainly

1. Cold lime process
2. Hot lime process
3. Zeolite or permutit process
4. Use of resins for softening

1. Cold soda: lime process

The basic principle involved in softening of hard water, preferably, removal of salts of calcium and magnesium of temporary hardness, is to neutralize and precipitate as their insoluble salts by the addition of adequate amounts of Na_2CO_3 (soda ash) and $Ca(OH)_2$ (lime) which get deposited at the bottom.

When requisite amounts of soda ash and lime are treated, the soluble calcium and magnesium salts of hard water chemically interact and get transformed to insoluble salts of calcium and magnesium. These stoichiometric reactions are depicted below which provides an idea for the calculation of the required amounts of either soda ash (S) or lime (L) or both depending on the type of hardness. The stoichiometric reactions indicate in terms of their equivalent amounts needed for reactions and neutralization.

1. Permanent hardness

$$CaCl_2 + Na_2CO_3 \rightarrow CaCO_3\downarrow + 2NaCl \text{ [Treat with soda ash (S) only]}$$

$$MgSO_4 + Ca(OH)_2 \rightarrow Mg(OH)_2\downarrow + Ca^{2+}$$

$$Ca^{2+} + Na_2CO_3 \rightarrow CaCO_3\downarrow + Na_2SO_4 \text{ [Treat with lime (L) + soda ash (S)]}$$

2. Temporary hardness

$$Ca(HCO_3)_2 + Ca(OH)_2 \rightarrow CaCO_3\downarrow + 2H_2O \text{ [Treat with lime only (L)]}$$

$$Mg(HCO_3)_2 + 2Ca(OH)_2 \rightarrow 2CaCO_3\downarrow + Mg(OH)_2 + 2H_2O \text{ [lime only (2L)]}$$

Water drawn from natural sources have temporary hardness that is conveniently and economically removed by lime treatment, as lime is cheap and removes efficiently without adding soluble salts in water. Thus, net outcome of soda-lime treatment is

- It eliminates both temporary and permanent hardness by precipitating dissolved impurities as insoluble carbonates which is filtered off.
- Sodium salts of the acid radicals (NaCl, Na_2SO_4, etc.) are formed after the treatment.
- Since the reactions are slow during mixing of soda ash and lime, there may be the possibility deposition/corrosion of the metals due to the deposition of products like $Mg(OH)_2$ and fine $CaCO_3$.

Calculation of lime and soda ash requirements for water softening

The stoichiometric relationship observed during the soda-lime additions to eliminate/minimize the dissolved salts of calcium and magnesium helps to find out the adequate requirement of both soda and lime for the process.

The stoichiometry in terms of equivalents is as follow:

100 parts by mass of $CaCO_3$ are equivalent to (i) 74 parts by mass of $Ca(OH)_2$ and (ii) 106 parts by mass of Na_2CO_3.

Therefore,

1. No of parts by mass of lime required for softening = $\frac{74}{100}$ [Temporary Ca^{2+} + 2 × Temporary Mg^{2+} + (Permanent $Mg^{2+} + Fe^{2+} + Al^{3+}$) + {$CO_2 + H^+$ (from acids) + HCO_3^- – $NaAlO_2$} All in terms of $CaCO_3$ equivalents.
2. No of parts by mass of Soda required for sofetening $\frac{106}{100}$ [Permanent ($Ca^{2+} + Mg^{2+} + Al^{3+}$ Fe^{2+}) H^+ (acids) – HCO_3^-] All in terms of $CaCO_3$ equivalents.

Process

The most common method of water softening used in water plants is lime softening, or lime/soda ash softening. The reason for this is that lime and soda ash are readily available and cost efficient for softening the large amounts of water processed through a water plant.

Raw water containing the dissolved salts of calcium and magnesium is fed through the water inlet as shown in Fig. 8.9. Water fills the inner cylindrical vessel completely. A hopper with a connecting tube is attached to the inner cylinder containing the raw water. Requisite quantities of both soda ash and lime as desired in equivalents are fed through the hopper into the inner cylinder. A stirrer attached to a motor swirls the raw water continuously along with the soda ash and lime. As a result of chemical interaction, carbonates of the metal ions are precipitated as a fine powder that settles down very slowly. A coagulant like $NaAlO_2$, aluminium sulphate, sodium aluminate are introduced to grow the size of the carbonates of calcium and magnesium particles and to settle quickly at the bottom, as sludge that can be easily removed. The soft water moves to the outer cylinder and passes through a filter that filters out unwanted materials. Finally, soft water can be withdrawn through the outlet at the top of the outer cylinder.

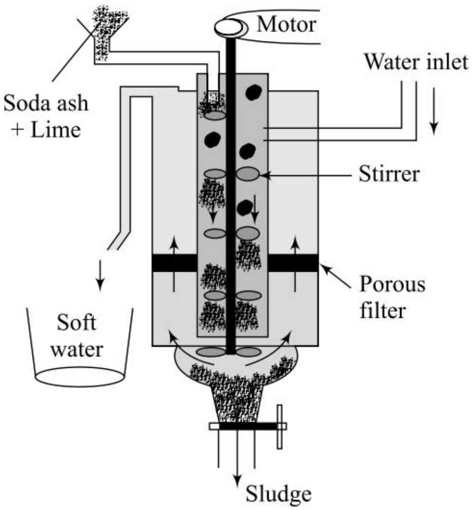


Figure 8.9 *Cold soda-lime softener.*

2. Hot lime-soda process

This method is similar to cold soda-lime process except the temperature is maintained at 80–110°C and there is no need for coagulants. The advantages of this method are

- Soda ash and lime interacts with calcium and magnesium ions of hard water quickly at this temperature and settles down easily and hence there is no need for the addition of coagulants.
- Softening rate of water is faster.
- Calcium and magnesium contents in water are reduced to 15–25 ppm.
- Sludge settles rapidly and can be removed.
- Filtration process of water is easier.

The hot soda-lime processing unit for softening of raw water is described in Fig. 8.10. It consists of two chambers. Raw water is sent to the conical chamber along with the addition of both lime and soda ash through the inlet having the hopper. Super-heated steam is also sent simultaneously at 110°C into this reaction chamber. Reaction occurs at a faster rate at this temperature yielding soft water and the carbonates of metal ions settles rapidly as sludge and collects at the bottom and can be removed, as shown in Fig. 8.10. The soft water thus got passes through another, known as filtration unit. The filtration unit consists of layers of fine sand, coarse sand and gravel which is employed to remove other unwanted impurities of water. Softened water is collected and stored in tank for industrial use.

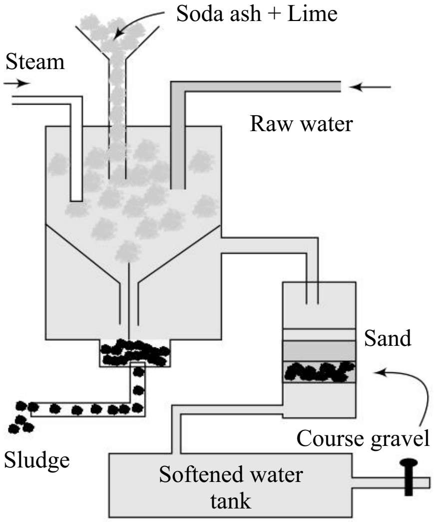


Figure 8.10 *Hot lime-soda method.*

16. When a sample of water was analysed, it is observed to have the following mineral contaminants: They are
    - $CaCO_3 = 50$ mg/L
    - $MgCO_3 = 100$ mg/L
    - $CaCl_2 = 111$ mg/L
    - $MgCl_2 = 47.6$ mg/L
    - NaCl = 5.86 mg/L

    If this hard water is to be softened by lime-Soda process, calculate the amount of lime and soda needed for 20,000 litres of hard water.

Salt	Quantity, mg/L	Conversion factor	$CaCO_3$ equivalent
$CaCO_3$	50	100/100	50 × 100/100 = 50
$MgCO_3$	100	100/84	100 × 100/84 = 119
$CaCl_2$	111	100/111	111 × 100/111 = 100
$MgCl_2$	47.5	100/95	47.5 × 100/95 = 50
NaCl	86	100/58.6	—

(i) Amount of lime required $= \dfrac{74}{100}\,[2 \times MgCO_3 + CaCO_3 + MgCl_2]$

$$= \frac{74}{100} \times [2 \times 119 + 50 + 50] \times 20{,}000$$

$$= 50{,}02{,}400 \text{ mg} = 5.0024 \text{ kg}$$

(ii) Amount of soda required $= \dfrac{106}{100}\,[MgCO_3 + MgCl_2 + CaCl_2]$

$$= \frac{106}{100}\,[119 + 50 + 100] \times 20000$$

$$= 1.06 \times 269 \times 20{,}000 = 5702800 \text{ mg}$$

$$= 5.7028 \text{ kg.}$$

17. Calculate the amount of lime and soda required for softening of 75,000 L of hard water. The analytical data obtained are as follows:

$Ca^{2+} = 20$ ppm; $Mg^{2+} = 45$ ppm; $HCO_3^{-} = 200$ ppm.

Salt	Quantity, ppm	Conversion factor	$CaCO_3$ equivalent
Ca^{2+}	20	100/24	$20 \times 100/24 = 83.34$
Mg^{2+}	45	100/40	$45 \times 100/40 = 112.5$
HCO_3^-	30	100/61	$30 \times 100/61 = 49.2$

(i) Amount of lime required $= \dfrac{74}{100}[HCO_3^- + Mg^{2+}] \times$ vol of water

$$= 0.74[49.2 + 112.5] \times 75{,}000 = 0.74 \times 161.5 \times 75{,}000$$

$$= 119.5 \times 75{,}000 = 8963{,}250 \text{ mg}$$

$$= 8.96\text{kg}$$

(ii) Amount of soda required $= 106\ [Mg^{2+} + Ca^{2+} - HCO_3^-] \times$ Volume of water

$$= 1.06[(112.5 + 83.34) - 49.2] \times 75{,}000 =$$

$$= 1.06 \times 146.6 \times 75000 = 11{,}654{,}700 \text{ mg}$$

$$= 11.65 \text{ kg}$$

18. Calculate the temporary and permanent hardness of water whose analysis is as follows:

$Ca(HCO_3)_2 = 36$ mg/L
$Mg(HCO_3)_2 = 46$ mg/L
$MgSO_4 = 30$ mg/L
$CaSO_4 = 24$ mg/L
$CaCl_2 = 28$ mg/L
NaCl =10 mg/L

Also, calculate the amount of lime (84 percent pure) and soda (90 percent pure) required for the treatment of 25,000 L of this hard water.

	Quantity, mg/l	M. w	$CaCO_3$ equivalent
$Ca(HCO_3)_2$	36	162	$36 \times 100/162 = 22.2$
$Mg(HCO_3)_2$	46	146	$46 \times 100/146 = 31.5$
$MgSO_4$	30	120	$30 \times 100/120 = 25$
$CaSO_4$	24	136	$24 \times 100/136 = 17.6$
$CaCl_2$	28	111	$28 \times 100/111 = 25.2$

Note: NaCl do not contribute to hardness.

Temporary hardness = $[Ca(HCO_3)_2 + Mg(HCO_3)_2]$
= 22.2 + 31.5 = 53.7 mg/L
Permanent hardness = $[MgSO_4 + CaSO_4 + CaCl_2]$
= 25 + 17.6 + 25.2 = 67.8 mg/L.

Lime requirement:

Lime is required for $Ca(HCO_3)_2$, $Mg(HCO_3)_2$ and $MgSO_4$.

Hence, 84 percent lime required for treating 25,000 L of water is

$$= \frac{74}{100}\,[Ca(HCO_3)_2 + 2x\,Mg(HCO_3)_2 + MgSO_4] \times \text{purity factor} \times \text{kg factor for } 25{,}000$$

$$= 0.74[22.2 + 2 \times 31.5 + 25] \times \frac{100}{84} \times \frac{25000}{1000 \times 1000}$$

$$= 0.74 \times 110.2 \times 1.19 \times 0.025$$

$$= 2.43 \text{ kg.}$$

Soda (90 percent) required for softening 25,000 L of water sample

$$= \frac{106}{100}\,[CaSO_4 + CaCl_2 + MgSO_4] \times \text{Purity factor} \times \text{kg factor for } 25000$$

$$= 1.06[17.6 + 25.2 + 25] \times \frac{100}{90} \times \frac{25000}{1000 \times 1000}$$

$$= 1.06 \times 67.8 \times 1.11 \times 0.025 = 1.994 \text{ kg}$$

3. Calgon process

Calgon is a trade name of a complex salt, sodium hexametaphosphate $(NaPO_3)_6$. It is used for softening hard water. Calgon ionizes to give a complex anion:

$$(NaPO_3)_6 \text{ or } Na_2(Na_4P_6O_{18}) \rightarrow 2Na^+ + \underset{\text{(complex ion)}}{Na_4P_6O_{18}^{2-}}$$

The addition of calgon to hard water causes the calcium and magnesium ions of hard water to displace sodium ions from the anions of calgon.

$$Ca^{2+} + \underset{\text{Anion of calgon}}{Na_4P_6O_{18}^{2-}} \rightarrow 2Na^+ + \underset{\text{goes into solution}}{CaNa_2P_6O_{18}^{2-}}$$

This results in the removal of calcium and magnesium ions from hard water in the form of a complex with Calgon. The water is softened and sodium ions are released into water.

4. Zeolite or permutit process

Permutit or sodium aluminium silicate is a complex chemical compound which occurs as a natural mineral called Zeolite. Permutit or zeolites are insoluble in water and have the property of exchanging ions present in them with the ions present in the solution.

A cylindrical column is initially packed with a layer of course gravel followed by a layer of fine gravel. Above this, permutit or zeolites layer is packed as shown in Fig. 8.11. A slow stream of hard water is passed through this material. As a result, calcium and magnesium ions present in hard water are exchanged with sodium ions in the permutit (Na^+Al-silicate$^-$). The outgoing water contains sodium salts that do not cause hardness.

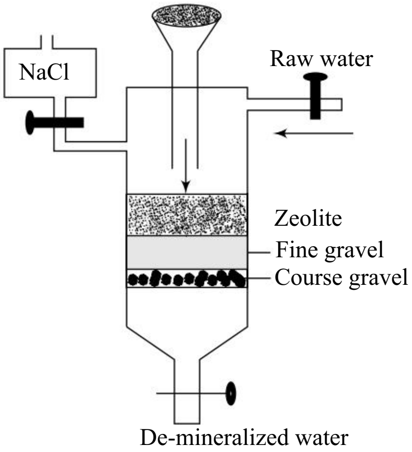


FIGURE 8.11 *Zeolite method of softening.*

$$\underset{\text{(Permutit)}}{CaCl_2 + 2Na^+ (\text{Al-silicate}^-)} \rightarrow Ca(\text{Al-silicate})_2 + 2NaCl$$

$$\underset{\text{(Permutit)}}{MgSO_4 + 2Na^+ (\text{Al-silicate}^-)} \rightarrow Mg(\text{Al-silicate})_2 + Na_2SO_4$$

Regeneration

Over a period of time permutit gets converted into a mixture of calcium and magnesium aluminium-silicates and permutit has to be regenerated for further use. This is done by packing brine—a concentrated solution of sodium chloride, through the column packed with spent permutit. The following reactions take place.

$$Ca(\text{Al-silicate})_2 + 2NaCl \rightarrow 2Na^+ (\text{Al-silicate}^-) + CaCl_2$$

$$Mg(\text{Al-silicate})_2 + 2NaCl \rightarrow 2Na^+ (\text{Al-silicate}^-) + MgCl_2$$

The resulting calcium chloride and magnesium chloride produced are washed out through a tap at the bottom. The regenerated permutit is reused for softening water.

19. 1000 L of raw hard water is taken for softening. A solution of NaCl containing 120 g/L of NaCl was passed through the exhausted resin. Volume of NaCl needed for regeneration of resin is 50 L. Calculate the hardness of water.

Data:
Concentration of NaCl solution = 120 g per 1000 mL.
Volume of hard water taken = 1000 L
Volume of NaCl needed for regeneration = 50 L
Calculation:
Amount of NaCl present in 50 L of NaCl solution = 50 × 120 g of NaCl.= 6000 g

$$= \frac{6000 \times 50}{58.5} \text{ equivalents of } CaCO_3$$

$$= 5128.2 \text{ g equivalent of } CaCO_3$$

Hence,
1000 L of water = 5128.2 g equivalent of $CaCO_3$

$$\text{Therefore, 1 lire of raw water} = \frac{5128.2}{1000} = 5.1282 \text{ g equivalent of } CaCO_3$$

$$= 5128.2 \text{ mg/L (ppm)}$$

20. For regeneration of an exhausted zeolite needed 100 L of NaCl solution. The strength of the NaCl solution used is 130 g/L. How many litres of raw hard water sample, having 200 mg/L can be softened, using this softener?

100 L of NaCl solution contains = 100 × 130 = 13,000 g

$$= \frac{13000 \times 50}{58.5} = 11{,}111.1 \text{ g equivalent of } CaCO_3$$

Now, 1 L of water contains 200 mg/L = 0.2 g of equivalent $CaCO_3$
Therefore, the volume of hard water that can be softened by this softener is

$$\frac{13000 \times 50}{0.2 \times 58.5} = 55{,}555 \text{ L}$$

5. Softening water by ion-exchange resins

Ion exchange is not nearly as cost-efficient as lime softening when producing large amounts of water. Therefore, water plants typically do not use ion exchange to soften the finished product water. However, some water plants do need to soften smaller amounts of water used in some treatment process, for example fluoridation, which can be sufficiently handled with ion-exchange methods. Likewise, most in-home water softeners are ion-exchange units.

Ions can exist with either a positive charge or a negative charge. A positive charge ion is called a cation. A negative charge ion is called an anion. Calcium (Ca) and magnesium (Mg), the principle elements causing hardness, are typically positive charged cations. Therefore, the basic premise of which ion-exchange softening uses is trading one ion for another.

Giant organic molecules having acidic or basic groups are known as *Ion-exchange resins.* Acid resins contain the acid group (–COOH). Basic resins contain the basic group $[(-NH_3^+)OH^-]$, that is substituted ammonium hydroxide group. Acid and basic ion exchange resins are represented as $RCOO^-H^+$ and $RNH_3^+ + OH^-$, respectively.

An ion-exchange water softener has two main units: one containing cation exchange resin and the other containing anion exchange resin.

Hard water is sent to the first compartment containing a layer of gravel and cation exchange resin as shown in Fig. 8.12. When water percolates through the resin, cations like Ca^{2+}, Mg^{2+}, Na^+, K^+, etc. present in the hard water get exchanged with H.

$$2RCOO^-H^+ + Ca^{2+} \rightarrow (RCOO)_2Ca + 2H^+$$

$$2RCOO^-H^+ + Mg^{2+} \rightarrow (RCOO)_2Mg + 2H^+$$

$$RCOO^-H^+ + Na^+ \rightarrow RCOONa + 2H^+$$

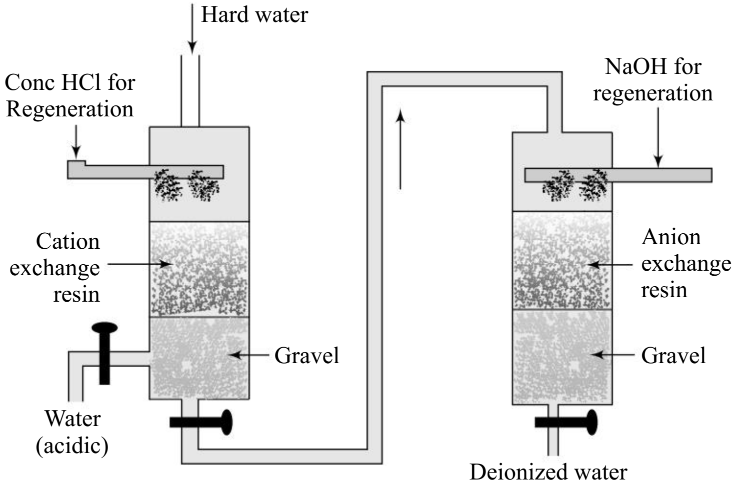


FIGURE 8.12 *Ion-exchange process for water softening.*

It is observed that the water emerging out from the cationic compartment is acidic in nature. In order to eliminate both H^+ and anions from water that is received from the first column, the water is made to pass through the second anionic resin column as shown in figure above. The resin interacts with anions present in hard water.

$$RNH_3^+OH^- + Cl^- \rightarrow R\,NH_3Cl + H^+$$

$$2RNH_3^+OH^- + SO_4^{-2} \rightarrow (RNH_3)_2SO_4 + 2OH^-$$

The water that is collected from the second resin column is free from both cations and anions. The H^+ and OH^- ions combine during the softening process and give water molecules.

The water obtained after passing through the ion-exchangers is called deionized water or demineralized water. This is as good as distilled water. The water becomes soft after this process.

Regeneration

After some time, the resins require regeneration as they become ineffective. Adding a strong acid regenerates cation resin and the anion resin is regenerated by treating it with a solution of a strong base.

$$(RCOO)_2 + H_2SO_4 \rightarrow 2RCOOH + Ca^{2+} + SO_4^{2-}$$

$$RNH_3Cl + NaOH \rightarrow RNH_3OH + Na^+ + Cl^-$$

The regenerated resins are washed to remove Ca^{2+}, Na^+, SO_4^{2-}, Cl^-, etc. The resins can be used again. The water treated by ion-exchange process not only becomes soft but also free from all dissolved mineral impurities. It gets completely demineralized.

8.11 Boiler

A boiler is an enclosed vessel that provides a means for combustion heat to be transferred into water until it becomes heated water or steam. The hot water or steam under pressure is then usable for transferring the heat to a process. Water is a useful and cheap medium for transferring heat to a process. When water is boiled into steam its volume increases about 1600 times, producing a force that is almost as explosive as gunpowder. This causes the boiler to be extremely dangerous equipment that must be treated with utmost care.

Boiler system

The boiler system (Fig. 8.13) comprises a feed water system, steam system and fuel system. The feed water system provides water to the boiler and regulates it automatically to meet the steam demand. Various valves provide access for maintenance and repair. The steam system collects and controls the steam produced in the boiler. Steam is directed through a piping system to the point of use. Throughout the system, steam pressure is regulated using valves and checked with steam pressure gauges. The fuel system includes all equipment used to provide fuel to generate the necessary heat. The equipment required in the fuel system depends on the type of fuel used in the system. A typical schematic boiler room is shown in Fig. 8.13.

The water supplied to the boiler which is converted into steam is called feed water. The two sources of feed water are:

- Condensate or condensed steam returned from the processes and
- Makeup water (treated raw water) that must come from outside the boiler room and plant processes. For higher boiler efficiencies, the feed water is preheated by economizer, using the waste heat in the flue gas.

8.11.1 Nature of Natural Water

All natural waters contain varying amounts of suspended and dissolved matter as well as dissolved gases. The type and amount of impurities in fresh water vary with the source (lake, river, well) and with the area of location. Impurities in water become an important consideration when water is to be used for steam generation. With the trend towards higher-pressure boilers, pretreatment has become the key to successful operation of industrial power plants. Feedwater must be pretreated to remove impurities to control deposition, carryover and corrosion in the boiler system. Poor quality water gives poor quality steam. The first step in any treatment is

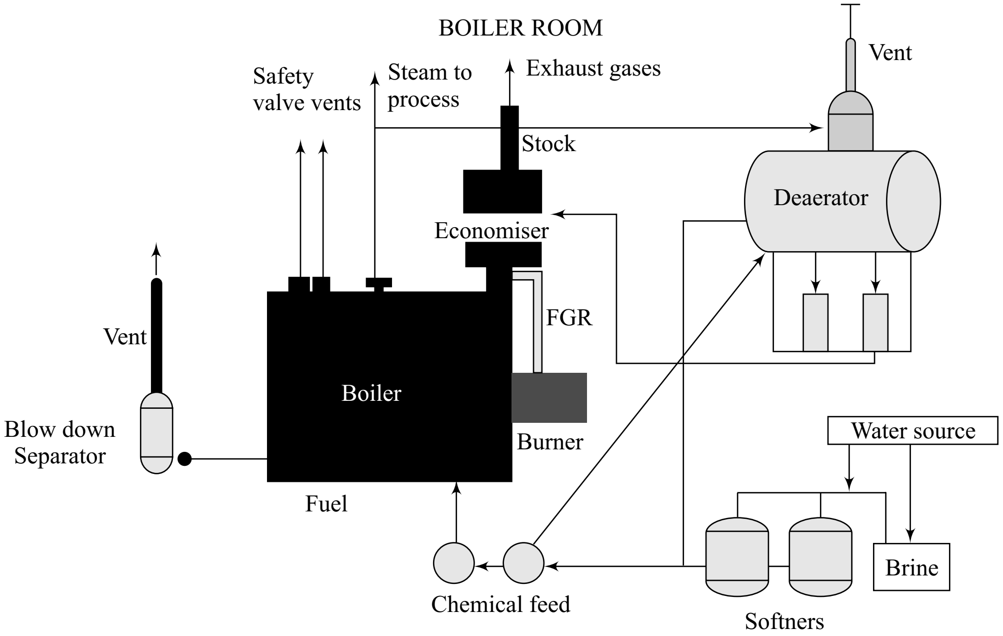


Figure 8.13 *Schematic boiler room.*

filtration of suspended solids. On the basis of proven satisfactory performance, cost and other considerations, cartridge filters are a practical solution to most problems of water clean-up.

- **Dissolved minerals**
  Dissolved minerals picked up by the water consist mainly of calcium carbonate (limestone), calcium sulphate (gypsum), magnesium carbonate (dolomite), magnesium sulphate (epsom salts), silica (sand), sodium chloride (common salt), hydrated sodium sulphate (Glauber salt) and smaller quantities of iron, manganese, fluorides, aluminium and other substances. The nitrates and phosphates found in water are usually due to sewage contamination.
- **Water hardness**
  Water containing high amounts of calcium and magnesium minerals is hard water. The amount of hardness in natural water can vary from a few ppm to 500 ppm. Calcium and magnesium compounds are relatively insoluble in water and tend to precipitate out. This causes scale and deposit problems. Such water must be treated to make it suitable for steam generation.
- **Other impurities in water**
  Natural waters contain varying levels of soil, sand, turbidity, colour, precipitated minerals, oil, industrial wastes and other suspended solid particles. Turbidity is due to very fine organic materials and micro-organisms, as well as suspended clay and silt. Colour is due to the decaying vegetable matter.

Problems caused by impurities in water and their remedial treatments for feedwater are shown below:

Impurity (Chemical Formula)	Problems	Common Chemical Treatment Methods
Alkalinity (HCO_3^-, CO_3^{2-} and $CaCO_3$)	Carryover of feedwater into steam, produce CO_2 in steam leading to formation of carbonic acid (acid attack)	Neutralizing amines, filming amines, combination of both, and lime-soda.

(*Contd.*)

(*Contd.*)

Impurity (Chemical Formula)	Problems	Common Chemical Treatment Methods
Hardness (calcium and magnesium salts, $CaCO_3$)	Primary source of scale in heat exchange equipment	Lime softening, phosphate, chelates and polymers
Iron (Fe^{3+} and Fe^{2+})	Causes boiler and water line deposits	Phosphate, chelates and polymers
Oxygen (O_2)	Corrosion of water lines, boiler, return lines, heat exchanger equipments, etc. (oxygen attack)	Oxygen scavengers, filming amines and deaeration
pH	Corrosion occurs when pH drops below 8.5	pH can be lowered by addition of acids and increased by addition of alkalies
Hydrogen Sulfide (H_2S)	Corrosion	Chlorination
Silica (SiO_2)	Scale in boilers and cooling water systems	Lime softening

8.11.2 Importance of Boiler Feedwater

Maintaining good feedwater is an important and fundamental aspect of any steam boilers. A plant that maintains good feedwater achieves the following three benefits:

1. Help to ensure maximum life out of its boilers, steam turbines, condensers and pumps.
2. Reduce maintenance expenses.
3. Maintain optimal thermal performance.

There is a definite legal and moral responsibility on the part of the management to ensure reliable, continuous and efficient operation of the steam boiler and to ensure that no damage to equipment or physical hazard occurs.

8.11.3 Boiler Feedwater

Boiler feedwater is the water supplied to the boiler. Often, steam is condensed and returned to the boiler as a part of the feedwater. The water needed to supplement the returned condensate is termed as make-up water. Make-up water is usually filtered and treated before use. Feedwater composition therefore depends on the quality of the make-up water and the amount of condensate returned. Sometimes people think that there is a great deal of similarity between the requirements for potable (drinking) water and the requirements for boiler feedwater. The minerals in drinking water are considered desirable and are absorbed by the body. On the other hand, minerals in water cannot be handled as well by boilers. Although a boiler is a big mass of steel, it is more sensitive to water impurities than the human stomach. For this reason, a lot of care is needed in filtration and treatment of the boiler water supply.

8.11.4 Problems Arise due to Impurities in Water for Boilers

In treatment of water, complete elimination of all the impurities is not possible. The impurity that gives rise to certain troubles will be removed to certain extent. In modern pressure boilers and laboratories, water purer than the distilled water is required. Some of the boiler troubles caused by the use of unsuitable water are:

- Carryover: priming and foaming
- Scale formation
- Boiler corrosion
- Caustic embrittlement
- Carryover: priming and foaming

(a) Priming

Priming is the carryover of varying amounts of droplets of water in the steam (due to high level of water in the boiler), which lowers the energy efficiency of the steam and leads to the deposition of salt crystals on the super heaters and in the turbines. This process of wet steam formation is called priming. Priming is common cause of high levels of boiler water carryover. These conditions often lead to super heater tube failures as well. Priming is related to the viscosity of the water and its tendency to foam. These properties are governed by alkalinity, the presence of certain organic substances and by total salinity or TDS. The degree of priming also depends on the design of the boiler and its steaming rate.

The cause for priming may be due to the following:

- Dissolved solids in water
- High steam velocities
- Sudden increase in steam production rate
- Sudden fluctuations in steam demand
- Due to improper boiler design

Priming in a boiler can be controlled by the following

- Maintaining the concentration of solids in the boiler water at reasonably low levels
- Avoiding high water levels
- Proper boiler design
- Fitting mechanical steam purifier
- Avoiding rapid change in steam-rate
- Uniform distribution of fuel and providing anti-priming pipes
- Keeping the water level low and avoid sudden steam demands and
- Efficient softening and filtration of the boiler feed water

(b) Foaming

The contamination of the steam in the boiler carryover is the impurities of water. Bubbles or froth actually build up on the surface of the boiler water and pass out with the steam. This is called foaming and it is caused by high concentration of any solids in the boiler water. It is generally believed, however, that specific substances such as alkalis, oils, fats, greases, certain types of organic matter and suspended solids are particularly conducive to foaming. In theory, suspended solids collect in the surface film surrounding a steam bubble and make it tougher. The steam bubble therefore resists breaking and builds up foam. It is believed that the finer the suspended particles the greater their collection in the bubble.

Foaming in boilers can be controlled by the following:

- Using anti-foaming chemicals
- Removal of concentrated boiler water containing impurities and replacing it by fresh feed water
- Removing oil from boiler water by adding compounds like sodium aluminate

Impurities present in water, problems and their remedial treatment methods are shown below:

Impurity (Chemical Formula)	Problems	Common Chemical Treatment Methods
Alkalinity (HCO_3^-, CO_3^{2-} and $CaCO_3$)	Carryover of feedwater into steam, produce CO_2 in steam leading to formation of carbonic acid (acid attack)	Neutralizing amines, filming amines, combination of both, and lime-soda.
Hardness (calcium and magnesium salts, $CaCO_3$)	Primary source of scale in heat exchange equipment	Lime softening, phosphate, chelates and polymers

(*Contd.*)

(*Contd.*)

Impurity (Chemical Formula)	Problems	Common Chemical Treatment Methods
Iron (Fe^{3+} and Fe^{2+})	Causes boiler and water line deposits	Phosphate, chelates and polymers
Oxygen (O_2)	Corrosion of water lines, boiler, return lines, heat exchanger equipments, etc. (oxygen attack)	Oxygen scavengers, filming amines and deaeration
pH	Corrosion occurs when pH drops below 8.5	pH can be lowered by addition of acids and increased by addition of alkalies
Hydrogen sulfide (H_2S)	Corrosion	Chlorination
Silica (SiO_2)	Scale in boilers and cooling water systems	Lime softening

8.11.5 Scale and Sludge Deposits in Boiler

Water evaporating in the boiler causes impurities to concentrate. Boiler scale is caused by impurities being precipitated out of the water directly on heat transfer surfaces or by suspended matter in water settling out on the metal and becoming hard and adherent. As minerals are deposited out from water, they form many types of crystalline and rock-like structures. The most common scale in boilers is due to carbonate deposits caused by hardness.

Carbonate scale is usually granular and sometimes very porous. Sulphates scales are harder and denser. Silica scales resemble porcelain and are very brittle. Iron deposits are very dark coloured that is either due to corrosion or iron contamination in the water. Deposits are seldom composed of one constituent alone, but are generally a mixture of various types of solid sediments, dissolved minerals, corrosion products like rust and other water contaminants.

(i) Scale formation

Scales are hard, adhering precipitates formed on the inner walls of the boilers. They stick very firmly onto the inner wall surface and are difficult to remove with chisel and hammer. Generally, scales are formed due to decomposition of calcium bicarbonates, decomposition of calcium sulphates, presence of silica and hydrolysis of magnesium salts. Scales formation can be prevented by external treatment of boiler feed water, internal treatment of water in boiler and blowdown process (Fig. 8.14).

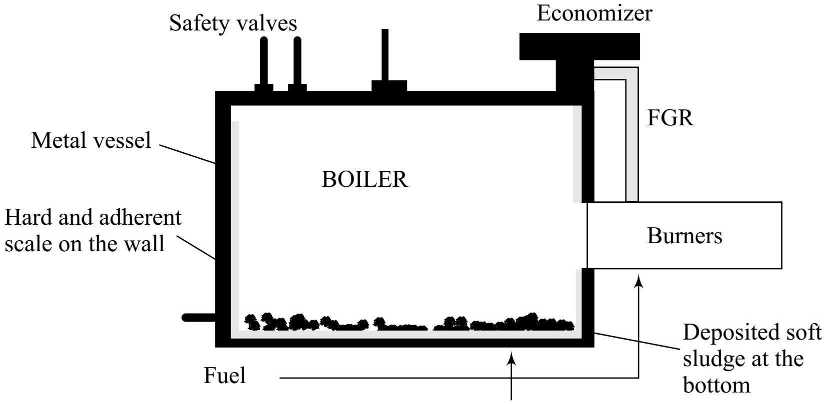


Figure 8.14 *Scale and sludge deposit in boiler.*

Reasons of scale formation

(a) Following are the causes

- decomposition of calcium bicarbonate:

$$Ca(HCO_3)_2 \rightarrow CaCO_3 + H_2O + CO_2$$

- In low pressure boilers, $CaCO_3$ causes scale formation.
- In high pressure boilers, $CaCO_3$ becomes soluble.

$$CaCO_3 + H_2O \rightarrow Ca(OH)_2 + CO_2$$

(b) Decomposition of calcium sulphate:
The solubility of $CaSO_4$ in water decreases with rise of temperature. In super-heated water, $CaSO_4$ is insoluble. This is the main cause in high-pressure boilers.

(c) Hydrolysis of magnesium salts**:**
Dissolved magnesium salts undergo hydrolysis forming $Mg(OH)_2$ precipitate.

$$MgCl_2 + 2H_2O \rightarrow Mg(OH)_2 + 2HCl$$

$Mg(OH)_2$ so found by hydrolysis of magnesium salts is a soft type of scale.

(d) Presence of silica:
Silica is present in small quantity deposits as silicates like $CaSiO_3$ and $MgSiO_3$. These are very difficult to remove.

8.11.6 Problems Caused by Scale in Boilers

Scale is one of the most common deposit related problems. Scale is a build-up of solid material from the reactions between impurities in water and tube metal, on the water-side tube surface. The biggest problem caused by scale is overheating and failure of boiler tubes. The thermal conductivity of porous boiler scale is similar to insulating brick. The scale acts as an insulating layer and prevents an efficient transfer of heat through the tubes to the circulating water. The reduction in thermal conductivity means lower boiler efficiency which in turn leads to overheating and may result in the softening, bulging or even fracturing of the boiler tubes. Boiler scale can also cause plugging or partial obstruction of circulating tubes in a water tube boiler, which again causes starvation and overheating of the tubes.

Another important aspect is that corrosion may occur under the boiler scale. In general, boiler scale causes:

- increased fuel bill by decreasing the operating efficiency
- thermal damage
- unscheduled down-time
- increased cleaning time and cleaning costs
- reduced working life of a boiler

Prevention

(a) External treatment: Efficient softening of water is to be carried out.

(b) Internal treatment: Suitable chemicals are added to the boiler water either to precipitate or to convert the scale into compounds.

8.11.7 Sludge Formation in Boilers

In boilers, because of continuous evaporation of water, the concentration of salts increase progressively and after the saturation point is reached, precipitate forms on the inner walls of boiler.

Sludge is a soft, loose and slimy precipitate formed within the boiler. It is formed at comparatively colder portions of the boiler and collects in the area where flow rate is slow.

These are formed by substances which have greater solubilities in hot water than in cold water.

Example $MgCO_3$, $MgCl_2$, $CaCl_2$, $MgSO_4$, etc.

Disadvantages

- As the sludge is poor conductor of heat, it causes loss of heat.
- The working of the boiler is disturbed because of chocking of pipes by the sludge.

Prevention

(a) By using well softened water.
(b) By drawing off a portion of concentrated water frequently.

8.11.8 Boiler Corrosion

The process of corrosion is actually not so simple, it is a complex electro-chemical reaction. Corrosion may generally be over a large metal surface but sometimes it results in pinpoint penetration of metal. Though basic corrosion is usually due to reaction of the metal with oxygen, other factors including stresses produce different forms of attack. Corrosion may occur in the feedwater system as a result of low pH water and the presence of dissolved oxygen and carbon dioxide. Corrosion in the boiler itself normally occurs when boiler water's alkalinity is too low or too high or when the metal is exposed to oxygen-bearing water during either operation or idle periods. High temperatures and stresses tend to accelerate the corrosion. Corrosion in steam condensate system and pipelines is caused by carbon dioxide and oxygen carried into the system by steam. Dissolved carbon dioxide in condensed steam forms corrosive carbonic acid. If oxygen is present with carbon dioxide, the corrosion rate is much higher, and is likely to produce localized pitting. Ammonia, in combination with oxygen, attacks copper alloys.

Corrosion problems

Uniform corrosion of boiler metal surfaces is bound to occur and is not of much concern as all boilers experience a small amount of general corrosion. Corrosion, however, takes many forms and deep pitting that causes only a small amount of total iron loss causes penetration and leakage in boiler tubes. Corrosion beneath certain types of boiler deposits can weaken the metal and cause tube failure. Likewise, corrosion in steam condensate system can damage pipelines and equipment.

Prevention of steam condensate corrosion

Generally, corrosion prevention is by removing oxygen from the feedwater by mechanical (deaerator) means, by use of suitable chemicals and pretreatment of the make-up water to minimize potential carbon dioxide formation in the boiler. Further, boiler water treatment is done by use of volatile amines to neutralize carbon dioxide or volatile filming inhibitors to form a barrier between the metal and the corrosive condensate. Mechanical conditions need to be checked and corrected, like poor trapping and draining of lines. Deaerator can reduce oxygen to as low as 0.007 ppm. Since very small amounts of oxygen can cause boiler and steam condensate system corrosion, chemical treatment is needed to assure complete oxygen removal. Sodium sulphite and hydrazine chemicals are commonly used for this purpose. Catalysts are sometimes also used to speed up the reaction.

Prevention of deposits and water corrosion in feedwater systems

Deposits in feedwater systems are usually caused by hardness precipitation as the water goes through feedwater heaters or as the feed lines enter the boilers. Deposits can also occur from premature reaction of treatment chemicals with hardness in the feedwater. Prevention is by means of the use of stabilizing chemicals fed continuously to retard hardness precipitation. The corrosion of feedwater system is due to the low alkalinity or dissolved oxygen in the water. Raising the pH of the water with caustic or amines and feed of catalysed sodium sulphite minimizes this problem.

Deaeration of water

Dissolved oxygen in water is a major cause of boiler system corrosion. It should be removed before the water is put in the boiler. Feedwater deaeration removes oxygen by heating the water with steam in a deaerating heater. Part of the steam is vented, carrying with it the bulk of the dissolved oxygen.

8.11.9 Caustic Embrittlement

The formation of brittle and incrystalline cracks in the boiler shell is called caustic embrittlement. It is a type of boiler corrosion and the main reason for this is the presence of alkali-metal carbonates and bicarbonates in feed water and also the presence of sodium sulphate. In lime-soda process, it is likely that, some residual Na_2CO_3 is still present in the softened water. This Na_2CO_3 decomposes to give NaOH and CO_2, due to which the boiler water becomes 'Caustic'.

$$Na_2CO_3 + H_2O \rightarrow 2NaOH + CO_2$$

This caustic water flows inside the boiler and causes some minute hair-cracks by capillary action. On evaporation of water, the dissolved caustic soda increases its concentration which attacks the surrounding area, thereby dissolving iron of boiler as sodium ferroate. This causes embrittlement of boiler parts such as bends, joints, reverts, etc., due to which the boiler gets fail. Caustic cracking can be explained by considering the following concentration of cell:

Iron at bends, iron at plane surfaces joints, reverts etc., NaOH solution

+|Concentrated| Dilute NaOH solution |–

Therefore, caustic embrittlement can be prevented

(a) By maintaining the pH value of water and neutralization of alkali
(b) By using sodium phosphate as softening reagents, in the external treatment of boilers.
(c) Caustic embrittlement can also be prevented by adding tannin or lignin or sodium sulphate which prevents the infiltration of caustic-soda solution blocking the hair-cracks.

Prevention of caustic embrittlement

Organic materials like lignins applied to the boiler water are effective in preventing caustic embrittlement of boiler metal. Sodium nitrate inhibits embrittlement at low concentrations up to 0.4 part of sodium nitrate per part of caustic soda in boiler water. Maintaining an organic content of 50–100 ppm and a sodium nitrate content of 50 ppm is a commonly used embrittlement prevention program.

8.11.10 Oil Contamination in Boiler, Problem and Remedies

Main problems caused by oil in the boiler water are:

- Oil can coat metal surfaces, cut down heat transfer and produce metal overheating
- Oil can cause sludge to become sticky and adhere to heat transfer surfaces
- Oil can produce foaming and boiler water carryover

Oil contamination should be completely eliminated whenever possible. Organic chemicals help counteract the effects of small amounts of oil contamination, but not of gross contamination. When sudden boiler water oil contamination is experienced, normal procedure is to blow down heavily to remove oil and to check for the source of contamination. In case of severe contamination, the boiler needs to be taken off the line and cleaned out to remove the oil from the boiler surfaces. When oil contamination is continuous and unavoidable, some of the methods used are:

- Free oil can be reduced by passing the water through absorbent cartridge filters
- Emulsified oil is broken down by chemical additives and filtered
- Special filters are used with aids like diatomaceous earth
- Flotation method
- Coalescence method

8.11.11 Purity Requirements of Feedwater

Feedwater is a matter both of quantity of impurities and the nature of impurities. Hardness, iron and silica, for example, are of more concern than sodium salts. The purity requirements of feedwater depend on how much feedwater is used as well as toleration of the particular boiler design (pressure, heat transfer rate, etc.). In today's high-pressure boilers, practically all impurities must be removed. The feedwater (make-up water) from outside needs to be treated for the reduction or removal of impurities by the filtration first, and then followed by softening, evaporation, deaeration, ion exchange, etc. Internal treatment is also required for the conditioning of impurities within the boiler system, to control corrosion, as reactions occur in the boiler itself and the steam pipelines.

Boiler feedwater treatment

Water is an essential medium for steam generation. Conditioning it properly can increase the efficiency of boiler as well as extended boiler's operating life. Treating boiler water also ensures safe and reliable operation.

The treatment and conditioning of boiler feed water must satisfy three main objectives:

- Continuous heat exchange
- Corrosion protection
- Production of high quality steam
- Removing impurities from boiler feedwater

Feedwater is filtered to remove suspended matter and if the suspended solids are very fine, a flocculation step may be needed to enable effective filtration. The water is then subjected to other treatments to make it suitable for the boiler. Depending on the quality of water, it may be subjected to one or more treatments like chemical precipitation, lime-soda softening, ion exchange, deaeration and reverse osmosis.

- **Filtration**

  Filtration is the essential first step before the chemical treatment and conditioning of the boiler feedwater. Filtration removes or minimizes all types of suspended solid impurities. If rust, sand (silica), etc. are not filtered out, they lead to severe scale formation, then it is difficult to clean and reduce boiler efficiency. Even the condensate feedwater must be filtered before returning to the boiler. The boiler itself and the steam piping produce rust particles, etc. due to corrosion and other reactions. Filtration is also necessary for any water treatment process to work properly. For example, softening resins get coated with suspended matter, losing their effectiveness and capacity to regenerate. Reverse osmosis membranes get fouled up leading to reduced efficiency and shorter life. If the water is very dirty, sand filtration is initially done followed by cartridge filtration.

- **Coagulation and flocculation**

  Sometimes the suspended particles in water are so fine that even cartridge filters are unable to remove them. In such a situation, before cartridge filtration, the water is first treated with coagulants. Coagulation is charged neutralization of finely divided and colloidal impurities in water into masses that can be filtered. In addition, particles have negative electrical charges that cause them to repel each other and resist adhering together. Coagulation, therefore, involves neutralizing the negative charges and providing a nucleus for the suspended particles to adhere to. Flocculation is the bridging together of coagulated particles.

8.11.12 Types of Coagulants

Iron and aluminium salts such as ferric sulphate, ferric chloride, aluminium sulphate (alum) and sodium aluminate are the most common coagulants. Ferric and alumina ions each have three positive charges and therefore their effectiveness is related to their ability to react with the negatively charged colloidal particles. These coagulants form a floc in the water which serves like a net for collecting suspended matter. Polyelectrolytes, which are synthetic materials, have been developed for coagulation purposes. These consist of long chain like molecules with positive charges. In some cases, organic polymers and special types of clay are used in the coagulation process to serve as coagulant aids. These assist in coagulation by making the floc heavier.

- **Chemical precipitation**
  Internal treatment of boiler water is carried out by adding proper chemicals to precipitate the scale forming impurities in the form of sludge and to convert the scale forming chemicals into compounds which will stay in dissolved form in water. This process is mainly used as a corrective treatment to remove the slight residual hardness and also sometimes to remove the corrosive tendencies in water. Some of the internal treatment methods used for the removal of scale formation in boilers are as follows:
- **Colloidal conditioning**
  The addition of organic substances such as kerosene, tannin, gel, etc. to the surface in low pressure boilers may prevent the scale formation. These substances gets coated over the scale forming precipitates and give a loose and nonsticky precipitates that can be removed by using blowdown operation.
- **Phosphate conditioning**
  The addition of sodium phosphate in hard water reacts with the hardness causing agents and gives calcium and magnesium phosphates that are soft and nonadhere and can be removed easily by blowdown operation. In this way, scale formation is removed in high-pressure boilers.

$$3CaCl_2 + 2Na_3PO_4 \rightarrow Ca_3(PO_4)_2 + 6NaCl$$

- **Carbonate conditioning**
  Chemical precipitation is a process in which chemical added reacts with dissolved minerals in the water to produce a relatively insoluble reaction product. Precipitation methods are used in reducing dissolved hardness, alkalinity and silica. The most common example is lime-soda treatment.

(a) Reaction of lime and soda in softening process

Calcium hydroxide (hydrated lime) reacts with soluble calcium and magnesium carbonates to form insoluble precipitates. They form a sludge that can be removed by settling and filtration. Lime, therefore, can be used to reduce hardness present in the bicarbonate form (temporary hardness) as well as decrease the amount of bicarbonate alkalinity in water. Lime reacts with magnesium sulphate and chloride and precipitates magnesium hydroxide, but in this process soluble calcium sulphate and chlorides are formed. Lime is not effective in removing calcium sulphates and chlorides (permanent hardness). Soda ash is used primarily to reduce nonbicarbonate hardness (permanent hardness). The calcium carbonate formed by the reaction precipitates as sludge and can be filtered out. The resulting sodium sulphate and chloride are highly soluble and nonscale forming.

8.11.13 Methods of Lime-Soda Softening

The older method of intermittent softening consists of mixing the chemicals with the water in a tank, allowing time for reaction and forming of sludge and filtering and drawing off the clear water. The modern method of continuous lime-soda softening involves use of compartmented tanks with provision for (a) proportioning chemicals continuously to the incoming water, (b) retention time for chemical reactions and sludge formation and (c) continuous draw-off of softened water. Lime-soda softening is classified as hot or cold, depending on the temperature of the water. Hot process softeners increase the rate of chemical reactions, increase silica reduction and produce over-all better quality water.

Coagulants used in lime-soda process

In the initial clarification process, coagulants are used to agglomerate fine suspended particles that can then be filtered out. Likewise, in the softening process, coagulants speed up settling of sludge by 25–50 percent. Sodium aluminate used as a coagulant in lime-soda softening being alkaline, also contributes to the softening reactions, particularly in reducing magnesium. Proper uses of coagulants help in removing silica in the softening process. Silica tends to be adsorbed on the floc produced by coagulation of sludge.

Disadvantages of lime-soda softening

The main disadvantage is that the hardness is not completely removed. Variations in raw water composition and flow rate also make control of this method difficult since it involves adjusting the amounts of lime and soda ash being fed.

Advantages of lime-soda softening

The main advantage is that while reducing hardness; alkalinity, total dissolved solids and silica are also reduced. Prior clarification of the water is not usually necessary with the lime-soda process. Another advantage is that with continuous hot process softening, some removal of oxygen and carbon dioxide can be achieved. Fuel savings can be realized with the hot process softening because of solids reduction. This reduction decreases the conductivity of the feedwater, thereby decreasing blow down and conserving heat.

In low-pressure boilers, scale-formation can be avoided by adding sodium carbonate to the boiler water, when $CaSO_4$ is converted into calcium carbonate in equilibrium.

$$CaSO_4 + Na_2CO_3 \rightarrow CaCO_3 + Na_2SO_4$$

Consequently, deposition of $CaSO_4$ as scale does not take place and calcium is precipitated as loose sludge of $CaCO_3$ which can be removed by blowdown operation.

- **Calgon conditioning**

  Involving of adding calgon to boiler water prevents the scale and sludge formation by forming soluble complex compound with $CaSO_4$.

  $$\text{Calgon} = \text{Sodium hexameta phosphate} = (NaPO_3)_6$$

  $$Na_2\,[Na_4(PO_3)_6] \rightarrow 2Na+ + [Na_4P_6O_{18}]^{2-}$$

  $$2CaSO_4\,[Na_4P_6O_{18}]^{2-} \rightarrow [Ca_2P_6O_{18}]^{2-} + 2Na_2SO_4$$

- **Sodium aluminate conditioning**

  Sodium aluminate gets hydrolysed yielding NaOH and a gelatinous precipitate of aluminium hydroxide.

  Thus, $NaAlO_2 + 2H_2O \rightarrow NaOH + Al(OH)_3$

  The sodium hydroxide, so formed, precipitates some of the magnesium as $Mg(OH)_2$, that is

  $$MgCl_2 + 2NaOH \rightarrow Mg(OH)_2 + 2NaCl$$

The flocculent precipitate of $Mg(OH)_2$ plus aluminium hydroxide, produced inside the boiler, entraps finely suspended and colloidal impurities, including oil drops and silica. The loose precipitate can be removed by predetermined blowdown operation.

8.11.14 Chemical Treatment of Boiler Feedwater

The primary function of a boiler is to transfer heat from hot gases generated by the combustion of fuel into water until it becomes hot or turns to steam. The steam or hot water can then be used in building or facilitate processes. Boiler feedwater often contains impurities that impair boiler operation and efficiency. Chemical additives can be used to correct the problems caused by these impurities. To improve feedwater quality and steam purity, these chemicals can be injected directly into the feedwater or steam.

Benefits of chemical treatments

- Increase boiler efficiency;
- Reduce fuel, operating and maintenance costs;
- Minimize maintenance and downtime; and
- Protect equipment from corrosion and extend equipment lifetime.

The feedwater is composed of makeup water (usually city water from outside boiler room/process) and condensate (condensed steam returning to the boiler). The feedwater normally contains impurities that can cause deposits and other related problems inside the boiler. Common impurities in water include alkalinity, silica, iron, dissolved oxygen and calcium and magnesium (hardness). Blowdown, a periodic or continuous water removal process, is used to limit the concentration of impurities in boiler water and to control the build-up of dissolved solid levels in the boiler. Blowdown is essential in addition to chemical treatments.

Generally, boiler feed water are subjected to both internal and external treatments for conditioning of impurities of water. Internal treatments are the conditioning of impurities within the boiler system. The reactions occur either in the feed lines or in the boiler properly.

- 'Internal treatment' may be used alone or in conjunction with external treatment. Its purpose is to properly react with feed water hardness, condition sludge, scavenge oxygen and prevent boiler water foaming.
- 'External treatment' is the reduction or removal of impurities from water outside the boiler. In general, external treatment is used when the amount of one or more of the feed water impurities is too high to be tolerated by the boiler system in question. There are many types of external treatment (softening, evaporation, deaeration, membrane contractors, etc.) which can be used to tailor-make feed-water for a particular system.

8.11.15 Internal Treatment of Boiler Feedwater

Internal treatment of water inside the boiler is essential whether or not the feedwater has been pretreated. Internal treatment compliments external treatment and is required regardless of whether the impurities entering the boiler with the feedwater are large or small in quantity. In some cases, feedwater supply needs to be only filtered without the need for any other external treatment. Internal treatment can constitute the sole treatment when boilers operate at low pressure, large amounts of condensed steam are used for feedwater, or the raw water available is of good quality. However, in moderate or high-pressure boilers, external treatment of the makeup water is mandatory for good results. With today's boilers having higher heat transfer rates, even a small deposit can cause tube failure or fuel wastage.

The purpose of an internal water treatment program is:

- To react with incoming feedwater hardness and prevent it from precipitating on the boiler metal as scale
- To condition any suspended matter such as hardness sludge in the boiler and make it nonadherent to the boiler metal

- To provide antifoam protection to permit a reasonable concentration of dissolved and suspended solids in the boiler water without foaming
- To eliminate oxygen from the feedwater
- To provide enough alkalinity to prevent boiler corrosion
- To prevent scaling and protect against corrosion in the steam-condensate systems
- Chemical used for internal treatment of boiler water

There are three major classifications of chemicals used in internal treatment:

- phosphates,
- chelants, and
- polymers.

These chemicals may be fed either separately or in combination; in most balanced treatment programmes, two or three chemicals are fed together. The preferred feed point varies with the chemicals specified. For example, when caustic soda is used to maintain boiler water alkalinity, it is fed directly to the boiler drum. When caustic is used to adjust the feedwater pH, it is normally injected into the storage section of the deaerating heater.

1. Phosphate

Mono-, di- or tri-sodium phosphate and sodium polyphosphate can be added to treat boiler feedwater and buffer the water to minimize pH fluctuation. It also precipitates calcium or magnesium into a soft deposit rather than a hard scale. Additionally, it helps to promote the protective layer on boiler metal surfaces. However, phosphate forms sludge as it reacts with hardness; blowdown or other procedures should be established to remove the sludge during a routine boiler shutdown. Phosphates are usually fed directly into the steam drum of the boiler, although they may be fed to the feedwater line under certain conditions. Treatments containing orthophosphate may produce calcium phosphate feedline deposits; therefore, they should not be fed through the boiler feedline. Orthophosphate should be fed directly to the boiler steam drum through a chemical feedline. Polyphosphates must not be fed to the boiler feedwater line when economizers, heat exchangers, or stage heaters are a part of the preboiler system. If the preboiler system does not include such equipment, polyphosphates may be fed to the feedwater piping provided that total hardness does not exceed 2 ppm. In all cases, feed rates are based on feedwater hardness levels. Phosphates should be fed neat or diluted with condensate or high-purity water. Mild steel tanks, fittings and feedlines are appropriate. If acidic phosphate solutions are fed, stainless steel is recommended.

Phosphates used to be the main conditioning chemical, but nowadays chelate and polymer type chemicals are mostly used. These new chemicals have the advantage over phosphates of maintaining scale-free metal surfaces. All internal treatment chemicals, whether phosphate, chelate or polymer, condition the calcium and magnesium in the feedwater.

2. Chelants

Nitrilotriacetic acid (NTA) and ethylenediamine tetraacetic acid (EDTA) are the most commonly used chelants. Chelants combine with hardness in water to form soluble compounds. The compounds can then be eliminated by blowdown.

- Chelants treatment is not recommended for feedwater with high hardness concentration
- Chelants should not be fed if the feedwater contains a significant level of oxygen
- Chelants should never be fed directly into a boiler

The preferred feed location for chelants is downstream of the feedwater pump. All chelant treatments must be fed to the boiler feedwater line by means of a stainless steel injection nozzle at a point beyond the discharge of the boiler feed pumps. If heat exchangers or stage heaters are present in the boiler feedline, the injection point should be at their discharge. Care should be exercised in the selection of metals for high-temperature injection quills.

At feed solution, strength and elevated temperatures, chelating agents can corrode mild steel and copper alloys; therefore, 304 or 316 stainless steel is recommended for all feed equipment. Equipment specifications are the same as those for regular sulphite. Chelant products may be fed neat or diluted with condensate. Chelant feed rates must be carefully controlled based on feedwater hardness because misapplication can have serious consequences.

3. Polymeric dispersants

In most applications, polymeric dispersants are provided in a combined product formulation with chelants and/or phosphates. Dilution and feed recommendations for chelants should be followed for chelant-dispersant and chelant-phosphate-dispersant programs. Dilution and feed recommendations for phosphates should be followed for phosphate-dispersant programs. This combination of programs typically has the best results with respect to boiler cleanliness.

4. Neutralizing amines

Neutralizing amines are high pH chemicals that neutralize the carbonic acid formed in the condensate (acid attack). The three most commonly used neutralizing amines are morpholine, diethyleminoethanal (DEAE) and cyclohexylamine. Neutralizing amines cannot protect against oxygen attack; however, it helps to keep oxygen less reactive by maintaining an alkaline pH. Neutralizing amines may be fed to the storage section of the deaerating heater, directly to the boiler with the internal treatment chemicals, or into the main steam header. Some steam distribution systems may require more than one feed point to allow proper distribution. An injection quill is required for feeding into a steam distribution line. Neutralizing amines are usually fed based on condensate system pH and measured corrosion rates. These amines may be fed neat, diluted with condensate or demineralized water, or mixed in low concentrations with the internal treatment chemicals. A standard packaged pump skid and tank can be used for feeding.

5. Filming Amines

Filming amines are various chemicals that form a protective layer on the condensate piping to protect it from both oxygen and acid attack. The two most common filming amines are octadecylamine (ODA) and ethoxylated soya amine (ESA). Combining neutralizing and filming amine is a successful alternative to protect against both acid and oxygen attack.

The filming amines should be continuously fed into steam headers at points that permit proper distribution. A single feed point is satisfactory for some systems. In every case, the steam distribution should be investigated and feed points established to ensure that all parts of the system receive proper treatment. Filming amines must be mixed with condensate or demineralized water. Water containing dissolved solids cannot be used, because the solids would contaminate the steam and could produce unstable amine emulsions.

The use of stainless steel tanks is recommended. Equipment specifications are the same as those for regular sulphite, except that a vapour-type injection nozzle or quill is required.

(a) Internal treatment for hardness

At boiler operating temperature, calcium carbonate in the feedwater breaks down to form calcium carbonate. Since it is relatively insoluble, it precipitates. Sodium carbonate in the water partially breaks down to sodium hydroxide and carbon dioxide. Internal treatment with phosphates transforms calcium bicarbonate to calcium phosphate and sodium carbonate. In the presence of hydroxide alkalinity, magnesium bicarbonate precipitates as magnesium hydroxide or reacts with silica to form magnesium silicate. These minerals are precipitated from solution in the form of sludge that must be conditioned to prevent its sticking to the boiler metal. The conditioned sludge is then removed from the boiler by blowdown. When chelate is used for internal treatment, it reacts with calcium and magnesium salts to form soluble complexes. These complexes are in the form of dissolved solids and are removed by blowdown. Dispersant polymers used in conjunction with chelate produces reaction products, which are better conditioned. They do not precipitate and are more easily removed by blowdown. Use of polymers further aids in conditioning any suspended solid contamination that may have entered with the boiler feed water.

(b) Internal treatment for sulphates

The boiler temperature makes the calcium and magnesium sulphates in the feedwater insoluble. With phosphates used as internal treatment, calcium reacts with the phosphate producing hydroxyapatite, which is much easier to condition than calcium sulphate. With chelates or polymer used as internal treatment, calcium and magnesium react with these materials producing soluble complexes that are easily removed by blowdown.

(c) Internal treatment for silica

If silica is present in the feedwater, it tends to precipitate directly as scale at hot spots on the boiler metal and or combines with calcium forming a hard calcium silicate scale. In the internal treatment for silica, the boiler water alkalinity has to be kept high enough to hold the silica in solution. Magnesium, present in most waters, precipitates some of the silica as sludge. Special organic materials or synthetic polymers are used to condition magnesium silicate from adhering to the boiler metal.

(d) Internal treatment for sludge conditioning

Internal treatment for hardness results in insoluble precipitates in the boiler that form sludge. In addition, insoluble corrosion particulates (metal oxides) are transported to the boiler by condensate returns and from preboiler feedwater corrosion resulting in suspended solids. Suspended solids, carried to the boiler by feedwater or subsequently formed within the boiler, adversely affect both boiler cleanliness and steam purity. These solids have varying tendency to deposit on the boiler metal. Conditioners prevent these solids from depositing and forming corrosive or insulating boiler scale.

Some of the principal types of sludge conditioners are:

- Starches: effective on high silica feedwater and where oil contamination is a problem
- Lignins: effective on phosphate type sludge
- Tannins: fairly effective on high hardness feedwater
- Synthetic polymers: highly effective sludge conditioners for all types of sludges

(e) Internal treatment advantages

Internal treatment is basically simple and with the help of a qualified consultant an effective program is easily established. Scales or deposits, corrosion and carryover are minimized thereby improving efficiency and reducing energy consumption, preventing tube failures and unscheduled costly repairs, and reducing deposits, corrosion and contamination in the downstream equipment or processes.

8.11.16 Internal Treatment Chemical Dosage

Chemical dosages are based on the amount and type of impurities in the feedwater. For example, the amount of boiler treatment chemicals depends on the feedwater hardness; the amount of sodium sulphite depends on the amount of dissolved oxygen in the feedwater. In addition, a certain amount of extra chemicals are added to provide a residual in the boiler water. Feeding methods include chemical solution tanks and proportioning pumps with the chemicals being added directly before entrance to the boiler. Ortho-type phosphates are fed through a separate line directly into the steam drum of the boiler. Chemicals used to prevent condensate system corrosion can be fed directly to the feedwater, boiler or steam, depending on the type of chemical used. Continuous feeding is the preferred method but intermittent feeding may also suffice in some cases.

Tests for treatment control

Boiler water control tests include tests for alkalinity, phosphate, polymer, hydrazine, chelate, sulphite, pH, organic colour and total dissolved solids. If sulphite test shows adequate residual is present (this test is not valid with use of uncatalysed sodium sulphite), the feedwater oxygen has been removed. Testing for organic colour gives an indication for both sludge conditioner and anti-foam level. Chelate testing can be either for total chelate or residual chelate.

Tests for checking contaminants

These tests depend on what type of contaminants is suspected. Common checks are for iron, oil and silica. Total iron tests serve as a check on corrosion products brought back by the condensate return. This test is also used to check iron present in the makeup water. Laboratory facility is required for oil test but visual inspection can reveal gross contamination. Periodic checking should be done to detect unusual silica contamination and to determine when blowdown is needed.

8.11.17 Blowdown

Blowdown is the discharge of boiler water containing concentrated suspended and dissolved feedwater solids. As the blowdown water is replaced with lower solids feedwater, the boiler water is diluted. With proper regulation of blowdown, the amount of solids in the boiler water can be controlled. The amount of blowdown needed depends on how much feedwater impurities a given boiler can tolerate. For example, if a particular boiler can tolerate 500 ppm maximum dissolved solids, and the feedwater contains 50 ppm, it can be concentrated only about 10 times. This means that for every 100 L of water fed to the boiler about 10 L of boiler water must be blown down to keep the dissolved solids from exceeding 500 ppm. Total dissolved solids are not the only limiting factors in determining blowdown, other considerations include suspended solids, alkalinity, silica and iron.

8.12 Flash Evaporation

Flash evaporation is the most widely used method of water desalination. *The seawater is heated and then pumped into a low-pressure tank, where the water is partially vaporized. The water vapour is then condensed and removed as pure water. This process is repeated many times (three stages are shown).* The remaining liquid, called brine, contains a large amount of salt and is removed and often processed for minerals. Note that the incoming seawater is used to cool the condensers in each evaporator. This design conserves energy since the heat released when the condensed vapour is used to heat the next batch of seawater.

It is an advantage to perform flash-distillation in several steps. This is called multi-stage flash evaporation as shown in Fig. 8.15. This is where the preheated liquid passes through a series of stages or chambers with each successive stage at a lower vapour pressure so some of the liquid will flash at each stage. Flash distillation begins when the salt water enters a bundle of tubes that are located in the vapour space of the preheat chamber. The water then flows into a heater consisting of a bundle of tubes that are heated externally by steam. Here, the water is heated to 100°C, but it does not boil because the pressure is above 1 atmosphere. The hot sea water then enters a flash chamber that is kept under reduced pressure. The vapours which are flashed off are then condensed on the tubes carrying the incoming flow of cold sea water. The distillate and the remaining salt water are then restored to atmospheric pressure by pumps. Condensing the water by heat exchange with the incoming flow is one of the economical advantages of this process.

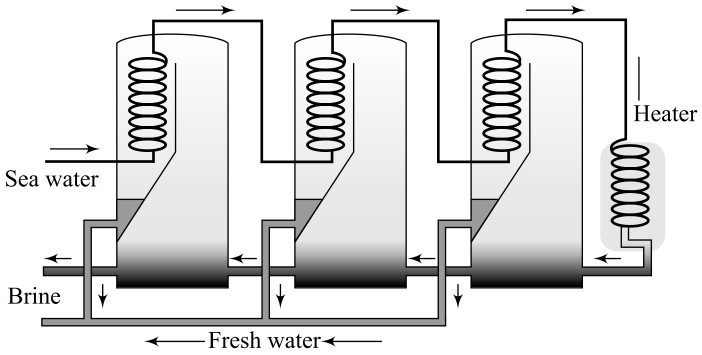
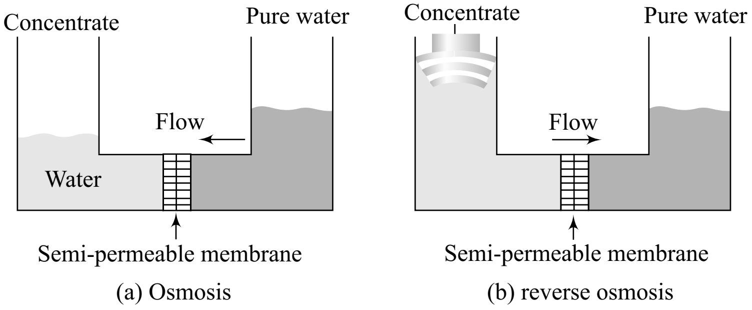


FIGURE 8.15 *Flash evaporation technique for desalination of water.*

8.13 Water Purification by Reverse Osmosis

Basic principle of Reverse Osmosis (RO)

Reverse osmosis is a complicated process that uses a membrane under pressure to separate relatively pure water (or other solvent) from a less pure solution. *When two aqueous solutions of different concentrations are separated by a semi-permeable membrane, water passes through the membrane in the direction of more concentrated solution as a result of osmotic pressure as shown in Fig. 8.16a.* If enough counter pressure is applied to the concentrated solution to overcome the osmotic pressure, the flow of water will be reversed (Fig. 8.16b).



(a) Osmosis

(b) reverse osmosis

FIGURE 8.16 *(a) and (b) A typical reverse osmosis cell.*

Reverse osmosis is known as the finest hyper-filtration technique. This process will allow the removal of particles as small as dissolved individual ions from a solution as shown in Fig. 8.17.

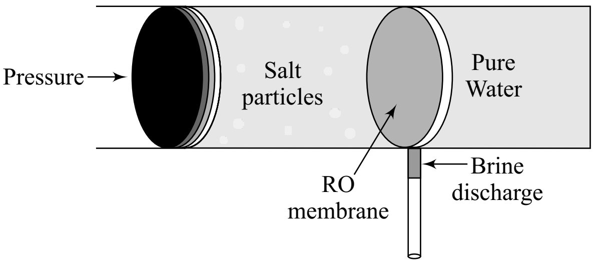


FIGURE 8.17 *Function of RO membrane and purification of water by reverse osmosis.*

Reverse osmosis is used to purify water and remove ions and dissolved organic molecules. The most common use for reverse osmosis is in purifying water. It is used to produce water that meets the most demanding specifications that are currently in place.

A water treatment process whereby dissolved salts such as sodium, chloride, calcium carbonate and calcium sulphate may be separated from water by forcing the water through a semi-permeable membrane under high pressure. The water diffuses through the membrane and the dissolved salts remain behind on the surface of the membrane.

Water molecules can form hydrogen bonds in the RO membrane and fit into the membrane matrix. The water molecules that enter the membrane by hydrogen bonding can be pushed through under pressure. Most organic substances with a molecular weight over 100 are sieved out, i.e. oils, pyrogens and particulates including bacteria and viruses. Salt ions, on the other hand, are rejected by a mechanism related to the valence of the ions. Ions are repelled by dielectric interactions; ions with higher charges are repelled to a greater distance from the membrane surface. The nominal rejection ratio of common ionic salts is 85–98 percent.

Membrane for reverse osmosis

The majority of the commercially manufactured RO membranes are usually made from *cellulose acetate, polysulphonate and polyamide*. The membrane consists of a skin about 0.25 microns and a support layer about 100 microns. The skin is the active barrier and primarily allows water to pass through. The process of reverse osmosis requires a driving force to push the fluid through the membrane, and the most common force is pressure from a pump. The higher the pressure the larger the driving force as shown in Fig. 8.17.

Advantages

- Reverse osmosis is capable of rejecting bacteria, salts, sugars, proteins, particles, dyes and other constituents that have a molecular weight of greater than 150–250 daltons.
- The maintenance cost is due to the replacement of semi-membrane only that can be done within few minutes for continuous supply of water.
- Low capital and low cost for converting sea water into drinking water.

8.14 Water Filtration by Electrodialysis

Electrodialysis depends on the following general principles:

- Most salts dissolved in water are ionic, being positively (cationic) or negatively (anionic) charged. These ions are attracted to electrodes with an opposite electric charge.
- Membranes can be constructed to permit selective passage of either anions or cations.

The dissolved ionic constituents in a saline solution such as Na^+, Ca^{2+} and CO_3^{2-} are dispersed in water, effectively neutralizing their individual charges. When electrodes connected to an outside source of direct current like a battery are placed in a container of saline water, electrical current is carried through the solution, with the ions tending to migrate to the electrode with the opposite charge.

Electrodialysis is an electro-membrane process in which ions are transported through ion permeable membranes from one solution to another under the influence of a potential gradient.

Electrodialysis is an electrically driven membrane separation process that is capable of separating, concentrating and purifying selected ions from aqueous solutions (as well as some organic solvents). The process is based on the property of ion exchange membranes to selectively reject anions or cations as shown in Fig. 8.18.

Electrodialysis (ED) uses an electric current in the water to remove contaminants. The electric current pulls ionic impurities through a membrane that filters out these impurities from water. The result is purified water.

- Electrodialysis requires a large amount of electricity to produce a current in the water strong enough to manipulate the impure ions.
- Electrodialysis also has a limited range of contaminates with which it can filter. It cannot remove organics, pyrogens and elemental metals.
- Electrodialysis can remove salts from food, dairy and other products, waste streams and other solutions, as well as concentrate salts, acids or bases.

The system is a useful tool to remove unwanted total dissolved solids that can build up in product streams.

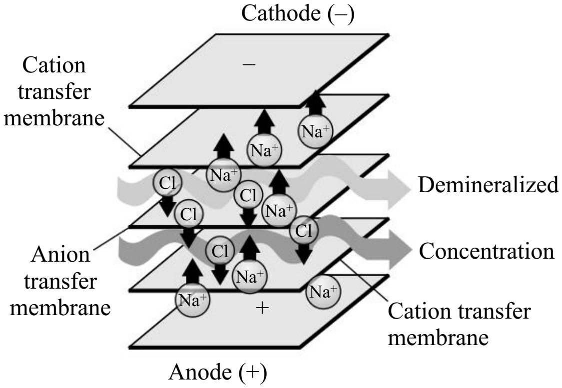


FIGURE 8.18 *Process of electrodialysis of impure water.*

8.14.1 Electrodialysis Unit

The basic electrodialysis unit consists of several hundred cell pairs bound together with electrodes on the outside and is referred as a membrane stack as shown in Fig. 8.19. Feedwater passes simultaneously in parallel paths through all of the cells to provide a continuous flow of desalted water and brine to emerge from the stack.

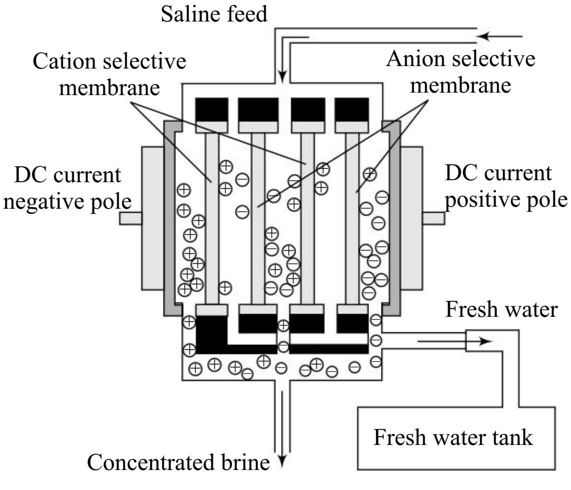


FIGURE 8.19 *A typical stack of electrodialyses for desalination process.*

For these phenomena to desalinate water, membranes that will allow either cations or anions (but not both) to pass are placed between a pair of electrodes. These membranes are arranged alternately with an anion-selective membrane followed by a cation-selective membrane. A spacer sheet that permits water to flow along the face of the membrane is placed between each pair of membranes. One spacer provides a channel that carries feed (and product) water, while the next carries brine. As the electrodes are charged and saline feed water flows along the product water spacer at right angles to the electrodes, the anions in the water are attracted and diverted towards the positive electrode as shown in Fig. 8.19. This dilutes the salt content of the water in the product water channel. The anions pass through the anion-selective membrane, but cannot pass any farther than the cation-selective membrane, which blocks its path and traps the anion in the brine. Similarly, cations under the influence of the negative electrode move in the opposite direction through the cation-selective membrane to the concentrate channel on the other side. Here, the cations are trapped because the next membrane is anion-selective and prevents further movement towards the electrode. By this arrangement, concentrated and diluted solutions are created in the spaces between the alternating membranes. These spaces, bounded by two membranes (one anionic and the other cationic) are called cells. The cell pair consists of two cells, one from which the ions migrate (the dilute cell for the product water) and the other in which the ions concentrate (the concentrate cell for the brine stream).

The raw feed water must be pretreated to prevent materials that could harm the membranes or clog the narrow channels in the cells from entering the membrane stack. The feed water is circulated through the stack with a low-pressure pump with enough power to overcome the resistance of the water as it passes through the narrow passages. A rectifier is generally used to transform alternating current to the direct current supplied to the electrodes on the outside of the membrane stacks.

8.14.2 Ion Permeable Membranes

The ion permeable membranes used in electrodialysis are an essentially sheets of ion-exchange resins. They usually also contain other polymers to improve mechanical strength and flexibility. The resin component of a cation-exchange membrane would have negatively charged groups (e.g. $-SO_3^{2-}$) chemically attached to the polymer chains (e.g. styrene/divinyl benzene co-polymers). Ions with a charge opposite to the fixed charge (counter ions) are freely exchanged at these sites. The concentration of counter ions (e.g. Na^+) is relatively high; therefore, counter ions carry most of the electric current through the membrane. The fixed charges attached to the polymer chains repel ions of the same charge (co-ions), in this case the anions. Since their concentration in the membrane is relatively low, anions carry only a small fraction of the electric current through a cation permeable membrane. Attachment of positive fixed charges (e.g. $-NR_3^+$ or $C_5H_5N^+R$ where commonly $R = CH_3$) to the polymer chains forms anion permeable membranes, which are selective to transport of negative ions, because the fixed $-NR_3+$ groups repel positive ions.

8.15 Hazardous Chemicals

Chemicals are considered highly *hazardous* for many reasons. They may cause cancer, birth defects, induce genetic damage, cause miscarriage or otherwise interfere with the reproductive process. They may be a cholinesterase inhibitor, cyanide, or other highly toxic chemical that, after a comparatively small exposure, can lead to serious injury or even death. Working with compounds like these generally necessitates implementation of additional safety precautions. Below are definitions of the classes of chemicals that are considered highly hazardous.

8.15.1 Characteristic Hazardous Wastes

If any chemicals or materials possess one of the four characteristics, it is advisable to handle them as a hazardous waste. The following are the four characteristics and a few examples of common wastes.

1. Ignitable

Flammable liquids (Flashpoint < 140°F):
alcohols, benzene, toluene, xylene, etc.

Oxidizers:
nitrates, perchlorates, bromates, permanganates, peroxides, iodates
Organic Peroxides: *benzoyl peroxide, cumene hydroperoxide, methyl ethyl ketone peroxide*

2. Corrosive

Aqueous liquids with pH > 2 or pH < 12.5

or other liquids capable of corroding steel at a rate of > 6.35 mm (0.250 inches)/year at a test temperature of 55°C.

Inorganic acids: *hydrochloric acid, sulphuric acid, nitric acid, perchloric acid, phosphoric acid*
Organic acids: formic acid, lactic acid
Bases: *hydroxide solutions, amines*

3. Reactive

Materials that can react violently or create toxic fumes:

Sulphides and cyanides
Peroxide formers: Ethers, potassium amide, sodium amide, vinyl acetate, tetrahydrofuran
Alkali metals: *Sodium, Potassium, Lithium*
Dinitro- and trinitro-compounds: *Picric Acid*
Carbonyl compounds
Isocyanates
Perchlorate crystal formers: *Perchloric Acid*

4. Toxic

Substances with a high degree of acute toxicity are those that can cause death, disability or serious injury after a single, relatively low-level exposure. A selected group of few heavy metals, pesticides and organic chemicals are classified as hazardous due to their toxicity characteristic.

5. Ill Effects of Heavy Metals on Water

Various heavy metals get dissolved in water and cause ill effects. The important heavy metals that cause maximum damage are as follows:

1. Cadmium

Source: From Ni–Cd battery and gold ornaments industries.

Ill effects: Get absorbed on suspended matter in the water which when ingested, cause 'itai-itai' disease (increased salivation, acute gastritis, liver and kidney damage).

2. Lead

Source: From lead–acid battery industries, galvanized iron (GI) pipe connections and plastic industries.

Ill effects: Lead is cumulative poison, causing constipation, loss of appetite, abdominal pain and anaemia, gradual paralysis of muscles, mental retardation, a blue line at the junction of teeth and gums and finally nervous and skeletal disorders and brain damage.

3. Mercury

Source: From alkali, thermometer industries and certain pharmaceutical industries.

Ill Effects: When present in water in amounts exceeding 1–10 mg/L, are converted into the highly toxic methyl mercury. It gets concentrated in fish and shell-fish in the sea, which when ingested cause nerve and brain damage and paralysis followed by death in human beings and other animals.

Exercises

1. 100 mL of a raw water sample on titration with N/50 sulphuric acid required 12.4 mL of the acid to phenolphthalein end-point and 15.2 mL of the acid to methyl orange end-point. Determine the type and extent of alkalinity present in the water. [*Ans:* OH^- alkalinity = 96 mg/L; CO_3^- alkalinity = 56 mg/L and total alkalinity = 152 mg/L].
2. After titration, 100 mL of raw water needed 5.8 mL of N/50 sulphuric acid for neutralization to phenolphthalein end-point and 11.6 mL to methyl orange end-point. Calculate the alkalinity of the water sample in terms of $CaCO_3$ and comment on the alkalinity present.
   [*Ans:* Alkalinity of water due to CO_3^- = 116 mg/L].
3. Water analysis of a sample of raw water contains the following:
   Ca^{2+} = 160 mg/L Mg^{2+} = 72 mg/L
   CO_2 = 88 mg/L HCO_3^- = 488 mg/L
   For coagulation, $FeSO_4.7H_2O$ containing 139 mg/L is used. Calculate the amount of lime and soda required to soften 10,000 L of water.
   [*Ans:* Lime: 70.3 kg; Soda = 37.1 kg].
4. 100 mL of water sample needed 16.9 mL of 0.02N N/50 HCl with methyl orange. It was that the water sample has no phenolphthalein alkalinity. Calculate the alkalinity present in water sample. [*Ans:* 169 mg/L]
5. 0.28 g of $CaCO_3$ was dissolved in 1 L of water. 100 mL of this water on titration required 28 mL of EDTA. 100 mL of unknown hard water sample required 33 mL of same EDTA. After boiling and cooling, 100 mL of this sample required 10 mL of EDTA. Calculate temporary and permanent hardness.
   [*Ans:* Permanent Hardness 10 mg/L; temporary hardness: 230 mg/L]
6. Water analysis gave the following data for a sample of water
   $MgSO_4$ = 120 mgL^{-1}, $CaCO_3$ = 40 mg/L; $CaCl_2$ = 5.5 mg/L; $Mg(NO_3)_2$ = 37 mg/L; KCl = 20 mg/L. Calculate the amount of lime (88 percent) and soda (80 percent) required for the treatment of 60,000 L of water.
7. Calculate the amount of lime and soda required for the treatment of 90,000 L of raw water. On analysis, water is found to have the following: $Mg(HCO_3)_2$ = 146 ppm; $Ca(HCO_3)_2$ = 243 ppm; $CaSO_4$= 10.2 ppm; $MgCl_2$ = 95 ppm and NaCl = 100 ppm.
   Calculate the temporary and permanent hardness of water.
8. An exhausted zeolite softener was regenerated by passing 100 L of NaCl solution having strength of 125 g/L of NaCl. If the hardness of water is 425 ppm, calculate the total volume of water that is softened by this softener.
9. 50 mL of standard hard water containing 1.5 mg of pure $CaCO_3$ per mL consumed 44.0 mL of EDTA. 40 mL of the same water sample needed 25 mL EDTA using EBT indicator. Calculate the total hardness of water sample in ppm.
10. A sample of water contains the following impurities:
    $Mg(HCO_3)_2$ = 146 ppm.; $Ca(HCO_3)_2$ = 243 ppm; $MgSO_4$ = 90 ppm; $CaCl_2$ = 111 ppm and $Ca(NO_3)_2$ = 120 ppm.
    Calculate the amount lime (90 percent) and soda (85 percent) needed for softening 6000 L of water.

Review Questions

1. What are the sources of drinking water?
2. What are impurities present in water?
3. Mention inorganic impurities in water sample.
4. Mention the names of few organic impurities.
5. Name few dissolved organics in water? How do they originate in water.

6. What are dissolved solids? Mention.
7. Write a note on the importance of temperature on water.
8. What is salinity of water?
9. A note on dissolved oxygen (DO) in water.
10. Write a note on ill effects of lead, mercury and cadmium containing water. What if all the constituents can be analysed in water?
11. What is hard water? Mention the causes for hardness of water.
12. Mention the types of hard water.
13. What are the units expressed for hard water?
14. What is temporary hard water? Which mineral causes hardness for this type of hardness?
15. How to eliminate temporary hardness of water? Explain.
16. What do you mean by Permanent hardness? What are the causes?
17. How do you estimate total hardness of water by EDTA method? Give the detailed procedure.
18. How do you analyse alkalinity of water?
19. What is phenolphthalein alkalinity?
20. What do you mean by total alkalinity?
21. Give brief procedure to estimate chloride of water.
22. How to estimate nitrate by Colorimetric method?
23. How do you estimate sulphate of water gravimetrically?
24. How can fluoride of water be estimated by SPANDS method colorimetrically?
25. How do you estimate total dissolved solids in water?
26. What is dissolved oxygen (DO) of water? Enumerate the factors that affects DO of water.
27. Write a note on DO of water and the importance of temperature.
28. Give an experimental procedure. How is DO of water determined?
29. What is BOD of waste water sample?
30. How is BOD of waste water determined by Winkler's method?
31. What is chemical oxygen demand (COD) of waste water?
32. Give a brief procedure to estimate the COD of waste water sample?
33. What is sewage? Mention the characteristics.
34. Give an account of sewage treatment.
35. What is primary treatment of sewage water?
36. Write a note on secondary treatment of sewage water.
37. Write a note on activated sludge process of sewage water treatment.
38. Give a brief account of tertiary treatment of sewage water.
39. What is potable water?
40. Mention the methods adopted for the purification of municipal water.
41. Give an account of purification of municipal water.
42. Write a note on primary treatment of municipal water purification.
43. Write a note on secondary treatment of municipal water purification.
44. Explain tertiary treatment of municipal water purification for domestic purpose.
45. What do you mean by water softening?
46. Name the methods adopted for water softening for domestic and industrial use.
47. Write a note on cold lime-soda process of water softening.
48. Explain hot soda-lime process for softening of raw hard water.
49. Write a note on Calgon process of water softening.
50. Explain with a neat figure zeolite process of water softening.
51. Explain ion exchange method of water softening with a neat figure.
52. What is reverse osmosis? How do you purify water by reverse osmosis?
53. What is an electrodialysis?
54. What is flash evaporation? Explain.
55. What are hazardous chemicals? Name few such contaminants of water.
56. Write a note on hazardous chemicals.
57. Write a note on ill effects of lead, mercury and cadmium containing water.
58. (a) Discuss the ion exchange process for softening water.
    (b) Write a note on scales and sludges.
    [JNTU June, 2014]

59. (a) What is hardness of water? Give causes for hardness.
    (b) Calculate the amount of temporary and permanent hardness of water sample that contains the following impurities. $Ca(HCO_3)_2$ = 12.5 ppm, $Mg(HCO_3)_2$ = 1168 ppm, $MgCl_2$ =79.8 ppm, $CaSO_4$ = 102 ppm. [JNTU June, 2014]
60. A water sample is nonalkaline to phenolphthale in indicator. 100 mL of the same sample on titration with 0.02 N H_2SO_4 requires 14.5 mL of acid to obtain the end-point using methyl orange indicator. Identify the type of alkalinity and determine its content. [Pune University, Dec, 2014]
61. (a) Describe ion exchange method for softening water.
    (b) Define caustic embrittlement. Give causes and prevention of caustic embrittlement in boiler.
    (c) A water sample is not alkaline to phenolphthalein. However, 100 mL of water sample on titration with N/50 HCL required 16.9 mL of acid to get methyl orange end-point. Identify the type and determine the extent of alkalinity. [Pune University, May, 2014]
62. (a) Explain caustic corrosion and caustic embrittlement as ill effects of water in boilers. State their causes and its preventive measures.
    (b) A 50 mL of hard water containing 1.2g of $CaCO_3$ per L required 15 mL of EDTA solution for the end-point. Whereas 50 mL of water sample required 19 mL of EDTA solution and 50 mL of boiled sample of water required 11 mL of EDTA solution for the end-point. Calculate, temporary and permanent hardness of sample water in ppm. [Pune University, April, 2013]
63. (a) Explain the formation in boilers, give their disadvantages and methods of removal.
    (b) A zeolite softener gets exhausted on softening 4000 L of hard water. Calculate the hardness of water, if the exhausted zeolite requires 10 L of 10 percent NaCl for regeneration. [Pune University, Dec, 2013]
64. 50 mL of water required 12.1 mL of N/50 HCL to reach phenolphthalein end-point and 18.5 mL of the same acid for the methyl orange end-point. Calculate types and amount of alkalinities present. [Pune University, Dec, 2015]
65. (a) What are zeolites? Explain zeolite process of softening of water. Give regeneration reactions, advantages and disadvantages of the process.
    (b) 50 mL of water requires 18 mL of 0.05 M EDTA during titration. Whereas 50 mL of boiled water requires 12.5 mL of same EDTA in the titration. Calculate total, temporary and permanent hardness of water sample. [Pune University, June, 2015]
66. (a) Explain the methods for internal treatment of Boiler feed water. [Pune University, Oct, 2012]
67. State Zeolite process for the removal of hardness of water. Discuss its merit over soda lime. [UPTU, 2012]
68. (a) What are ion exchange resins? How will you purify water using the resins?
    (b) Calculate the lime and soda needed for softening 50,000 L of water containing the following:
    $CaSO_4$ = 13.6mg/L, $MgCl_2$ = 9.5 mg/L, $Mg(HCO_3)_2$ = 7.3 mg/L, $Ca(HCO_3)_2$ = 16.2 mg/L.
    Given the molar mass of $Ca(HCO_3)_2$ is 162 and that of $MgCl_2$ is 95.
    (c) Discuss the different methods of internal treatment of boiler feed water. [UPTU, 2012]
69. (a) What is boiler feed water? Explain the priming and foaming in boilers.
    (b) Define COD? Discuss the experimental determination of waste water.
    (c) Explain the activated sludge treatment of sewage water.
    (d) Discuss the desalination of seawater by reverse osmosis. [VTU, June/July, 2015]

70. (a) Explain the gravimetric estimation of sulphate.
    (b) Define COD and BOD? 25 cm^3 of an effluent sample requires for oxidation 8.3 cm^3 of 0.001 M $K_2Cr_2O_7$. Calculate the COD of the effluent.
    (c) Mention the impurities present in natural water. Explain the reverse osmosis process of desalination of water.
    [VTU, June/July, 2016]
71. Explain the treatment of sewage sludge by activated sludge process.
    [VTU, CBSC, June/July, 2016]
72. (a) What is boiler feed water? Explain the corrosion of boilers due to dissolved oxygen, CO_2 and Cl_2.
    (b) Explain the Winkler's method of determining dissolved oxygen. Give the reaction involved. [VTU, June/July, 2016]
73. Define COD? If 20 cm^3 of waste water sample consumes 30 cm^3 of 0.01 N $K_2Cr_2O_7$ for oxidation of impurities, calculate the COD of water sample. [VTU, June/July, 2016]

9

Instrumental Methods of Analysis

Chapter Outline

Introduction. Colorimetry-estimation of copper in copper sulphate solution. Titrimetric methods of analysis. Potentiometry-electrochemical analysis, potentiometric titration with a reference electrode, electrochemical cells with ion sensitive electrode, determination of equivalent point in potentiometry. Conductometry-solution conductance, molar conductance, Kohlrausch law, applications of conductometry, acid-base titrations, titrations involving precipitation. Flame photometry. Basics few other instrumental techniques, atomic absorption spectroscopy (AAS), fluorometry, ultraviolet visible spectroscopy (UV/VIS), infrared spectroscopy (IR), microwave spectroscopy (MW), nuclear magnetic resonance spectroscopy (NMR), electron spin spectroscopy (ESR), chromatography, liquid chromatography (LC/HPLC), gas liquid chromatography (GLC).

9.1 Introduction

The chemical analysis of substances in solution using instrumentation techniques are fascinating and exciting; which undoubtedly interacts with all the areas of chemistry; and with many other fields of pure and applied science. Analytical instrumentation also plays an important role in the production and evaluation of new products and in the protection of consumers and the environment. The areas of instrumentation techniques are so vast and definitely, not possible to discuss every technique in detail in this chapter. However, an introductory attempt is made here to elucidate the theory and concept of techniques of the following:(i) Colorimetry, (ii) potentiometry, (iii) conductometry, (iv) flame photometry, (v) fluorometry, (vi) UV spectroscopy, (vii) NMR, (viii) IR, (ix) microwave, (x) ESR, (xi) gas chromatography HPLC and few other instrumental techniques.

9.2 Colorimetry

Colorimetric measurements are made using a white light source which is passed through a colour filter or alternative wavelength selection device. This incident light then passes through a cuvette containing a coloured chemical compound in solution. The intensity of the light leaving the sample will be less than the light entering the cuvette. *The loss of light or absorption is proportional to the concentration of the compound.*

Colorimetry, however, only applies to measurements made in the visible region of the electromagnetic spectrum e.g. (380–780 nm). The extent to which light is absorbed by a sample is dependent upon many factors. The main general contributors are the wavelength of the incident light and the colour and the thickness of the solution.

Colorimetry is concerned with the determination of the concentration of a substance by measurement of the relative absorption of light with respect to a known concentration of a substance. A schematic working of a colorimeter is illustrated in Fig. 9.1a.

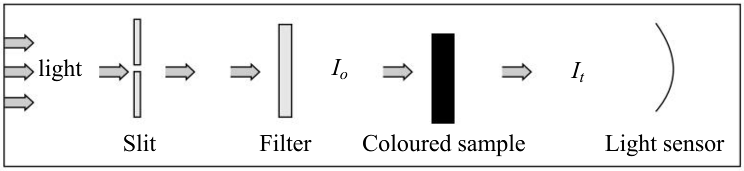


FIGURE 9.1 *(a) Schematic working of a colorimeter.*

When a monochromatic light of intensity I_0 is incident on a transparent liquid medium, a part of it is absorbed, I_a, a part I_r is reflected and the remaining I_t is transmitted, i.e. $I_o = I_a + I_r + I_t$, as shown in Fig. 9.1b.

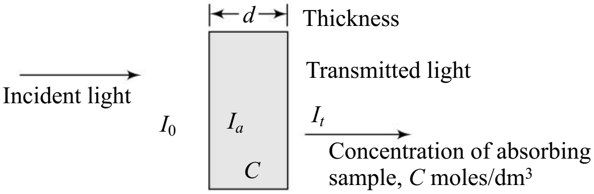


FIGURE 9.1 *(b) Absorption of light by a sample solution.*

For a glass–air interface (covet), I_r is negligible. Therefore,

$$I_o = I_a + I_t$$

The relation between absorbance '*A*', concentration '*c*' (mol/dm^3) and path length '*d*' (cm) is given by Beer-Lambert's law; that is, the intensity of a monochromatic light decreases exponentially as the concentration of the absorbing substance increases arithmetically.

$$I_t = I_o e^{-\varepsilon cd} \quad \text{or} \quad A = \log I_0/I_t = \varepsilon cd,$$

where 'ε' is the molar extinction coefficient, '*d*' is the path length and '*c*' is the sample concentration (mol/dm^3). therefore, $A \alpha C$.

Therefore, *absorbance 'A' of the coloured solution is proportional to the concentration 'c' (mol/dm^3) of the substance in solution.*

In colorimetry, the transmission of light and its absorbance have very specific meanings,

where: 'ε' is a constant—the ability of a given molecule to absorb a particular wavelength of light,
'd' is the path length—the longer the path, the less light gets through.
'c' is the concentration—the more molecules in the solution, the more light is absorbed.

The ratio $I_t/I_o = T$ is called the *transmittance*, and $A = \log(1/T) = \log(I_o/I_t)$ is called the *'absorbance'* (A) or '*optical density'* (OD).

That is

$$\text{transmittance } (T) = \frac{\text{Transmitted light}}{\text{Incident light, } I_0} \quad \text{Absorbance, } A = -\log T = \varepsilon c d$$

In colorimetry, many substances that do not absorb or have a very low absorbance in the visible region of the spectrum will react quantitatively with specific reagents to give a coloured product that can be measured in the spectrophotometer. The absorbance (A) of such solutions can be plotted against the quantity or concentration (c) of the test substance producing the colour. This graph is known as the calibration curve (or standard curve).

To get a straight line calibration curve, absorbance versus concentration needs to be plotted as shown in Fig. 9.2.

Since it is often possible to produce comparatively high absorbance with relatively small amounts of materials, colorimetry is widely used in biochemistry to assay a wide range of biologically important molecules.

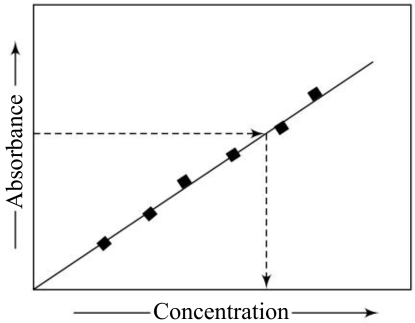


Figure 9.2 *Calibration curve.*

In colorimetry, it is a fact that maximum absorption occurs for a particular wavelength of light wherein the efficiency of estimation is reliably higher than any other wavelengths of the source of light. This necessitates the use of a colour filter which has a complementary colour to that of the solution. For example, for copper complex solution, it is a 620-nm filter, while for reddish coloured iron salt complexes a 480-nm filter should be used.

9.3 Estimation of Copper in Copper Sulphate Solution

A series of standard solutions of copper sulphate is treated with ammonia and diluted to the definite volume (say, 25 cm^3). The absorbance of each of these copper sulphate solutions is measured using a 620-nm filter; since the coloured complex of copper absorbs maximum light at this wavelength. The absorbance, A, values are plotted against concentration of copper to get a linear calibration graph passing through the origin confirming Beer-Lambert's law.

An unknown solution of copper sulphate is similarly treated with ammonia and diluted to the same volume (25 cm^3). The absorbance, A, of this solution is noted at 620 nm and its concentration (mg/cm^3) is found out from the linear calibration graph.

Some important points to bear in mind when carrying out colorimetry are as follows:

(i) Unlike straightforward spectrophotometry, colorimetry is a destructive technique; that is once reacted, the sample cannot be recovered.

(ii) A chromophore reflects the complementary colour(s) that it absorbs, that is a yellow compound appears yellow because it absorbs blue light and therefore it must be estimated in the blue region of the spectrum.

(iii) Colorimetric assays are usually most sensitive at λ (maximum) of the chromophore produced. If this is not known, the spectrum needs to be determined before commencing assay.

(iv) Reference cuvette (i.e. blank) should contain everything except the substance being assayed, that is all the reagents in the same concentrations as in the test cuvette.

(v) The reference cuvette and its contents require, if anything, more care because any error in these will be reflected in all of the values obtained.

(vi) The 'line of best fit' should be drawn through the data points, not necessarily the line that passes through the origin and the other points.

(vii) Calibration curves may vary as batches of reagents and standards vary. Therefore, new calibration curves should be prepared each time an assay is run.

(viii) Calibration curves should never be extrapolated beyond the highest absorbance value measured. It is always more accurate to repeat an assay at a concentration which falls within the most accurate region of the calibration curve (i.e. you may need to dilute a solution to accomplish this).

9.4 Titrimetric Methods of Analysis

Titrimetric analysis a method of determining the number of moles of reagent (titrant) required to react quantitatively with the substance being determined. The titrant can be added (a) volumetrically, with a glass burette (b) electrochemical method. Various methods are available for end-point determination: spectrophotometry, potentiometry, amperometry, conductometry, etc. The potentiometric end-point determination is the most widely used.

9.4.1 Potentiometry

Potentiometry is the field of electroanalytical chemistry in which potential is measured in an electrochemical cell.

Potential is measured under the conditions of no current flow. The measured potential is proportional to the concentration of some component of the analyte.

The potential that develops in the *electrochemical cell* is the result of the *free energy change* that would occur if the chemical phenomena were to proceed until the *equilibrium condition* has been satisfied. The measured potential may then be used to determine the analytical quantity of interest, generally, the concentration of some component of the analyte solution.

A typical electrochemical cell for potentiometric studies is constructed by coupling two half-cells; one is a reference electrode whose reaction and potential is constant and the other is known as the indicator electrode where the potential of the analyte solution is observed with the variation of the concentration of the ionic species/oxidation states during the titration. Generally, saturated calomel electrode (SCE- 0.2422 V) is coupled with inert platinum electrode which is the indicator electrode of the cell during titration. This is shown in Fig. 9.3.

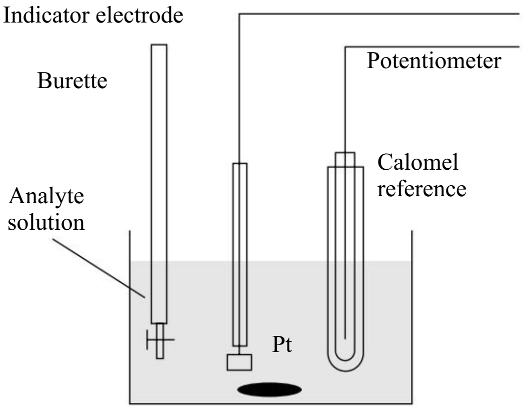


FIGURE 9.3 *An electrochemical cell.*

The measured EMF of a cell is

$$E_{\text{cell}} = E^{\text{o}}_{\text{cathode}} - E^{\text{o}}_{\text{anode}}$$

$$E_{\text{cell}} = E_{\text{indi}} - E_{\text{ref}}$$

$$E_{\text{indi}} = E_{\text{cell}} + E_{\text{ref}} = E_{\text{cell}} + 0.2422 \text{ V}$$

For any redox analyte species at the indicator electrode.

When a platinum electrode is used as the indictor electrode in potentiometric studies and the temperature is maintained constant, the potential of the indicator electrode depends on the ratio of the redox ionic concentrations, to which the electrode is dipped.

The potential of an electrode depends upon the concentration of the ions to which it is reversible in accordance with *Nernst equation* as given.

$$E = E^{\circ} + \frac{0.0591}{n}\log_{10}[M^{+n}]$$

where n is the change in the oxidation number/no of moles of the electrons involved.

In general, the EMF of the cell (E_{cell}) is given by

$$E_{\text{cell}} = E^{\circ}_{\text{cell}} - \frac{0.0591}{n}\log_{10}\frac{[M_1^{+n}]}{[M_2^{+n}]}$$

9.4.2 Electrochemical Analysis

Electrochemical analysis is the group of chemical analytical methods in which a potential change is generated by an electrochemical reaction and is used for quantitative determination of a substance.

These methods depend on electrode reactions. Two of these methods: (a) potentiometry using a reference calomel electrode and an indicator electrode and (b) potentiometric titrations using ion-selective electrodes (ISEs) as a sensor.

The possibilities of modern electronics and data processing make these methods powerful and comfortable analytical tools for a wide range of applications.

9.4.3 Potentiometric Titration

9.4.3.1 The Potentiometric Titration of Ferrous Ammonium Sulphate against Standard Potassium Dichromate in the Presence of a Mineral Acid

The electrode reaction of one of the half-cells of an electrochemical cell in which charge transfer takes place at the interface between the electrode and the electrolyte is coupled with a calomel reference electrode. The reference electrode should have a known, or at least, a constant potential value under the prevailing experimental conditions. Such a cell is represented as

$$\text{Pt/M}^{+2}\,;\,\underline{\text{M}}^{+3}(\text{C}_1)//\text{Cl}^{-1}/\text{Hg}_2\text{Cl}_2/\text{Hg}$$

where Pt is an inert metal in contact with a redox ion system and is referred to as an indicator electrode. The potential of the indicator electrode is sensitive to the concentration of the analyte in the solution and changes with analyte concentration, and the reference electrode (typically the saturated calomel or the silver/silver chloride electrode) provides a stable reference potential for measurement of the potential of the indicator electrode. Therefore, the potential of this electrochemical cell depends upon the analyte concentration. The simplest one is an inert metal such as platinum, which can be used to measure the ratio of the concentrations of the oxidized and reduced species (M^{+2}/M^{+3}).

$$\text{The EMF of the cell} = E_{\text{cathode}} - E_{\text{anode}} = +\text{ve (volts)}$$

Important: If the cell includes a reference electrode than the cell potential is always calculated as

$$E_{\text{cell}} = E_{\text{indicator}} - E_{\text{reference}}$$

(when writing cell reactions we always assume an oxidation reaction at the reference electrode).

Since the potential of the reference calomel electrode remains fixed throughout the experiment, the variation of potential of the indicator electrode for a particular redox system could be ascribed as

$$E_{\text{cell}} = E^{\text{o}}_{\text{cell}} - \frac{0.0591}{1}\log 10\frac{[M^{+2}]}{[M^{+3}]} \text{ or } E = E^{\text{o}}_{\text{cell}} + \frac{0.0591}{1}\log 10\frac{[M^{+3}]}{[M^{+2}]}$$

$$E = E^{\text{o}}_{\text{cell}} + \frac{0.0591}{n}\log_{10}\frac{[\text{Oxidized state}]}{[\text{reduced state}]}$$

During the potentiometric titration of ferrous ammonium sulphate against standard potassium dichromate in the presence of a mineral acid, the potential variation observed before the equivalence point of titration is at the inert redoxelectrode (Pt/Fe^{+2}; Fe^{+3}) of the cell and is shown in terms of Nernst equation as

$$E = E^{\text{o}}_{\text{cell}} + \frac{0.0591}{n}\log_{10}\frac{[\text{Fe}^{+3}]}{[\text{Fe}^{+2}]}$$

and thereafter, could be attributed to the potential contribution due to Pt/Cr^{+6}; Cr^{+3} at the indicator electrode.

9.4.4 Determination of the Equivalence Point in Potentiometry

There are three methods by which the end point of potentiometric titration can be determined as shown in Fig. 9.4a–c.

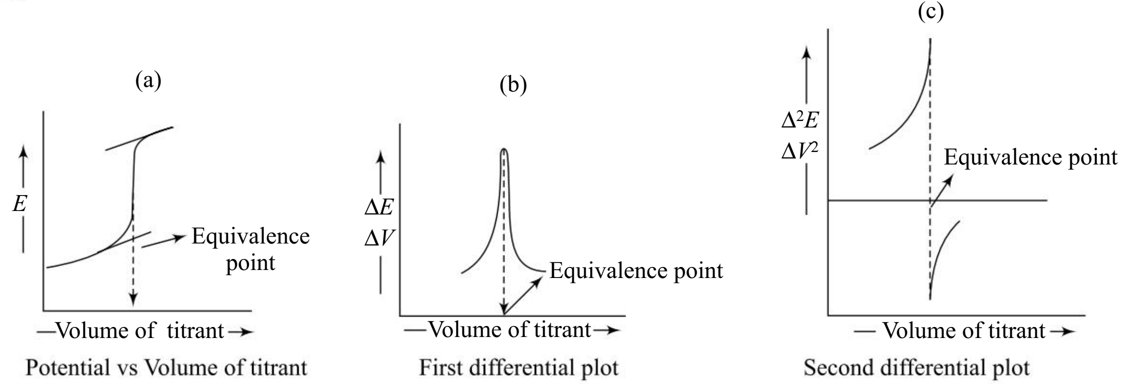


FIGURE 9.4 *(a–c) End point curves of potentiometric titrations.*

Direct plot of the potential *E* against the volume of titrant are given in Fig. 9.4a. The end point of the titration is where there is an abrupt change in the potential value at the indicator electrode.

The first derivative values of the potential and the volume of titration are taken and are plotted against the volume of titrant as shown in Fig. 9.4b; and the values being taken as the second derivative for the potential and volume gives the best result. The point at which the second derivative crosses zero is the inflection point which is taken as the end point of titration (Fig. 9.4c.)

9.4.5 An Electrochemical Cell with Ion Sensitive Electrode

Example: Determination of pH of a solution

Two half-cells coupled form an electrochemical cell that is one solution is in contact with an ISE (combined glass electrode) and the other with a reference electrode. The EMF of the cell (E) can be measured across the two terminals of the potentiometer/pH meter. The pH observed through the ion selective electrode(ISE) is directly proportional to the concentration of the H^+ ion in the analyte solution.

The cell construction of the electrochemical cell is as given for the pH determination of an unknown solution.

Calomel electrode (SCE)//glass/H^+ (unknown)/Cl^-/AgCl/Ag
(Combined glass electrode)

And the EMF of the cell,

$$E = E^{\mathrm{o}}_{\mathrm{cathode}} - E^{\mathrm{o}}_{\mathrm{anode}} = E^{\mathrm{o}}_{\mathrm{G}} - E^{\mathrm{o}}_{\mathrm{SCE}}$$

Therefore, $\mathrm{pH} = \dfrac{E^{\mathrm{o}}_{\mathrm{G}} - E^{\mathrm{o}}_{\mathrm{SCE}} - E_{\mathrm{cell}}}{0.0591}$

Other applications:

A number of different titrations are commonly used. These are reviewed in an instrumental analysis and quantitative analysis textbooks.

- *Acid–base titrations*: These titrations are based on the neutralization reaction between acids and bases. The precise shape of the titration curve depends upon the number of ionisable groups present, and the degree of ionization.
- *Redox titrations*: These titrations involve electron transfer reactions between two redox species; for example, $Ce^{+4} + Fe^{+2} = Ce^{+3} + Fe^{+3}$. The shape of the titration curve can depend upon the reversibility of the redox couple.
- *Precipitation titrations*: This method can be used if the analyte ion forms a low solubility salt (e.g. silver chloride).
- *Complexation titrations*: This method is used for the titration of transition metal ions and requires that the metal ion form a soluble complex with the titrating ligand (e.g. EDTA).

Advantages of potentiometric titrations:

Potentiometric titrations have the great advantage, in common with conductometric titrations, of being possible in turbid, coloured and very dilute solutions. Further advantages of potentiometric titrations are the generally very sharp end points and the ease of automation, and a large number of commercial rigs are available. The range of applications is enormous, and accurate methods have been developed for many electroanalytical processes.

9.5 Conductometry

Measurement of solution conductance is a classical electroanalytical technique that finds application in a variety of chemical and biochemical studies. *Conductometry means measuring the conductivity of an ionic solution*. The electrical conductivity of ionic solutions can be measured by using a conductometer. This is done by applying an electric field between two electrodes. The ions wander in this field. The anions migrate to the anode and the cations to the cathode. In order to avoid substance conversions and the formation of diffusion layers at the electrodes (polarization), high frequency AC source is used.

9.5.1 Solution Conductance

It is customary to *measure the conductance* of an ionic solution rather than its resistance. The conductivity is an intrinsic property of a solution.

Theory:

Electrolyte solutions obey Ohm's law just as metallic conductors do. Thus, the current (I) passing through a given body of a solution is proportional to the applied potential difference. The resistance R of the body of the solution in ohms (Ω) is given by $R = V/I$, where the potential difference (V) is expressed in volts and the current(I) in amperes.

The *reciprocal of solution resistance is called conductance*, symbolized by the letter C. Conductance is expressed in SI units called Siemens (S or Ω^{-1}).

The *measured conductance* for a solution is related to the distance between the electrodes (L) and the surface area (A) of each electrode, as well as the ionic concentration (c). These parameters are interrelated as

$$C = 1/R \quad \text{where } R = \rho \cdot L/A$$

where R is the resistance offered by the solution in a cell with a dimension of L/A and ρ is the proportionality constant known as *specific resistance.*

For ionic solutions, the term *specific conductance* is exclusively used to define the property of the solutions. *Specific conductance* (κ) *of a solution is reciprocal of specific resistance at a temperature and is defined as the conductance of one centimetre cube of a solution.*

$$= 1/\rho$$

The unit of specific conductivity, κ, in SI system is Sm^{-1}

$$\kappa = 1/\rho = [L/A] \cdot (1/R) = [L/A] \times \text{conductance.}$$

The quotient of the length and area is the cell constant, $X = [L/A]$, (unit m^{-1}).

The cell constant, X, is determined by measuring the resistance of a cell filled with a solution of known specific conductance, which is invariable KCl. Once the cell constant X has been determined, specific conductances (κ) of any solution can be determined by multiplying it with the conductance of the solution.

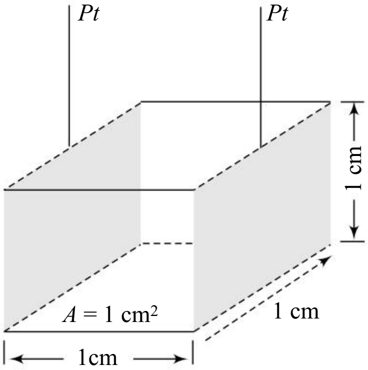


FIGURE 9.5 *Conductivity cell.*

9.5.2 Molar Conductance

The specific electrical conductivity and the electrical conductance are a measure of the ability of a solution to conduct an electrical current.

Generally, molar conductance of ionic solutions is frequently measured. *It is defined as the conducting power of all the ions produced by one mole of an electrolyte in 1dm³ of water*.

The specific conductance, κ, decreases as the concentration decreases in solution as a result of dilution and evidently, the molar conductance increases at a particular temperature. A more fundamental unit of an electrolytic conductance is the equivalent conductance, Λ, which can be though as the value of κ contributed by one mole ions contained in 1 dm^3 (1000 cm^3) of solvent. It is defined as

$$\Lambda_M = \frac{1000 \times \kappa}{C} \qquad \text{(unit is S m}^2\text{ mol}^{-1}\text{)}$$

where the unit of concentration, C, is mol/m^3.

In order to compare the conductances of the electrolytes differing in the ionic composition, the equivalent conductance is defined as

$$\Lambda_{\text{eq}} = \frac{1000 \times \kappa}{C_{\text{eq}}}$$

where c is the equivalent concentration, that is expressed in equiv/dm^{-3}.

9.5.3 Kohlrausch's Law

When the dissociation is complete at infinite dilution, each ion contributes a definite value towards the molar conductance of the electrolyte irrespective of the nature of the other ion with which it is associated and that the molar conductance for any electrolyte at infinite dilution is the algebraic sum of the ionic conductances of the cation and the anion.

$$\Lambda_M = \lambda_+^0 + \lambda_-^0$$

where λ_+^0 is the ionic conductance of cation and λ_-^0 is the ionic conductance of an anion towards the molar conductance at infinite dilution.

It is important to realize that conductance (and conductivity) values contain more information than simple ion concentrations, that is the measured conductivity for a solution of 1 molar HCl will be substantially different than for 1 molar KCl. This fact arises because protons are much more mobile in solution than the potassium ions. Such differences may be quantified by considering the parameter called the *mobility* (u) of anion. The *mobility* of an ion is essentially that part of the conductivity that is independent of concentration.

It may be emphasized that the *ionic conductance is not the same value as ionic mobility*. These are, however, proportional to each other. Thus, at infinite dilution

$$\lambda_+^0 = k \cdot u_+^0 \quad \text{and} \quad \lambda_-^0 = k \cdot u_-^0$$

where λ_+^0 and λ_-^0 are the ionic mobilities of cation and anion, respectively.

Here, mobilities are expressed in terms of m/second under unit potential gradient. The value of F is known to be 96,500, that is 1 Faraday.

$$u_\pm^0 = \frac{\lambda_+^0}{F} = \frac{\lambda_+^0}{96500}$$

9.5.4 Applications of Conductometry

Conductometry is used for *direct measurements* and in *titration*. The theory is identical for both methods. Whereas in direct measurements, it is the *absolute value* that is of interest in titrations, it is *the change in the measured value*. Direct measurement is often used for monitoring surface waters, waterworks, water desalination plants and in the preparation of ultrapure water, where particular limits must not be exceeded. Conductivity detection is mostly used for the precipitation titrations, where the equivalent point is recognized by the conductivity reaching a minimum value.

There are many applications of conductance measurements in chemical analysis.

They are

- degree of dissociation of weak electrolyte
- ionic product of water
- solubilities and solubility product of sparingly soluble salts and
- acid–base and precipitation titrations.

Examples:

Conductance measurements are frequently employed to find the end points of alkali and other titrations. *The principle involved is that electrical conductance depends upon the number of ions and the mobility of ions.*

9.5.4.1 Acid Base Titrations

As mentioned above that the measured conductance of an electrolyte solution depends primarily on the concentration and types of the ions. Conductivity measurement can thus provide a sensitive measure of the changes taking place in an ionic composition in the course of chemical reaction occurring in the solution during conductometric titration.

Consider a solution of a strong acid, hydrochloric acid, HCl for instance, to which a solution of a strong base, sodium hydroxide NaOH, is added. The reaction

$$H^+ + OH^- \rightarrow H_2O \text{ (unionized water)}$$

occurs. For each amount of NaOH added an equivalent amount of hydrogen ions is removed. Effectively, the *faster moving H^+ cation* is replaced by the *slower moving Na^+ ion*, and the conductivity of the titrated solution as well as the measured conductance of the cell decreases. This continues until the equivalence point is reached, at which we have a solution of sodium chloride, NaCl. If more base is added, an increase in conductivity or conductance is observed, since more ions are being added and the neutralization reaction no longer removes an appreciable number any of them.

Consequently, in the titration of a strong acid with a strong base, the conductance has a minimum at the equivalence point. This minimum can be used instead of an indicator dye to determine the end point of the titration. The conductometric titration curve, that is a plot of the measured conductance or conductivity values against the number of millilitres of NaOH solution, is shown in Fig. 9.6.

The position of the equivalence point may be localized precisely as the point of intersection of two straight lines, both determined using readings obtained before and after the minimum observed, respectively. It makes the conductometric titration more objective and independent of a nature of an indicator used in the end-point method.

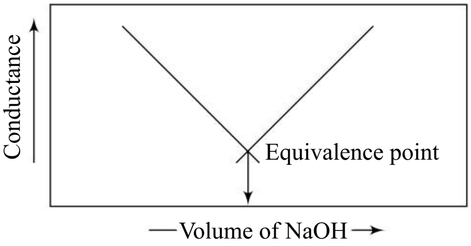


Figure 9.6 *Conductometric titration curve for hydrochloric acid titrated using solution of sodium hydroxide.*

This is one of the advantages of the instrumental method. The same reaction of neutralization takes place when a solution of strong base is titrated using a solution of strong acid. Thus, analogous effects and very similar shape of conductometric titration curve are observed.

Consider the titration of solution of weak acid, such as acetic acid CH_3COOH, using a solution of strong base, NaOH. As we know, the weak acids, as well as other weak electrolytes, are dissociated into very small extent and they exist in solution essentially in the form of neutral acid molecules. When a solution of NaOH is added the reaction occurs

$$CH_3COOH + NaOH \rightarrow Na^+ + CH_3COO^- + H_2O$$

and, as is seen, the undissociated molecules of acetic acid are transformed into dissociated molecules of potassium acetate. The changes are accompanied by increase in conductivity of the solution (Fig. 9.7).

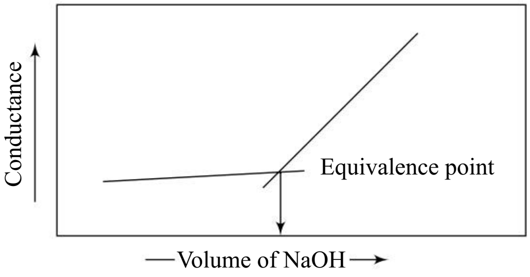


Figure 9.7 *Titration of a weak acid versus NaOH.*

It should be noted, however, that an initial decrease in a conductivity of the solution may be observed after addition of the first drop of a titrant. This minor importance effect is related to the neutralization reaction of the protons resulting from a dissociation and existing even in a solution of the weak acid, $H^+ + OH^- \rightarrow H_2O$.

Thus, a mild increase in conductivity of a titrated solution is observed until the equivalence point is reached, at which we have a solution of sodium acetate, CH_3COONa. If an excess of titrant, that is the potassium hydroxide solution, is added a sharp increase in conductivity is observed.

This distinct difference in the rate of increase is related to the fact that the excess OH^- anions, as well as the protons, exhibit particular mechanisms of charge migration. More detailed inspection of the conductometric titration curve presented in Fig. 9.7 indicates that the equivalence point is less sharp than that observed for the strong acid. Thus, it should be localized as the intersection point of two lines determined by the two sections of the conductometric curve. The slope of the first part of the conductometric curve is dependent on the strength of the acid. It means that it is positive for very weak acid only.

The method of conductometric titration is thus well adapted to the estimation of mixtures of acids of differing strengths. When a mixture of strong and weak acid is titrated, a plot of conductance against alkali added takes form of Fig 9.8. As shown in the figure, the conductometric titration curve is a combination of the diagrams obtained during the titration of strong and weak acid, respectively, where the first end point corresponds to the neutralization of the strong acid present in the sample and the second one is associated with the neutralization of the weak acid in the solution under investigation. The volume of the alkali consumed by the latter is given by a difference of their respective volumes.

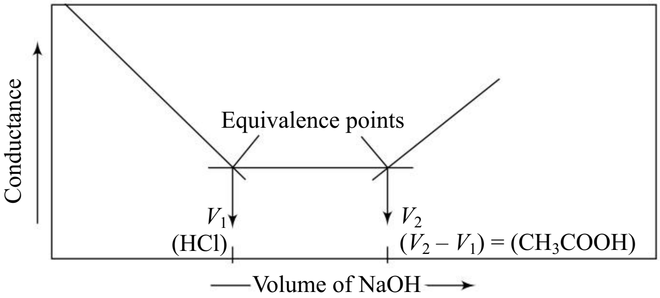


Figure 9.8 *Conductometric titration curve for the hydrochloric acid–acetic acid mixture titrated using solution of sodium hydroxide.*

9.5.4.2 Titrations Involving Precipitation

The titration of $AgNO_3$ against KCl can also be carried out by this method. The reaction involved may be represented as

$$Ag^+ + NO_3^- + K^+ + Cl^- \rightarrow K^+ + NO_3^- + AgCl\ (\text{solid})$$

Since the mobility of potassium ion is nearly the same as that of silver ion which replaces, the conductance will remain more or less constant and will begin to increase only after the end point.

9.6 Flame Photometry

(i) Principles

Flame photometry is an *atomic emission* method for the routine detection of metal salts, principally Na, K, Li, Ca and Ba. Quantitative analysis of these species is performed by measuring the flame emission of solutions containing the metal salts.

Solutions are *aspirated* into the flame. The hot flame *evaporates* the solvent, *atomizes* the metal and excites a *valence electron* to an upper state. Light is emitted at characteristic wavelengths for each metal as the electron returns to the ground state. Optical filters are used to select the emission wavelength monitored for the analyte species. Comparison of emission intensities of unknowns to either that of standard solutions, or to those of an internal standard, allows quantitative analysis of the analyte metal in the sample solution.

Sample solution sprayed or aspirated as a fine mist into flame.

Conversion of sample solution into an aerosol by an atomizer (scent spray) principle leads to no chemical change in the sample at this stage. Heat of the flame vapourizes sample constituents. Still there is no chemical change.

By heat of the flame, +ve actions of the reducing gas (fuel), molecules and ions of the sample species are decomposed and reduced to give *atoms*.

e.g. $Na^+ + e \rightarrow Na$

Heat of the flame causes excitation of some atoms into higher electronic states. Excited atoms revert to the ground state by emission of light energy, $h\nu$, of characteristic wavelength; measured by a detector.

(ii) Instrumentation

A simple flame photometer consists of burner, nebulizer, monochromator, detector and recorder, which are shown in Fig. 9.9.

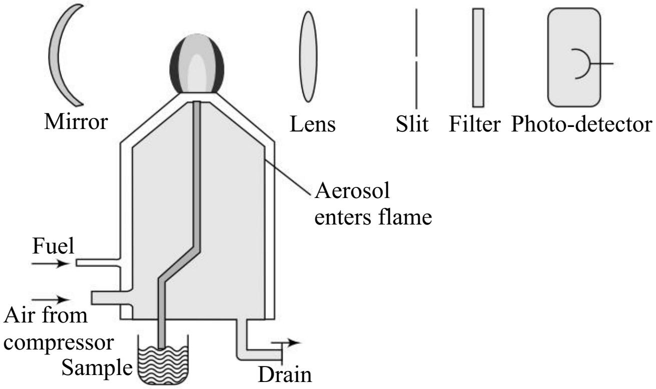


FIGURE 9.9 *Flow diagram of a flame photometer.*

Flame photometry is a simple, relatively inexpensive, high sample throughput method used for clinical, biological and environmental analysis. The low temperature of the natural gas and airflame, compared to other excitation methods such as arcs, sparks and rare gas plasmas, limit the method to easily ionized metals. Since the temperature is not high enough to excite transition metals, the method is selective towards the detection of alkali and alkali earth metals. On the other hand, the low temperatures render this method susceptible to certain disadvantages, most of them related to interference and the stability (or lack thereof) of the flame and aspiration conditions. Fuel and oxidant flow rates and purity, aspiration rates, solution viscosity, concomitants in the samples, etc. affect these. It is therefore very important to measure the emission of the standard and unknown solutions under conditions that are as nearly identical as possible.

This experiment will serve as an introduction to sodium analysis by flame emission photometry and will demonstrate the effects of cleanliness and solution viscosity on the observed emission intensity readings. The instrument is *calibrated* with a series of *standard solutions* that covers the range of concentrations expected of the samples. Standard calibrations are commonly used in instrumental analysis. They are useful when sample concentrations may vary by several orders of magnitude and when the value of the analyte must be known with a high degree of accuracy. This experiment does not produce hazardous waste.

(iii) Applications

It is used exclusively in the quantitative determination of metals in the solution, especially alkali and alkaline earth metals in the given samples. The principle is like that described for atomic absorption. Qualitative determination is also possible as each element emits its own characteristic line spectrum.

(iv) Disadvantages

Intensity of emission is very sensitive to changes in flame temperature. Usually, spectral interference and self-absorption are also encountered which affects the precision of the measurement. Furthermore, a linear plot of absorbance versus concentration is not as always expected.

(v) Experimental procedure for the estimation of sodium

Standard sodium stock solution, 100.0 ppm:

1. Accurately (to 0.1 mg) weigh out by difference 0.1271 g of reagent grade NaCl into a small plastic weighing boat. It is very difficult and time consuming to weigh out this amount exactly. Get it as close as you can, record the exact weight and correct your concentrations accordingly.
2. Carefully transfer the salt *quantitatively* into a 500-mL volumetric flask. Use a few squirts of deionized water from your wash bottle on the weighing boat and the sides of the flask to wash all of it down into the flask. [0.100 g Na/L = 100 mg/L = 100 μg/mL = 100 ppm Na).
3. Add about 100 mL of deionized water to the flask, swirl several times, and dissolve all of the salt before diluting to the volume with deionized water. This is critical.

Sodium standard calibration solutions:

1. Pipette 10.00, 20.00, 30.00, 40.00 and 50.00 mL of the standard 100-ppm sodium solution into the first, second, third, fourth and fifth 100-mL volumetric flasks, respectively.
2. Dilute carefully to the mark with deionized water and mix thoroughly.
3. Use deionized water for the 'blank'.

Unknown solution:

Obtain the unknown from the instructor and carefully dilute to the 100-mL mark with deionized water and mix thoroughly.

Procedure:

Carefully follow the instructions provided to you for the use of the instrument and determine the emission intensity for each standard, the blank (deionized water) and the unknown.

1. When ready to take emission readings, call the teaching assistant to light the flame, stabilize the flame photometer and instruct you in its proper and safe use.
2. Aspirate deionized water from a small glass container that is designed to fit into the instrument for at least 2–3 minutes to clean out and stabilize the unit. While aspirating deionized water, set the zero knob on the digital readout unit to 0.0 units. There will be some 'bounce' or fluctuation in the meter reading.
3. Aspirate the most concentrated standard solution, and use the gain knob to set the reading to 100.0 units.

4. Recheck the zero setting with deionized water and the 100 setting again with the most concentrated standard.Repeat if necessary. The instrument may drift substantially while its temperature is stabilizing, and more slowly throughout the experiment.
5. Now aspirate all the standards and unknowns in turn, starting with the least concentrated standard solution. Aspirate deionized water in between each Na solution to clean the unit out. The burner displays a 'memory effect.' It takes a while to clean out the previous solution aspirated.

9.7 Basics of Few Other Instrumental Techniques of Analysis

The need of the sophisticated analytical instruments and determinations using them is almost a routine and inevitable nowadays for the modern chemical laboratories. It has been a vast expanding area of knowledge as the instrument and computer manufacturers are producing analytical machines, which are in everincrease of power and scope. Furthermore, all the manual techniques in the line of the analytical studies had steadily been transferred to the instrumental techniques.

9.7.1 Atomic Absorption Spectroscopy (AAS)

(i) Introduction

Atomic absorption spectroscopy (AAS) is the measurement of absorption of optical radiation by atoms in the gaseous state.

(ii) Principles

An electrically heated surface is used to vapourize through the aspiration of solution of the analyte sample into free atoms of a particular element in an atomic absorption spectroscopy. When the analyte is of free atoms, a beam of light is passed through them. The atoms will absorb in the visible and ultraviolet region resulting in changes in electronic structure (excited state). *Only a particular wavelength of light is absorbed by the electrons of the atom to go to excited state, which is a measure of the characteristics of the sample.*

There are three basic components for every AA spectrophotometer as shown in Fig. 9.10.

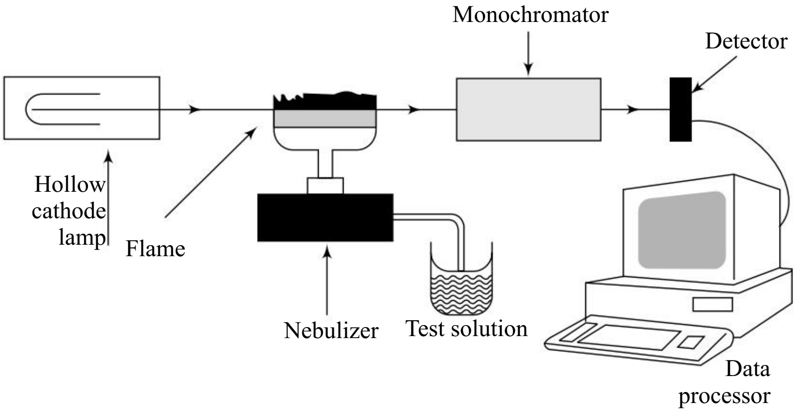


FIGURE 9.10 *Components of atomic absorption spectrophotometer.*

1. Light source

The light source emits the atomic spectrum of a particular element. To detect particular element, specific lamps are used in each case. The *hollow cathode lamps* (HCL) or electrodeless lamps (EDL) are widely used.

2. Sample cell

An atomic sample vapour is generated in the sample cell in the light beam from the source by introducing the sample into a *burner system* (flame AAS) or electrically heated furnace or platform, aligned in the optical path of the spectrophotometer.

3. Specific light measurement

It Includes several components:

(a) Amonochromator to disperse several wavelengths of lights that are emitted from the light source to isolate a particular line of interest,

(b) A detector to produce an electrical current that is dependent on the light intensity. This electrical current is amplified and processed by the instrument electronics to produce a signal, which is a measure of the light attenuation occurring in the sample cell and

(c) This signal is further processed to generate instrument readout in concentration units.

(iii) Applications

This is the most widely used technique for the quantitative determination of metals at trace levels (0.1–100 ppm), which is presented in various materials. It utilizes Beer-Lambert law for the analysis and a standard curve is obtained by plotting absorbance versus concentration of the samples taken. The usual procedure is to prepare a series of standard solutions over a concentration range suitable for the sample to be analysed. Then, the standards and the samples are separately aspirated into the flame, and the absorbance's read from the instrument. The plot will give the useful linear range and the concentrations of the samples that can be found out from the plot.

(iv) Disadvantages

Sample must be in the solution or at least volatile. Individual source lamp and filters needed for each element, since, each and every metal has its own characteristic absorption.

9.7.2 Fluorometry

A fluorometer is an instrument that measures the amount of fluorescent radiation produced by a sample exposed to monochromatic radiation.

(i) Principle

When an adequate amount of energy is supplied to the electrons of an atom of the element, they will be excited to higher energy levels of the atom and sometimes they fall back to the different lower energy states of atom with the emission of energy before reaching the actual ground state. This fall through the electronic energy levels is accompanied by the emission of light which is an 'atomic fluorescence'.

The intensity of this emitted light is measured at right angles to the incident light and related to concentration.

The sample to be analysed passes through a cell of a fluorometer where it is exposed to ultraviolet light from a special light source. The molecules of the analyte sample absorb the energy from this radiation, and then give back this energy by becoming fluorescent and emitting light. The light detected by a sensitive detector in the fluorometer is converted into an electrical current, and is transmitted to the instrument's display. The amount of fluorescent light that is produced depends on the amount of fluorescing molecules, so a high reading above the background under a dispersant-treated slick indicates that molecules of analyte were dispersed into the solvent medium column.

(ii) Applications

- The application of this technique is limited and it offers quantitative estimations of those compounds like benzene and fused benzene ring systems.
- Inorganic metals can also be analysed by the ability of them to form complexes with the ligands.
- It finds uses in the analysis of foods for vitamin content, since vitamins like riboflavin, niacin, etc. exhibit fluorescence, it is a good tool for this area.

Since, only limited compounds show the fluorescence, this technique is relatively free of any interference and is very sensitive.

(iii) Disadvantages

The application is very limited, as only a few of the substances exhibit fluorescence.

9.7.3 Ultraviolet-Visible Spectroscopy (UV/VIS)

(a) Principles

The absorption of an electromagnetic radiation by the substances in the visible and ultraviolet regions of the spectrum, (i.e. 200–700 nm) changes in the electronic structure of ions and molecules. Ultraviolet and visible lights are energetic enough to promote outer electrons to higher energy levels. The uv-vis spectroscopy is usually applied to molecules and inorganic ions or complexes in solution. The uv-vis spectra have broad features that are of limited use for sample identification but are very useful for quantitative measurements. *The concentration of an analyte in solution can be determined by measuring the absorbance at some wavelength and applying the Beer-Lambert law.*

Many molecules absorb ultraviolet or visible light. The absorbance of a solution increases as attenuation of the beam increases. Absorbance is directly proportional to the path length, b, and the concentration, c, of the absorbing species. *Beer's law* states that

$A = ebc$, where e is a constant of proportionality, called the *absorptivity*.

Different molecules absorb radiation of different wavelengths. An absorption spectrum will show a number of absorption bands corresponding to the structural groups within the molecule. For example, the absorption that is observed in the UV region for the carbonyl group in acetone matches with the same wavelength as the absorption from the carbonyl group in diethyl ketone.

(i) Electronic transitions

The absorption of UV or visible radiation corresponds to the excitation of outer electrons of atoms of analyte specimen. Three types of electronic transition are possible:

- Transitions involving p, s and n electrons
- Transitions involving charge-transfer electrons
- Transitions involving d and f electrons (not covered in this Unit)

When an atom or molecule absorbs energy, electrons are promoted from their ground state to an excited state. In a molecule, the atoms can rotate and vibrate with respect to each other. These vibrations and rotations also have discrete energy levels, which can be considered as being packed on the top of each electronic level (Fig. 9.11).

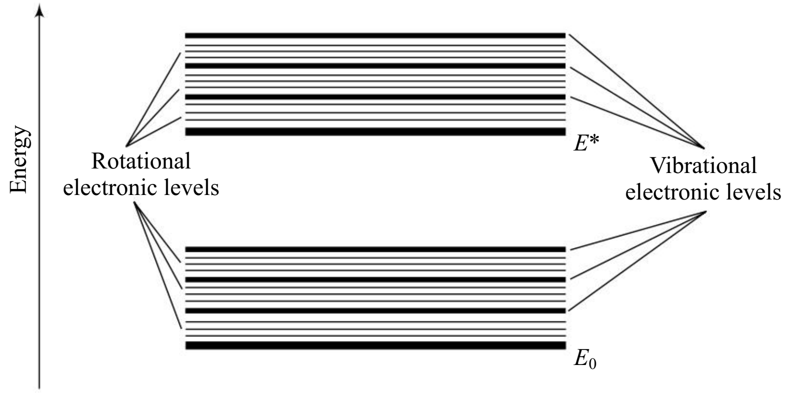


FIGURE 9.11 *Rotational and vibrational electronic levels of an atom or molecule.*

(ii) Absorbing species containing *p*, *s* and *n* electrons

Absorption of ultraviolet and visible radiation in organic molecules is restricted to certain functional groups (*chromophores*) that contain valence electrons of low excitation energy. The spectrum of a molecule containing these chromophores is complex. This is because the superposition of rotational and vibrational transitions on the electronic transitions gives a combination of overlapping lines. This appears as a continuous absorption band.

Possible *electronic* transitions of *p*, s and *n* electrons are shown in Fig. 9.12.

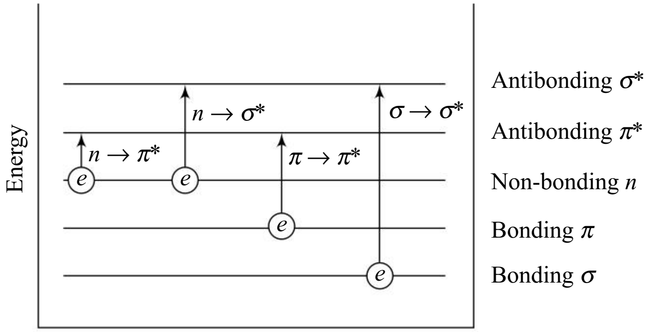


FIGURE 9.12 *Possible electronic transitions in atom or molecule.*

1. $s \rightarrow s^*$ transitions

An electron in a bonding s orbital is excited to the corresponding antibonding orbital. The energy required is large. For example, methane (which has only C–H bonds, and can only undergo $s \rightarrow s^*$ transitions) shows an absorbance maximum at 125 nm. Absorption maxima due to $s \rightarrow s^*$ transitions are not seen in typical uv-vis. spectra (200–700 nm).

2. $n \rightarrow s^*$ transitions

Saturated compounds containing atoms with lone pairs (non-bonding electrons) are capable of $n \rightarrow s^*$ transitions. These transitions usually need less energy than $s \rightarrow s^*$ transitions. They can be initiated by light whose wavelength is in the range 150–250 nm. The number of organic functional groups with $n \rightarrow s^*$ peaks in the UV region is small.

3. $n \rightarrow p^*$ and $p \rightarrow p^*$transitions

Most absorption spectroscopy of organic compounds is based on the transitions of *n* or p electrons to the p^* excited state. This is because the absorption peaks for these transitions fall in an experimentally convenient region of the spectrum (200–700 nm). These transitions need an unsaturated group in the molecule to provide the p electrons.

Molar absorptivities from $n \rightarrow p^*$ transitions are relatively low, and range from 10 to 100 L mol^{-1} cm^{-1}. $p \rightarrow p^*$ transitions normally give molar absorptivities between 1000 and 10,000 L mol^{-1} cm^{-1}. The solvent in which the absorbing species is dissolved also has an effect on the spectrum of the species. Peaks resulting from $n \rightarrow p^*$ transitions are shifted to shorter wavelengths (*blue shift*) with increasing solvent polarity. This arises from increased solvation of the lone pair, which lowers the energy of the *n* orbital. Often (but *not* always), the reverse (i.e. *red shift*) is seen for $p \rightarrow p^*$ transitions. This is caused by the attractive polarization forces between the solvent and the absorber, which lower the energy levels of both the excited and unexcited states. This effect is greater for the excited state, and so the energy difference between the excited and unexcited states is slightly reduced resulting in a small red shift. This effect also influences $n \rightarrow p^*$ transitions but is overshadowed by the blue shift resulting from solvation of the lone pairs.

(iii) Charge–transfer absorption

Many inorganic species show charge-transfer absorption and are called *charge-transfer complexes*. For a complex to demonstrate charge-transfer behaviour, one of its components must have electron donating properties and another component must be able to accept electrons. Absorption of radiation then involves the transfer of an electron from the donor to an orbit associated with the acceptor. Molar absorptivities from charge-transfer absorption are large (greater than 10,000 L mol^{-1} cm^{-1})

(b) Instrumentation

Spectra are obtained by scanning the wavelength separator and quantitative measurements can be made from a spectrum or at a single wavelength. It consists of a dual light source viz., tungsten lamp for visual range measurements and deuterium lamp for the measurements at ultraviolet regions, grating monochromator, photo detector, mirrors and glass or quartz cells.

[For measurements to be made under visible region, both glass and quartz cells can be used. For the measurements under ultraviolet region, only quartz cell should be used, since, glass cells absorb ultraviolet rays.]

There are two types of instruments for this technique as single beam (Fig. 9.13) and double beam (Fig. 9.14) spectrophotometers. However, nowadays, double beam spectrophotometers are widely used and following is the outline of the instrument.

Instruments for measuring the absorption of UV or visible radiation are made up of the following components:

- Sources (UV and visible)
- Wavelength selector (monochromator)
- Sample containers
- Detector
- Signal processor and readout
- Each of these components will be considered in turn.

(i) Schematic of a single beam uv-vis spectrophotometer:

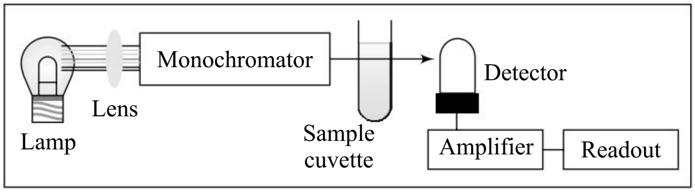


FIGURE 9.13 *A single beam uv-vis spectrophotometer.*

(ii) The schematic diagram of a double-beam uv-vis. Spectrophotometer:

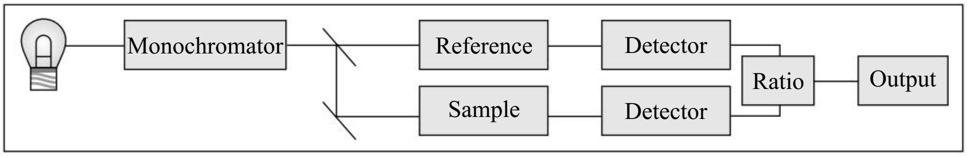


Figure 9.14 *A double beam uv-vis spectrophotometer.*

(i) Sources of UV radiation

It is important that the power of the radiation source does not change abruptly over its wavelength range. The electrical excitation of deuterium or hydrogen at low pressure produces a continuous UV spectrum. The mechanism for this involves formation of an excited molecular species, which breaks up to give two atomic species and an ultraviolet photon. This can be shown as

$$D_2 + \text{electrical energy} \rightarrow D_2^* \rightarrow D' + D'' + hv$$

Both deuterium and hydrogen lamps emit radiation in the range 160–375 nm. Quartz windows and quartz cuvettes must be used in these lamps, because glass absorbs radiation of wavelengths less than 350 nm.

(ii) Sources of Visible Radiation

In order to get a source of visible light, tungsten filament lamp is used and its utility is in the wavelength range of 350–2500 nm. The energy emitted by a tungsten filament lamp is proportional to the fourth power of the operating voltage. This means that for the energy output to be stable, the voltage to the lamp must be very stable indeed. Electronic voltage regulators or constant-voltage transformers are used to ensure this stability.

Tungsten/halogen lamps are very efficient, and their output extends well into the ultraviolet. They are used in many modern spectrophotometers.

(iii) Wavelength selector (monochromator)

All monochromators contain the following component parts:

- An entrance slit
- A collimating lens
- A dispersing device (usually a prism or a grating)
- A focusing lens
- An exit slit

Polychromatic radiation (radiation of more than one wavelength) enters the monochromator through the entrance slit. The beam is collimated, and then strikes the dispersing element at an angle. The beam is splitted into its component wavelengths by the grating or prism. By moving the dispersing element or the exit slit, radiation of only a particular wavelength leaves the monochromator through the exit slit.

(iv) Cuvettes

The containers for the sample and reference solution must be transparent to the radiation which will pass through them. Quartz or fused silica cuvettes are required for spectroscopy in the UV region. These cells are also transparent in the visible region. Silicate glasses can be used for the manufacture of cuvettes for use between 350 and 2000 nm.

(v) Detectors

The photomultiplier tube is a commonly used detector in uv-vis spectroscopy. It consists of a photo-emissive cathode (a cathode which emits electrons when struck by photons of radiation), several dynodes (which emit several electrons for each electron striking them) and an anode.

Photomultipliers are very sensitive to UV and visible radiation. They have fast response times. Intense light damages photomultipliers; they are limited to measuring low power radiation.

(c) Applications

- It is the most widely used technique for quantitative trace analysis, for this Beer-Lambert law is applied.
- Sometimes, it is used in conjunction with other techniques in the identification and structural analysis of organic materials.
- For qualitative analysis, it provides a valuable information, the so-called 'molecular absorption spectrum' is obtained, which exactly tells the nature of the compound, since no two compounds can have the same absorption and hence the spectra.

(d) Disadvantages

Samples should be in the solution. Mixture of substances poses difficult to analyse and requires prior separation. Interference sometimes makes the measurement difficult, but these disturbances are quite common with these types of techniques.

9.7.4 Infrared Spectroscopy (IR)

(A) Introduction

The term 'infrared' covers the range of the electromagnetic spectrum between 0.78 and 1000 mm. In infrared spectroscopy, wavelength is measured in 'wave numbers', which have the units cm^{-1}.

Wave number = 1/wavelength (in centimetres)

It is useful to divide the infrared region into three sections; *near*, *mid* and *far* infrared:

Region	**Wavelength range** (mm)	**Wave number range** (cm^{-1})
Near	0.78–2.5	12800–4000
Middle	2.5–50	4000–200
Far	50–1000	200–10

The most useful IR region lies between 4000 and 670 cm^{-1}.

(B) Theory of infrared absorption

The vibrational energy of molecule changes when the molecule absorbs electromagnetic radiation in the infrared region of the spectrum. All the molecules have vibrations in the form of stretching, bending, etc. at the normal set of conditions and as a result of absorption these energy radiations and the energy levels of the molecules will change. It is a valuable and formidable tool in identifying compounds (Table 9.1).

IR radiation does not have enough energy to induce electronic transitions as seen with UV. *Absorption of IR is restricted to compounds with small energy differences in the possible vibrational and rotational states.*

For a molecule to absorb IR, the vibrations or rotations within a molecule must cause a net *change in the dipole moment* of the molecule. The alternating electrical field of the radiation interacts with fluctuations in the dipole moment of the molecule. If the frequency of the radiation matches the vibrational frequency of the molecule then radiation will be absorbed, causing a change in the amplitude of molecular vibration.

(a) Molecular rotations

Rotational levels are *quantized*, and absorption of IR by gases yields line spectra. However, in liquids or solids, these lines broaden into a continuum due to the molecular collisions and other interactions.

(b) Molecular vibrations

The positions of atoms in a molecule are not fixed; they are subject to a number of different vibrations. Vibrations fall into the two main categories of *stretching* and *bending*.

(i) Stretching

Change in inter-atomic distance along bond axis (Fig. 9.15)

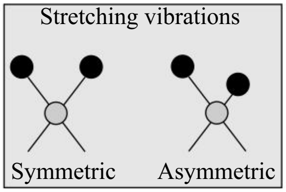


Figure 9.15 *(a) Symmetric and asymmetric stretching vibrations.*

(ii) Bending

Change in angle between two bends. There are four types of bend:

- Rocking
- Scissoring
- Wagging
- Twisting

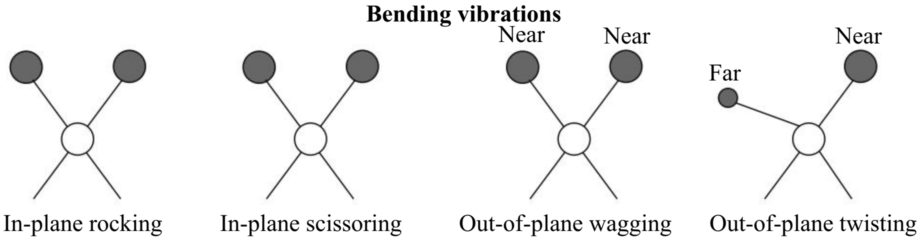


Figure 9.16 *Bending vibrations of a molecule.*

(iii) Vibrational coupling

In addition to the vibrations mentioned above, interaction between vibrations can occur (*coupling*) if the vibrating bonds are joined to a single, central atom. Vibrational coupling is influenced by a number of factors:

1. Strong coupling of stretching vibrations occurs when there is a common atom between the two vibrating bonds
2. Coupling of bending vibrations occurs when there is a common bond between vibrating groups
3. Coupling between a stretching vibration and a bending vibration occurs if the stretching bond is one side of an angle varied by bending vibration
4. Coupling is greatest when the coupled groups have approximately equal energies
5. No coupling is seen between the groups separated by two or more bonds

(C) Instrumentation

A double beam IR spectrophotometer consists of the following:

- an IR source (usually a red hot ceramic material)
- grating monochromator
- thermocouple detector
- cells made up of either sodium chloride or potassium bromide materials, etc.

In this process, the light is dispersed by the monochromator. But, this type of basic design for IR measurements has been outdated. Today, *Fourier Transform Infrared* (FTIR) spectroscopy replaces the old one; for it is better, convenient and efficient technique. This technique utilizes a single beam of undispersed light and has the instrumentation essentially similar to the previous one.

In FTIR, the undispersed light beam is passed through the sample and the absorbance and all wavelengths are received at the detector simultaneously. A computerized mathematical manipulation (known as 'Fourier Transform') is performed on this data, to obtain absorption data for each and every wavelength. To perform this type of calculations, interference of light pattern is required for which the FTIR instrumentation contains two mirrors, one fixed and one moveable with a beam splitter in between them. Before running with the sample, run the instrument with a reference or a blank.

(a) Group frequencies

The following table provides a collection of such data for the most common functional groups. More detailed descriptions for certain groups (e.g. alkenes, arenes, alcohols, amines and carbonyl compounds) may be viewed by clicking on the functional class name. Since most organic compounds have C–H bonds, a useful rule is that absorption in the 2850–3000 cm^{-1} is due to sp^3 C–H stretching; whereas, absorption above 3000 cm^{-1} is from sp^2 C–H stretching or *sp* C–H stretching if it is near 3300 cm^{-1}.

(D) Applications

It finds extensive use in:

The identification and structural analysis of organic compounds, natural products, polymers, etc.

The presence of particular functional group in a given organic compound can be identified. Since, each and every functional group has its own vibrational energy, the IR spectra can be seen as their fingerprints.

(E) Disadvantages

Samples containing mixture of substances cannot be analysed. Since the sample holders and beam splitter, are made of moisture sensitive materials like sodium chloride or potassium bromide (KBr), special cells are required for aqueous samples (KRS-5, ZnSe).

9.7.5 Microwave Spectroscopy

This technique is an extension to IR spectroscopy. Microwave region lies at the far infrared region of the electromagnetic spectrum and its absorption by molecules give rises to changes in the rotational energies of the molecules. In IR spectroscopy, the molecules are subjected for changes in vibrational energies; the energy required for making changes at rotational levels is lesser than that for vibrational levels. Though the principles are same to that of IR, the instrumentation is slightly different in IR as it requires samples in gaseous state for the analysis. Its applications are limited to smaller and simpler molecules since, larger molecules will have interactions between the rotational energy levels within the molecule through various bonds they have. In certain cases this technique can be considered as a good alternate to IR spectroscopy. Besides qualitative analysis, this technique can be applied for conformational analysis of simpler compounds.

9.7.6 Nuclear Magnetic Resonance Spectroscopy (NMR)

(i) Principles

UV, IR and microwave techniques utilize the absorption of light in the visible, ultraviolet and infrared ranges involving changes in electronic states as well as in molecular vibrational and rotational levels. In NMR, the absorption by substances occurs in the radio frequency region of the electromagnetic spectrum resulting in

Functional		Stretching vibrations			Bending vibrations	
Class	Range (cm^{-1})	Intensity	Assignment	Range (cm^{-1})	Intensity	Assignment
Alkanes	2850–3000	Str	CH_3, CH_2& CH two or three bands	1350–1470 1370–1390 720–725	med med wk	CH_2& CH_3 deformation CH_3 deformation CH_2 rocking
Alkenes	3020–3100 1630–1680 1900–2000	med var str	=C–H and=CH_2 (usually sharp) C=C (symmetry reduces intensity) C=C asymmetric stretch	880–995 780–850 675–730	str med med	=C–H and=CH_2 (out-of-plane bending) cis-RCH=CHR
Alkynes	3300 2100–2250	str var	C-H (usually sharp) C≡C (symmetry reduces intensity)	600–700	str	C–H deformation
Arenes	3030 1600 and1500	var med-wk	C-H (may be several bands) C=C (in ring) (two bands) (three if conjugated)	690–900	str-med	C–H bending and ring puckering
Alcohols & Phenols	3580–3650 3200–3550 970–1250	var str str	O–H (free), usually sharp O–H (H-bonded), usually broad C–O	1330–1430 650–770	med var-wk	O–H bending (in-plane) O–H bend (out-of-plane)
Amines	3400–3500 (dil. soln.) 3300–3400 (dil. soln.) 1000–1250	wk wk med	N–H (1°-amines), 2 bands N–H (2°-amines) C–N	1550–1650 660–900	med-str var	NH_2 scissoring (1°-amines) NH_2& N–H wagging (shifts on H-bonding)
Aldehydes & Ketones	2690–2840(2 bands) 1720–1740 1710–1720 1690 1675 1745 1780	med str str str str str str	C–H (aldehyde C–H) C=O (saturated aldehyde) C=O (saturated ketone) aryl ketone α, β-unsaturation cyclopentanone cyclobutanone	1350–1360 1400–1450 1100	str str med	α-CH_3 bending α-CH_2 bending C–C–C bending

(*contd.*)

(*contd.*)

Functional	Stretching vibrations			Bending vibrations		
Class	Range (cm^{-1})	Intensity	Assignment	Range (cm^{-1})	Intensity	Assignment
Carboxylic Acids & Derivatives	2500–3300 (acids)	str	O–H (very broad)	1395–1440	med	C–O–H bending
	overlap C–H	str	C=O (H-bonded)			
	1705–1720 (acids)	med-str	O–C (sometimes 2-peaks)			
	1210–1320 (acids)					
	1785–1815 (acyl	str	C=O			
	halides)	str	C=O (two bands)			
	1750 and1820	str	O-C			
	(anhydrides)	str	C=O			
	1040–1100	str	O–C (two bands)	1590–1650	med	N–H (1¡-amide) II band
	1735–1750 (esters)	str	C=O (amide I band)	1500–1560	med	N–H (2¡-amide) II band
	1000–1300					
	1630–1695(amides)					
Nitriles	2240–2260	med	C≡N (sharp)			
Isocyanates, Isothiocyanates, Di-imides, Azides & Ketenes	2100–2270	med	–N=C=O, –N=C=S –N=C=N–,–N_3, C=C=O			

Table 9.1

Typical infrared absorption frequencies

changes in the orientation of the spinning nuclei. This is observed by applying a magnetic field. It is a well-known fact that the nuclei of the atoms bonded to each other in molecules spin on an axis like a top. Since nuclei are positively charged, this spin will create a small magnetic field around it. If an external magnetic field, suppose, applied to these nuclei, then the spin of the nuclei will align to the magnetic field deviating from the original axis of spin. In this situation if radio waves are applied to the system, the nuclei will absorb this energy and re-align back to their original axis of spin. This will give the information regarding the nature of the compounds and the presence of various functional groups and their environment.

Since this technique mostly measures the spinning of the hydrogen nuclei (almost all the organic compounds contain hydrogen atoms), it is sometimes referred as proton magnetic resonance (PMR) spectroscopy.

(ii) Instrumentation

The instrumentation for this technique include powerful magnet, radiofrequency signal generator, amplifier, detector, etc. The following is the outline of the instrument as shown in Fig. 9.17:

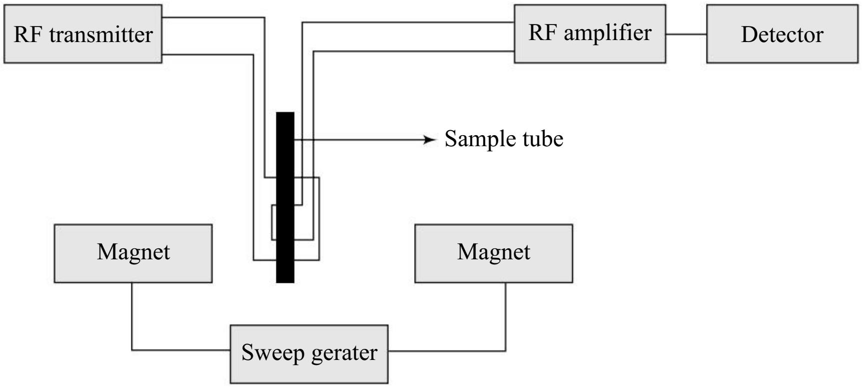


FIGURE 9.17 *A schematic components diagram of NMR spectroscopy.*

(iii) Applications

The application lies mostly in the identification and structural analysis of organic compounds and thus, it is mostly a tool for qualitative analysis.

It gives valuable information regarding the position of the functional groups in a molecule and provides distinguished spectra for the isomer.

Much precise information on the structure of the compounds can be obtained using the same technique with other magnetic nuclei like C^{13}, O^{17}, etc. The instrumentation being the same except that the sweep by the magnetic field is varied.

(iv) Disadvantages

Very expensive and the instrumentation is complex and needs exceptional skills to operate. Its sensitivity ranges from moderate to poor, however, can get clear information using C^{13} or O^{17} NMR. The usage of the solvents is limited and in most of the situations deuterated solvents are required.

9.7.7 Electron Spin Resonance Spectroscopy (ESR)

The basic principle of electron spin resonance spectroscopy is that, electrons always have a spin and thus have a magnetic moment. Thus, the magnetic resonance theory applies to 16 electrons too like that of nuclei, as in NMR. Especially, this technique is of high value when it comes to the compounds which contain odd electrons, that is, those substances which have paramagnetic behaviour (if electrons are paired as in bonded orbital then their mutual spinning will cancel each other and there will be no response for the applied magnetic field, whereas, if it is unpaired then it can align with the applied magnetic field and the feasibility of getting ESR spectra is higher). Thus, the principle and the instrumentation are much similar to that of NMR technique. It is also referred as, electron magnetic resonance.

9.7.8 Chromatography

The chemical components present in a mixture are separated, identified and determined using an efficient technique known as chromatography. This technique is widely used like spectroscopy and is a very powerful tool. This technique is utilized for the analytical purposes and also used for preparative methods. Compounds of high-grade purity can be obtained by this way. The chromatographic technique is defined as follows:

'It is the technique in which the components of a mixture are separated based upon the rates at which they are carried or moved through a stationary phase by a gaseous or liquid mobile phase'.

Based on the *mobile phase*, this technique can be simply classified into two categories as:

- Liquid chromatography and
- Gas chromatography.

The so called 'column' that holds the stationary phase (which in the form of small particles of the diameter of the order in microns) plays a unique role in these processes. Usually, silica is the base material for producing this phase.

9.7.8.1 Liquid Chromatography (LC/HPLC)

Principles

Liquid chromatography was carried out in long glass columns with wide diameter. The diameters of the stacked particles inside the column were of the order of 150–200 microns range. Even then, the flow rates (eluent time) of the mobile phase with the analyte were very slow and separation times were long often several hours. With the advent of latest technology, the particle diameters were reduced as small as to 10 microns with replacement of glass columns with steel ones. The speed of the flow rates was improved by applying high pressure to the column using pumps and hence the performance was improved. This development led the instrument to be mostly called as 'High-performance liquid chromatography' or 'High-pressure liquid chromatography (HPLC)'. Though HPLC retains major of the credits to the analytical side, the earlier one of simple liquid chromatography still finds applications in the preparative purposes.

9.7.8.2 Gas Liquid Chromatography (GLC)

In gas liquid chromatography, it is the interaction between the gaseous sample (the mobile phase) and a standard liquid (the stationary phase), which causes the separation of different molecular constituents. The stationary phase is either a polar or non-polar liquid, which in the case of capillary column, coats the inside of the column, or is impregnated onto an inert solid that is then packed into the GC column.

A schematic layout of a GC instrument is shown in Fig. 9.18. The basic components are inert carrier gases, most commonly helium, nitrogen or hydrogen, a GC column packed or coated with an appropriate stationary phase, an oven that allows for precise temperature control of the column and some type of detector capable of detecting the sample as it exits or elutes from the column.

Gas liquid chromatography works because the molecules in the samples are carried along the column in the carrier gas, but with the partition between the gas phase and the liquid phase. Because this partitioning is

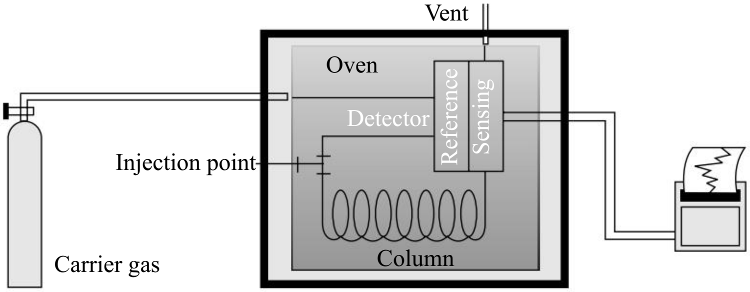


Figure 9.18 *Gas chromatography Instrument.*

critically dependent on the solubility of the sample in the liquid phase, different molecular species travel along the column and elute at different times. Those molecules that have a greater solubility in the liquid phase take longer to elute and thus are measured at a longer interval. Solubility is dependent on the physical and chemical properties of the solute; therefore, separation between different components of the sample occurs based on molecular properties such as relative polarity (like ethylene glycol versus base oil) and boiling point (like, fuel versus diesel engine base oil). For example, using a polar stationary phase, with a mixture of polar and non-polar compounds will generally result in longer elution times for the polar compounds, because they will have greater solubility in the polar stationary phase.

(i) Flame ionization

There are many methods used to detect molecules as they elute. However, the most commonly employed method is flame ionization. In flame ionization, the eluting sample is passed through a hydrogen gas flame and the ion flux measured. As the sample passes through the flame, any molecules present are ionized, resulting in an increased ion flux. The total increase in an ion flux is proportional to the amount of species present allowing the area under the increasing ion flux peak to be directly related to the concentration of the eluting species. GC is often also coupled with Fourier transform infra red (FTIR) or mass spectrometric (MS) detectors.

GLC: This is the most important and widely used methods for separating and determining the chemical components of complex mixtures. Here the stationary phase is a liquid that is immobilized on the surface of a solid support by adsorption or by chemical bonding. The separation of the mixture into individual components is by distribution ratio (partition) of these anayte components between the gaseous phase and the immobilized liquid phase. Because of its wide applications most of the GCs are configured for the GLC technique.

(ii) Recorder

A powerful tool for qualitative analysis as very accurate and precise information like mass or IR spectrum of the individual sample components are readily obtained as they elute from the GC column. It saves time and reduces the steps involved for a component to be separated and analysed.

(iii) Disadvantages

Samples must be volatile and thermally stable below about 4000°C. No single universal detector is available and most commonly used detectors are non-selective. One should take much care in the analytical steps starting from the selection of the column and the detector and must define the temperatures of all the three ports viz., injection port, column oven and detector. An improper calculation on these will lead to insensitive results.

Review Questions

1. What is colorimetry?
2. State Beer-Lambert's law. Explain how this law can be used to determine the concentration of coloured solutions.
3. What do you mean by absorbance of a coloured solution in colorimetry.
4. Define the term 'transmittance'.
5. Give a brief procedure for the estimation of copper in copper sulphate solution colorimetrically.
6. What are titrimetric analytical methods?
7. What is potentiometry?
8. Construct an electrochemical cell for potentiometry and elucidate the function of each electrode involved in the measurement.
9. Give a proper experimental set-up for the estimation of FAS with standard potassium dichromate solution potentiometrically.
10. Construct a cell for the determination of pH of an acid solution, you are provided with a combined glass electrode. Write an expression for the pH of a solution.
11. Show how the equivalence point of titration can be determined potentiometrically. Mention the merits of each method.
12. What is conductometry?
13. What do you mean by solution conductance? What is the unit of conductance?
14. How can the conductance of a solution be measured?
15. Explain (a) specific conductance, and (b) molar conductance.
16. Explain, how the solution conductance is related to the mobility of ions.
17. Give an experimental procedure to estimate the amount of an acid titrated with a standard solution of NaOH conductometrically.
18. What is flame photometry?
19. Explain the components of a flame photometer and their functions.
20. How do you estimate Na of a solution adopting flame photometric technique?
21. Write a brief note on atomic absorption spectroscopy.
22. What is fluorometry? Discuss its principle of working and mention few applications.
23. Mention the working principle of UV spectroscopy.
24. Discuss the possible permitted electronic transitions of UV spectroscopy.
25. Describe UV spectroscopic instrumentation and mention its applications.
26. What is an IR spectroscopy? Discuss the theory of IR spectroscopy.
27. Discuss the role of common functional group frequencies and its importance to the determination of structure of molecules.
28. A note on microwave spectroscopy?
29. Discuss the principle, instrumentation and applications of NMR spectroscopy.
30. A note on ESR spectroscopy.
31. A note on chromatography.
32. Write a note on gas liquid chromatography.
33. (a) What is the purpose of IR spectrometer?
    (b) Discuss the principle and instrumentation of UV-visible spectrometer.
    [Annamalai University, 2014]
34. Explain the instrumentation of UV-visible spectrometer. [Annamalai University, 2013]
35. (a) Draw the block diagram of IR spectrometer and explain the function of various components.
    (b) Explain in detail rational and rotational transitions.
    [Annamalai University, May, 2016]
36. (a) What is absorption of radiation?
    (b) Explain the principle and instrumentation of UV-visible spectroscopy with a neat diagram.
    [Annamalai University, Dec, 2013]
37. (a) What is meant by a wave number?
    (b) Explain the various electronic transitions occur in spectroscopy.
    [Annamalai University, May, 2014]
38. (a) Explain the conductometric curve for reaction between weak acid and strong base.

(b) Explain different types of transitions occurring in organic molecules on absorption of UV-Visible radiation.
[Pune University, May, 2014]

39. (a) Explain the curve for conductometric titration in case of a strong acid–weak base.
[Pune University, April, 2013]

40. (a) Explain the principle instrumentation and applications of UV-visible spectroscopy.
[Pune University, Dec, 2013]

41. Explain the titration curve of conductometric titration in case of strong acid and weak base.
[Pune University, June, 2015].

42. Explain the pH metric titration a mixture of weak acid and strong acid against standard alkali, giving reactions, procedure, titration curve and calculations.
[Pune University, Oct, 2012]

42. Explain the theory and applications of colorimetry in the estimation of copper.
[VTU, June/July 2016]

10

The Solid State

Chapter Outline

Introduction. The solid state-crystalline and amorphous. Types of crystalline solids. Classification of crystalline solids, ionic solids, covalent crystals, molecular crystals, metallic crystals. Structure of simple ionic solids, ionic crystals, AB type ionic crystals, AB_2 type ionic crystals, ABO_3 type compounds, AB_2O_4 type compounds, close packing in solids, types of close packing in solids, types of cation sites available in close packed anion arrays, 3D close packed crystal structure of solids-hexagonal close packing (hcp), coordination in hcp structure, cubic close packing (ccp), coordination in ccp, voids in close packing. Determination of no of tetrahedral and octahedral sites in close packed structure. Ionic crystals and radius ratio rule. Stable bonding configuration in ionic solids, examples of cubic close packing (ccp) structures, example of ccp structures, examples of hcp structures. The crystal structure—the basic concepts of crystallographic axes, the unit cell, lattice parameters. The crystallography, geometric crystallography. The law of constancy of angles, the law of rationality of indices. Interpretation of planes, Miller's indices, Miller's indices d_{hkl} of a plane of a crystal, symmetry equivalent surfaces, inter-planar distances of a cubic crystal. Basic symmetry operations and space groups of crystals, symmetry, symmetry operations and elements, combinations of symmetry operations, glide planes and screw axes. Elements of symmetry of a cubic crystal, crystal systems, Braivais lattice, Bravais lattices of a cube. Introduction to powder X-ray diffraction, X-ray diffraction from crystalline solids, indexing X-ray patterns of a cubic crystal, mathematical approach of indexing. Calculation of no of atoms/ions per unit cell, calculation of no of atoms/ions per unit cell of a cube, some basic aspects of unit cells.

10.1 Introduction

The basic constituents (atoms/ions/molecules) of materials have a natural tendency for agglomeration as 'solids', depending on the type of cohesive force acting on them. The nature of chemical bonding of these basic structural units gives rise to different kinds of solids; and these materials exhibit a wide variety of 'solid state properties' such as semiconductivity, magnetic properties, superconductivity, dielectrics, optical, and so on.

A study of the structure of solids and the assessment of their 'solid state properties' are essential; and therefore, is the subject matter of solid state chemistry.

The materials in the solid state find extensive application in modern technology. Solid state materials are used widely as semiconductors, thyristors, superconductors, magnetic ferrites, photoconductors, etc.

'Material science' has recently developed enormously to such an extent, that, today it is possible for a material scientist to plan and synthesize a new material of desired solid state properties for suitable applications.

A knowledge of 'crystal structure' is, therefore, quite essential before the assessment of its solid state properties; and hence, a brief discussion of structure of solids is aimed at in the foregoing section.

10.2 The Solid State

'*A solid is always characterized by its definite shape, volume, in-compressibility, rigidity and mechanical strength*'.

The constituent atoms, ions or molecules of materials in the solid state are always bound together by a cohesive force in such a way that these constituents are in fixed equilibrium position to each other and describe a definite crystal lattice.

Solids can be classified on their characteristic properties as (i) crystalline and (ii) amorphous.

10.2.1 Crystalline Solids

These 'crystalline solids' have well defined geometric structures or shapes due to definite and orderly arrangement of ions or atoms or molecules in three-dimensional network with a long range order.

Example NaCl (FCC), CsCl (BCC), ZnS (Sphalerite), diamond, quartz, etc.

Generally, crystalline solids possess (a) sharp melting points and (ii) their physical properties such as electrical, thermal and mechanical, etc. are 'anisotropic'. These properties of materials depend on the direction along which they are measured.

10.2.2 Amorphous Solids

The 'amorphous solids' do not exhibit any repetitive pattern or arrangement of atoms (or molecules) of substance and has in fact, no regular geometrical structure in three-dimensional space.

Example Glass, plastic, sugar, etc.

By virtue of amorphous nature of solids, they have (i) no sharp melting points and (ii) that their measured physical properties are 'isotropic'.

10.3 Types of Crystalline Solids

The solids are classified based on the nature of chemical bonding in crystalline solids.

They are as follows:

(a)	Ionic crystals	NaCl, CsCl, etc.
(b)	Covalent crystals	Diamond, SiC,
(c)	Molecular crystals	Ice, solid $CO_{2,}$ etc.
(d)	Metallic crystals	Cu, Zn, Ag, etc.

10.4 Classification of Crystalline Solids

10.4.1 Ionic Crystals

- Ionic crystals are hard and brittle solids.
- They possess high melting points.
- They are poor conductors of electricity, but their ability to conduct increases drastically in melt.

The packing of spheres in ionic crystals is complicated by two factors:

(i) Charged species are present

(ii) Anions and cations are generally quite different in size.

Some general conclusions can be drawn from ionic radii:

(i) within the same period the anions always have larger radii than the cations

(ii) the radius of the trivalent cation is smaller than that of the divalent cation, which is smaller than that of the monovalent cation.

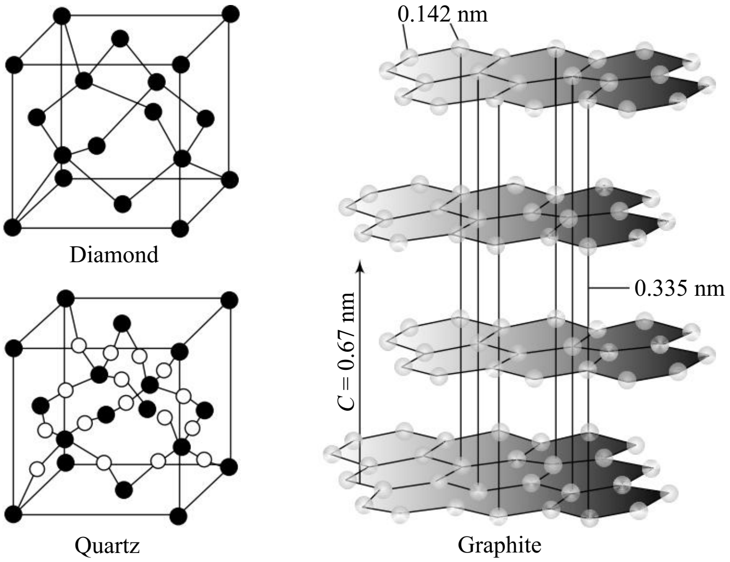
10.4.2 Covalent Crystals

- Covalent crystals are hard solids that possess very high melting points.
- They are poor conductors of electricity.
- In covalent crystals, atoms are held together by covalent bonds.
- Well-known examples are two allotropic forms of carbon, diamond and graphite.

The structure of diamond is based on an fcc lattice (Fig. 10.1). There are eight carbon atoms at the centre of the cube, six carbon atoms in the face centre and four more within the unit cell. Each carbon is tetrahedrally bonded to four others. This tightly bound lattice contributes to diamond's unusual hardness. In graphite, each carbon is bonded to the three others and the layers are held together only weakly.

10.4.3 Molecular Crystals

- Molecular crystals are very soft solids that possess low melting points.
- They are poor conductors of electricity.
- Molecular crystals consist of such substances as N_2, CCl_4, I_2 and benzene.
- Generally, the molecules are packed together as closely as their size and shape will allow.
- The attractive forces are mainly van der Waals interactions.



Diamond

Quartz

Graphite

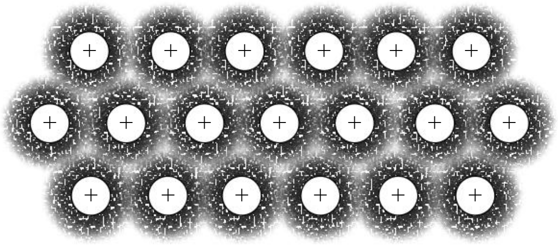
FIGURE 10.1 *Structure of diamond, quartz and graphite.*

10.4.4 Metallic Crystals

Figure 10.2 depicts that electrons are not bound to any particular atom and are free to move about in the solid.

- Metals are characterized by their tensile strength and the ability to conduct electricity.
- Both properties are the result of the special nature of the metallic bond. Bonding electrons in metals are highly delocalized over the entire crystal.
- Metal atoms can be thought of as an array of positive ions immersed in a sea of delocalized valence electrons.
- The great cohesive force resulting from the delocalization is responsible for the great strength noted in metals.
- The mobility of the delocalized electrons accounts for electrical conductivity.
- Metal crystals all have a high density which means that they usually have the hcp or fcc structure.

Metallic Sea of Electrons



Electrons are not bonded to any particular atom
and are free to move about in the solid

FIGURE 10.2 *Sea of electrons in metallic crystals.*

- Magnesium, scandium, titanium, cobalt, zinc and cadmium have the hcp structure
- Aluminium, calcium, nickel, copper, palladium, silver, platinum, gold and lead have the fcc structure.
- Alkali metals, iron, chromium, barium and tungsten have the bcc structure.

10.5 Structures of Simple Ionic Compounds

10.5.1 Ionic Crystals

'Ionic crystals' have well defined geometric shapes or structures due to definite and orderly arrangement of both +ve and –ve ions in the three-dimensional space.

Examples

NaCl	Face-centred cubic
CsCl	Body-centred cubic
ZnS	Sphalerite

In ionic crystals, the repeating units at the lattice sites are +ve and –vely charged constituents which are bound to each other by coulombic forces. The bonding in these crystals is predominantly 'ionic' in character. These cohesive forces in ionic crystals are nondirectional and hence, the final structure of the crystal would be mainly governed by the 'relative sizes of the constituent ions'.

The bonding between the ions is mostly electrostatic and rather strong (binding energies around 1000 kJ/mol); it has no directionality.

- Ionic crystals, thus, can be described as an ensemble of hard spheres that try to occupy a minimum volume while minimizing electrostatic energy at the same time (i.e. having charge neutrality in small volumes, too).
- There are no free electrons, ionic crystals are insulators.

Ionic crystals come in simple and more complicated lattice types; the latter is true in particular for oxides which are often counted among ionic crystals. Some prominent lattice types are discussed in the next section.

10.6 AB Type Ionic Crystals

1. The NaCl structure

The lattice is face-centred cubic (fcc), with *two* atoms in the base: one at (0, 0, 0) and the other one at (½, 0, 0) (Fig. 10.3).

The AB type structure is observed in many salts and oxides, e.g. KCl, AgBr, KBr, PbS or MgO, FeO.

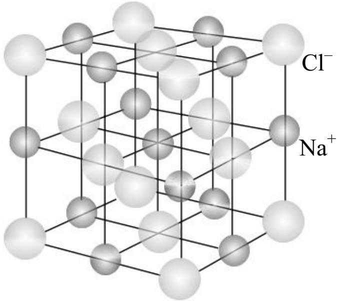


Figure 10.3 *Structure of NaCl lattice.*

2. The CsCl structure

The lattice is *cubic primitive* (Fig. 10.4) with *two* atoms in the base at (0,0,0) and (½, ½, ½). It is a common error to mistake it for a bcc lattice.

Intermetallic compounds (not necessarily ionic crystals), but also common salts assume this structure, e.g. CsCl, TlO or AlNi, CuZn.

3. The ZnS (diamond or sphalerite) structure

The 'zinc blende' lattice is *face-centred cubic* (fcc) (Fig. 10.5) with two atoms in the base at (0,0,0) and (¼, ¼, ¼).

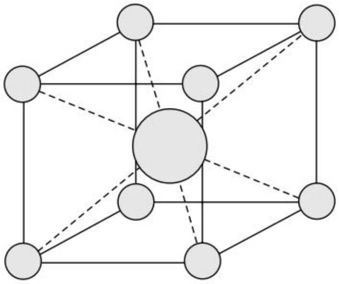


Figure 10.4 *CsCl lattice structure.*

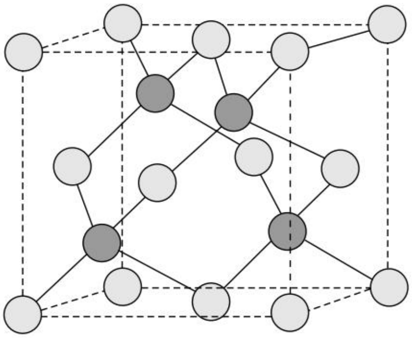


Figure 10.5 *Sphalerite lattice structure.*

It is not only an important lattice for other ionic crystals like ZnS, which gave it its name, but also the typical lattice of *covalently bonded group-IV semiconductors* (C (diamond form), Si, Ge) or III–V compound semiconductors (GaAs, GaP, InSb and InP).

The ZnS lattice is easily confused with the ZrO_2 lattice.

10.7 AB_2 Type Ionic Compounds

The CaF_2 or ZrO_2 structure

The lattice is *face-centred cubic* (fcc) (Fig. 10.6) with *three* atoms in the base, one kind (the cations) at (0,0,0) and the other two (anions of the same kind) at (¼, ¼, ¼) and (¼, ¾, ¼).

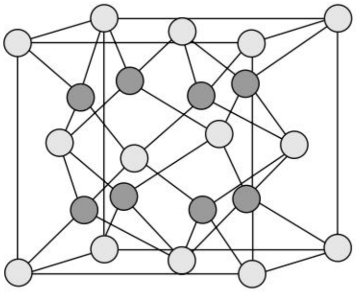


Figure 10.6 *Fluorite lattice structure.*

10.8 ABO_3 Type Compounds

Perovskite Structure

The lattice is essentially *cubic primitive*, but may be distorted to some extent and then becomes *orthorhombic* or worse. It is also known as the $BaTiO_3$ or $CaTiO_3$ lattice (Fig. 10.7) and has *three* different atoms in the base. In the example, there would be Ba at (0, 0, 0), O at (½, ½, ,0) and Ti at (½, ½, ½).

A particular interesting perovskite (at high pressures) is $BaTiO_3$, $MgSiO_3$, etc.

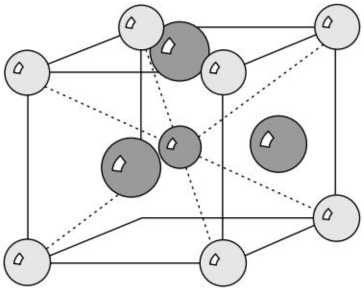


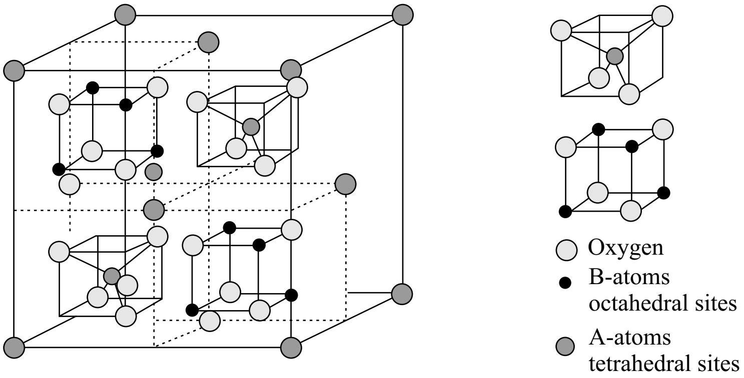
Figure 10.7 *Perovskite lattice structure.*

10.9 AB_2O_4 Type Compounds

Spinel structure

The spinel structure is named after the mineral *spinel* ($MgAl_2O_4$); the general composition is AB_2O_4. It is essentially *cubic*, with the O-ions forming an fcc lattice. The cations (usually metals) occupy 1/8 of the *tetrahedral sites* and 1/2 of the *octahedral sites* and there are 32 O-ions in the unit cell (Fig. 10.8).

There are two types of cubic building units inside a big fcc O-ion lattice, filling all eight octants.



AB_2O_4 spinel The red cubes also contained in the back half of the unit cell

Figure 10.8 *A spinel lattice structure.*

A few examples:

Magnetite:	$Fe^{3+}(Fe^{2+}\ Fe^{3+})O_4$
Spinel:	$Mg^{2+}(Al_2{}^{3+})O_4$
Chromite:	$Fe^{3+}(Cr_2{}^{3+})O_4$
Jacobsite:	$Fe^{3+}(Mn^{2+}\ Fe^{3+})O_4$

10.10 Close Packing in Solids

In order to understand the packing of the constituent particles in a crystal, it is assumed that these particles are hard spheres of identical size. The packing of these spheres takes place in such a way that they occupy the maximum available space and hence the crystal has maximum density. This type of packing is called *close packing*.

The forces binding the fundamental units (atoms, ions or molecules) together may be essentially nondirectional or directional. *Nondirectional* forces result in structures based on close packing, while *directional forces* lead to more open structures, often based on linked polyhedra or groups of polyhedral.

'*A close-packed structure is a way of arranging identical atoms in space such that the available space is filled efficiently*'.

10.10.1 Types of Close Packing in Solids

The packing of spheres of equal size takes place as follows:

(a) When the spheres are placed in a horizontal row, touching each other, an edge of the crystal is formed (Fig. 10.9).

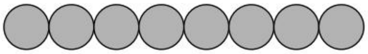


FIGURE 10.9 *Formation of an edge of a crystal.*

(b) Two-dimensional packing:
When the rows are combined, touching each other, *crystal plane* is obtained (Fig. 10.10).

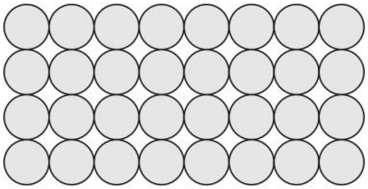


FIGURE 10.10 *Plane of crystal in one direction.*

The rows can be combined in two different ways (Fig. 10.11):

1. The particles when placed in the adjacent rows show a horizontal as well as vertical alignment and forms square or cubic pattern. This type of packing is called *cubic close packing* (ccp).
2. The particles in every next row are placed in the depressions between the particles of the first row. The particles in the third row will be vertically aligned with those in the first row. This type of packing gives a hexagonal pattern and is called *hexagonal close packing* (hcp).

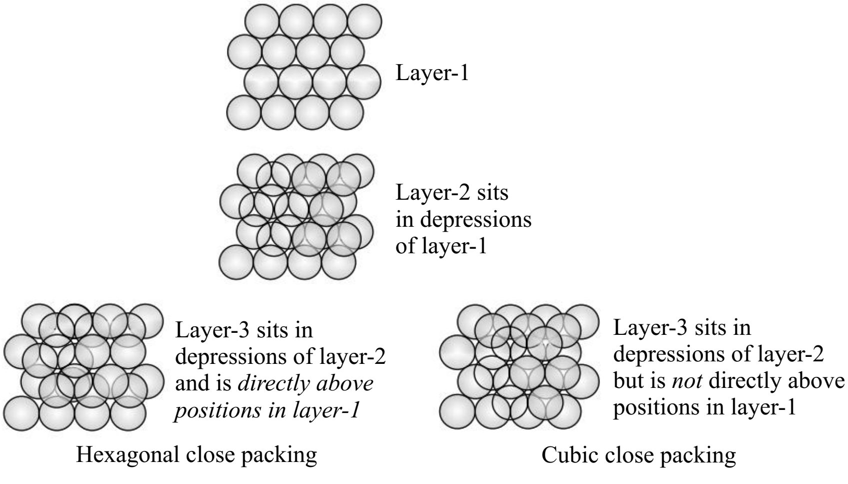


FIGURE 10.11 *Hexagonal and cubic close packing of atoms in an array.*

Hence, there are two types of close-packed structures in solids:

- Cubic close-packed (ccp) structures and
- Hexagonal close-packed (hcp) structures.

The hcp is more efficient as more space is occupied by the spheres in this arrangement. In cubic close packing, a central sphere is in contact with four other spheres whereas in hcp, a central sphere is in contact with six other spheres.

10.11 Types of Cation Sites Available in Close-Packed Anion Arrays

10.11.1 3D Close-Packed Crystal Structure of Solids

The close-packed layers of atoms are stacked on top of each other to get a three-dimensional metal structure. There are different ways of doing this. The most efficient space saving way is to have the spheres in one layer fit into the 'holes' of the layer below.

To identify the stacking of layers of atoms, fix the first layer as 'A', the second layer over the first as 'B' as shown in Fig. 10.12. The third layer can then be added in two ways: in the first way, the third layer fits into the holes of the B layer such that the atoms lie above those in layer A. By repeating this arrangement, one obtains ABABAB stacking or hexagonal close packing (hcp) (Fig. 10.12).

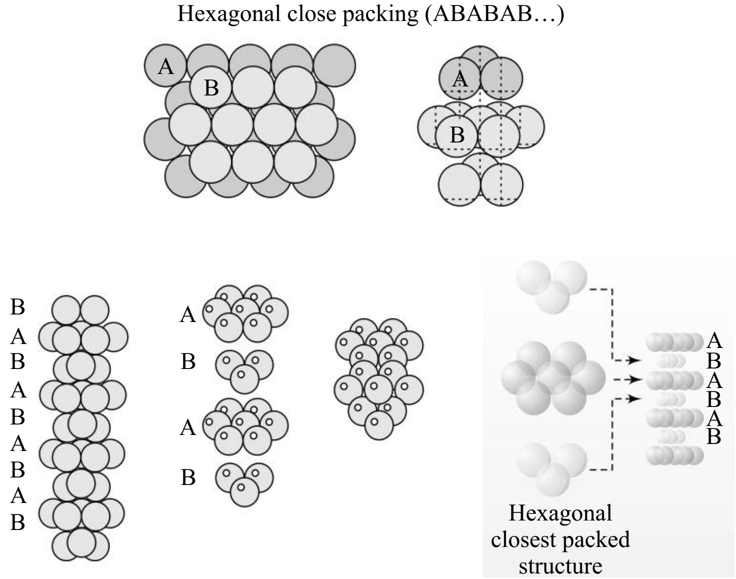

Hexagonal close packing (ABABAB…)

Figure 10.12 *Stacking of hcp close packing of atoms.*

10.11.1.1 Hexagonal Close Packing (hcp)

HCP is a very common type of structure for elemental metals. Examples are Be, Mg, Ti, Zr, etc.

10.11.1.2 Coordination in hcp Structure

Each host atom in an hcp lattice is surrounded by and touches 12 nearest neighbours, each at a distance of $2r$:

The coordination of the central atom of an hcp is shown in Fig. 10.13.

- There are six atoms in the planar hexagonal array (the central A layer);
- There are three atoms in the B layer above the A layer;
- There are three atoms in the B layer below the A layer.

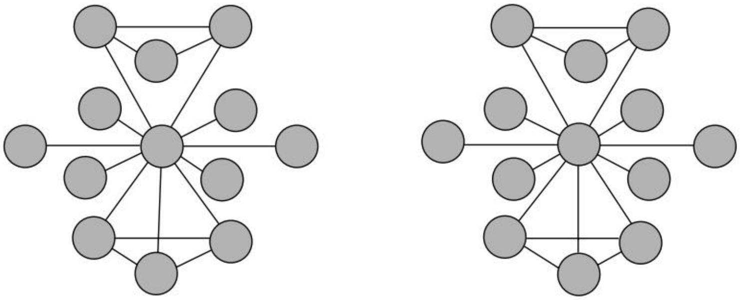


Figure 10.13 *Coordination of the central atom of hcp.*

10.11.2 Cubic Close Packing (ccp)

In contrast to the hcp type of close packing of layers as described above, an alternate method of stacking the third layer is packed in such manner that it lies in a unique position, in this way an 'ABCABC' close-packed layer sequence can be created. This method of stacking is called cubic close packing (ccp) (Fig. 10.14).

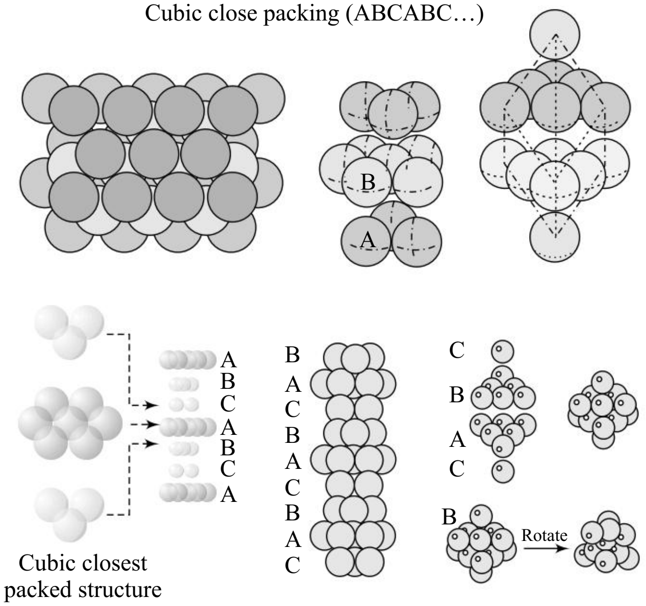


Figure 10.14 *Stacking of cubic close-packed structure.*

This is called cubic close packing because the smallest unit cell that can describe this arrangement is face-centred cubic. *Two sides of the unit cell* are apparent in Fig. 10.15.

By re-orienting the structure the entire face-centred cubic cell and its contents are apparent in Fig. 10.16.

Because the ccp structure is still close packed it is as efficient in its packing as the hcp structure (74 percent).

Examples of ccp (fcc) (ABCABC) metal structures are Al, Ni, Cu, Ag, Pt, etc.

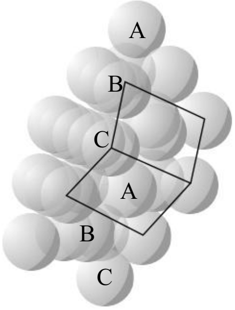


FIGURE 10.15 *A side view of stacking of an fcc lattice.*

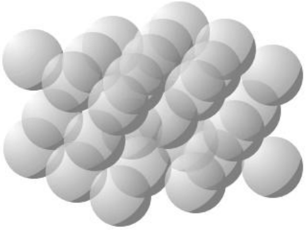


FIGURE 10.16 *A typical ccp lattice from another angle.*

10.11.2.1 Coordination in CCP

Each host atom in a CCP lattice is surrounded by and touches 12 nearest neighbours, each at a distance of $2r$:

The coordination of the central atom in a ccp lattice is shown in Fig. 10.17.

- There are six atoms in the central A layer;
- There are three atoms in the B layer above the A layer;
- There are three atoms in the C layer below the A layer.

The six atoms in the B and C layers form a *trigonal antiprism* (also known as a distorted octahedron) around the central atom.

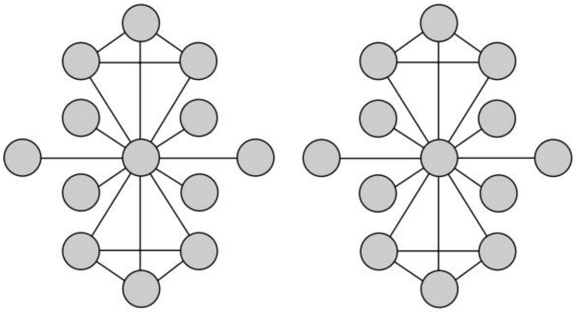


FIGURE 10.17 *Coordination of central atom in ccp.*

10.12 Voids in a Close-packing

The stacking of two close-packed anion layers produces two types of voids or holes as depicted in Fig. 10.18. One set of holes are *octahedrally* coordinated by six anions, the second set are *tetrahedrally* coordinated by four anions. One octahedral site and two tetrahedral sites are created by each anion in the close-packed layer.

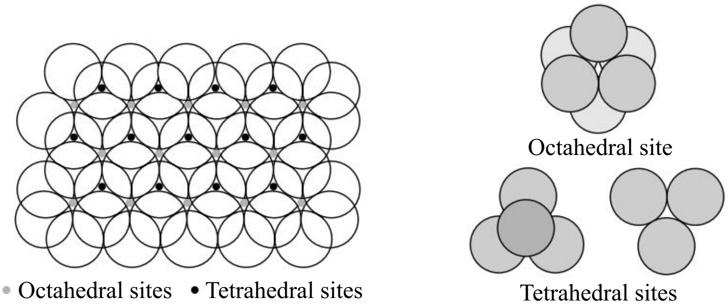


FIGURE 10.18 *Types of voids in close-packed structure.*

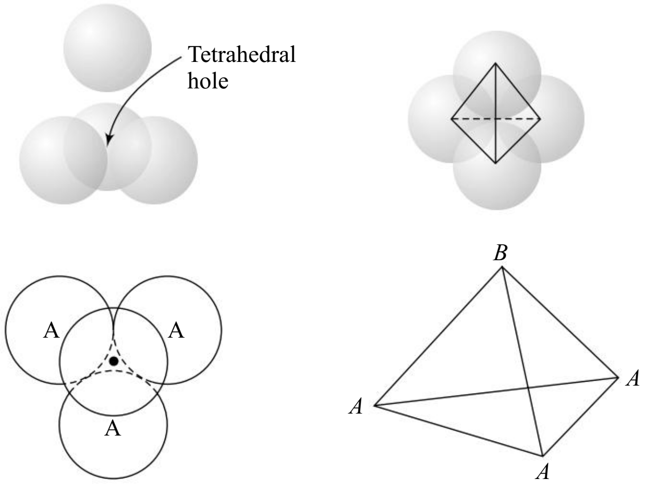
Having determined what types of holes/voids are available, we must now decide:

(a) Which sites are occupied by a given cation? This is determined by the *radius ratio* (= $r_{\text{cation}}/r_{\text{anion}}$).

(b) How many sites are occupied? This is determined by the stoichiometry.

In the case of close-packed inorganic compounds, the larger atoms or ions occupy positions approximately corresponding to those of equal spheres in a close-packing while the *smaller atoms* are distributed among the voids. Three-dimensional close-packing of spheres has two kinds of voids:

(i) If the triangular void in a close-packed layer has a *sphere directly over it*, there results a *void* with four spheres around it, as shown in Fig. 10.19(a).



(a) Void site (b) Tetrahedral site

Figure 10.19 *(a)–(b) A tetrahedral void/site.*

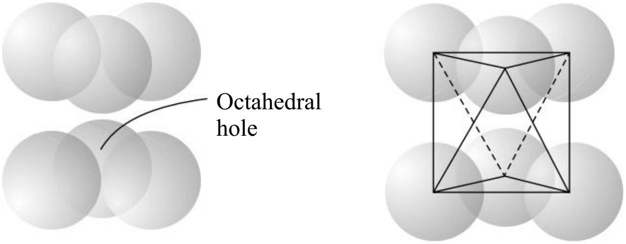
Such a void is called a *tetrahedral void* since the four spheres surrounding it are arranged on the corners of a regular tetrahedron [Fig. 10.19(b)]. If '*R*' denotes the radius of the four spheres surrounding a tetrahedral void, the radius of the spheres that would just fit into the void is given by 0.225 *R*.

(ii) If a triangular void pointing up in one close-packed layer is *covered by a triangular void pointing down in the adjacent layer*, then a void surrounded by six spheres results [Fig. 10.20(c)]. Such a void is called an octahedral void since the six spheres surrounding it lie at the corners of a *regular octahedron* [Fig. 10.20(d)]. The radius of the sphere that would just fit into an octahedral void in a close-packing is given by 0.414 *R*.

10.12.1 Determination of Number of Tetrahedral and Octahedral Sites in Close-Packed Structure

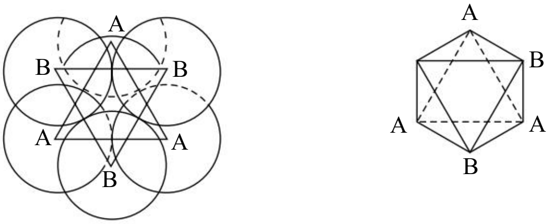
To determine the number of *tetrahedral* and *octahedral* voids in a three-dimensional close-packing of spheres, we note that a sphere in a hexagonal close-packed layer A is surrounded by three B voids and three C voids.

When the next layer is placed on the top of this, the three voids of one kind (say B) are occupied and the other three (say C) are not. Thus, the three B voids become *tetrahedral voids* and the three C voids become *octahedral voids*. A single sphere in a three-dimensional close-packing will have similar voids on the lower



(*a*) (*b*)

FIGURE 10.20



(c) Octahedral void (d) Octahedral site

FIGURE 10.20 *(c) and (d) An octahedral void/site.*

side as well. In addition, the particular sphere being considered covers a triangular void in the layer above it and another in the layer below it. Thus, two more tetrahedral voids surround the spheres. This results in 2 × 3 + 1 + 1 = 8 tetrahedral voids and 2 × 3 = 6 octahedral voids surrounding the sphere. Since a tetrahedral void is shared by four spheres, there are twice as many tetrahedral voids as there are spheres. Similarly, since an octahedral void is surrounded by six spheres, there are as many octahedral voids as there are spheres.

10.13 Ionic Crystals and Radius Ratio

10.13.1 Radius Ratio Rule

In ionic solids, each ion tends to surround itself with as many ions of opposite charge as possible in order to reduce the potential energy and obtain maximum stability. The difference in size of the two layers of ions forming the ionic crystal lattice is also taken into consideration while interpreting the geometry of this close packing.

For stability of an ionic compound, each cation should be surrounded by maximum number of anions and vice versa. *The 'coordination number' is defined as the number of opposite charge ion surrounding a given ion.*

The ratio of the radius of the cation (r_+) *to that of anion* (r_-) *is called radius ratio*.

Thus, $R = (r_+)/(r_-)$

Evidently, greater is the radius ratio, the larger the size of the cation and hence greater is the coordination number.

The crystals are classified according to the coordination number of the constituent ions.

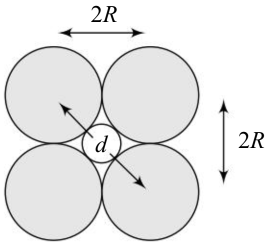
Thus the 'radius ratio' of ions in crystals can be used to predict the geometry (or shape) of ionic crystals.

Radius ratio	Coordination number	Geometry of crystals
0–0.155	2	Linear
0.155–0.225	3	Trigonal planer
0.225–0.414	4	Tetrahedral
0.414–0.732	4	Square planer
0.414–0.732	6	Octahedral
0.732–1.00	8	B.C.C
1.0	12	Close packing (hcp or ccp)

Table 10.1

Prediction of crystal geometry from radius ratio and coordination number

The relative sizes of the anions and cations required for a perfect fit of the cation into the octahedral sites in a close-packed anion (Fig. 10.21) array can be determined by simple geometry:



$$d = 2R_{\text{anion}} + 2r_{\text{cation}}$$

also

$$d = \sqrt{[(2R_{\text{anion}})^2 + (2R_{\text{anion}})^2]}$$

Therefore: $r_{\text{cation}}/r_{\text{anion}} = 0.414$

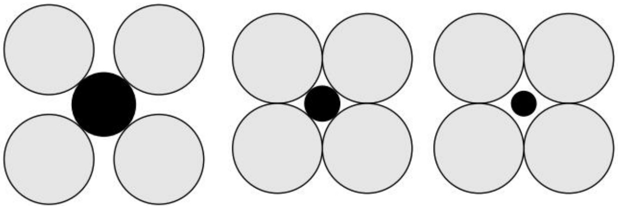
Figure 10.21 *The relative sizes of the anions and cations for perfect fit in lattice.*

Similarly, for a perfect fit of a cation into the tetrahedral sites it can be shown that $r_{\text{cation}}/r_{\text{anion}} = 0.225$. For these two 'ideal fit' radius ratios, the anions remain close-packed.

10.14 Stable Bonding Configurations in Ionic Solids

An ideal fit of a cation into the close-packed anion arrangement is rarely observed. Consider the placing a cation that is (a) larger than the ideal and (b) smaller than the ideal, into the cation sites.

If the bonded cation and anion are in contact with each other results in a stable coordination (Fig. 10.22).



$r_{\text{cation}}/r_{\text{anion}} >$ ideal
Stable

$r_{\text{cation}}/r_{\text{anion}} =$ ideal
Stable

$r_{\text{cation}}/r_{\text{anion}} <$ ideal
Unstable

Figure 10.22 *Stable bonding configuration in a lattice.*

If the cation is larger than the ideal radius ratio value the cation and anion remain in contact, however the cation forces the anions apart. If the cation is too small for the site then the cation would 'rattle' and would not be in contact with the surrounding anions. This is an unstable bonding configuration.

However, in a few rare cases solids do contain cations that are too small for their sites, in these cases the cation moves off the centre of the site and adopt a distorted octahedral coordination. These solids typically exhibit novel properties, such as, for example, ferroelectricity and piezoelectricity.

By including the possibility of a cubic coordination, we can now complete our table for predicting cation coordinations from the radius ratio rules (Fig. 10.23).

$$r_{\text{cation}}/r_{\text{anion}} = \xleftrightarrow[\text{tetrahedral}]{0.225}\xleftrightarrow[\text{octahedral}]{0.414}\xleftrightarrow[\text{cubic}]{0.73}$$

10.14.1 Examples of Cubic Close Packing (CCP) Structures

In these structures (Fig. 10.24), the anions are not close packed, but occupy just the corners of a cube. In this case the centre of the cube (surrounded by eight anions) can be occupied by a suitably sized cation. This site is larger than the *tetrahedral* or *octahedral* positions in the close-packed structures. The radius ratio for a perfect fit of a cation in a cubic site can again be calculated using radius ratio and is 0.73. This structure should therefore be adopted when the $r_{\text{cation}}/r_{\text{anion}}$ is equal to or greater than 0.73.

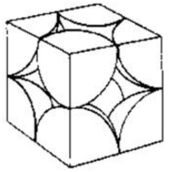
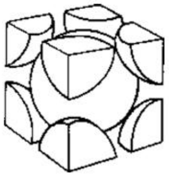
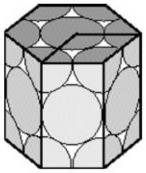
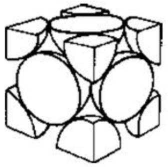
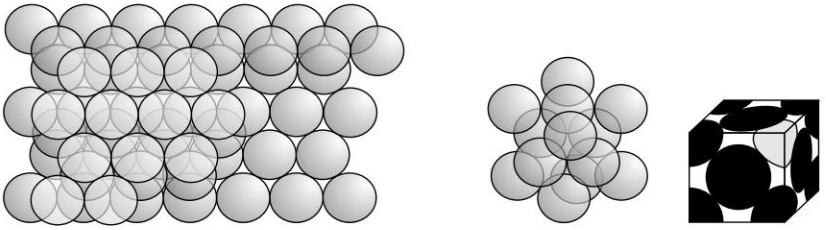
Type of Packing	Packing Efficiency	Coordination Number
Simple cubic (sc) 	52%	6
Body-centered cubic (bcc) 	68%	8
Hexagonal close packed (hcp) 	74%	12
Cubic close-packed (ccp or fcc) 	74%	12

FIGURE 10.23 *Type of Close packing, efficiency and coordination number.*



ABC layer stacking

Cubic unit cell

FIGURE 10.24 *ABC layer packing and ccp unit cell.*

The ABCABC structure is called face-centred cubic (fcc). It also has each atom with 12 nearest neighbours and the atoms fill 74.04 percent of the available space. The difference in the structure is in the different long ranged order and the unit cell.

The (fcc) structure is just one of the structures that is derived from a cubic unit cell (right angles, equal length edges).

The cubic cells are shown in Fig. 10.25.

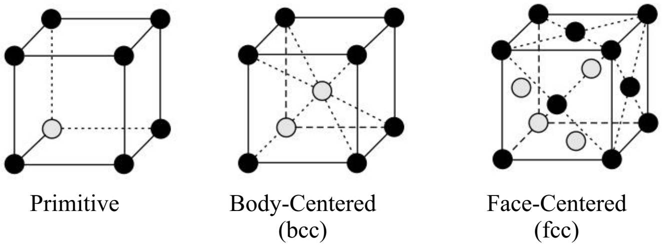


FIGURE 10.25 *Primitive, bcc and fcc cubic unit cells.*

The number of atoms in the unit cell is not the same as the coordination number (number of nearest neighbours).

In the body-centred cubic (bcc) structure above the number of atoms in the unit cell is two but the number of nearest neighbours is eight. (The number of gray atoms in the above gives the number of atoms in the unit cell).

The (bcc) structure is not as tightly packed as the (hcp) or (fcc) structures, with the atoms occupying only 68.02 percent of the available space.

10.14.2 Few Examples

(i) Caesium Chloride Structure (CsCl)

Radius of Cs^+ = 1.74 Å, radius of Cl^- = 1.81 Å: Radius ratio = 0.96 = reduction: cubic coordination.

All cubic sites are filled by Cs^+ cations (Fig. 10.26).

Coordination numbers: $Cs^+ = 8$; $Cl^- = 8$

Note: Cs^+ and Cl^- ions are in contact along the body diagonal.

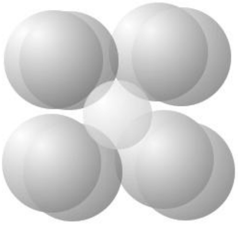


FIGURE 10.26 *A bcc lattice of CsCl.*

(ii) Fluorite structure (CaF_2)

Simple cubic arrangement of anions: 50 percent cubic sites filled.

In CaF_2, ionic radius of Ca^{2+} = 1.12Å; radius F^- = 1.31 Å;

Therefore, radius ratio = 0.85, and Ca^{2+} cubic coordination. One cubic site per F^- anion.

From stoichiometry, only 50 percent cubic sites are filled by Ca^{2+} cations (Fig. 10.27).

Arrangement of the filled cubic sites is such that the Ca–Ca distances are as large as possible (compare the Ca distribution to that of Zn in ZnS).

Coordination numbers: Ca^{2+} surrounded by eight F^- ions; and F^- surrounded by four Ca^{2+} ions.

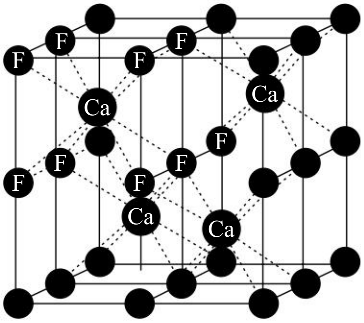


FIGURE 10.27 *A fluorite structure.*

Filling of cation sites from proper stoichiometry

Figure 10.28 shows a view of the octahedral and tetrahedral interstices that are available in the fcc cell of a ccp anion arrangement. By filling these to differing degrees, a number of very common types of crystal structures (Fig. 10.28) can be produced.

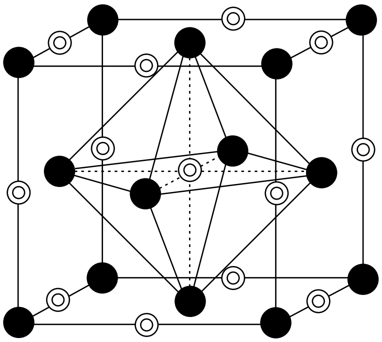


FIGURE 10.28 *A view of the octahedral and tetrahedral interstices.*

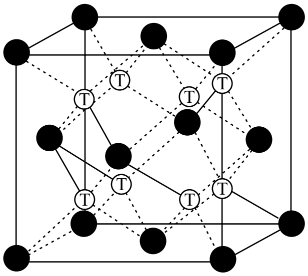


FIGURE 10.29 *'O' represent centres of the octahedral interstices.*

Circles labelled O represent *centres of the octahedral interstices* in the ccp arrangement of anions (fcc unit cell). The cell 'owns' four octahedral sites (Fig. 10.29).

Circles labelled T represent the *centres of the tetrahedral interstices* in the ccp arrangement of anions. The cell 'owns' eight tetrahedral sites (Fig. 10.29).

(iii) Sodium chloride (NaCl) Structure

The sodium chloride has an anion cubic close-packed structure—(fcc).

Radius Na^+ = 1.02 Å, radius Cl^- = 1.81 Å; therefore radius ratio = 0.563

Therefore, Na^+ and Cl^- ions have octahedral coordinations.

Therefore, octahedral sites are 100 percent filled.

Coordination of Na^+ = 6; coordination of Cl^- = 6

Figure 10.30 shows the filling of the octahedral sites by Na^+ (green) within the ccp (ABCABC) Cl^- anion (marked A, B and C) array.

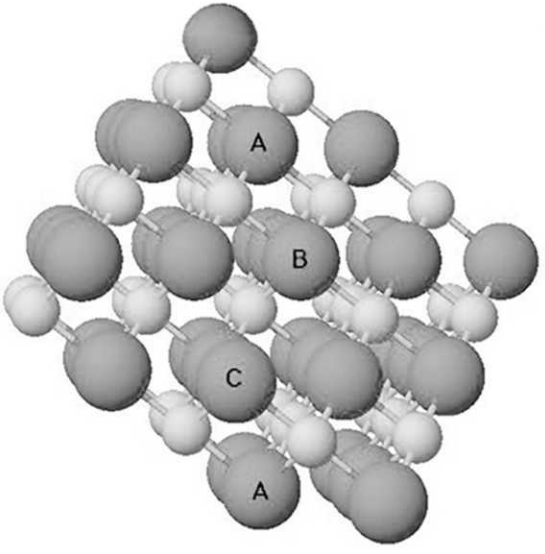


FIGURE 10.30 *ABC packing of NaCl.*

Figure 10.31 shows the face-centred cubic unit cell of the NaCl structure.

For an fcc lattice there are four lattice points per cell, the motif in this case is a Na^+ cation and a Cl^- anion. Therefore, the cell contents are four Na+ and Cl^- anions.

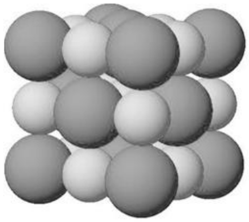


Figure 10.31 *An fcc lattice of NaCl.*

(iv) Zinc blende (Zinc sulphide, ZnS) structure

The ZnS crystal structure has an anion cubic close packing (fcc),

Radius Zn^{2+} = 0.6Å, radius S^{2-} = 1.84 Å;

Therefore, radius ratio = 0.33

Zn occupies tetrahedral sites of (fcc) cubic structure = refers two tetrahedral sites/anion, therefore from formula of ZnS only 50 percent of the tetrahedral sites can be filled (Fig. 10.32).

Coordination number of Zn^+ ion = 4; coordination No of S^{2-} = 4

The filled sites are shown in Fig. 1.32. Note the filling of diagonally opposite sites to maximize the cation-cation separations.

Again the lattice is fcc and the motif consists of one S and Zn.

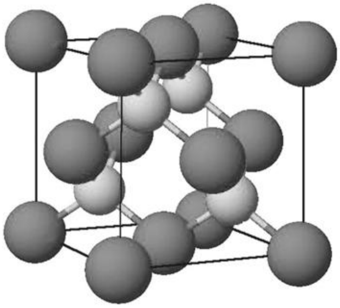


Figure 10.32 *A Zinc blend lattice of ZnS.*

10.14.3 Examples of Hexagonal Close Packing (hcp) Structures

The filling of the octahedral and tetrahedral sites in an hcp (ABAB) array of anions follows simple methods. Many common structural types can be generated.

The close packing of atoms leads to possibility of two unique structures, depending on how planes of 2D closest packed spheres are layered.

If every other layer is exactly the same then we have a so, called ABABA structure.

If not, then the structure is ABCABCABC. Figure 10.33 shows the difference between these two structures:

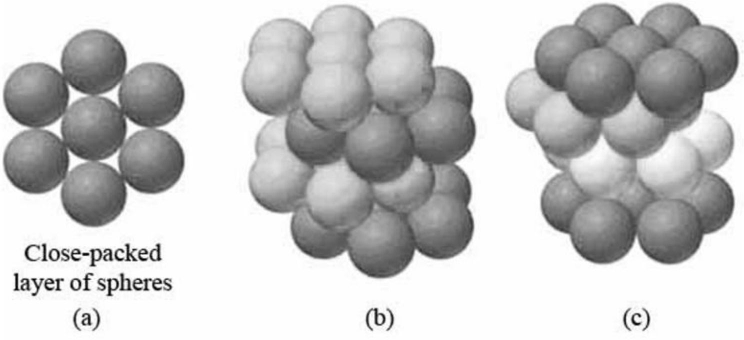
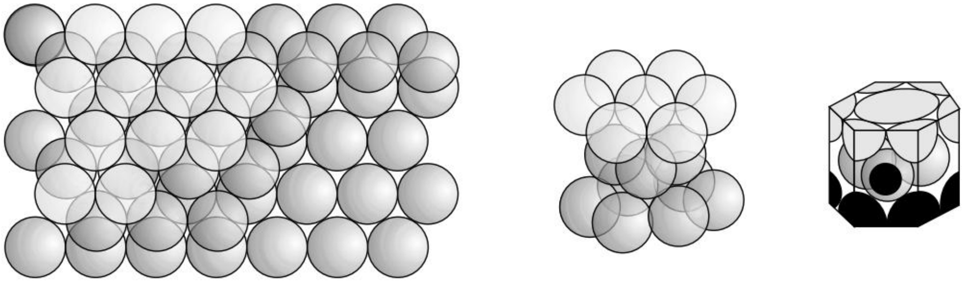


Figure 10.33 *Stacking of hcp close packing of atoms.*

The ABABAB structure as shown in the above Fig. 10.33(b) is called the *hexagonal closest packed (hcp)* structure. In this structure (Fig. 10.34), each atom has 12 nearest neighbours and the volume of the spheres fills the maximum possible space: 74.04 percent.



ABAB layer stacking Hexagonal structure

FIGURE 10.34 *ABAB stacking of lattice and hcp structure.*

10.15 Crystal Structure

1. The concept of crystalline lattice

It is desirable at this stage to understand few terms and concepts involved in the study of ionic crystal structures.

A 'crystal' is a solid substance having a definite geometric shape with flat faces and share edges.

A cubic crystal is an example of one of the crystal systems.

'Ionic crystals' are those which exist in a well defined crystalline form with their ions arranged in certain definite geometric patterns in the entire three-dimensional network of the crystals.

2. Crystal lattice

'Crystal lattice' is a geometrical concept in crystalline structure. The symmetry observed in crystals as exhibited by their crystal faces is due to the ordered internal arrangement of atoms in a crystal structure. This arrangement of atoms in crystals is called a *lattice*.

Crystal lattice is a highly ordered three-dimensional structure formed by its constituent ions (or atoms). An idealized crystal lattice is shown in Fig. 10.35.

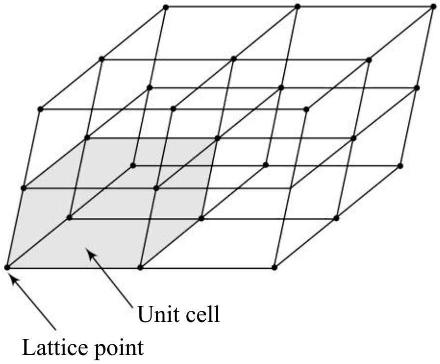


FIGURE 10.35 *Lattice points and unit cell.*

In two-dimensions, a plane lattice consists of an orderly array of points (Fig. 10.36). The array is defined by the spacing and the directions (or angles) between the points. Thus, the array can be reproduced by specifying the distance and angle to move from point to point. This is referred to as *translational symmetry*. In the example here, the array is reproduced by moving down a distance *a* and moving to the right a distance *b*. The angle between the two directions of translation in this case is 90°.

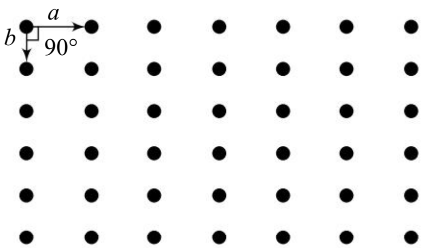


FIGURE 10.36 *Two plane lattice.*

In the example to the right, the translation distances 'a' and 'b' are not equal and the translation angle is not 90° (Fig. 10.37).

Crystals, of course, are made up of three-dimensional arrays of atoms. Such three-dimensional arrays are called *space lattices*.

3. Lattice point

If the constituents of crystal are denoted by points, then the 'lattice point' is a point in an array, all of which have identical surroundings and may also be regarded as an array of indefinite set of points, repeated regularly throughout space is shown in Fig. 10.36.

4. Space lattice

'Space lattice' is an array of points, as shown in the above Fig. 10.36, indicating how ions (atoms/molecules) are arranged at different sites in three-dimensional space.

5. Unit cell lattice

A 'unit cell lattice' is the smallest repeating unit in space lattice which when repeated over and over again results in a crystal structure of the substance.

A unit cell lattice of a particular crystal type (Fig. 10.38) indicates lattice site (ions/atoms) positions, as well as the lattice parameters.

6. Crystallographic axes

The crystallographic axes are *imaginary lines* (Fig. 10.39) that we can draw within the crystal lattice. These will define *a coordinate system* within the crystal. For three-dimensional space lattices, we need three crystallographic axes *that define directions within the crystal lattices*. Depending on the symmetry of the lattice, the directions may or may not be perpendicular to one another, and the divisions along the coordinate axes may or may not be equal along the axes.

The lengths of the axes are in some way proportional to the lattice spacing along an axis and this is defined by the smallest group of points necessary to allow for translational symmetry to reproduce the lattice.

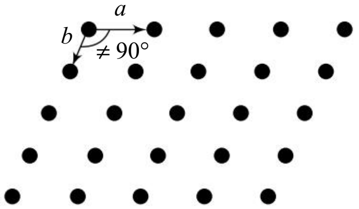


FIGURE 10.37 *Two plane lattice with translation.*

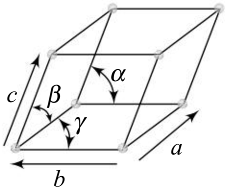


FIGURE 10.38 *A unit cell lattice.*

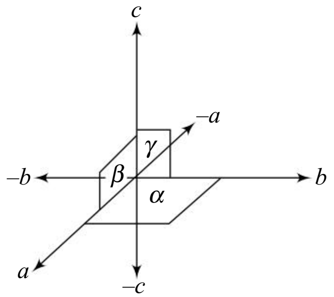


FIGURE 10.39 *Imaginary crystallographic axes that define directions within the crystal lattices.*

10.15.1 The Basic Concepts of the Crystallographic Axes

The axes are defined based on the symmetry of the lattice and the crystal. Each crystal system has different conventions that define the orientation of the axes (Fig. 10.40), and the relative lengths of the axes.

10.15.2 The Unit Cell

The 'lengths' of the various crystallographic axes are defined on the basis of a *unit cell*. When arrays of atoms or molecules are laid out in a space lattice we define a group of such atoms as the unit cell. *This unit cell contains all the necessary points on the lattice that can be translated to repeat itself in an infinite array*. In other words, the unit cell defines the basic building blocks of the crystal, and the entire crystal is made up of repeatedly translated unit cells.

The simplest repeating unit in a crystal is called a *unit cell*. Each unit cell is defined in terms of *lattice points*—the points in space about which the particles are free to vibrate in a crystal.

In 1850, Auguste Bravais showed that crystals could be divided into 14 unit cells, which meet the following criteria.

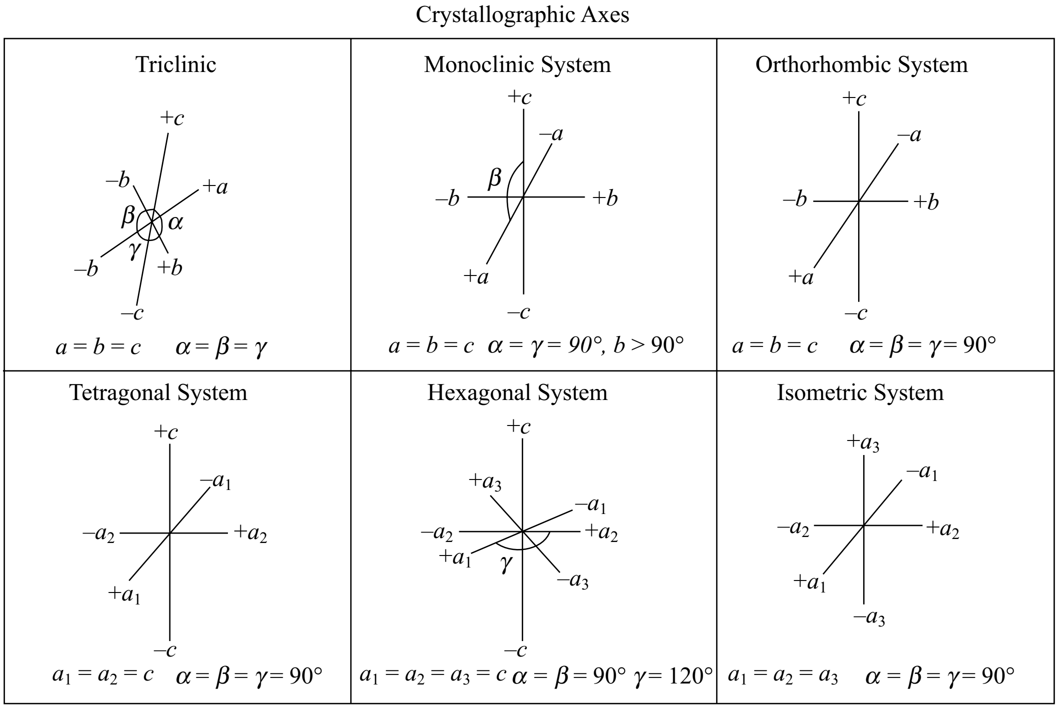


Figure 10.40 *The orientation of the axes, and the relative lengths of the axes.*

The unit cell is the simplest repeating unit in the crystal.

- Opposite faces of a unit cell are parallel.
- The edge of the unit cell connects equivalent points.

In defining a unit cell for the crystal the choice is somewhat arbitrary. But, the best choice is one where:

- The edges of the unit cell should coincide with the symmetry of the lattice.
- The edges of the unit cell should be related by the symmetry of the lattice.
- The smallest possible cell that contains all elements should be chosen.

Once we have chosen a unit cell for the crystal, then it can be oriented on the crystallographic axes to define the angles between the axes and to define the axial lengths. This will allow us to define directions within the crystal that become important when we realize that many properties of crystals depend on direction in the crystal. Properties that depend on direction in the crystal are called *vectorial properties*.

Another important point is that the relative lengths of the crystallographic axes, or unit cell edges, can be determined from measurements of the angles between the crystal faces.

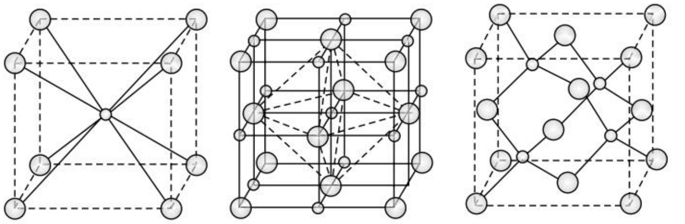
The structures of the unit cell for a variety of salts are shown below (Fig. 10.41).

10.15.3 Lattice Parameters

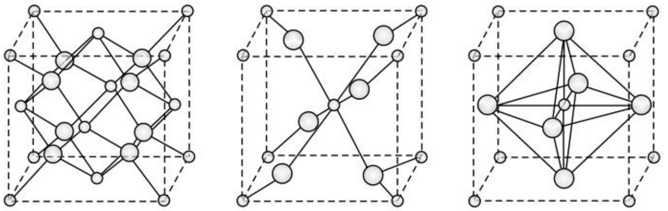
A unit cell lattice, as shown in Fig. 10.42, is described in terms of (1) lengths of the edges of unit cell (a, b and c) and (2) the interfacial angles (α, β and v). These are collectively known as the 'unit cell parameters'.

Example The unit cell parameters of a simple cubic system are

$$a = b = c;\ \alpha = \beta = v = 90^\circ$$



CaCl NaCl Zinc blende (cubic ZnS)



Flourite (CaF_2) Rutile (TiO_2) Perovskite ($CaTiO_2$)

○ = Ca^{2+} ○ = Ti^{4+} ○ = Ti^{4+} ○ = Ca^{2+} ◯ = O^{2-}

Figure 10.41 *Examples Unit cells of variety of salts.*

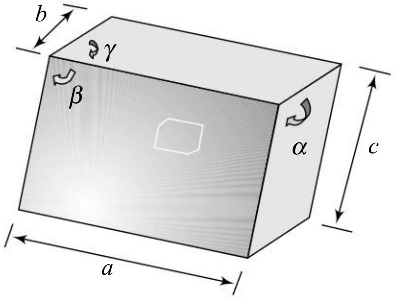


Figure 10.42 *Lattice parameters of a unit cell.*

10.16 Crystallography

It is the branch of science which deals with the geometry, properties and structure of crystalline substances.

The geometry of a crystal can be studied in terms of the following.

(i) Crystallographic axes

The geometry of a crystal can be best described in terms of its crystallographic axes (Fig. 10.43). These are generally three lines meeting at a point at right angles to each other. The rule of crystallographic orientation, according to the priority, is that:

- the c-axis is straight up,
- the b-axis is towards the right side and
- the a-axis goes toward yourself.

As per the convention, the angles between the crystallographic axes are represented by α, β and γ, respectively, Fig. 10.44,

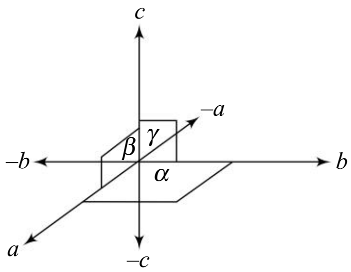


Figure 10.43 *Crystallographic axes.*

(ii) Lattice plane

The face of a crystal is described in terms of 'lattice plane' or 'unit plane' as shown in Fig. 10.45(a).

'Lattice plane' is described as an imaginary plane passing through maximum number of lattice points as

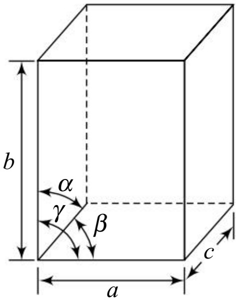


Figure 10.44 *The angles between the crystallographic axes are represented by α, β and γ.*

shown in Fig. 10.45(a) of a space lattice of a crystal. A crystal has number of such equidistant parallel planes, Fig. 10.45(b).

These lattices planes are described by a set of integers (*hkl*).

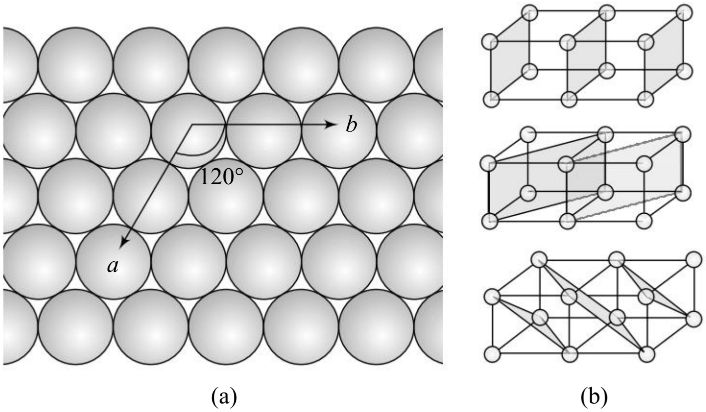


Figure 10.45 *Lattice plane and parallel equidistant lattice planes.*

10.17 Geometric Crystallography

A crystal of a substance has geometric shape with flat planes and sharp edges.

'Geometric Crystallography' is concerned with geometric shapes, properties and arrangement of planes of crystals.

The geometric crystallography is based on the three fundamental laws, they are:

(a) The law of constancy of interfacial angles,
(b) The law of rationality of indices, and
(c) The law of symmetry.

10.17.1 The Law of Constancy of Interfacial Angles

Consider a cubic crystal as shown in figure.

The angle between any two lattice planes of a crystal is called an interfacial angle and is always constant (Fig. 10.46). The interfacial angle between two planes of a cubic crystal is always 90°.

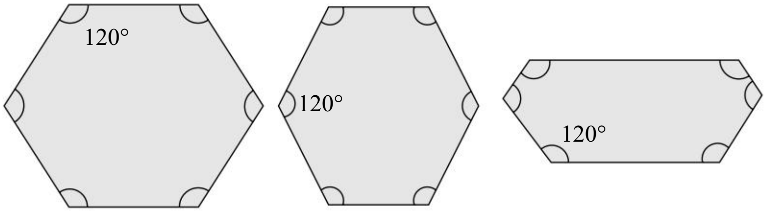


Figure 10.46 *The interfacial angle between two planes.*

'The law of constancy of interfacial angles states that for a given crystal, the corresponding planes forming the external plane of a crystal intersect a definite angle, called interfacial angle and is constant no matter how the planes develop.'

10.17.2 The Law of Rationality of Indices

A 'space lattice' of a particular crystal system can have different intercepts along the crystallographic axes of the crystal.

'The law of rationality of indices of a crystal states that the intercepts of various planes of a crystal along the crystallographic axes are simple integral multiples of the intercepts made by the unit plane.'

It is therefore desirable to characterize and index the lattice planes of a crystal for a better understanding of the geometry and the arrangement of these planes in a crystal. The following procedures have been adopted to index a particular 'lattice plane' of a crystal.

Weiss indices

The geometry of a crystal is always described in terms of the crystallographic axes as shown in Fig. 1. Let the intercepts of a unit lattice plane on the three crystallographic axes are a, b and c, while the intercepts of a parallel plane KLM is *la*, *mb* and *nc*, where *l*, *m* and *n* are small integers.

This set of integers; *l*, *m* and *n* are called the Weiss indices of the corresponding KLM plane. This particular plane of the crystal is described as (*lmn*) plane.

For reasons, Weiss method of indexing a plane has been discarded.

10.17.3 Interpretation of the Planes

In crystallography, *a crystal is described in terms of a regular, identically repeating array of points called a* ***lattice***. The points of the lattice form lines, which need not be perpendicular to each other. Along any one line, each adjacent pair of points is separated by a constant spacing called the *lattice constant* in that direction. To obtain the crystal structure, it is necessary to associate with each *lattice point an atom or group of atoms*, called the *basis*. Repetition of the basis at each lattice point produces the crystal (Fig. 10.47).

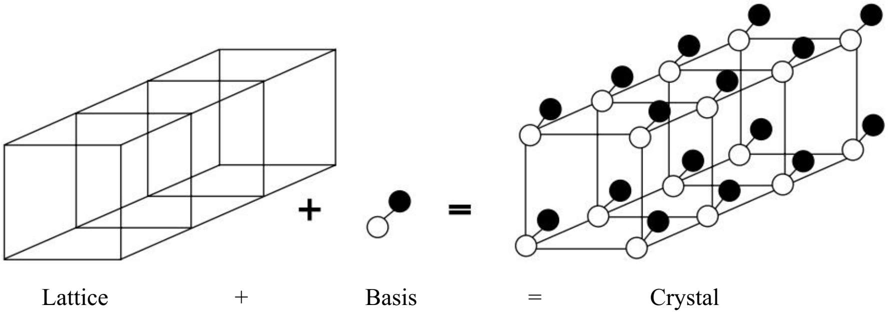


Figure 10.47 *Description of a crystal in terms of lattice and basis.*

It is somewhat arbitrary how the lattice and basis are chosen to describe a given crystal. It is common to describe a crystal in terms of the most conveniently visualized lattice suggested by the symmetry of the crystal.

10.18 Miller Indices

Various planes of a crystal are to be characterized and indexed for a better understanding of the crystal.

Miller introduced a set of integers (*hkl*) to specify a plane of the crystal. This set of three numbers, (*hkl*), is known as 'Millers indices' of a particular plane of a crystal.

'The Miller indices (hkl) of a plane of a crystal are inversely proportional to the intercepts of that plane on the three crystallographic axes' (Fig. 10.48).

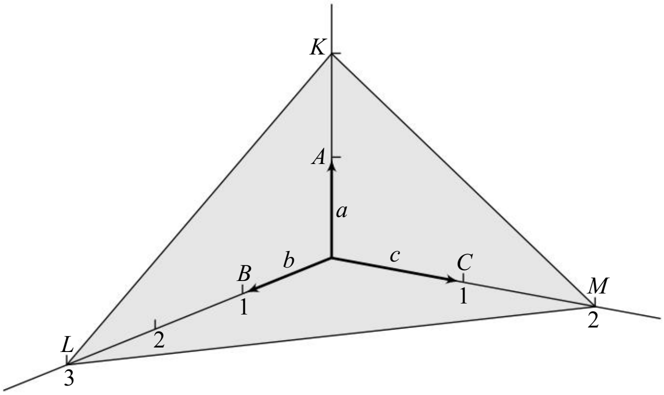


FIGURE 10.48 *Crystallographic axes and intercepting planes.*

It is customary to describe crystal planes by a set of **Miller indices**. In brief, Miller indices of a plane of crystal can be obtained, thus:

1. *Find the intercepts on the crystal axes as multiples of the lattice constants* a_1, b_2, c_3.
2. *Take reciprocals of these numbers and*
3. *Using an appropriate multiplier, convert the 1/intercept set to the smallest possible set of whole numbers.*
4. *Represent the above as a set of integers (hkl) for a given plane of a crystal.*

10.18.1 Miller Indices d_{hkl} of a Plane of a Crystal

The *orientation of a surface or a crystal* plane may be defined by considering how the plane (or indeed any parallel plane) intersects the main crystallographic axes of the solid. The application of a set of rules leads to the assignment of the Miller Indices, (*hkl*); a set of numbers which quantify the intercepts and thus may be used to uniquely identify the plane or surface.

The following treatment of the procedure used to assign the Miller Indices is a simplified one and only a cubic crystal system (Fig. 10.49).

The procedure is most easily illustrated using an example so we will first consider the following surface/plane (Fig. 10.50):

Step 1: Identify the intercepts on the x-, y- and z-axes.
In this case, the intercept on the x-axis is at $x = a$ (at the point (a, 0, 0)), but the surface is parallel to the y- and z-axes—strictly, therefore, there is no intercept on these two axes but we shall consider the intercept to be at infinity (∞) for the special case where the plane is parallel to an axis. The intercepts on the x-, y- and z-axes are thus

Intercepts: a, ∞ and ∞

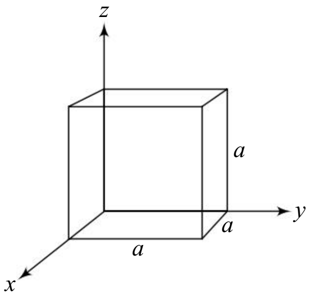


FIGURE 10.49 *A simple cubic lattice and crystallographic axes.*

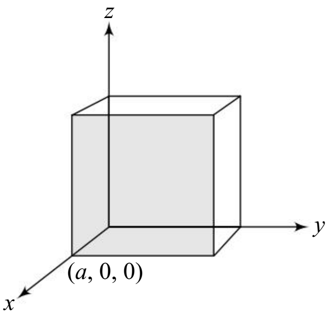


FIGURE 10.50 *(100) Plane of a cubic lattice.*

Step 2: Specify the intercepts in fractional coordinates.

The coordinates are converted to fractional coordinates by dividing by the respective cell-dimension, for example, a point (x, y, z) in a unit cell of dimensions $a \times b \times c$ has fractional coordinates of (x/a, y/b, z/c). In the case of a cubic unit cell, each coordinate will simply be divided by the cubic cell constant, a. This gives:

Fractional Intercepts: a/a, ∞/a, ∞/a, i.e. 1, ∞, ∞

Step 3: Take the reciprocals of the fractional intercepts.

This final manipulation generates the Miller Indices which (by convention) should then be specified without being separated by any commas or other symbols. The Miller Indices are also enclosed within the standard brackets (….) when one is specifying a unique surface which is being considered here.

The reciprocals of 1 and ∞ are 1 and 0, respectively, thus yielding Miller Indices: (100).

Other Examples

1. The (110) surface (Fig. 1.51)

   Assignment

   (i) Intercepts: a, a, ∞

   (ii) Fractional intercepts:1, 1, ∞

   (iii) Miller indices: (110)

2. The (111) surface (Fig. 10.52)

   Assignment

   (i) Intercepts: a, a, a

   (ii) Fractional intercepts: 1, 1, 1

   (iii) Miller Indices: (111)

   The (100), (110) and (111) surfaces considered above are the so-called low index surfaces of a cubic crystal system.

   These surfaces have a particular importance but there is an infinite number of other planes that may be defined using Miller index notation.

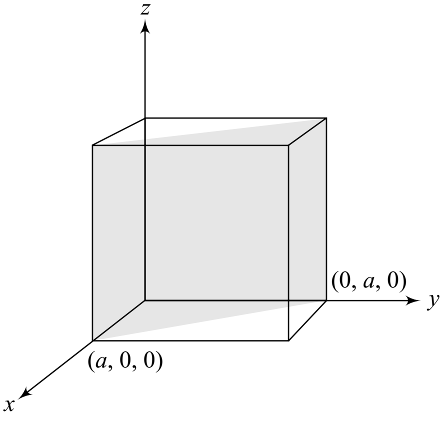


Figure 10.51 *(110) Plane of a cubic lattice.*

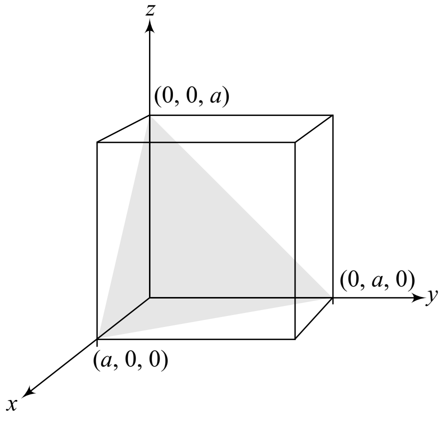


Figure 10.52 *(111) Plane of a cubic lattice.*

3. The (210) surface (Fig. 10.53)

   Assignment

   (i) Intercepts: ½ a, a, ∞

   (ii) Fractional intercepts: ½, 1, ∞

   (iii) Miller indices: (210)

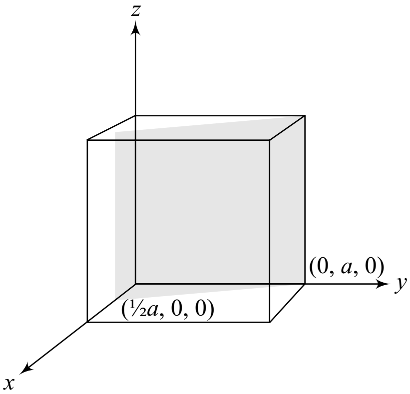


FIGURE 10.53 *(a) (210) Plane of a cubic lattice.*

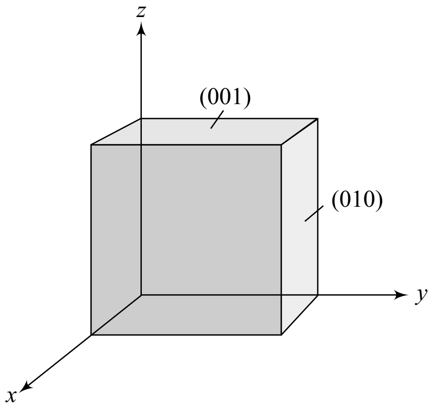


FIGURE 10.53 *(b) (100), (110) and (111) Planes of a cubic lattice.*

Note:

(i) In some instances, the Miller indices are the best *multiplied or divided* through by a common number in order to simplify them by, for example, removing a common factor. This operation of multiplication simply generates a parallel plane which is at a different distance from the origin of the particular unit cell being considered. e.g. (200) is transformed to (100) by dividing through by 2.

(ii) If any of the intercepts are at negative values on the axes then the negative sign will carry through into the Miller indices; in such cases the negative sign is actually denoted by over striking the relevant number.

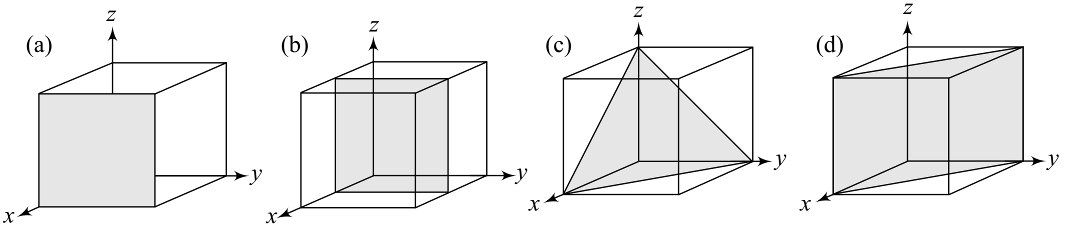
e.g, (00–1) is instead denoted by $(00\bar{1})$

10.18.2 Symmetry—Equivalent Surfaces

In the following diagram (Fig. 10.53) the three highlighted surfaces are related by the symmetry elements of the cubic crystal, they are entirely equivalent.

In fact there are a total of six faces related by the symmetry elements and equivalent to the (100) surface—any surface belonging to this set of symmetry related surfaces may be denoted by the more general notation {100} where the Miller indices of one of the surfaces is instead enclosed in curly-brackets.

So the surface/plane illustrated (Fig. 10.54) is the (100) plane of the cubic crystal.



(100) (200) (111) (110)

FIGURE 10.54 *Different planes of a cubic lattice.*

Note that all planes parallel to each other have the same set of Miller indices. Hence the indices (*hkl*) may denote a single plane or a set of parallel planes.

10.19 Interplanar Distance of a Cubic Crystal

The crystals have number of parallel planes described by their (*hkl*) values (Fig. 10.55).

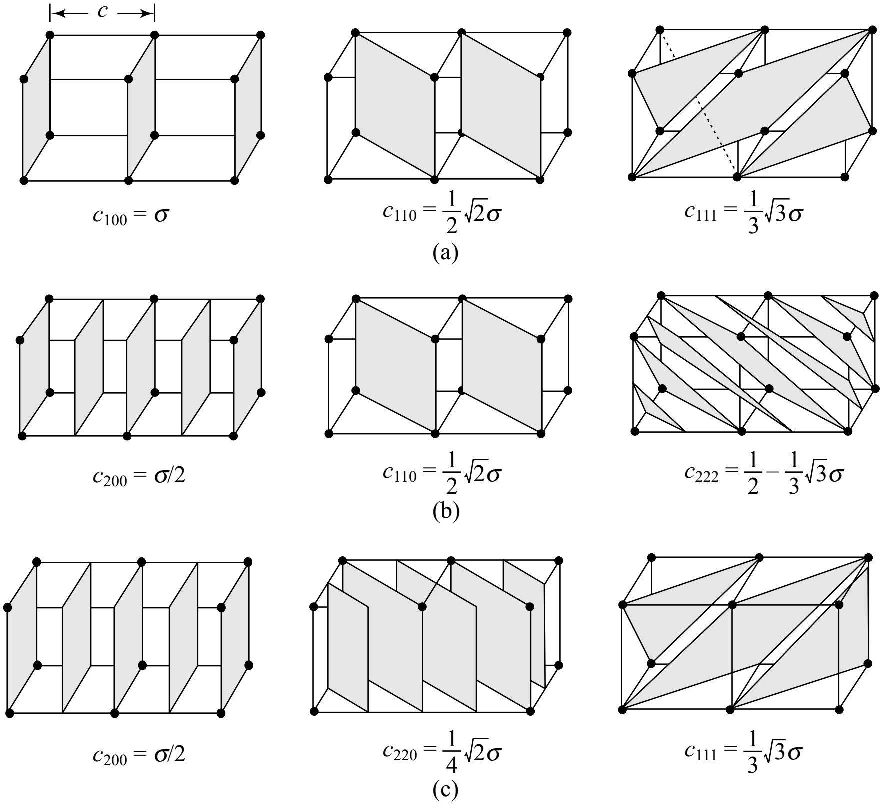


FIGURE 10.55 *Interplanar spacing of simple cubic, bcc and fcc crystals.*

The interplanar distance of any two parallel (*hkl*) planes of a cubic crystal is designated as d_{hkl}.

$$d_{(hkl)} = a/(h^2 + k^2 + l^2)^{1/2}$$

where '*a*' is the edge length of a cube, while (*h*, *k*, *l*) are the Miller indices of a particular plane.

The ratios of interplanar distances of three planes (110), (110) and (111) of cubic crystals are:

(a) Simple cubic : $d_{100} : d_{110} : d_{111} = a : \dfrac{a}{\sqrt{2}} : \dfrac{a}{\sqrt{3}}$

(b) F.C.C : $d_{200} : d_{220} : d_{222} = \dfrac{a}{2} : \dfrac{a}{2\sqrt{2}} : \dfrac{a}{2\sqrt{3}}$

(c) B.C.C : $d_{200} : d_{110} : d_{222} = \dfrac{a}{2} : \dfrac{a}{\sqrt{2}} : \dfrac{a}{2\sqrt{3}}$

10.20 Basics of Symmetry Operations and Space Groups of Crystals

The *Space groups* in three-dimensions are made from combinations of the 32 crystallographic *point groups* with the *14 Bravais lattices* which belong to one of seven crystal systems. This results in a space group being some combination the translational symmetry of a unit cell including lattice centring, and the point group symmetry operations of reflection rotation and rotoinversion. Furthermore, one must consider the screw axis and glide plane symmetry operations. These are called compound symmetry operations and are combinations of a rotation or reflection with a translation less than the unit cell size. *The combination of all these symmetry operations results in a total of 230 unique space groups describing all possible crystal symmetries.*

10.20.1 Symmetry

Crystals have an ordered internal arrangement of atoms. This ordered arrangement shows symmetry, i.e. the atoms are arranged in a symmetrical fashion on a three-dimensional network referred to as a ***lattice****.*

If there are no impediments to crystal its growth, crystal faces form as smooth planar boundaries that make up the surface of the crystal. These crystal faces reflect the ordered internal arrangement of atoms and thus reflect the symmetry of the crystal lattice.

Let us follow the rotation of a two-dimensional crystal composed of atoms in an ordered internal arrangement as shown below (Fig. 10.56).

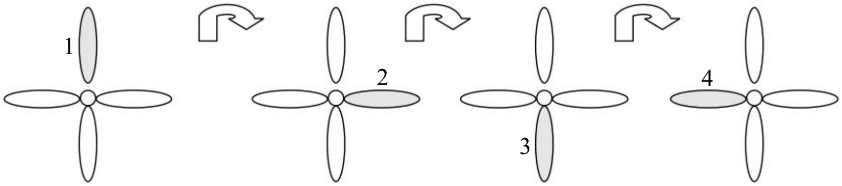


Figure 10.56 *An illustrated example of rotation through 90° to exhibit symmetry operation.*

If we rotate the simple crystals by 90° notice that the lattice and crystal look exactly the same as what we started with. Rotate it another 90° and again it is the same. Another 90° rotation again results in an identical crystal, and another 90° rotation returns the crystal to its original orientation. Thus, in one 360° rotation, the crystal has repeated itself, or looks identical four times. We thus say that this object has fourfold rotational symmetry.

10.20.2 Symmetry Operations and Elements

A symmetry operation is an operation that can be performed either physically or imaginatively that results in no change in the appearance of an object.

It can be clearly stated that in crystals, the symmetry is internal, i.e. it is an ordered geometrical arrangement of atoms and molecules on the crystal lattice. It is evident that the internal symmetry is reflected in the external form of perfect crystals.

There are three types of symmetry operations:

1. rotation,
2. reflection, and
3. inversion.

(i) Rotational Symmetry

If an object can be rotated about an axis and repeats itself *n* times of rotation then it is said to have an axis of *n*-fold rotational symmetry. *The axis along which the rotation is performed is an element of symmetry referred to as a rotation axis*. The following types of rotational symmetry axes are possible in crystals.

(a) 1-fold rotation axis

If crystal face or an object is rotated full 360° in order to restore it to its original appearance, has no rotational symmetry. Since it repeats itself once every 360° it is said to have onefold axis of rotational symmetry (Fig. 10.57).

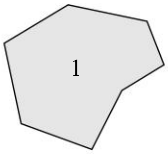


Figure 10.57 *1-fold rotation axis.*

(b) 2-fold rotation axis

A twofold rotation axis arises when a crystal or an object appears identical after a rotation of 180° i.e. twice in a 360° rotation (Fig. 10.58).

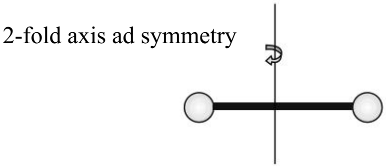


Figure 10.58 *2-fold rotation axis.*

(c) 3-fold rotation axis

A crystal face or an object that repeat themselves upon rotation of 120° are said to have a threefold axis of rotational symmetry (360/120 = 3), and they will repeat three times in a 360° rotation about an axis. The location of a threefold rotation axis is represented as a filled triangle (Fig. 10.59).

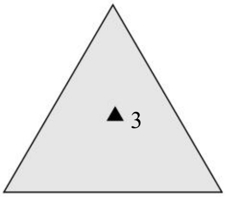


Figure 10.59 *3-fold rotation axis.*

(d) 4-fold rotation axis

If a crystal face or an object repeats itself after 90° of rotation about an axis, it will repeat four times in a 360° rotation, as illustrated in Fig. 10.60.

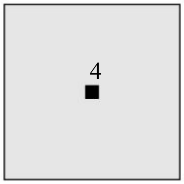


Figure 10.60 *4-fold rotation axis.*

(e) 6-fold rotation axis

If rotation of 60° about an axis causes the object to repeat itself, then it has sixfold axis of rotational symmetry (360/60 = 6). A filled hexagon is used as the symbol for a sixfold rotation axis (Fig. 10.61).

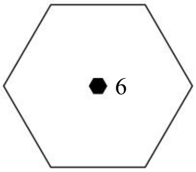


Figure 10.61 *6-fold rotation axis.*

(ii) Mirror Symmetry

A mirror symmetry operation is an imaginary operation that can be performed to reproduce an object. The operation is done by imagining that you cut the object in half, and then place a mirror next to one of the halves of the object along the cut. If the reflection in the mirror reproduces the other half of the object, then the object is said to have mirror symmetry. *The plane of the mirror is an element of symmetry referred to as a* ***mirror plane***, and is symbolized with the letter '*m*'.

The rectangles shown below (Fig. 10.62) have two planes of mirror symmetry.

These rectangles shown above have two planes of mirror symmetry. Three-dimensional and more complex objects could have more. For example, the hexagon shown above, not only has a sixfold rotation axis, but has six mirror planes.

Note that a rectangle does not have mirror symmetry along the diagonal lines as shown below.

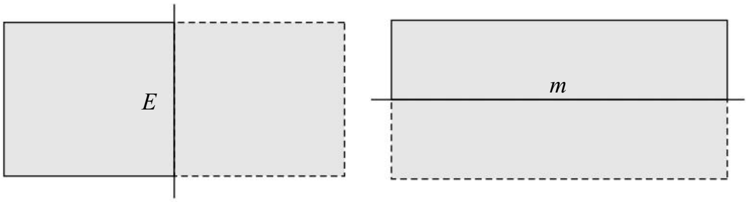


Figure 10.62 *Two types of mirror plane symmetry.*

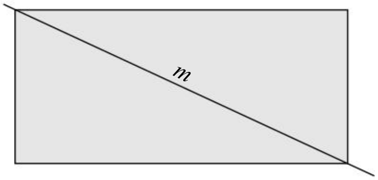


Figure 10.63 *A rectangle does not have mirror symmetry.*

(iii) Centre of Symmetry

Another operation that can be performed is inversion through a point. In this *operation lines are drawn from all points on the object through a point in the centre of the object, called a symmetry centre* (Fig. 10.64). The lines each have lengths that are equidistant from the original points.

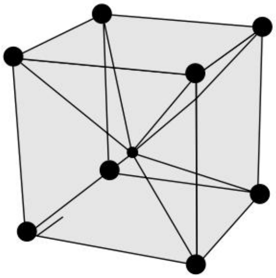


Figure 10.64 *Centre of symmetry of a cube.*

(iv) Rotoinversion

The symmetry operation of rotoinversion is the combinations of rotation with a centre of symmetry.

Any crystals that have rotoinversion symmetry have an element of symmetry called a rotoinversion axis. The centre symmetry of a crystal is also known as one-fold rotoinversion axis.

(a) Two-fold rotoinversion

The twofold rotoinversion operation involves first rotating the object by 180° then inverting it through an inversion centre. *This operation is equivalent to having a mirror plane perpendicular to the two-fold rotoinversion axis* (*Fig. 10.65*). A twofold rotoinversion axis is symbolized as a $\overline{2}$ with a bar over the top. But, since this the equivalent of a mirror plane, '*m*', the bar 2 is rarely used.

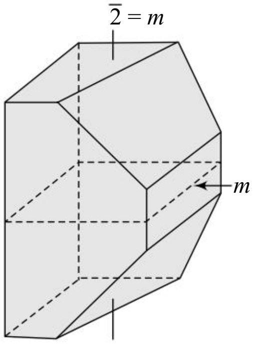


Figure 10.65 *Two-fold rotoinversion.*

(b) Three-fold rotoinversion

A three-fold rotoinversion operation involves a rotation of crystal by 120° (360/3 = 120), and inverting through a centre. A cube is a good example that possesses threefold rotoinversion axes (Fig. 10.66). A threefold rotoinversion axis is denoted as $\overline{3}$. There are actually four $\overline{3}$ axes in a cube, one running through each of the corners of the cube. If one holds one of the $\overline{3}$ axes vertical, then note that there are three faces on top, and three identical faces upside down on the bottom that are offset from the top faces by 120°.

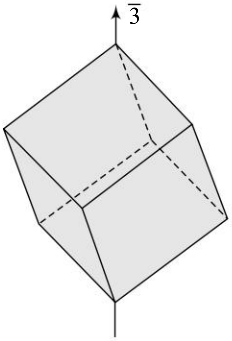


Figure 10.66 *Three-fold rotoinversion.*

(c) Four-fold Rotoinversion

This involves rotation of the object by 90° then inverting through a centre. A four-fold rotoinversion axis is symbolized as $\overline{4}$ (Fig. 10.67). Note that an object possessing a fourfold rotoinversion axis will have two faces on top and two identical faces upside down on the bottom, if the $\overline{4}$ axis is held in the vertical position.

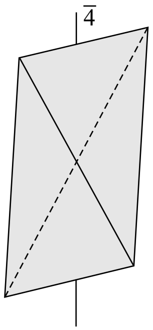


Figure 10.67 *Fourfold rotoinversion.*

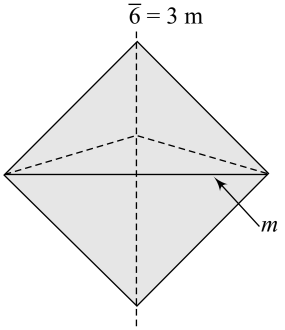


Figure 10.68 *Sixfold rotoinversion.*

(d) Sixfold rotoinversion

A sixfold rotoinversion axis ($\bar{6}$) involves rotating the object by 60° and inverting through a centre (Fig. 10.68). Note that this operation is identical to having the combination of a threefold rotation axis perpendicular to a mirror plane.

10.21 Combinations of Symmetry Operations

The symmetry elements for three-dimensional crystals may have different combinations.

In fact, in crystals there are 32 possible combinations of symmetry elements. These 32 combinations that define the *32 crystal classes*. Every crystal must belong to one of these 32 crystal classes.

It can be emphasized that, *if two kinds of symmetry elements are present in the same crystal, then they will operate on each other to produce other symmetrical symmetry elements.*

Consider a crystal of rectangular-shaped sides with a square-shaped top and bottom. The square-shaped top indicates that there must be a four-fold rotation axis perpendicular to the square shaped face. This is shown in the diagram (Fig. 10.69).

Also, the rectangular shaped face on the left side of the crystal must have a two-fold rotation axis that intersects it. Again, the two-fold axis runs through the crystal and exits on the left-hand side, so that both the left and right-hand sides of the crystal are perpendicular to a two-fold rotation axis.

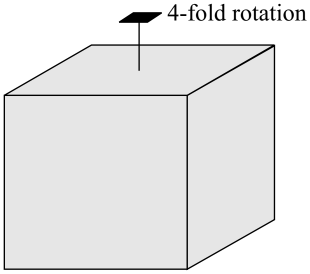


Figure 10.69 *Four-fold rotation axis perpendicular to the square shaped face.*

Since the top face of the crystal has a four fold rotation axis, operation of this four-fold rotation must reproduce the face with the perpendicular fourfold axis on a 90° rotation. Thus, the front and back faces of the crystal will also have perpendicular twofold rotation axes, since these are required by the fourfold axis (Fig. 10.70).

The square-shaped top of the crystal also suggests that there must be a twofold axis that cuts diagonally through the crystal. This two-fold axis is shown here in the left-hand diagram. But, again operation of the fourfold axis requires that the other diagonals also have two-fold axis, as shown in the right-hand diagram (Fig. 10.71).

Furthermore, the square-shaped top and rectangular-shaped front of the crystal suggest that a plane of symmetry is present as shown by the left-hand diagram here. But, again the operation of the four-fold axis requires that a mirror plane is also present that cuts through the side faces, as shown by the diagram on the right (Fig. 10.72).

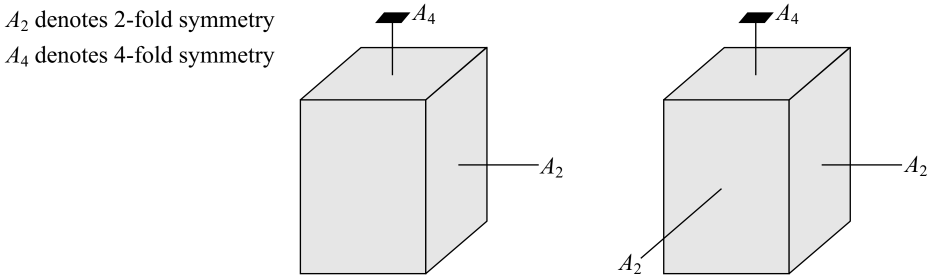


FIGURE 10.70 *The rectangular shaped face on the left side of the crystal must have a two-fold rotation axis.*

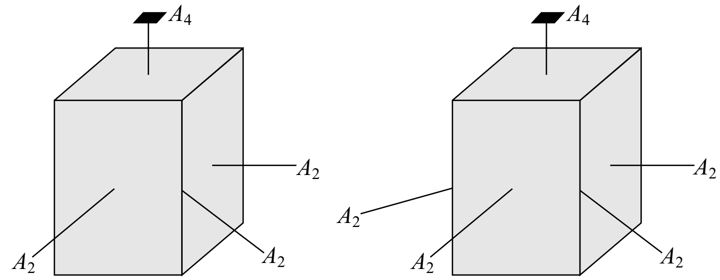


FIGURE 10.71 *Illustration of two-fold axis cuts diagonally through crystal.*

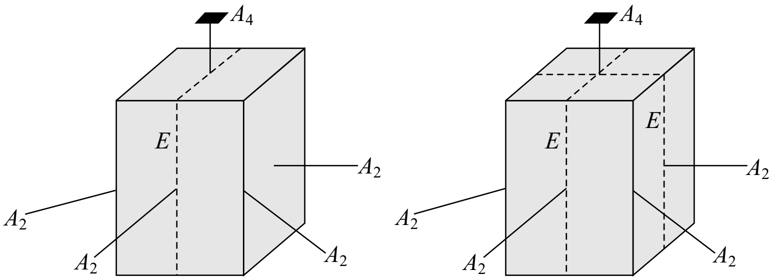


FIGURE 10.72 *A symmetry operation of 4-fold axis requires operation of the 4-fold axis requires that a mirror plane a mirror plane.*

The square top further suggests that there must be a mirror plane cutting the diagonal through the crystal. This mirror plane will be reflected by the other mirror planes cutting the sides of the crystal, or will be reproduced by the fourfold rotation axis, and thus the crystal will have another mirror plane cutting through the other diagonal, as shown by the diagram on the right (Fig. 10.73).

Finally, there is another mirror plane that cuts through the centre of the crystal parallel to the top and bottom faces.

Thus this crystal has the following symmetry elements (Fig. 10.74):

- 1—4-fold rotation axis (A_4)
- 4—2-fold rotation axes (A_2), two cutting the faces and two cutting the edges.
- Five-mirror planes (*m*), two cutting across the faces, two cutting through the edges and one cutting horizontally through the centre.
- There is a centre of symmetry (i).

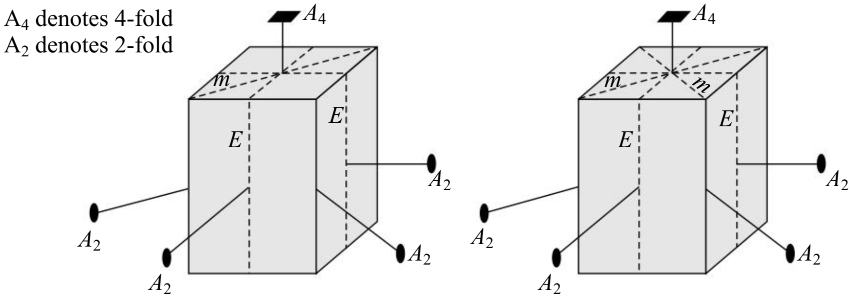


FIGURE 10.73 *A mirror plane cutting the diagonal through the crystal and the symmetry operations.*

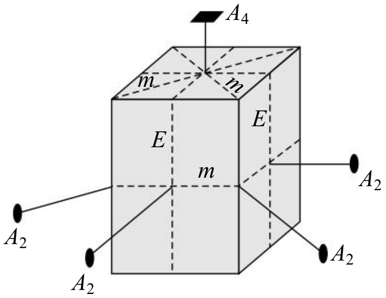


FIGURE 10.74 *Total symmetry elements of a crystal.*

10.21.1 Glide Planes and Screw Axes

Two of the symmetry operations involved in the space groups are not contained in the corresponding point group or Bravais lattice. These are the compound symmetry operations called the *glide plane* and the *screw axis*.

A glide plane is a reflection in a plane followed by a translation parallel with that plane. This is noted by *a*, *b* or *c*, depending on which axis the glide is along. There is also the '*n*' glide, which is a glide along the half of a diagonal of a face, and the *d* glide, which is along the fourth of the diagonal of either a face or space diagonal of the unit cell. The latter is often called the diamond glide plane as it features in the diamond structure.

A screw axis is a rotation about an axis followed by a translation along the direction of the axis. These are noted by a number, *n*, to describe the degree of rotation, where the number is how many operations must be applied to complete a full rotation (e.g. 3 would mean a rotation one third of the way around the axis each time). The degree of translation is then added as a subscript showing how far along the axis the translation is, as a portion of the parallel lattice vector. So, 2_1 is a 2-fold rotation followed by a translation of 1/2 of the lattice vector.

In 3D crystal, this gives a grouping of the *230 space group types into 32 crystal classes*, one for each associated crystallographic point group. *A space group with a screw axis is in the same crystal class as one with a corresponding pure axis of rotation*. Similarly *a space group with a glide plane is in the same crystal class as one with a corresponding pure reflection*.

In addition to translations, and the point operations of reflection, rotation and improper rotation, there are combinations of reflections and rotations with translation: the *screw* axis and the *glide plane*.

10.22 Elements of Symmetry of a Cubic Crystal

The law of crystallography states that 'all the crystals of the same substance possess the same elements of symmetry'.

Though, 32 elements of symmetry are required to describe and characterize all crystal systems, only three types of symmetry elements are needed for a 'cubic crystal system'.

They are (1) plane of symmetry, (2) axis of symmetry, and (3) centre of symmetry.

'The total number of planes, axis and centre of symmetry possessed by a crystal is termed as elements of symmetry of a crystal'.

Any cubic crystal (like NaCl, CsCl, KCl, etc.) system possesses a total of 23 elements of symmetry.

Let us have a brief description of three symmetry operations of a cubic crystal.

1. Plane of symmetry

'If a crystal can be divided by an imaginary plane passing through the centre into two identical portions, each of which is a mirror image of the other', then the crystal is said to possess a plane of symmetry. Consider a cubic crystal shown in Fig. 10.75.

A cubic crystal possesses three rectangular plane of symmetry, each of which is at right angles to the plane as shown in Fig. 10.76.

A cubic crystal has six diagonal type of plane of symmetry.

Therefore, all cubic crystals have a total of nine plane of symmetry.

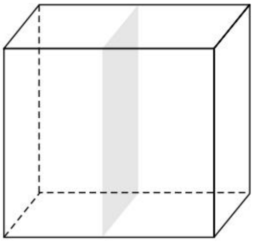


Figure 10.75 *Plane of symmetry.*

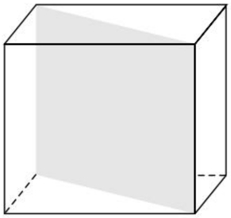


Figure 10.76 *Diagonal type of plane of symmetry.*

2. Axis of symmetry

The axis of symmetry operation for a cubic crystal is as follows:

'An axis of symmetry is an imaginary line about which the crystal may be rotated such that it presents the same appearance more than once during the complete rotation'.

A cubic crystal system can have only the fourfold, threefold and twofold axis of rotation.

When a cubic crystal as shown in Fig. 10.77, is rotated about an imaginary axis passing through the centre, presents exactly identical appearance four times in one complete revolution.

Therefore, the cubic crystal is said to have fourfold (or C_4) axis of symmetry.

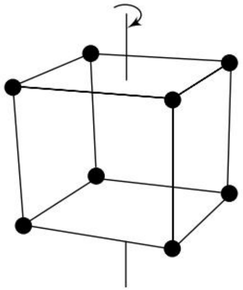
A cubic crystal has three fourfold axis of symmetry at right angles to each other.

A 3-fold axis of symmetry for a cubic crystal is described in Fig. 10.78.

If an imaginary axis passes through the opposite corners of a cube, the crystal can be rotated about this axis and is said to have a 3-fold (C_3) axis of symmetry.

Since, there are 8-corners for a cubic, there can be a total of four such threefold axis of symmetry operations. Thus, 3-fold (or C_3) axis of symmetry for a cube = 4.

If the axis of a crystal possesses through the centre of two opposite edges and the crystal is rotated about the axis, it presents two identical appearances. The crystal is said to have 2-fold (or C_2) axis of symmetry (Fig. 10.79). Thus, 2-fold axis of symmetry of a cubic crystal = 6.



4-fold symmetry

Figure 10.77 *Fourfold axis of symmetry of a cubic crystal.*

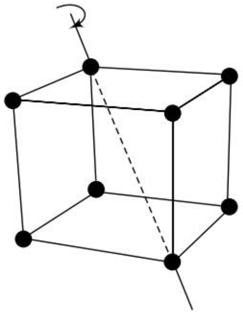


Figure 10.78 *3-fold axis of symmetry.*

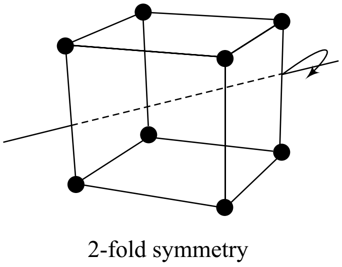


Figure 10.79 *Twofold axis of symmetry.*

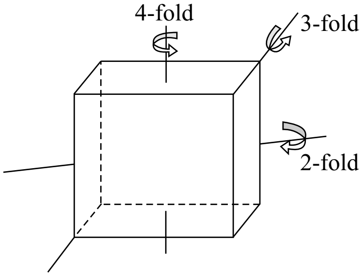


Figure 10.80 *Total axis of symmetry of a cube.*

A cubic crystal has 13 axis of symmetry (Fig. 10.80).

3. Centre of symmetry

A cubic crystal has one centre of symmetry. 'Centre of symmetry of a cubic crystal is such a point that any line drawn through it intersects the crystal at equal distances in all directions' (Fig. 10.81).

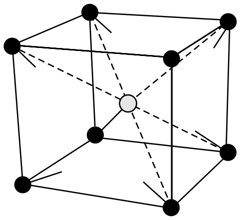


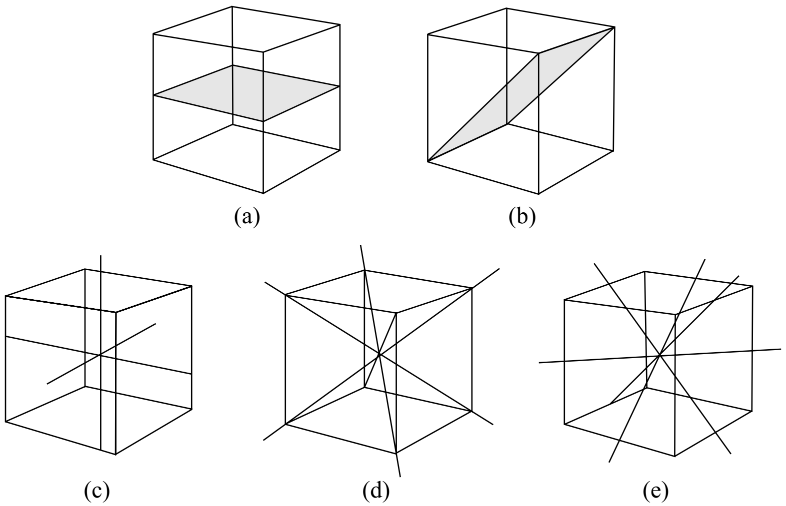
Figure 10.81 *centre of symmetry of a cube.*

Thus, a cubic crystal has maximum number of elements of symmetry compared other crystal systems (Fig. 10.82).

They are

1. plane of symmetry = 9
2. axis of symmetry = 13
3. centre of symmetry = 1

$\overline{23}$



(a) horizontal plane of symmetry = 3
(b) diagonal plane of symmetry = 6
(c) 4-fold axis of symmetry = 3
(d) 3-fold axis of symmetry = 4
(e) 2-fold axis of symmetry = 6
(f) centre of symmetry = 1

$\overline{23}$

Figure 10.82 *Total elements of symmetry of a cube.*

10.23 Crystal Systems

The internal structure and the geometry of a crystal are described in terms of their 32 elements of symmetry operations. Based on these aspects of crystals, it is possible to describe 14 space lattices and a total number of seven crystal systems.

The crystals are classified into seven crystal systems. They are (1) cubic (2) tetragonal (3) orthorhombic (4) monoclinic (5) triclinic (6) rhombohedral and (7) hexagonal.

The lengths of the edges, interfacial angles and their symmetry elements of seven crystal systems are given below:

Sl. No.	System	Axial Characteristics	Symmetry Elements
1.	Cubic	$a = b = c$ $\propto = \beta = v = 90^\circ$	9 planes, 13 axes, 1 centre = 23
2.	Tetrogonal	$a = b \neq c$ $\propto = \beta = v = 90^\circ$	5 planes, 5 axes = 10
3.	Orthorhombic	$a \neq b \neq c$ $\propto = \beta = v = 90^\circ$	3 planes, 3 axes = 6
4.	Monoclinic	$a \neq b \neq c$ $\propto = v = 90^\circ, \beta \neq 90^\circ$	1 plane, 1 axis = 2
5.	Triclinic	$a = b = c$ $\propto \neq \beta \neq v = 90^\circ$	None
6.	Rhombohedral	$a = b = c$ $\propto = v = 90^\circ, \beta \neq 90^\circ$	None
7.	Hexagonal	$a = b \neq c$ $\propto = \beta = 90^\circ, v = 120^\circ$	None

10.24 Bravais Space Lattices

The internal structures, geometry and the 32 elements of symmetry operations of crystals revealed that there are seven crystal systems; and it is possible to describe 14 different space lattices in three-dimensional lattice (Fig. 10.83).

These 14 space lattices of seven crystal systems are known as 'Bravais space lattices'.

S. No.	Crystal system	Number of space lattices	Lattice types
1.	Cube	3	Simple, B.C.C., F.C.C
2.	Tetragonal	2	Simple, B.C tetragonal
3.	Orthorhombic	4	Simple, B.C. orthorhombic, F(ABC) central, face centred
4.	Monoclinic	2	Simple, face centred
5.	Triclinic	1	Simple
6.	Rhombohedral	1	Rhombohedral
7.	Hexagonal	1	Simple
	Total	14	

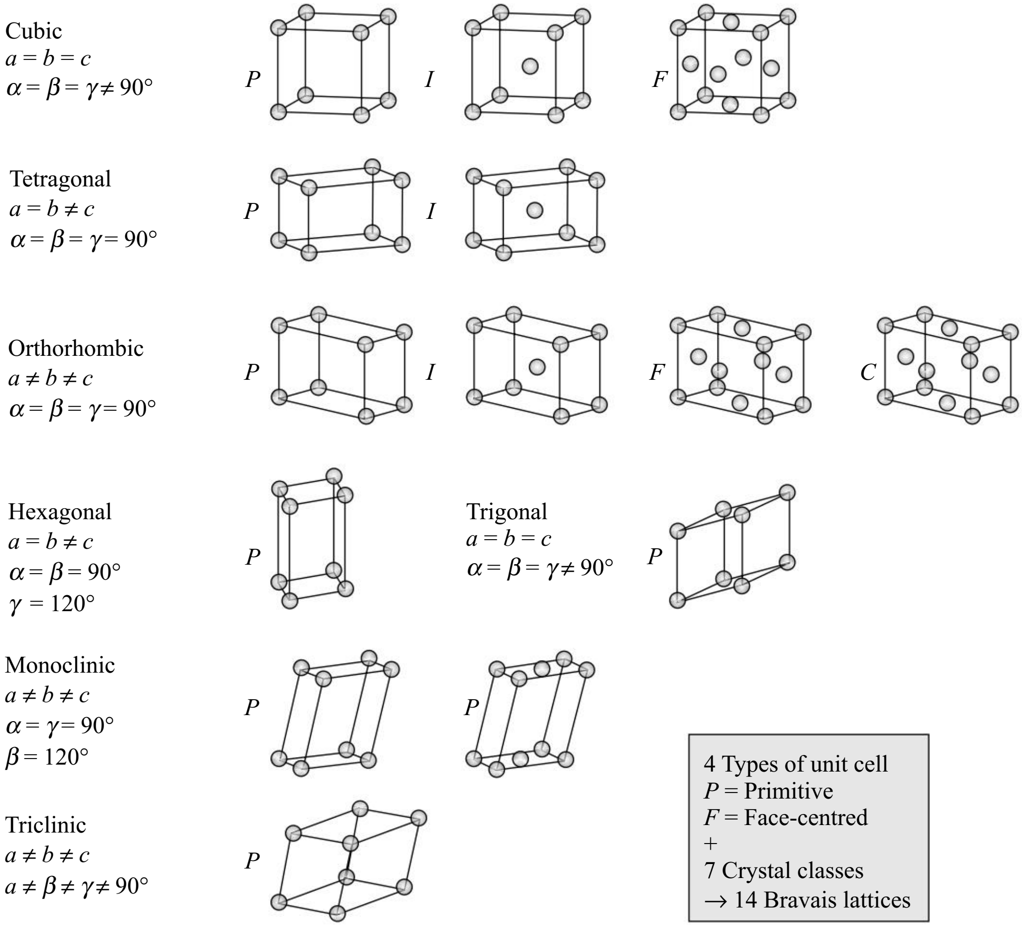


Figure 10.83 *Bravais lattices of crystal.*

10.25 Bravais Lattices of a Cube

A crystal belonging to a cubic crystal system has three bravais lattices. They are (1) simple (2) body-centred cubic, and (3) face-centred cubic space lattices.

1. Simple cubic space lattice

In simple cubic space lattice type (*p*-type)(Fig. 10.84), there are ions (or atoms) at the corner lattice sites of unit cell of a cube.

2. Body-centred cubic space lattice

The ions (or atoms) are situated at the corners of a cubic space lattice and as well, at the centre of a cube, Fig. 10.85.

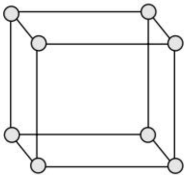


Figure 10.84 *Simple cubic space lattice.*

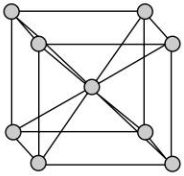


Figure 10.85 *Body-centred cubic space lattice.*

3. Face-centred cubic space lattice

In the face-centred cubic lattice, the ions (or atoms) are positioned at the corners; as well, at the centre of six faces of the cube as shown in Fig. 10.86.

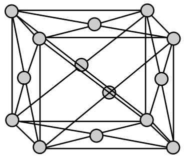


Figure 10.86 *Face-centred cubic space lattice.*

10.26 Introduction to X-ray Powder Diffraction

(i) Introduction

The characterization of crystalline materials is done through a nondestructive technique of *x*-ray powder diffraction. Traditionally, this method has been used for phase identification, quantitative analysis and the determination of structure imperfections. This technique has been extended to new areas, such as the determination of crystal structures and the extraction of three-dimensional micro-structural properties.

(ii) Applications of powder diffraction

The most widespread use of x-ray powder diffraction is for the identification of crystalline compounds by their diffraction pattern. Listed below are some specific uses

- peak position unit-cell parameter refinement
- pattern indexing
- space group determination (2θ/absent reflections)
- anisotropic thermal expansionphase identification (*d*/*I*)
- intensity phase abundance
- crystal structure analysis (whole pattern)

A schematic diagram (Fig. 10.87) of a diffractometer system is shown in Fig. 10.87.

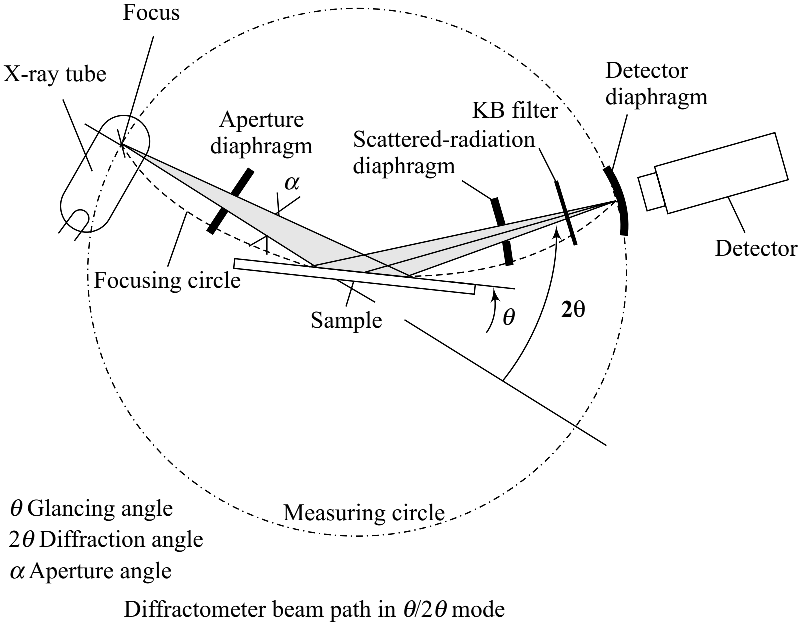


Figure 10.87 *A typical X-ray diffractometer.*

(iii) Sample preparation

A statistically infinite amount of randomly oriented fine powder (10 μm) is mounted in a manner in which there is no preferred crystallite orientation.

(iv) X-ray source

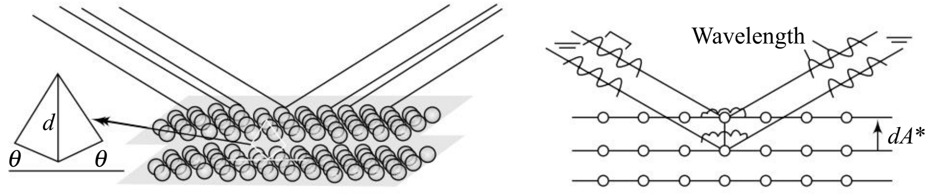
XRD analysis of powder sample needs a coherent beam of monochromatic x-rays of known wavelength. Copper (Cu) x-ray tubes are most commonly used for x-ray diffraction of inorganic materials. The wavelength of the strongest Cu radiation (Kα) is approximately 1.54 angstroms (Å). Other anodes commonly used in x-ray generating tubes include Cr (Kα 2.29 Å), Fe (Kα 1.94 Å), Co (Kα 1.79 Å) and Mo (Kα 0.71 Å).

10.27 X-ray Diffraction from Crystalline Solids

The Bragg Law

The Bragg Law is the cornerstone of x-ray diffraction analysis, it allows us to make accurate quantification of the results of experiments carried out to determine the crystal structure.

Figure below (Fig. 10.88) represents the scattering of radiation from two crystal planes of a solid.



(a) (b)

Figure 10.88 *(a) and (b) X-ray diffraction and the Bragg's law.*

The darker and lighter lines represent radiation reflected from the upper and lower set of planes, respectively. The reflection is treated as though the two planes of atoms behaved like half-reflecting mirrors. There is a difference of $2d \sin \theta$ in the path length travelled by the two beams of radiation where d is the perpendicular distance between the planes (the path difference is marked in green and the expanded construction alongside makes the geometry more clear). As the angle of reflection is changed so does the difference in path length travelled by the two beams. When the path difference is equal to an integer number of wavelengths the two beams will reinforce one another and when it is an integral number of half wavelengths the two waves will interfere destructively with one another. The intensity of the total reflected radiation will vary sinusoidally with θ. Interaction of x-rays with sample creates secondary 'diffracted' beams of x-rays related to interplanar spacings in the crystalline powder according to a mathematical relation called 'Bragg's Law':

$$n\lambda = 2d \sin \theta$$

where n is an integer

'λ' is the wavelength of the x-rays
'd' is the interplanar spacing generating the diffraction and
'θ' is the diffraction angle
λ and d are measured in the same units, usually angstroms.

For a powder specimen in a diffractometer having a statistically infinite amount of randomly oriented crystallites, the diffraction maxima (or peaks) are measured along the 2θ diffractometer circle.

Thus, the (100) planes of a simple cubic structure will give diffraction peaks when

$$\sin \theta = n\lambda/2a,\ n = 1, 2, 3, \text{ and so on.}$$

where 'a' is the lattice parameter.

An alternative way of stating this is to say *that there is diffraction from the (100), (200), (300), etc. planes* at

$$\sin \theta = \lambda/2d, \qquad d = a, a/2, a/3.$$

Since Bragg's Law applies to all sets of crystal planes, the lattice can be deduced from the diffraction pattern, making use of general expressions for the spacing of the planes in terms of their Miller indices.

For cubic structures

$$d_{(hkl)} = a/(h^2 + k^2 + l^2)^{1/2}$$

It is evident that *the smaller the spacing, the higher the angle of diffraction,* i.e. the spacing of peaks in the diffraction pattern is inversely proportional to the spacing of the planes in the lattice. The diffraction pattern will reflect the symmetry properties of the lattice.

A simple example is the difference between the series of (n00) reflections for a simple cubic and a body-centred cubic lattice. For the simple cubic lattice, all values of 'n' will give Bragg peaks. However, for the body-centred cubic lattice the (100) planes are interleaved by an equivalent set at the halfway position. At the angle where Bragg's Law would give the (100) reflection the interleaved planes will give a reflection exactly out of phase with that from the primary planes, which will exactly cancel the signal. There is no signal from (n00) planes with odd values of n. This kind of argument leads to rules for identifying the lattice symmetry from "missing'.

The interplanar spacing, d, for a given hkl reflection is given by the unit cell dimensions

Cubic:

$$1/d^2 = (h^2 + k^2 + l^2)/a^2$$

Tetragonal:

$$1/d^2 = \{(h^2 + k^2)/a^2\} + (l^2/c^2)$$

Orthorhombic:

$$1/d^2 = (h^2/a^2) + (k^2/b^2) + (l^2/c^2)$$

Hexagonal:

$$1/d^2 = (4/3)\{(h^2 + hk + k^2)/a^2\} + (l^2/c^2)$$

Monoclinic:

$$1/d^2 = (1/\sin^2\beta)\{(h^2/a^2) + (k^2 \sin^2\beta/b^2) + (l^2/c^2) - (2hl \cos \beta/ac)\}$$

10.28 Indexing Diffraction Patterns of Crystals

(a) Cubic crystal

- We need to know about crystal structures because, to a large extent, it determines the properties of materials.
- X-ray diffraction (XRD) is one of the method can use to identify the structures of crystalline solids.
- The XRD patterns are somewhat like *fingerprints* in that they are unique to each material.

The information in an XRD pattern is a direct result of two things:

1. The *size and shape* of the unit cells, which determine the *relative positions* of the diffraction peaks.
2. *Atomic positions* within the unit cell, which determine the *relative intensities* of the diffraction peaks. Hence,

(a) We can calculate the size and shape of a unit cell from the positions of the XRD peaks.
(b) We can determine the positions of the atoms in the unit cell from the intensities of the diffraction peaks.

10.28.1 Mathematical Approach for Indexing

- For cubic materials, interplanar spacing is given in Fig. 10.89

$$\frac{1}{d^2} = \frac{h^2 + k^2 + l^2}{a^2} \tag{1}$$

Recalling Bragg's law

$$\lambda = 2d \sin \theta$$

$$\lambda^2 = 4d^2 \sin^2\theta \text{ or } \sin^2\theta = \frac{\lambda^2}{4d^2} \tag{2}$$

Combining Eqns. (1) and (2) yields or

$$\frac{1}{d^2} = \frac{h^2 + k^2 + l^2}{a^2} = \frac{4 \sin^2 \theta}{\lambda^2} \text{ or}$$

$$\therefore \quad \sin^2\theta = \frac{\lambda^2}{4a^2} [h^2 + k^2 + l^2]$$

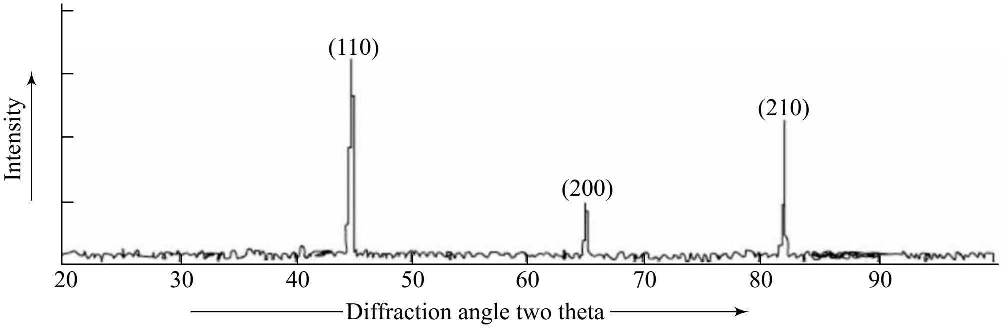


Figure 10.89 *XRD pattern of polycrystalline ion.*

- λ, a are constants, hence $\frac{\lambda^2}{4a^2}$ is a constant.
- $\sin^2\theta$ is proportional to $h^2 + k^2 + l^2$ i.e. planes with higher Miller indices will diffract at higher first XRD peak values of θ.
- In cubic systems, the in the XRD pattern will be due to diffraction from planes with the lowest Miller indices
- For the close-packed planes, i.e. simple cubic, (100), $h^2 + k^2 + l^2 = 1$
  body-centred cubic, (110), $h^2 + k^2 + l^2 = 2$, and
  face-centred, (111), $h^2 + k^2 + l^2 = 3$

Allowed reflection for cubic lattices

- Primitive $h^2 + k^2 + l^2 = 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14$ and 16.
- Body-centred $h^2 + k^2 + l^2 = 2, 4, 6, 8, 10, 12, 14$ and 16.

- Face-centred $h^2 + k^2 + l^2 = 3, 4, 8, 11, 12, 16, 19, 20, 24, 27$ and 32.
- Diamond cubic $h^2 + k^2 + l^2 = 3, 8, 11, 16, 19, 24, 27$ and 32.

Example Index this pattern and determine the lattice parameter

XRD pattern for Aluminium, which was collected using CuKα radiation

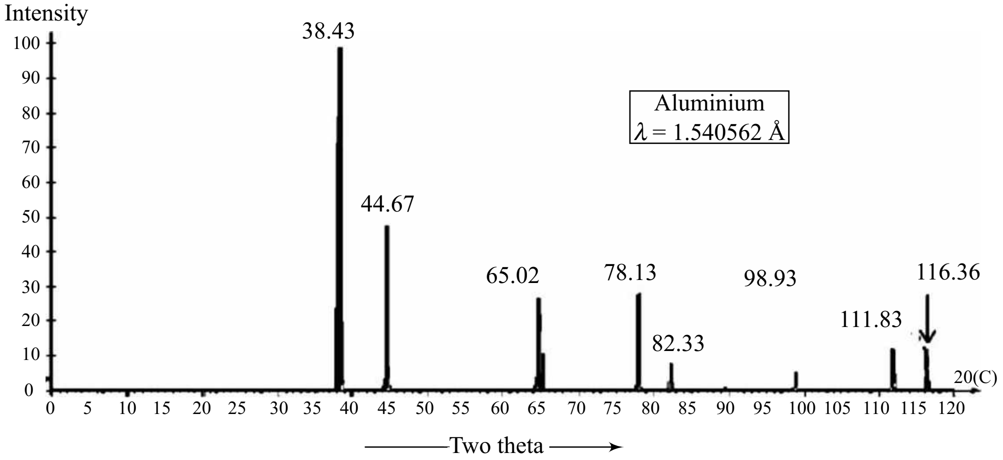


FIGURE 10.90 *XRD pattern of aluminium using Cu$_{k\alpha}$ radiation.*

An XRD pattern of aluminium using Cu$_{k\alpha}$ radiation is shown in Fig. 10.90.

Steps:

1. Identify the peaks.
2. Determine $\sin^2\theta$.
3. Calculate the ratio $\sin^2\theta/\sin^2\theta$ min and multiply by the appropriate integers.
4. Select the result from (3) that yields $h^2 + k^2 + l^2$ as an integer.
5. Compare results with the sequences of $h^2 + k^2 + l^2$ values to identify the Bravais lattice, e.g. bcc $h^2 + k^2 + l^2 = 2, 4, 6, 8$, and so on.
6. Calculate lattice parameters.

Step 1: Identify the peaks and their proper 2θ values. Eight peaks for this pattern

Peak No	2θ	$\sin^2\theta$	$1\times\frac{\sin^2\theta}{\sin^2\theta_{\min}}$	$2\times\frac{\sin^2\theta}{\sin^2\theta_{\min}}$	$3\times\frac{\sin^2\theta}{\sin^2\theta_{\min}}$	$h^2+k^2+l^2$	*hkl*	*aÅ*
1.	38.43							
2.	44.67							
3.	65.02							
4.	78.13							
5.	82.33							
6.	98.93							
7.	111.83							
8.	116.36							

Step 2: Determine of $\sin^2\theta$

Peak No	2θ	$\sin^2\theta$	$1\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$2\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$3\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$h^2+k^2+l^2$	*hkl*	*aÅ*
1.	38.43	0.1083						
2.	44.67	0.1444						
3.	65.02	0.2888						
4.	78.13	0.3972						
5.	82.33	0.4333						
6.	98.93	0.5776						
7.	111.83	0.6869						
8.	116.36	0.7220						

Step 3: Calculate the ratio $\sin^2\theta/\sin^2\theta_{\text{min}}$ and multiply by the appropriate integers

Peak No	2θ	$\sin^2\theta$	$1\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$2\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$3\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$h^2+k^2+l^2$	*hkl*	*aÅ*
1.	38.43	0.1083	1.004	2.000	3.000			
2.	44.67	0.1444	1.333	2.667	4.000			
3.	65.02	0.2888	2.667	5.333	8.000			
4.	78.13	0.3972	3.667	7.333	11.000			
5.	82.33	0.4333	4.004	8.000	12.000			
6.	98.93	0.5776	5.333	10.665	15.998			
7.	111.83	0.6869	6.333	12.665	18.998			
8.	116.36	0.7220	6.666	13.331	19.997			

Step 4: Select the result from (3) that yields $h^2 + k^2 + l^2$ as series of integers

Peak No	2θ	$\sin^2\theta$	$1\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$2\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$3\times\frac{\sin^2\theta}{\sin^2\theta_{\text{min}}}$	$h_2+k_2+l_2$	*hkl*	*aÅ*
1.	38.43	0.1083	1.004	2.000	3.000			
2.	44.67	0.1444	1.333	2.667	4.000			
3.	65.02	0.2888	2.667	5.333	8.000			
4.	78.13	0.3972	3.667	7.333	11.000			
5.	82.33	0.4333	4.004	8.000	12.000			
6.	98.93	0.5776	5.333	10.665	15.998			
7.	111.83	0.6869	6.333	12.665	18.998			
8.	116.36	0.7220	6.666	13.331	19.997			

Step 5: Compare results with the sequences of $h^2 + k^2 + l^2$ values to identify the Bravais lattice.

Peak No	2θ	$\sin^2\theta$	$1 \times \frac{\sin^2\theta}{\sin^2\theta_{min}}$	$2 \times \frac{\sin^2\theta}{\sin^2\theta_{min}}$	$3 \times \frac{\sin^2\theta}{\sin^2\theta_{min}}$	$h^2+k^2+l^2$	*hkl*	*aÅ*
1.	38.43	0.1083	1.004	2.000	3.000	3	111	4.0538
2.	44.67	0.1444	1.333	2.667	4.000	4	200	4.0539
3.	65.02	0.2888	2.667	5.333	8.000	8	220	4.0538
4.	78.13	0.3972	3.667	7.333	11.000	11	311	4.0538
5.	82.33	0.4333	4.004	8.000	12.000	12	222	4.0538
6.	98.93	0.5776	5.333	10.665	15.998	16	400	4.0541
7.	111.83	0.6869	6.333	12.665	18.998	19	331	4.0541
8.	116.36	0.7220	6.666	13.331	19.997	20	420	4.0541

Bravais lattice is face-centred cubic

Therefore $\sin^2\theta = \frac{\lambda^2}{4a^2}[h^2 + k^2 + l^2]$

Step 6: Calculate lattice parameters

Peak No	2θ	$\sin^2\theta$	$1 \times \frac{\sin^2\theta}{\sin^2\theta_{min}}$	$2 \times \frac{\sin^2\theta}{\sin^2\theta_{min}}$	$3 \times \frac{\sin^2\theta}{\sin^2\theta_{min}}$	$h^2+k^2+l^2$	*hkl*	*aÅ*
1.	38.43	0.1083	1.004	2.000	3.000	3	111	4.0538
2.	44.67	0.1444	1.333	2.667	4.000	4	200	4.0539
3.	65.02	0.2888	2.667	5.333	8.000	8	220	4.0538
4.	78.13	0.3972	3.667	7.333	11.000	11	311	4.0538
5.	82.33	0.4333	4.004	8.000	12.000	12	222	4.0538
6.	98.93	0.5776	5.333	10.665	15.998	16	400	4.0541
7.	111.83	0.6869	6.333	12.665	18.998	19	331	4.0541
8.	116.36	0.7220	6.666	13.331	19.997	20	420	4.0541

Average lattice parameter of face-centred cubic lattice is 4.0539 Å

(b) There also exists geometric equations for each crystal system that relate the *d*-spacing of any given (*hkl*) plane and the lattice parameters.

For the isometric (cubic) system:

$$\frac{1}{d^2} = \frac{h^2 + k^2 + l^2}{a^2} = \frac{4\sin^2\theta}{\lambda^2}$$

Let's work an example for NaCl crystal.

$a = 5.639$ Å. If $(hkl) = (111)$, then $a^2 = 31.8$ and $(h^2 + k^2 + l^2) = 3$

$1/d^2 = 0.0943.$ or $d_{111} = 3.26$ Å

Bragg's Law and the equation above can be combined to give:

$$\sin^2\theta = \frac{\lambda^2}{4a^2}\,[h^2 + k^2 + l^2]$$

For a particular wavelength of monochromatic radiation (e.g. $\text{CuK}_\alpha = 1.54059$ Å) and a particular crystal in the isometric system with a unit cell of length a (e.g. 5Å), then all possible Bragg angles of diffraction can be determined for every possible (*hkl)* plane.

Note: $\lambda^2/4a^2$ (e.g. 0.0237 using the above values) is always constant for any one diffraction pattern. Also note that $(h^2 + k^2 + l^2)$ will always be an integral value and certain combinations are *not* possible (i.e. 7, 15, 23, 28).

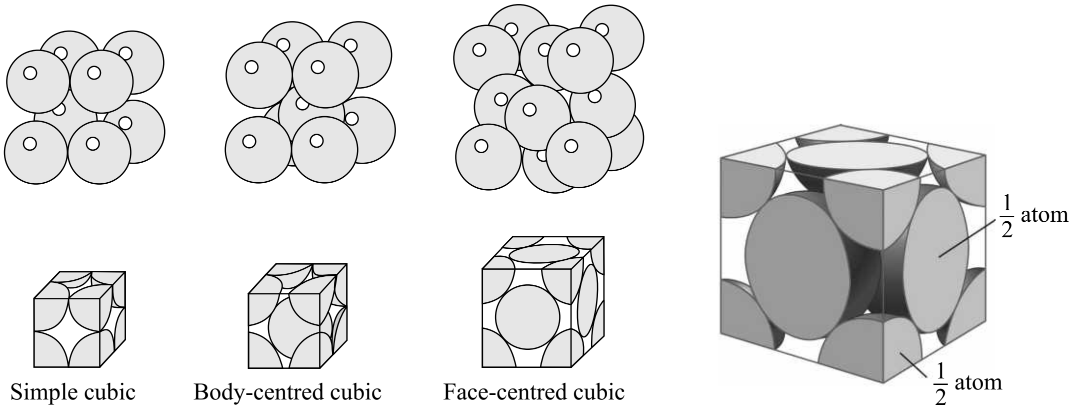
For all possible families of planes *hkl* in a unit cell, it is possible to calculate the angle at which reflection would occur.

Example

Line	$\sin^2\theta$	$(h^2 + k^2 + l^2)$	$\lambda^2/4a^2$	a(Å)	*hkl*
1	0.0712	3	0.0237	5.0	111
2	0.0949	4	0.0237	5.0	200
3	0.1899	8	0.0237	5.0	220
N	—	—	—	—	—

10.29 Calculation of Number of Ions/Atoms per Unit Cell of Different Space Lattices of a Cube Determining the Unit Cell of a Crystal

Atoms on the corners, edges and faces of a unit cell are shared by more than one unit cell, as shown in the figure (Fig. 10.91) below.



Simple cubic Body-centred cubic Face-centred cubic

FIGURE 10.91 *Number of ions per unit cell.*

- An atom on a face is shared by two unit cells, so only half of the atom belongs to each of these cells.
- An atom on an edge is shared by four unit cells, and

- An atom on a corner is shared by eight unit cells. Thus, only one-quarter of an atom on an edge and one-eighth of an atom on a corner can be assigned to each of the unit cells that share these atoms.

Example If nickel crystallized in a *simple cubic unit cell*, there would be a nickel atom on each of the eight corners of the cell. Because only one-eighth of these atoms can be assigned to a given unit cell, each unit cell in a simple cubic structure would have one net nickel atom.

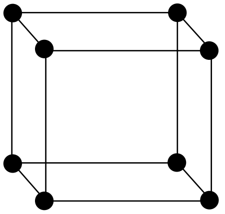
10.30 Calculation of the Number of Ions (or Atoms) per Unit Cell of a Cube

(i) Simple cubic lattice

In a simple cubic (P) lattice structure, an ion (or atom) situated at any corner of a unit cell is shared by a total of eight unit cells (Fig. 10.92).

Therefore, each unit cell has 1/8 share of every corner ion/atom.

The total contribution of all the eight ions (or atoms) of a unit cell = 1/8 × 8 = 1 ion/atom per unit cell.



∴ 8 corners × 1/8 = 1 atom

Figure 10.92 *A simple cubic lattice.*

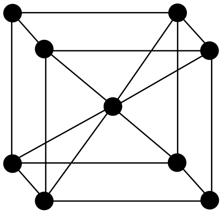
(ii) Body-centred cubic lattice

A body-centred cubic (I-type) lattice (Fig. 10.93) has one additional ion (or atom) at the centre of cube, besides having one ion (or atom) at the corners. The atom at the centre is independent of other unit cells; while each of the eight ions (or atoms) situated at the corners are shared by a total of eight unit cell lattices.

Thus, the total number of ions (or atoms) per unit cell

= 1 (at the centre) + × 8 (at the corners).
= 1 + 1 = 2 ions/atoms

If nickel formed a body-centred cubic structure, there would be two atoms per unit cell, because the nickel atom in the centre of the body wouldn't be shared with any other unit cell.

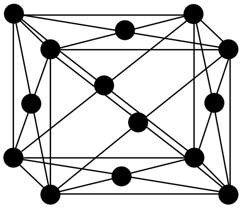


∴ (8 corners × 1/8) + 1 body
= 2 atoms

Figure 10.93 *A bcc lattice.*

(iii) Face-centred cubic lattice

In an F-type (FCC) lattice (Fig. 10.94), the ions (or atoms) are found at the eight corners of a unit cell, besides having one ion (or atom) at the centre of each face of a cube.



∴ (8 corners × 1/8)
+ (6 faces × 1/2)
= 4 atoms

Figure 10.94 *An fcc lattice.*

Thus, every ion (or atom) situated at the centre of a face of a unit cell is shared by two adjoining unit cells, and there are six faces for a cube.

The total number of ions (or atoms) per unit cell

= × 8 (at corners) + × 6 (at the centre of faces)
= 1 = 3 = 4 ions/atom.

If nickel crystallized in a face-centred cubic structure, the six atoms on the faces of the unit cell would contribute three net nickel atoms, for a total of four atoms per unit cell.

10.31 Some Basic Aspects Related to Unit Cell

Each type of the above-mentioned unit cells have different numbers of atoms and in fact, have different densities.

Illustrated example of a unit cell of Nickel

Let's therefore calculate the density for nickel based on each of these structures and the unit cell edge length for nickel given as 0.3524 nm. In order to calculate this, we need to know the volume of the unit cell in cubic centimetres and the mass of a single nickel atom.

(i) The volume (V) of the unit cell of nickel

Let the volume of a cubic unit cell is 'V' nm^3, which is equal to the cell edge length (a) cubed.

$$V = a^3 = (0.3524 \text{ nm})^3 = 0.04376 \text{ nm}^3$$

Since, there are 10^9 nm and 100 cm in a metre, there must be 10^7 nm in a cm.

$$\frac{10^9}{1 \text{ m}} \times \frac{1 \text{ m}}{100 \text{ cm}} = 10^7 \text{ nm/cm}$$

Now, converting the *volume* of the unit cell into cm^3.

$$4.376 \times 10^{-2} \text{ nm}^3 \times \left[\frac{1 \text{ cm}}{10^7 \text{ nm}}\right]^3 = 4.376 \times 10^{-23} \text{ cm}^3$$

(ii) Mass of an atom of Nickel in a unit cell

The mass of an atom can be calculated from the atomic weight of the metal and Avogadro's number.

$$\frac{58.69 \text{ g Ni}}{1 \text{ mol}} \times \frac{1 \text{ mol}}{6.022 \times 10^{23} \text{ atoms}} = 9.746 \times 10^{-23} \text{ g/atom}$$

The density of nickel, if it crystallized in a simple cubic structure, would, therefore, be 2.23 g/cm^3.

(i) Simple cubic structure

$$\frac{9.746 \times 10^{-23} \text{ g/unit cell}}{4.376 \times 10^{-23} \text{ cm}^3\text{/unit cell}} = 2.23 \text{ g/cm}^3$$

(ii) Body-centred cubic structure

Because there would be twice as many atoms per unit cell, if nickel crystallized in a body-centred cubic structure, the density of nickel in this structure would be twice as large.

$$\frac{2(9.746 \times 10^{-23} \text{ g/unit cell})}{4.376 \times 10^{-23} \text{ cm}^3\text{/unit cell}} = 4.45 \text{ g/cm}^3$$

(iii) Face-centred cubic structure

There would be four atoms per unit cell in a face-centred cubic structure and the density of nickel in this structure would be four times as large.

$$\frac{4(9.746 \times 10^{-23}\ \text{g/unit cell})}{4.376 \times 10^{-23}\ \text{cm}^3\text{/unit cell}} = 8.91\ \text{g/cm}^3$$

The experimental value for the density of nickel is 8.90 g/cm^3. The obvious conclusion is that nickel crystallizes in a *face-centred cubic unit cell* and therefore has a *cubic closest-packed structure*.

(iv) Calculation of atomic radius of nickel

Nickel atoms crystallize in a face-centred cubic unit cell (Fig. 10.95) with a cell edge length of 0.3524 nm. The radius of a nickel atom can be calculated thus:

One of the faces of a face-centred cubic unit cell is shown in Fig. 10.95.

It is evident from Fig. 10.95 that four times the radius of a nickel atom is equal to the diagonal across the face of this unit cell.

$$d_{\text{face}} = 4\, r_{\text{Ni}}$$

The diagonal across the face of the unit cell is therefore related to the unit-cell edge length by the following equation (Pythagoras theorem).

$$d^2_{\text{face}} = a^2 + a^2 = 2\, a^2$$

Taking the square root of both sides gives the following result.

$$d_{\text{face}} = \sqrt{2}a$$

This is equal to four times the radius of a nickel atom. Therefore,

$$4\, r_{\text{Ni}} = \sqrt{2}a$$

Solving for the radius of a nickel atom, we have

$$r_{\text{Ni}} = \frac{\sqrt{2}a}{4} = \frac{\sqrt{2}\,(0.3524\ \text{nm})}{4} = 0.1246\ \text{nm}$$

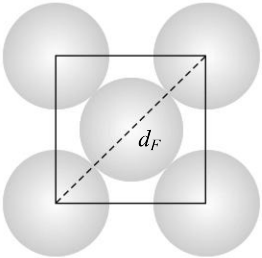


Figure 10.95 *Diagonal across the face of a unit cell.*

(v) Estimation of the radius of an ion

Example CsCl crystal

The size or radius of an ion can be determined similarly. The cell edge length in caesium chloride is 0.4123 nm. This is desired to calculate the distance between the centres of the Cs+ and Cl$^-$ ions in CsCl.

CsCl crystallizes in a simple cubic unit cell (Fig. 10.96) of Cl$^-$ions with a Cs$^+$ ion in the centre of the body of the cell, as shown in Fig. 10.96.

One of the simplest assumptions about ionic solids is that the positive and negative ions that form these crystals touch each other. We can therefore assume that the diagonal across the body of the CsCl unit cell is equivalent to the sum of the radii of two Cl$^-$ ions and two Cs+ ions.

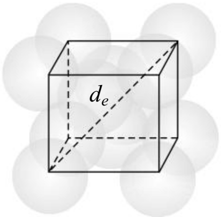


Figure 10.96 *A CsCl lattice.*

$$d_{\text{bcc}} = 2\, r_{\text{cs}} + 2\, r_{\text{cl}^-}$$

The three-dimensional equivalent of the Pythagorean theorem suggests that the square of the diagonal across the body of a cube is the sum of the squares of the three sides.

$$d_{\text{bcc}} = a^2 + a^2 + a^2$$

Taking the square root of both sides of this equation gives the following result.

$$d_{\text{bcc}} = \sqrt{3}a$$

If the cell edge length in CsCl is 0.4123 nm, the diagonal across the body in this unit cell is 0.7141 nm.

$$d_{\text{bcc}} = \sqrt{3}a = \sqrt{3}\ (0.4123\ \text{nm}) = 0.7141\ \text{nm}$$

The sum of the ionic radii of Cs^+ and Cl^- ions is half this distance, or 0.3571 nm.

$$r_{\text{cs}} + r_{\text{cl}} = \frac{d_{\text{bcc}}}{2} = \frac{0.7141}{2} = 0.3571\ \text{nm}$$

If the radius of either the Cs^+ or Cl^- ion is known, it is possible to calculate the radius of the other ion. The ionic radius of the Cl^- ion is 0.181 nm. Substituting this value into the last equation gives a value of 0.176 nm for the radius of the Cs^+ ion.

$$r_{\text{cs}} + r_{\text{cl}} = 0.3571\ \text{nm and therefore,}$$

$$r_{\text{cs}} + 0.181\ \text{nm} = 0.3571\ \text{nm}$$

Therefore, $r_{\text{cs}} = 0.176$ nm

The results of this calculation are in reasonable agreement with the value of 0.169 nm known for the radius of the Cs^+ ion.

Review Questions

1. What is a solid state?
2. What is a crystalline solid? Give examples.
3. What is an amorphous solid? Give examples.
4. Mention the types of crystalline solids with examples.
5. Write note on ionic crystals.
6. Write a note on covalent crystals.
7. Write a note on molecular crystals.
8. Write a note on metallic crystals.
9. What are the types of ionic crystals?
10. Discuss AB type of ionic crystals with any example.
11. Discuss AB_2 type of ionic crystals with any example.
12. A note on ABO_3 type compounds.
13. A note on AB_2O_4 ionic compounds.
14. What do you mean by close packing of solids?
15. Mention the types of close packing of solids and mention an example in each case.
16. A note on cubic close packing of solids.
17. A note on hexagonal packing of solids.
18. Discuss the types of voids in solid structures.
19. Mention the importance of radius ratio in crystal structure.
20. What do you mean by (a) crystal, (b) lattice, (c) a unit cell lattice, and (d) crystallographic axes?
21. A note on lattice parameters and crystal planes.
22. What is geometric crystallography?
23. Write a note on (i) the law of constancy of interfacial angles, and (ii) the law of rationality of indices.
24. What are Miller indices of a plane of a crystal? Explain with the help of crystallographic axes.
25. Discuss Miller indices of the following planes of a cubic crystal with a neat diagram.
    (i) 100, (ii) 110, (iii) 111, (iv), 202 and (v) 210
26. A note on symmetry elements of crystals.
27. What do you mean by symmetry operations?
28. Discuss the types of symmetry operations.
29. A note on rotational symmetry operation.
30. A note on reflection symmetry operation.
31. A note on rotoinversion symmetry.
32. What is mirror symmetry of a plane?
33. Write a note on Glide plane and screw axis of symmetry.
34. Discuss the elements of symmetry of a cubic crystal.

35. Name the crystal systems and mention their axial characteristics.
36. Name the Bravais lattices and mention the lattices possible in each case.
37. Discuss the types of Bravais lattices of a cube.
38. Discuss the x-ray diffraction of a crystalline solid and obtain Bragg's law.
39. How diffraction pattern of a cubic crystal can be indexed with an example?
40. Calculate the number of ions/atoms per unit cell of a cubic unit cells.
41. How do calculate the (i) volume of cubic unit cell, (ii) mass of an atom, and (iii) atomic radius of an atom?
42. Calculate the density of BCC crystal. Side of cube is 4Å and $M = 60$. [UPTU, 2012]
43. What is Bragg's Law? Derive Bragg's equation for diffraction of x-rays by crystals. [UPTU, 2014]
44. Density of sodium chloride is 2.166 g/cc. Calculate the edge of the cubic cell. Assuming that four molecules of NaCl are associated per unit cell. [UPTU, 2016]

11

Conductivity in Solids

Chapter Outline

Introduction. Electrical properties of solids. Band theory of solids—types of energy bands, overlapping energy bands, nonoverlapping energy bands, valence bands, forbidden energy gap. Application of band theory to solids, conductors, insulators, semiconductors. Elemental semiconductor, intrinsic semiconductors. Electrical conductivity of an intrinsic semiconductor, extrinsic semiconductor, *p*-type extrinsic semiconductor, the effect of temperature on *n*- or *p*-type extrinsic semiconductor. Calculation of carrier density and conductivity of a *n*- or *p*-type extrinsic semiconductor, nonelemental semiconductors. Nonstoichiometric *n*-type semiconductors, semiconductivity of FeO, *p*-type semi-conductivity of NiO, nonstoichiometric *n*-type semiconductor, semiconductivity of ZnO, semiconductivity of lithium doped NiO. Chalcogen semiconductors. Photocopying process. Preparation of pure germanium semiconductor by fractional distillation, preparation of pure germanium by crystal growth. Superconductors.

11.1 Introduction

An elementary account of the crystal structure of solids is discussed in the preceding topic, while a lucid attempt is made in this chapter to explain the conductivity behaviour of solids. It is well known that materials in the solid state exhibit a wide variety of interesting solid-state properties depending on their constituents, imperfections and the crystal structures. Many solid-state materials are metals (Cu, Zn, Ag, etc.), semiconductors (elemental silicon or germanium, intermetallics, oxide ceramic materials), superconductors ($YBa_2Cu_3O_{7-x}$, $Ba_2Ca_2Sr_2Cu_3O_{10}$, etc.), insulators (or dielectrics), photo conductors (chalcogens) and magnetic materials (ferrites, etc.). Materials in the solid state exhibit a wide range of electrical properties. 'Metals' have very high electrical conductivity, whereas the 'insulators' have very low conductivity. The electrical properties of semiconductors are between that of metals and the insulators.

Many of the interesting and unique properties of solids, especially electrical conductivity, arise from the nature of solids, history of synthesis and the presence of lattice defects or imperfections. Electrical imperfections also arise in solids because of deliberate additions (or doping) of the impurity.

11.2 Electrical Properties of Solids

Materials in the solid state exhibit a wide range of electrical properties. 'Metals' have very high electrical conductivity, whereas the 'insulators' have very low conductivity. The electrical properties of semiconductors are between that of metals and the insulators.

Purely on the basis of electrical conductivity, solids are classified into three types:

(i) Conductors, (ii) semiconductors, and (iii) insulators.

(i) Conductors

If any solid substance offers a very low resistance to the passage of electricity through them, then they are said to be 'conductors' or metals. They have electrical resistances ranging from 10^{-1} to 10^{-8} ohm cm or their electrical conductivities ranging from 10^1 to 10^8 ohm^{-1} cm^{-1} at room temperature.

The materials are conductors due to the presence of free or delocalized electrons at temperatures greater than 0°C.

(ii) Semiconductors

The electrical conductivity of semiconductors is found to be between metals and insulators. The materials have conductivity values ranging from 1 to 10^{-8} ohm^{-1} cm^{-1} or their resistance values are in the range of 1–10^8 ohm (or at the room temperature). It can be remarked that all semiconductors are basically insulators at 0 K. The electrical conductivity of such substance increases with the rise in temperature. The charge carriers in these semiconductors are electrons (*n*-type) or holes (*p*-type) or both (intrinsic type).

(iii) Insulators

These solid-state materials are generally known as bad conductors of electricity, due to the fact that they offer very high resistance to the flow of electricity. The resistance values of these substances are greater than 10^8 ohm cm at room temperature and called 'insulators' (or dielectrics).

11.3 Band Theory of Solids

The information of the band of energy levels of solids can be clearly illustrated in terms of the 'molecular orbital theory'.

The salient features of band theory of solids are as follows:

1. Materials in the solid state do have very large aggregation of atoms (or ions) and are arranged in regular close-packed structures.
2. The constituents (ions or atoms) of solids are bound together by a cohesive force and as a result, the atomic orbitals of valence electron shells interact and overlap with each other and give rise to 'molecular orbitals'.

When two atomic orbitals (half-filled) overlap, two molecular orbitals of different energy levels are formed as shown in Fig. 11.1.

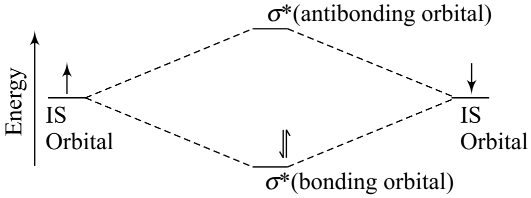


FIGURE 11.1 *Overlap of two molecular orbitals (MO).*

It can be visualized that one of the molecular orbitals (MO) gets stabilized due to pairing up of electrons and acquires lower energy state, which is called 'bonding molecular orbital' (BMO), while the other empty molecular orbital is destabilized (higher energy) and is known as 'antibonding molecular orbital' (ABMO).

When a $2N$ number of valence atomic orbitals of constituent atoms (or ions) overlap with each other in solids, they give rise to the 'N' number of closely spaced bonding molecular orbitals and as well to 'N' number of antibonding molecular orbitals.

For very large aggregation of atoms (or ions), a very closely spaced energy levels (σMOs and σ^*MOs) of extremely small energy difference results in the formation of 'bands' in solids and evidently, these 'energy bands' appear to be continuous as shown in Fig. 11.2.

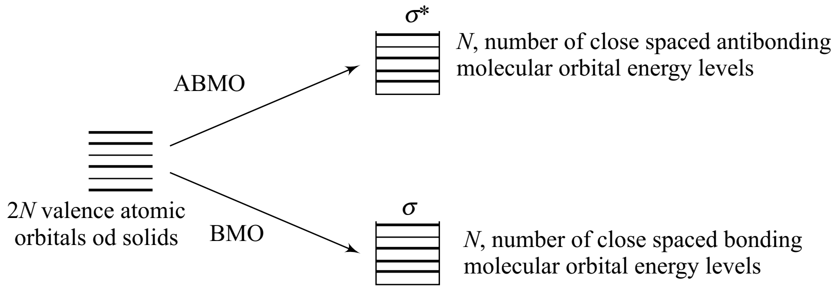


Figure 11.2 *Formation of energy bands in solids.*

The energy gap between the ABMO and BMO (i.e. σ^* and σ bands) or the magnitude of the separation of these bands depends on the nature of the constituents, the internuclear distances of atoms (ions) and the orientation of the valence electron orbitals in a particular crystal structure.

However, in few cases, the energy bands of atoms in solid may also overlap and gives rise to a continuous half-filled energy bands.

A 'band' is a group of infinitesimal energy levels in a solid.

11.4 Types of Energy Bands

The composition, electronic configuration of constituents and the structure of crystals/solids give rise to two kinds of energy band models.

11.4.1 Overlapping Energy Bands

For a particular crystal type, the higher energy band of the solid overlaps the lower energy band to some extent, as shown in Fig. 11.3.

Beryllium and magnesium are metals because the empty np° energy band overlaps the lower filled ns^2 energy band in both the cases.

11.4.2 Nonoverlapping Energy Bands

In many solids, the higher empty energy band does not overlap the lower filled (or half-filled) energy level band and evidently, there is an energy gap between the two bands of energy levels. This energy gap is known as 'forbidden energy gap' (Fig. 11.4).

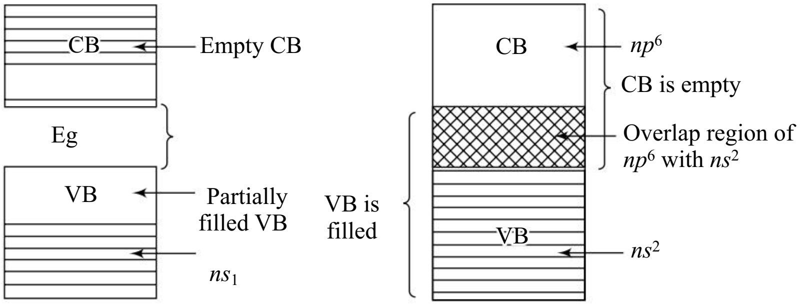


Figure 11.3 *Overlapping energy bands.*

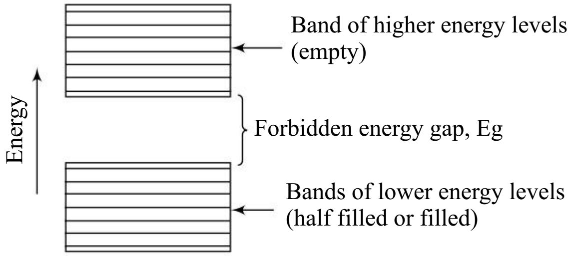


Figure 11.4 *Nonoverlapping energy bands.*

The energy level bands of Li, Na and Cu metals are shown in Fig. 11.5. They are good conductors due to half-filled (ns^1) lower energy band (Fig. 11.5).

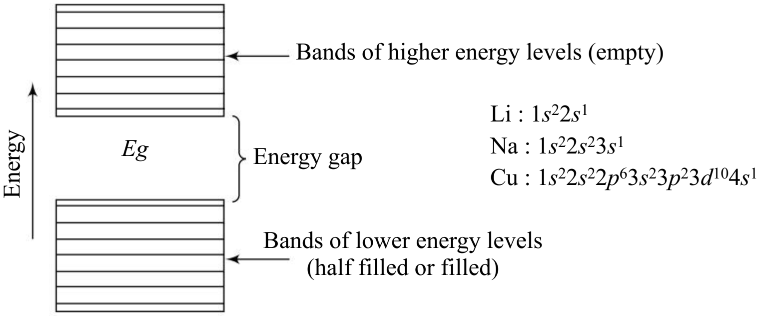


Figure 11.5 *Energy bands of Li, Na and Cu.*

11.4.3 Valence Band

A band of stabilized lower energy levels occupied by the valence electrons of the solid is called the 'Valence band', (VB) and obviously it is the highest filled state of a band. As shown in Fig. 11.6.

The valence band of a solid may be either (i) partially filled or (ii) completely filled with electron depending on the electronic configuration of the constituent in the solid.

(i) The 'valence band' of solids, especially in Be ($1s^2\ 2s^2$) and Mg ($1s^2\ 2s^2\ 2p^6\ 3s^2$), is completely filled with ns^2 electrons.

(ii) The valence shells of Na ($1s^2\ 2s^2\ 2p^6\ 3s^1$), Al ($1s^22s^22p^63s^23p^1$) and Cu ($1s^2\ 2d^{10}\ 4s^1$) are partially filled and therefore, the valence band (VB) of these solids are partially filled with electrons.

11.4.4 Conduction Band

A band of energy levels above the valence band in solids is called 'conduction band', (CB). The conduction band is, generally empty at absolute 0 K for solids (Fig. 11.6).

At higher temperatures, CB is occupied by electrons. The electrons occupying this band are called conduction (free or delocalized) electrons.

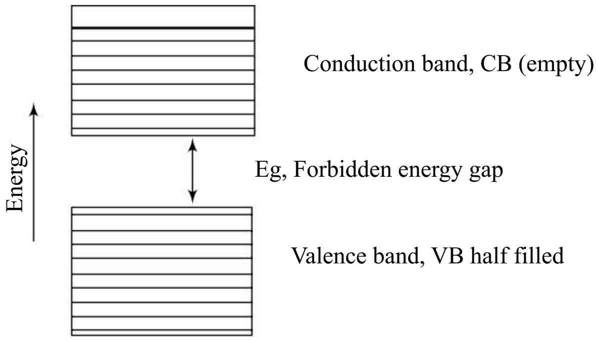


Figure 11.6 *Energy levels of solids indicating empty conduction band.*

11.4.5 Forbidden Energy Gap

An energy gap exists between the VB and CB in solids of particular types. This energy gap E_g (eV) is called the 'forbidden energy gap' as shown in Fig. 11.6.

11.5 Application of Band Theory to Solids

(i) Conductors

Many solid-state substances are good conductors of electricity. This particular behaviour of solids can be explained in terms of 'band theory of solids'.

'A material is a good conductor if it possesses either a partially filled valence band (VB) or an empty conduction band (CB) overlaps with filled VB' (Fig. 11.7).

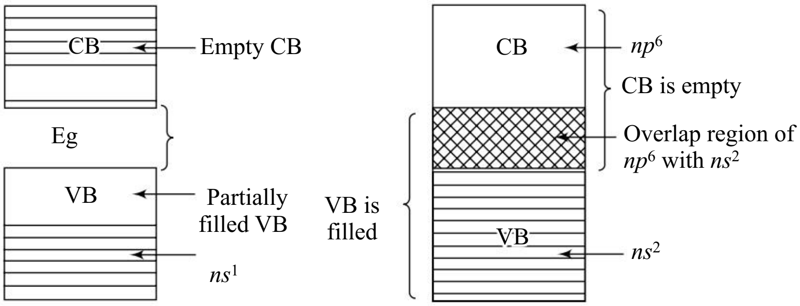


Figure 11.7 *Energy bands levels for conductors.*

In metals such as Na, Li, etc. the valence band (VB) is partially filled, with their valence electrons, ns^1 as shown in Fig. 11.7(a).

When these partially filled electrons of the VB acquire sufficient energy, they are promoted to the higher energy levels of the same VB and thus, the electrons are free to conduct electricity. Such solids behave as good conductors.

Be and Mg are metals. Both the solids have a completely filled VB with ns^2 valence electrons. The metallic properties of Be and Mg are due to the overlap of empty np^0 energy band (CB) with the completely filled VB (ns^2 electrons) as shown in Fig. 11.7(b).

A slight thermal excitation promotes these free electrons to move into different higher energy levels of the overlapped bands. Thus, Be and Mg exhibit metallic conduction.

11.5.1 Insulator

Generally, if the solids have very high value of electrical resistivities, which is greater than 10^{10} ohm cm at room temperature, then such materials is classified as 'insulators' (Fig. 11.8).

However, in terms of band theory of solids, the insulators are described in terms of the following.

(i) VB is completely filled with valence electrons.
(ii) CB is completely empty and
(iii) The forbidden energy gap, E_g between VB and CB is very large, e.g. 5–10 eV.

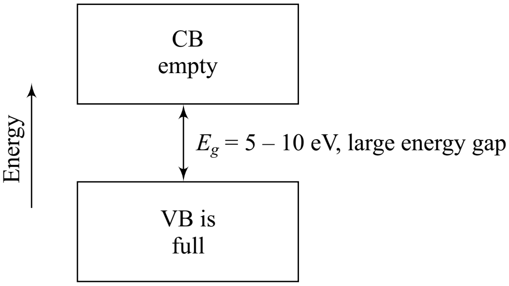


FIGURE 11.8 *Energy band levels with a wide energy gap.*

11.5.2 Semiconductors

It is found that the electrical conductivity of solid semiconductors is in between those of insulators and conductors. The electrical conductivity of these solids ranges between 1 and 10^{-8} ohm^{-1} cm^{-1} at room temperature. A solid is a semiconductor, if it satisfies the following characteristics in terms of the band theory of solids (Fig. 11.9).

(i) Almost filled VB,
(ii) Empty CB, and
(iii) A very narrow energy gap, E_g, of the order 1–2 eV

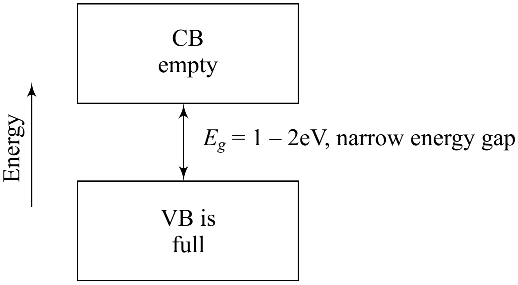


FIGURE 11.9 *Energy band levels of semiconductors with narrow energy gap.*

All semiconductors are insulators at 0 K. At temperatures above 0 K, thermal agitation will lift a predictable number of electrons from VB into CB, a higher energy band, since $E_g \simeq 1.0$ eV (Fig. 11.9). The electrical conductivity of semiconductors increases with the rise in temperature.

The best examples are pure elemental silicon (1.1 eV) and Ge (0.7 eV).

'Chemical impurities' and 'lattice defects' usually determine the electrical conductivity of semiconductors, which increases rapidly with temperature over certain ranges.

Oxide materials (or ceramic oxides) such as ZnO, NiO, V_2O_5 and host of others are good semiconductors with very narrow energy gaps (about 1.0 eV) due to their lattice defects.

Semiconductors are, generally, insulators at 0°C. As the temperature is raised above 0°C, thermal agitation will lift a predictable number of electrons from VB to a higher band, CB, if $E_g \sim 1.0$ eV. The electrical conductivity of semiconductors increases with the rise in temperature.

11.6 Elemental Semiconductors

Pure Silicon and Germanium are the best examples of elemental semiconductors.

Based on the electrical conductivity behaviour elemental semiconductors are classified as:

(i) Intrinsic semiconductors, and
(ii) Extrinsic semiconductors.

11.6.1 Intrinsic Semiconductors

Pure germanium and silicon exhibit semiconducting behaviour at temperature above 0 K.

The forbidden energy gaps, E_g, for Si and Ge are 1.0 and 0.7 eV, respectively. The conductivity of these semiconductors increases with rise in temperature. The semiconducting behaviour of pure Si (or Ge) can be described by two approaches as follows:

(a) Atomic lattice model

A two-dimensional lattice structure of silicon crystal is shown in Fig. 11.10.

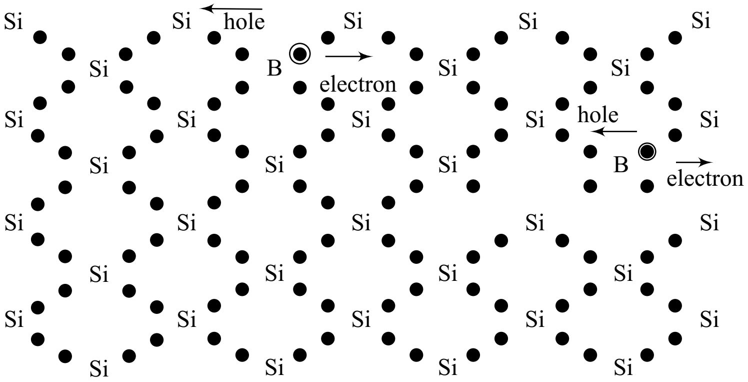


Figure 11.10 *Atomic lattice model for n-type semiconduction of silicon doped with boron.*

In the diamond cubic lattice structure of silicon, each atom of Si (sp^3 hybridized) forms four covalent bonds by sharing with four neighbouring silicon atoms.

In atomic terms, we would say that a pair of electrons is shared in the covalent bond between each pair of silicon atoms. Above 0 K, some of the electrons are thermally released from these bonds and are free to wander through the crystal. These are called ‘conduction electrons’. Thermally released electrons provide a mechanism for electrical conductivity in the solid. The ‘vacant position’ left behind in the covalent bonds is called a ‘positive hole’ or simply a ‘hole’.

Successive jumps of valence electrons into empty positions produce an opposite net motion of the hole in the direction of the electric field; hence the term ‘hole’ to indicate that the entity moves in a field as though it had a +ve charge, although actually what happens is a net shift of the valence bond electrons in the opposite direction.

Thus, the total conductivity of Si crystal is due to the transport of both electrons and holes.

$$\sigma = ne(\mu_e + \mu_h)$$

where ‘n’ denotes the total number of charge carriers $n = n_e = n_h$ for an intrinsic type of semiconductors. μ_e and μ_h represent the electron and hole mobilities, respectively.

(b) Intrinsic semiconductance of pure elemental silicon in terms of band theory

The ‘band theory of solids’ provide a good description of semiconductance of pure silicon (or Ge).

According to the band theory, these pure Si (or Ge) semiconductors are characterized by the presence of a completely filled valence band (VB) and an empty conduction band (CB), which are separated by a narrow forbidden energy gap, E_g.

The band of energy levels for pure elemental silicon crystal is shown in Fig. 11.11.

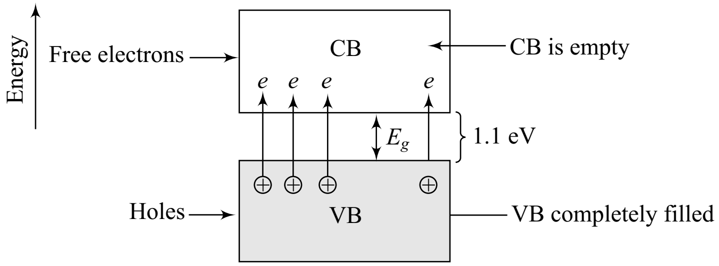


Figure 11.11 *The band of energy levels for pure elemental silicon crystal.*

At 0 K, all the valence electrons of silicon atoms lie in a completely filled VB, while the next higher conduction band is empty and, therefore, Si crystal behaves as an insulator. At temperatures above 0 K, thermal agitation will lift a predictable number of electrons into higher conduction band, depending on the distribution of energy of allowed status and the temperature. In other words, the number of electrons promoted to conduction band in the silicon crystal for semiconduction is, therefore, a function of temperature and the forbidden energy gap, E_g (~ 1.0 eV) of silicon. The thermally excited free electrons move in the conduction band, while an equal number of 'holes' created in the VB moves in the opposite direction under the applied field.

The total semiconductivity of pure intrinsic silicon is due to both electrons and holes at temperatures above 0 K and the number of electrons and holes increases with temperature.

This type of semiconductivity in solids is known as 'intrinsic semiconduction'.

11.6.2 Electrical Conductivity of an Intrinsic Semiconductor

The 'electrons and holes' contribute to the total conductivity of an intrinsic semiconductor.

The conductivity of a semiconductor is generally expressed as

$$\sigma = n_1 e\mu_e + n_h e\mu_h$$

where n = density of charge carriers
e = the charge of carrier
μ = the mobility of charge carriers

It is evident from the above expression that the conductivity of a semiconductor depends on (i) the density of charge carriers and (ii) the mobility of charge carriers.

It is well known that the density of charge carriers of an intrinsic semiconductor is proportional to exp$^{-Eg/2KT}$.

Therefore, $n = n_{o^-}\, e^{-E_g/2KT}$

where K = Boltzmann constant, T = temperature in absolute scale and E_g denotes the forbidden energy gap of a semiconductor.

If 'intrinsic' type semiconductor is observed, the charge carriers are both electrons and holes.

Hence, $= ne\mu_e + ne\mu_n = ne\,(\mu_e + \mu_n)$

where μ_e and μ_h are the mobilities of electrons and holes, respectively, and $n = n_e = n_h$ for intrinsic-type semiconductors.

Worked Examples

The electron and hole mobilities of silicon are 1900 and 500 cm^2/V/s, respectively. Calculate the number of charge carriers for silicon at 290 K. [VTU: Sept 2000]

Given; $e = 1.6 \times 20^{-19}$C, $\sigma = 5 \times 2^{-6}$ ohm^{-1} cm^{-1}

The conductivity of silicon at 298 K

$$= ne\,(\mu_e + \mu_h)$$

Given: $\mu_e = 1900\ \text{cm}^2/\text{V/s},\ \mu_h = 500\ \text{cm}^2/\text{V/s}$

$e = 1.602 \times 10^{-19}$ C

and $\sigma = 5 \times 10^{-6}\ \text{ohm}^{-1}\ \text{cm}^{-1}$

Therefore, $n = \dfrac{\sigma}{e(\mu_e + \mu_h)}$

$$n = \frac{5 \times 10^{-6}\ \text{ohm}^{-1}\ \text{cm}^{-1}}{1.6 \times 10^{-19}\ c[1900 + 500]\text{cm}^2/\text{V/s}}$$

$$= 3.125 \times 10^{13}\ \text{cm}^3[\text{ohm} = \text{VS/C}]$$

1. The electron and hole mobilities of germanium at 298 K are 3350 and 1855 $\text{cm}^2/\text{V/s}$, respectively. Calculate the number of charge carriers, given the conductivity of germanium = $2.08 \times 10^{-6}\ \text{ohm}^{-6}\ \text{cm}^{-1}$ and electronic charge = 1.602×10^{-19}C [VTU: Sept 1999]

   $$= ne\mu$$

   For intrinsic semiconductors, germanium.

   $$\sigma = ne\,\mu_e + ne\,\mu_n = ne\,(\mu_e + \mu_n)$$

   Therefore, $n = \dfrac{\sigma}{e(\mu_e + \mu_n)}$

   Given: $\sigma = 2.08 \times 10^{-6}\ \text{ohm}^{-1}\ \text{cm}^{-1}$

   $\mu_e = 3350\ \text{cm}^2/\text{V/s}$

   $\mu_h = 1855\ \text{cm}^2/\text{V/s}$

   $e = 1.602 \times 10^{-19}$ C

   Therefore, density of charge carriers of germanium is

   $$n = \frac{2.08 \times 10^{-6}/\text{ohm cm}}{1.6 \times 10^{-19}\ c[3350 + 1855]\ \text{cm}^2/\text{VS}}$$

   $$= 1.3 \times 10^{13}\ \text{cm}^3\ \text{at 298 K}$$

2. Calculate the electrical conductivity of Ge, if conduction electron density is $2 \times 10^{13}\ \text{cm}^3$, the electron and hole mobilities are 3600 and 1700 $\text{cm}^2/\text{V/s}$, respectively.

   Given: $e = 1.6 \times 10^{-19}$C

   $$\sigma = ne\mu$$

   For intrinsic semiconductivity,

   $$= n\,e\,(\mu_e + \mu_h)$$

   Given $\sigma = 2 \times 10^{13}\ \text{cm}^3$

   $\mu_e = 3600\ \text{cm}^2/\text{V/s}$

   $\mu_h = 1700\ \text{cm}^2/\text{V/s}$

   Therefore, conductivity of Ge semiconductor

   $$\sigma = 2 \times 10^{13}\ \text{cm}^3 \times 1.6 \times 10^{-19} e\ [3600 + 1700]\ \text{cm}^2/\text{V/s}$$

   $$= 1.69 \times 10^{-2}\ \text{ohm}^{-1}\ \text{cm}^{-1}$$

11.6.3 Extrinsic Elemental Semiconductors

'Chemical impurities' and 'lattice defects' (or imperfections) usually determine the electrical conductivity of semiconductors, which characteristically exhibit a conductivity that increases rapidly with temperature over certain ranges. Basically, all the pure elemental semiconductors (Si or Ge) are 'intrinsic type'. The conductivity of such intrinsic type of semiconductors can be altered by the deliberate addition of impurity atoms (known as dopants).

The conductivity of these doped materials is 'extrinsic' and the type of conductivity is decided by the excess electrons and holes in the host lattice due to impurity (dopant) atoms.

The process of addition of impurity atoms/ions to a semiconductor host lattice is known as 'doping'.

Based on the nature of impurity doped into the lattice of host, extrinsic semiconductors are classified as (a) *n*-type semiconductors and (b) *p*-type semiconductors. The chemical impurities that are generally used to dope the host lattice of Si or Ge are the III and V group elements of P.T.

Examples

	Host lattice	Impurities
(i) *n*-type semiconductor	Si & Ge	Ba, Ga, In & Al (trivalent)
(ii) *p*-type semiconductor	Si & Ge	P, As & Sb (Pentavalent)

(i) *n*-type extrinsic semiconductor

'If majority of charge carriers are electrons in an extrinsic type semiconductor, then it is called an *n*-type semiconductor'.

If a 'pentavalent atoms' are doped into an elemental Si (or Ge) semiconductor, the dopants create a large number of electrons as charge carriers. This type of extrinsic semiconductor is called as *n*-type semiconductor.

The *n*-type of semiconductor can be illustrated in terms of the following:
(a) Atomic lattice model and (b) Band theory model

(a) Atomic lattice model

Silicon atom in its crystal lattice is sp^3 hybridized and forms four covalent bonds with the neighbouring silicon atoms and crystallize in 'diamond cubic structure' (Fig. 11.12).

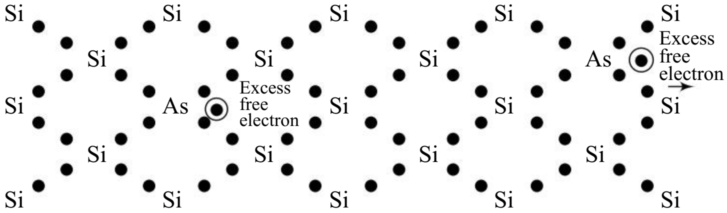


FIGURE 11.12 *Atomic lattice model for n-type semiconduction of doped silicon.*

When dopant atoms such as arsenic are added into the lattice, they substitute few of normal sites. Arsenic is pentavalent and forms four covalent bonds with the surrounding four atoms of silicon; while its fifth valence electron remains loosely bound to the donor atom itself. The presence of the excess electron on the donor atom causes electronic defects in the host lattice and is available for conduction as free electrons, which enhances the conductivity of the host silicon lattice. As a result of As doping, majority of charge carriers obtained are electrons and hence, the materials are *n*-type extrinsic semiconductors.

These thermally released electrons from As atoms provide a mechanism for electrical conductivity in solids.

Since majority of charge carriers are electrons in *n*-type semiconductor, the number of electrons exceeds the number of holes ($n_e > n_h$)

The conductivity of *n*-type semiconductor is

$$\sigma = n_d e \mu_e$$

where n_d denotes the density of the added donor impurity (dopant) atoms and the carrier density varies as $\exp^{-Ed/2KT}$.

(b) Band model

A dopant (or a donor) impurity atom forms four covalent bands with the neighbouring silicon atom of the host crystal lattice with the help of its four valence electrons; while the fifth 'excess electron' in the orbital remains at the donor atom unbonded (Fig. 11.13).

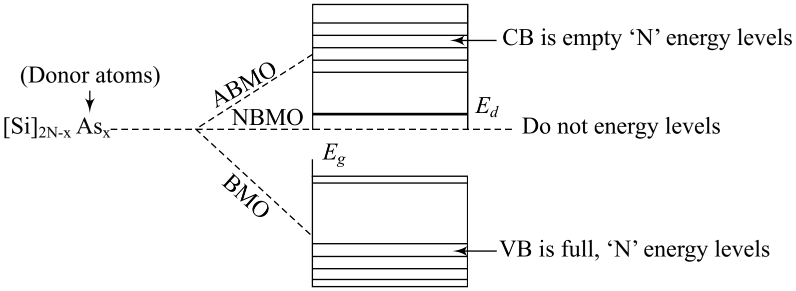


FIGURE 11.13 *Band model of n-type semiconductance of doped silicon.*

According to the band model, the substitution of Si atoms by donor As atoms (excess electron is in the nonbounded molecular orbital, NBMO) results in the creation of a new energy levels known as 'donor energy levels', just below the conduction band with an E_d value of ≃ 0.01 eV. The energy difference between the donor and conduction band energy levels is so narrow (0.01–0.05 eV) that the donor electrons can be easily promoted to the conduction band. This accounts for the observed *n*-type semiconductance of the material.

For *n*-type semiconductor

$$\sigma = n_d e \mu_e$$

where n_d denotes the density of electrons donated by the impurity of atoms and is dependent on the temperature. In the extrinsic conductivity zone, the increase of electrons density with temperature is given as

$$n_d = n_0 \exp^{-E_d/2KT}$$

where 'E_d' is the energy gap between the conduction band and the donor energy level.

(ii) *p*-type extrinsic semiconductors

'If majority of charge carriers are holes in an extrinsic type semiconductor, then it is known as p-type semiconductor'.

When host Si or Ge lattice is doped with trivalent impurities (B, Ga, In, etc.), the majority of the charge carriers are holes. This is known as *p*-type semiconductor.

(a) p-type extrinsic semiconductivity in terms of atomic lattice model.

Each silicon atom forms four covalent (Sp^3 hybridized) bonds that are arranged tetrahedrally in the crystalline lattice of silicon. Each hybrid orbital overlaps with a similar orbital on an adjacent silicon atom to form a covalent bond.

p-type extrinsic semiconductors are obtained by doping the host silicon crystal lattice with elements of III group. The addition of trivalent dopants, such as B, Ga, In, Al, etc., produces a number of 'holes' in the semiconductor. These trivalent impurities are called 'acceptors', because holes are created by them can accept electrons from the lattice.

A two-dimensional view of the silicon host lattice with impurity boron is shown in Fig. 11.14.

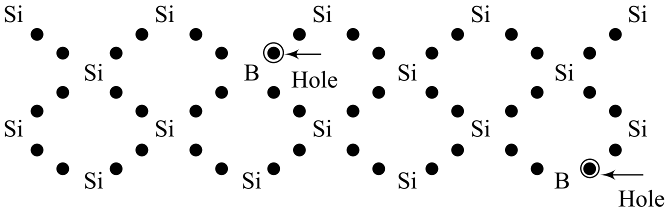


FIGURE 11.14 *Atomic lattice model for p-type semiconductance of doped silicon.*

All the three valence electrons of the acceptor boron atom forms three covalent bonds with the surrounding silicon atom in the host crystal lattice, as shown above (Fig. 11.14), resulting in the formation of +ve charge (a hole) in the silicon atom. These 'holes' created by the impurity acceptor atoms in the lattice of silicon, accounts for the conductivity in *p*-type semiconductor. During the conduction process, a hole travels to an adjacent atom by acquiring an electron and re-establishing a new covalent bond, by breaking an existing covalent bond in the adjacent silicon atom and create a new 'hole' at the site.

The majority of charge carriers in *p*-type extrinsic semiconductor are holes, since ($\mu_h > \mu_e$).
The conductivity of *p*-type semiconductor is.

$$\sigma = n_a e \mu_h$$

where n_a is the density of acceptor atoms or holes and μ_h is the mobility of holes.

(b) p-type extrinsic semiconductivity in terms band model.

Silicon atom in its crystalline lattice is sp^3 hybridized and forms four equivalent bonds with the neighbouring silicon atoms.

Trivalent boron (or acceptor) atom forms three covalent bonds with the surrounding silicon atom of the host lattice with the help of its three valence electrons and as a result a +ve charge is left behind at the boron atom (Fig. 11.15).

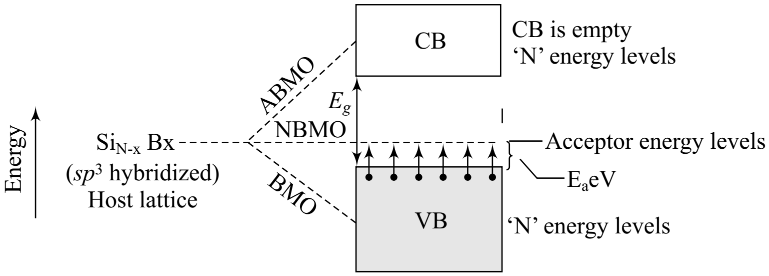


FIGURE 11.15 *Band model for p-type semiconductance of doped silicon.*

When trivalent boron atoms substitute tetravalent silicon atoms of the host silicon lattice, a +ve charge (a hole) is left behind uncompensated at the boron atom site. It is, therefore, evident that three valence electrons of boron atom overlap with three equivalent sp^3 orbitals of silicon atom. In the process, a deficiency of electrons is created. Such a deficiency may be regarded as a +ve charge. The formation of three covalent bonds between boron and silicon atoms results in the formation of conduction band (ABMO) and valence band (BMO) energy levels and both are separated by a narrow forbidden energy gap E_g, of 1.1 eV. A new 'acceptor energy levels' are created little above the VB due to the uncompensated +ve charge at the boron atom site (nonbonded molecular orbital, NBMO).

The energy difference between the 'acceptor energy levels' and VB is extremely small (0.01–0.05 eV) so that with little thermal excitation, electrons from VB band can be easily promoted to the acceptor energy levels. This leaves behind 'holes' in the VB. These 'holes' can move in an applied field and exhibit *p*-type semiconductance in the material.

Since, $n_h > n_e$

where n_a is the density of hole charge carriers and μ_h is the mobility of holes.

The temperature dependence of 'hole' charge carrier of a *p*-type extrinsic semiconductor is

$$n_a = n_0 \exp^{-E_a/2KT}$$

where 'E_a' is the energy difference between the acceptor VB energy levels.

11.6.4 The Effect of Temperature on *n* or *p*-type Extrinsic Elemental Semiconductors

The semiconductance of doped elemental silicon and germanium is extrinsic in nature. It can be observed that the majority of charge carriers in such doped silicon and germanium (having diamond cubic structure) are either electrons (*n*-type) or holes (*p*-type), depending on the nature of the dopants added to it.

The semiconductivity of such doped materials are described by,

$$\sigma = ne\mu.$$

It is evident, therefore, that the conductivity, σ, is thus dependent on the following factors:

1. the density of charge carriers, *n* and
2. the mobility of charge carriers, μ.

Both these factors are found to be dependent on temperature.

(a) Effect of temperature on the mobility of charge carriers, μ

The mobility, μ, of electrons and holes is influenced by 'scattering'. The chief source of scattering in a semiconductor is 'phonons' and 'ionized impurity atoms (donors or acceptors)'.

At low temperatures, it is found that the mobility of charge carriers in a semiconductor varies as $T^{3/2}$. At high temperatures, the converse holds. This signifies that 'ionic scattering' (due to dopants) tends to dominate in extrinsic semiconductors at low temperatures, and 'phonon scattering' at high temperatures. The variation of conductivity, σ, of these *n*- or *p*-type elemental semiconductor with temperature is due to thermally activated mobility, μ, of these localized electrons or holes in the host lattice.

(b) Effect of temperature on carrier density and conductivity

The carrier density, *n*, of a semiconductor is sensitive to temperature, since conduction electrons and holes are provided by thermal excitation. It can be shown that number of charge carriers, *n*, in a semiconductor is proportional to exp ($-E/2KT$) and the conductivity varies as $\sigma = ne\mu$.

Figure 11.16 shown above indicates how the carrier concentrations of *n*- or *p*-type extrinsic semiconductors vary with temperature.

It is evident that the following statements can be summed up regarding the effects of temperature on the variation of charge carrier density for *n* or *p*-type semiconductors.

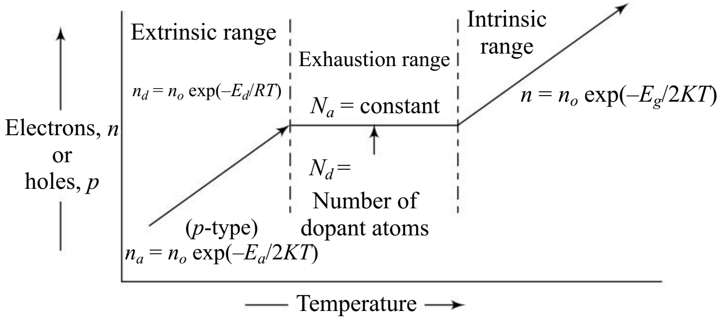


FIGURE 11.16 *Variation of charge carrier density and conductivity of a semiconductor.*

For germanium base extrinsic materials, the energy gap for intrinsic conduction is 0.72 eV and upper limit for extrinsic conduction is 100°C. For silicon base extrinsic materials, the gap E_g is 1.1 eV and extrinsic conduction up to 200°C is possible.

'The temperature at which intrinsic conduction becomes important sets an upper limit to the temperature at which a semiconductor device may be operated.'

11.6.5 Calculate the Carrier Density and Conductivity of *n* or *p*-Type Elemental Extrinsic Semiconductor

Elemental silicon and germanium semiconductors crystallize in 'diamond cubic structure' lattice as given (Fig. 11.17). The diamond cubic structure unit cell contains '8' atoms of silicon or germanium.

The volume of cubic unit cell of Si (or Ge)

$$= [\text{edge length of cube}]^3$$

$$= `a^3\text{' cm}^3$$

where 'a' is the edge length of a cubic unit lattice.

Now, 'a^3' cm^3 of cubic unit cell contains '8' atoms of Si (or Ge).

Therefore, 1 cm^3 of Si (or Ge) lattice contains $\frac{1 \times 8}{a^3}$ atoms of Si (or Ge).

Number of atoms of Si (or Ge) per cm^3

$$= \frac{8}{a^3} = \text{`}X\text{' atoms of Si (or Ge)}$$

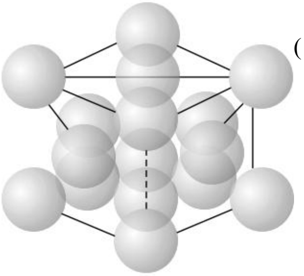


FIGURE 11.17 *Diamond cubic structure of Si or Ge.*

If, 'Z' atoms of Si (or Ge) contains 'Y' atoms of dopant,

Then, 'X' atoms of Si (or Ge) per cm^3 contains × Z atoms of dopant.

Thus, n_d, density of dopant atoms per cm^3 in Si (or Ge) lattice

$$= \left|\frac{Y}{Z}\right| \times X$$

$$n_d = \frac{\text{Number of dopant atoms}}{\text{Number of atoms of Si (or Ge)}} \times \text{Number of atoms of Si (or Ge) per cm}^3$$

In the 'exhaustion range' of semiconduction

'The density of dopant atoms, n_d = the density of charge carriers (electrons or holes) in the lattice of semiconductor'.

The conductivity of n (or p-type) elemental extrinsic semiconductor can be calculated thus,

$$\sigma = n_d\, e\, \mu_e.$$

where n_d = density of charge carriers,
e = charge of charge carriers,
μ_e = mobility of charge carriers.

Worked Examples

1. Calculate the number of charge carriers, n_d, and the conductivity, s, of an extrinsic semiconductor in the exhaustion range, when two atoms of phosphorus is doped for every 10^5 atoms of silicon in a host lattice.
   Given that: the edge of the cubic unit cell is 5.431×10^{-8} cm, the mobility m_e of electron at 25°C is 1900 cm^2/V/s and the electronic charge is 1.60×10^{-19} C.
   Volume of cubic unit cell of silicon lattice

   $$= [5.431 \times 10^{-8}]^3 = 160.20 \times 10^{-24}\ \text{cm}^3.$$

   160.20×10^{-24} cm^3 cubic silicon lattice contains 'eight' atoms of silicon.
   Therefore, number of atoms of silicon per cm^3

   $$= \frac{8}{[5.431 \times 10^{-8}]^3} = 5 \times 10^{22}$$

   If, 10^5 atoms of silicon in a host lattice contains '2' atoms of phosphorus

   Then, 5×10^{22} atoms of silicon per cm^3 contains $\dfrac{2}{10^5} \times \dfrac{8}{[5.431 \times 10^{-8}]^3}$ atoms of phosphorus.

   Therefore,

   $$n_d = \frac{\text{Number of dopant atoms}}{\text{Number of atoms Si (or Ge)}} \times \text{Number of atoms of Si (or Ge) per cm}^3$$

   $$n_d,\ \text{the density of dopant atoms per cm}^3 = \frac{2}{10^5} \times [5 \times 10^{22}]$$

   $$= 10 \times 10^{17}\ \text{at } 25°\text{C}.$$

   In the 'exhaustion range' of semiconduction,
   'The density of dopant atoms per cm^3 = the density of donor electrons, n_d per cm^3'
   The conductivity of extrinsic n-type semiconductor is

   $$\sigma = n_d\, e\mu_e$$

   $$\sigma = 10 \times 10^{17} \times 1.6 \times 10^{-19} \times 1900 = 304.0\ \text{ohm}^{-1}\ \text{cm}^{-1}.$$

2. For a particular purpose, an n-type extrinsic germanium semiconductor with conductivity of 158 ohm^{-1} cm^{-1} in the exhaustion range has to be obtained. What should be the value of the density of electrons, n_d, in the exhaustion range? Calculate the density of dopant [Sb or As] atoms per cm^3 in the germanium host lattice in such a case? Given: m_e = 2800 cm^2/V/s at 25°C, electronic charge is 1.60×10^{-19} C and edge length of the cubic lattice of germanium is 5.658 cm.
   In the exhaustion range, conductivity of an extrinsic n-type semiconductor is $\sigma = n_d e\mu$.

   Therefore, $n_d = \dfrac{\sigma}{e\mu_e} = \dfrac{158.0}{[1.60 \times 10^{-19} \times 2800]} = 3.57 \times 10^{17}$ electrons per cm^3 at 25°C.

   Thus, density of donor electrons per cm^3 in the exhaustion range,

   $$n_d = 3.57 \times 10^{17}\ \text{at } 25°\text{C}.$$

Now, the number of Ge atoms per cm^3 in the cubic lattice

$$= \frac{\text{Number of Ge atoms per unit cell}}{[\text{edge length of cube}]^3}$$

$$= \frac{8}{[5.658 \times 10^{-8}]^3}$$

$$= 4.4 \times 10^{22}$$

$$n_d = \frac{\text{Number of dopant atoms}}{\text{Number of atoms of Si (or Ge)}} \times \text{Number of atoms of Si (or Ge) per cm}^3$$

Therefore, the density of dopant antimony atoms

$$= \frac{n_d}{\text{Number of atoms of Ge per cm}^3}$$

$$= \frac{3.57 \times 10^{17}}{4.4 \times 10^{22}}$$

$$= 8.0 \times 10^{-6}$$

The density of dopant antimony atoms

$$= \frac{\text{Number of dopant atoms}}{\text{Number of atoms of germanium}}$$

$$= 8.0 \times 10^{-6}$$

Thus, ‘8’ atoms of antimony are to be doped for every 10^6 atoms of germanium in the host lattice to get *n*-type extrinsic semiconductor with a σ=158 ohm^{-1} cm^{-1}.

3. Calculate the conductivity of silicon in the exhaustion zone when the number of phosphorus atoms per m^3 of silicon is 9.86×10^{18}. Given, electronic charge, $e = 1.6 \times 10^{-19}$; the mobility of electrons = 0.19 m^2/Vs [VTU: Aug/Sept. 2001] (5 marks)

   In the exhaustion range, conductivity of an extrinsic *n*- type semiconductor is

   $$\sigma = n_d e \mu_e$$

   Given $n_d = 9.86 \times 10^{18}$ phosphorus atoms/m^3
   = number of donor electrons/m^3
   $e = 1.6 \times 10^{-19}$ C,
   $\mu_e = 0.19$ m^2/V/s
   $= 9.86 \times 10^{18} \times 1.6 \times 10^{-19} \times 0.19 = 0.30$ ohm^{-1} cm$^{-1.}$
   = 30.0 ohm^{-1} cm^{-1}

   The conductivity of silicon in the exhaustion zone = 0.30 ohm^{-1} m^{-1} or 30.0 ohm^{-1} cm^{-1}

11.7 Nonelemental Semiconductors

Host of nonelemental materials too exhibits semiconductivity above 0 K. These materials are known as nonelemental or ‘compound semiconductors’.

The compounds (or intermetallics) such as GaAs, InSb, PbTe, GaP, etc., and a host of oxide materials ZnO, NiO, FeO, V_2O_5, etc., are semiconductors. These compounds are the examples of nonelemental semiconductors.

The nonelemental or compound semiconductors are further classified as (i) stoichiometric, and (ii) nonstoichiometric semiconductors.

(a) Stoichiometric semiconductors

Many stoichiometric intermetallic compounds such as GaAs, GaP, InSb, CdTe, CdSe, etc., are semiconductors.

These intermetallic compounds are synthesized by combining the elements of groups III and V (or groups II and VI) in the 1:1 ratio; they are referred to as 'stoichiometric semiconductors'.

The crystal lattice structures of these stoichiometric intermetallic resemble the crystal structure lattice of pure Si and Ge. Like silicon and germanium, the different elements of these compounds also form four equivalent covalent bonds with the four neighbouring elements and crystallize in diamond cubic structure.

The mechanism of electrical conductivity is similar to intrinsic semiconductivity of pure elemental silicon or germanium crystal but these intermetallics have different forbidden energy, E_g values depending on the nature of constituents.

The intermetallic compounds can also be doped to get *n*-type or *p*-type extrinsic semiconductors. These stoichiometric semiconductors exhibit a wide range of intrinsic-type semiconductor and offers a wide range of E_g values ranging from 0.5 to 2.3 eV. These compounds find extensive application in electronics because of their wide range of E_g values.

(b) Nonstoichiometric semiconductors

Many oxides of metals, especially, transition metal oxides, are nonstoichiometric at temperatures above 0 K.

These 3*d*-transition metal oxides show a wide variety of electrical properties because of variable oxidation states of metal ions and defect lattice structures.

Many oxidic compounds are semiconductors, while few others exhibit semiconductor to metallic transition. By virtue of their wide variety of properties, they find extensive application in modern technology. Many such compounds show a tendency to nonstoichiometry. Metal oxides are nonstoichiometric showing either excess or deficient metal ions, compared to what one would expect from the chemical formula.

These defects (or imperfections) in the lattice leads to the semiconduction and the ionic solids are known as 'defect or nonstoichiometric semiconductors'.

The types of 'imperfections or lattice defects' present in these ionic solids give rise to two types of semiconductivity in materials.

They are

(i) *n*-type and (ii) *p*-type semiconductors.

Examples

(i) *n*-type semiconductors

$Zn_{1+x}O$, $V^{5+}{}_{2-x}V^{4+}{}_{x}O_{5-y}$.

(ii) *p*-type semiconductors

$Ni^{2+}{}_{1-x}Ni^{3+}{}_{x}O$, $Fe^{2+}{}_{1-x}Fe^{3+}{}_{x}O$.

where x and y are small for these semiconductors.

11.8 Nonstoichiometric *p*-type Semiconductor

A 'cation-deficient' lattice gives rise to *p*-type semiconductivity in ionic solids.

11.8.1 Semiconductivity of FeO

FeO is a *p*-type nonstoichiometric semiconductor because of 'cation-deficient' lattice structure above 0°C. In FeO lattice, few cation sites are vacant due to missing of Fe^{2+} ions, the excess –ve charges of anions of lattice are compensated by the neighbouring Fe^{2+} ions acquiring Fe^{3+} state.

An ionic crystal lattice of $Fe_{0.9}O$ can actually be represented with composition ${Fe^{2+}}_{0.7}{Fe^{3+}}_{0.2}O$.

The *p*-type semiconductivity of Fe_{1-x} O lattice can be described, thus

$Fe^{2+}\ O^{2-}\ Fe^{2+}\ O^{2-}\ Fe^{2+}\ O^{2-}\ Fe^{2+}$
$O^{2}\ Fe^{(3+)}\ O^{2-}\ Fe^{2+}\ O^{2-}\ Fe^{2+}\ O^{2-}$
$Fe^{(3+)}\ \square\ Fe^{(2+)}\ O^{2-}\ Fe^{2+}\ O^{2-}\ Fe^{(3+)}$
$O^{2-}\ Fe^{2+}\ O^{2-}Fe^{(3+)}\ O^{2-}\ \square\ O^{2-}$ - vacancy $\square$

[A cation-deficient lattice of $Fe_{1-x}O$]

Since Fe^{3+} ion is deficient of one electron with respect to Fe^{2+} lattice, so Fe^{3+} ion in the lattice of FeO can be considered as a 'hole'. The mechanism of conduction in such materials are due to hopping of holes between equivalent Fe^{2+} sites of the FeO lattice under the applied field.

11.8.2 Semiconductivity of *p*-type NiO

NiO is a *p*-type of nonstoichiometric semiconductor due to the presence of cation vacant lattice structure.

At absolute 0 K, NiO is an insulator. However, a cation vacant lattice structure occurs at temperature above 0 K. A defect lattice structure of NiO is as shown.

$Ni^{2+}\ O^{2-}\ Ni^{2+}\ O^{2-}\ Ni^{2+}$
$O^{2-}\ Ni^{(3+)}O^{2-}\ Ni^{2+}O^{2-}$
$Ni^{(3+)}\ \square\ Ni^{2+}O^{2-}\ Ni^{(3+)}$
$O^{2-}\ Ni^{2+}\ O^{2-}\ \square$ O2- - vacancy $\square$
$Ni^{2+}\ O^{2-}\ Ni^{(3+)}O^{2-}Ni^{(3+)}$

[A cation-deficient lattice site of ${Ni^{2+}}_{1-x}\ {Ni^{3+}}_{x}$ O]

Consider a lattice of NiO, in which all the Ni^{2+} ions are normally divalent. When few Ni^{2+} ions are missing due to cation vacant site; for charge compensation of the NiO lattice, two neighbouring Ni^{2+} sites changes to Ni^{3+} state or it is easy to introduce into the lattice a small concentration of Ni^{3+} (temperature effect). The cations are mainly Ni^{2+} but includes some $Ni^{3+}$and this effectively amounts to the introduction of few 'holes', since there is a valence electron missing at the Ni^{3+}sites in the lattice of a crystal. The 'holes' are highly 'localized in NiO'. These holes spend a considerable time on a particular cation before 'hopping' to adjacent sites, with such jumps possible whenever ions of different valence are adjacent.

A jump of this kind, for example, changes the sequence of ions.

$$Ni^{2+}\ Ni^{2+}\ Ni^{(3+)}\ Ni^{2+} \rightarrow Ni^{2+}\ Ni^{(3+)}\ Ni^{2+}\ Ni^{2+}$$

This provides a mechanism for *p*-type semiconduction in such materials.

11.9 Nonstoichiometric *n*-type Semiconductor

The *n*-type semiconductivity is observed in nonstoichiometric ionic solids, having either 'oxygen vacant' or 'cation interstitial' lattices. They are *n*-type semiconductors because the charge carriers are electrons.

Examples

(i) cation interstitial lattice : $Zn_{1+x}O$.

(ii) cation vacant lattice : $V^{5+}{}_{2-x}V^{4+}{}_{x}O_{5-y}$, $Fe^{2+}{}_{2-x}Fe^{3+}{}_{x}\ O_{3-y}$.

11.9.1 Semiconductivity of ZnO

ZnO is a nonstoichiometric *n*-type semiconductor. ZnO is nonstoichiometric due to the excess of Zn^{2+} ions in the lattice of ZnO. In ZnO, lattice defects occur with few Zn^{2+} ions occupying interstitial sites of the lattice, and simultaneously, two electrons occupies the interstitial position to balance the excess +ve charge of the lattice.

$$
\begin{array}{ll}
Zn^{2+}O^{2-}\ Zn^{2+}\ O^{2-}\ Zn^{2+} & \\
\qquad\qquad e \longleftarrow\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!- & \text{Electron} \\
O^{2-}\ Zn^{2+}\ O^{2-}\ Zn^{2+}\ O^{2-} & \\
\qquad\quad Zn^{2+} \longleftarrow\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!- & \text{Interstitial site} \\
Zn^{2+}\ O^{2-}\ Zn^{2+}\ O^{2-}\ Zn^{2+} & \\
O^{2-}\ Zn^{2+}\ O^{2-}\ Zn^{2+}\ O^{2-} &
\end{array}
$$

[A cation interstitial lattice of $Zn_{1+x}O$]

In this Zn^{2+} ion excess lattice, the electrons trapped at the interstitial sites are available for conduction under the applied field. This provides a mechanism for *n*-type semiconductor in ZnO lattice.

11.9.2 Semiconductivity of NiO

The conductivity of nonstoichiometric semiconductors depends on the (i) amounts of dopants added, (ii) temperature, and (iii) pressure of O_2.

Consider the case of Lithium doped NiO. NiO is an insulator at 0 K. However, black NiO of composition $Ni^{2+}{}_{1-x}Ni^{3+}{}_{x}$ O is a *p*-type semiconductor, as described earlier. The semiconductivity of such defect lattice $Ni^{2+}{}_{1-x}\ Ni^{3+}{}_{x}$ O can be altered by adding small amounts of dopants like lithium.

It is known that Li^+ ion can occupy the interstitial sites by virtue of its smaller ionic radius (0.6A) and as well, ions can substitute the normal Ni^{2+} sites of the host lattice of NiO for little higher concentration of lithium.

(i) Small Li^+ ion concentration

For extremely minute quantities of lithium addition to the NiO host lattice, the added Li^+ ions preferentially occupy the interstitial sites. For the charge compensation of the defect NiO lattice, for each Li^+ ion at the interstitial, an extra electron will be 'localized' at the Ni^{3+} site. As a result of this, for extremely minute concentration of Li^+ ion, the Ni^{3+} ion concentration decreases. The net effect observed for NiO is that the conductivity of *p*-type semiconductor decreases.

(ii) High Li^+ ion concentration

For moderately higher concentration of Li^+ ions in NiO lattice, the defect lattice enhances the *p*-type semiconduction.

In this case, Li^+ ions are forced to occupy the normal Ni^{2+} sites by its substitution, and for each Li^+ ion introduced, one additional Ni^{3+} (a hole) is created in the NiO lattice to maintain the electrical neutrality. As a result of this behaviour, the Ni^{3+} ion concentration increases and the material shows improved *p*-type semiconduction.

$Ni^{2+}\ O^{2-}Ni^{(3+)}\ O^{2-}Ni^{(2+)}O^{2-}$
$O^{2-}\square\ Ni^{2+}\ Ni^{2+}\ O^{2-}\ Ni^{2+}$
$Ni^{2+}\ O^{2-}\ Ni^{(3+)}\ O^{2-}\ Ni^{2+}\ O^{2-}$
$O^{2-}\ Li^{+}\ O^{2-}\ Ni^{2+}\ O^{2-}\ Ni^{2+}$
[Li^+ ion doped NiO lattice]

These are called ‘valence controlled semiconductors’ since the concentration of Ni^{3+} (defect structure) is controlled by the amounts of Li^+ (dopant).

The composition of the valence controlled *p*-type NiO semiconductor is $Li^{+}_{x}\ Ni^{2+}_{1-2x}Ni^{3+}_{x}O$ for small x values at a particular temperature.

11.10 Chalcogen Semiconductors

The elements of VI group, sulphur, selenium and tellurium are known as ‘chalcogens’.

They have characteristic electrical properties. However, in combination with other elements, they form compounds exhibiting a wide variety of interesting properties.

Among these chalcogen semiconductors, selenium (or As_2Se_3) is a good ‘photoconductor’.

Photoconductors are also generally known as ‘light sensitive semiconductors’.

In photoconductivity, the electrical conductivity of Si or As_2Se_3 is increased by the impact of a source of light on the sample. The source of energy from an optical beam causes increased charge carriers (electrons) to be excited across the energy gap, E_g of the photoconductor, as a result of which the conductivity enhances.

11.10.1 Photocopying Process

It is observed that when selenium is exposed to a source of light, its conductivity increases due to the release of electrons. Hence, this property of photosensitivity of selenium is adopted in photocopying (xeroxing) process.

The concept and brief process of photocopying is described below.

In a photocopying unit, the surface of a revolving drum acts as a ‘screen’, which is smeared with a thin coating of finely powdered selenium (Fig. 11.18).

1. In the first process, a corona discharge is induced on the screen of the drum by holding a high potential wire parallel to the rotating drum [see (1) in Fig. 11.18].
2. The document to be photocopied is exposed to a source of light. As the light falls on the screen of the drum, electron–hole pair is formed on the screen for bright areas of the document [see (2) in Fig. 11.8], while the dark areas of the documents do not affect the screen.
3. The drum is discharged under a potential and application of more light, so that the electron move towards the screen and holes in the opposite direction [see (3) in Fig. 11.8].
4. Since, light does not pass through the dark areas of the document, a pattern of +ve charges, i.e. an electrical image of the document, is formed on the screen [see (4) in Fig. 11.8]. As the drum rotates, carbon ink (toner) which has −ve polarity is captured by the electrical image on the screen of the drum.
5. In the last process, the induced discharge results in the transfer of image with ink from the screen to the plain paper [see (5) in Fig. 11.8].

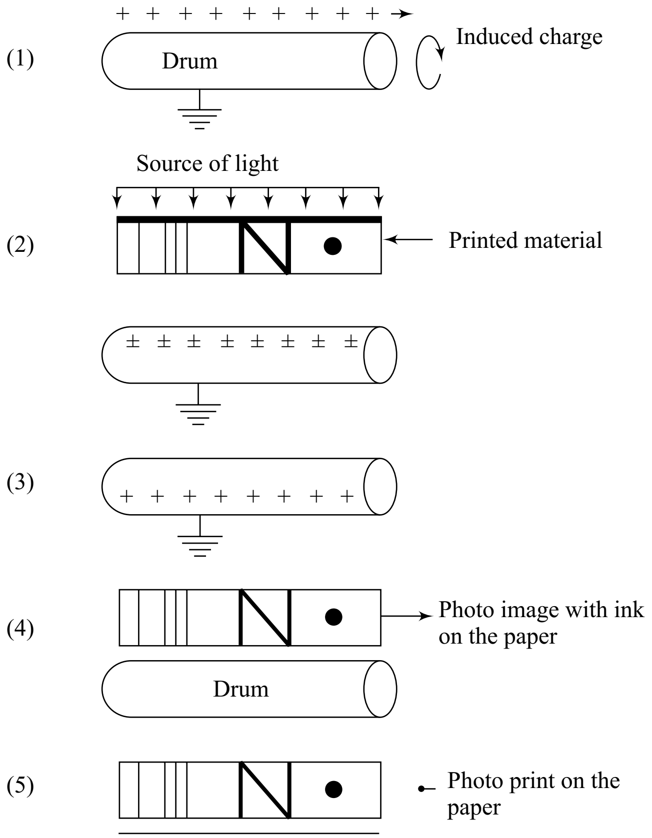


FIGURE 11.18 *Photocopying process.*

11.11 Preparation of Pure Germanium Semiconductor by Fractional Distillation Method

Pure germanium can be prepared by the fractional distillation process. The raw material used for fractional distillation is germanium tetrachloride, $GeCl_4$. In order to remove impurity like arsenic, a layer of HCl acid is kept over $GeCl_4$. The whole liquid is heated in an atmosphere of chlorine gas. The vapours are sent through fractionating column and the last fraction of pure $GeCl_4$ is then treated with water to get GeO_2. Finally, GeO_2 is reduced in a hydrogen atmosphere to obtain pure germanium and can be used as a semiconductor material. Similar process can be adopted to obtain pure silicon semiconductor. The raw materials in this case are $SiHCl_3$ and trichlorosilane.

11.12 Preparation of Pure Germanium by Crystal Growth

Mostly, polycrystalline germanium or silicon materials are obtained from other techniques. In crystal pulling (or growth) technique, a single crystal of germanium (or silicon) can be obtained.

In crystalline pulling technique, the temperature of the melt of polycrystalline materials (Si or Ge) is lowered to just above its melting points. A suitably oriented single crystal seed is attached to a puller rod and brought in contact with the melt. The puller rod is rotated slowly and pulled out simultaneously. Semiconductors (Si or Ge) of considerable size are obtained using the technique. These single crystals are cut into wafers of smooth surface.

11.13 Superconductors

In case of few metals, metallic alloys or ceramic oxides, a unique electrical behaviour at temperatures close to 0 K was observed due to the structural transformation in solids. These materials offer zero resistance to the flow of electricity at this critical temperature, T_c. (Fig. 11.19).

'If any materials in the solid state exhibit zero resistivity at a critical temperature, T_c, such materials are known as 'superconductors'.

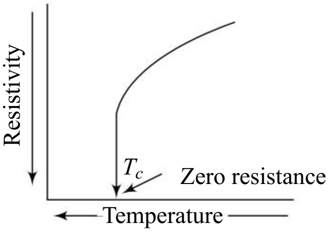


Figure 11.19 *Zero resistivity of a superconductor at critical temperature.*

1. Yttrium barium cuprate, with composition $YBa_2Cu_3O_{7-x}$ when x is small [$T_c = 90$ K].
   The constituents of this material yttrium, barium and copper are in the 1:2:3 molar stoichiometric ratio and hence, they are called as 1:2:3 superconductors.
2. $Bi_2Ca_2Sr_2Cu_3O_{10}$ (2223 compound) is a superconducting material at $T_c = 105$ K.

(i) Synthesis of 1:2:3 superconductor by the ceramic method

$YBa_2Cu_3O_{7-x}$, a 1:2:3 superconductor can be synthesized by the solid-state ceramic technique.

(a) Preparation of a homogeneous mixture of three oxides (Y_2O_3, $BaCO_3$ and CuO) in their molar ratio.
(b) Heating them to obtain oxygen-deficient superconductor in a muffle furnace.
(c) Annealing the above compound to room temperature to retain its composition, structure and superconducting properties.

High purity samples of Y_2O_3, $BaCO_3$ and CuO are taken in 0/5:2:3 molar proportions, mixed thoroughly till an intimate mixture is obtained. This is pressed into a pellet, placed in a Pt crucible and heated to about 950°C for about 2–3 days in a muffle furnace.

The solid-state reaction is

$$0.5Y_2O_3 + 2BaCO_3 + 3CuO \rightarrow YBa_2Cu_3O_{7-x} + 2CO_2$$

The yttrium barium cuprate, thus obtained is annealed to room temperature inside the furnace. The pallet is again crushed to fine powder, mixed well and heated to 950°C for 2–3 days; while maintaining a flow of O_2 gas over heated mass. This is necessary to increase the oxygen content to slightly less than seven and the material is annealed.

(ii) The crystal structure of 1:2:3 superconductor

The crystal structure of 1:2:3 superconductor ($YBa_2Cu_3O_{7-x}$) is described below:

Analysis of crystal structure of 1:2:3 superconductor ($YBa_2Cu_3O_{7-x}$) indicated that this crystal structure has resemblance to ABO_3 structural type units of mineral, perovskite ($CaTiO_3$).

Perovskites (ABO_3) have two basic types of lattice structures and are shown in Fig. 11.20.

The crystal structure of 1:2:3 ($YBa_2Cu_3O_{7-x}$) shown in Fig. 11.20. The 1:2:3 compound crystal structure consists of three perovskite units (two A-type and one B-type lattice) are stacked along the crystallographic C- axis, with copper atoms at the B-sites. Barium atoms are occupied at the A-lattices of perovskites (type-A) in the top and bottom units and yttrium atom is at the centre of B-type lattice at the middle perovskite unit (Fig. 11.21).

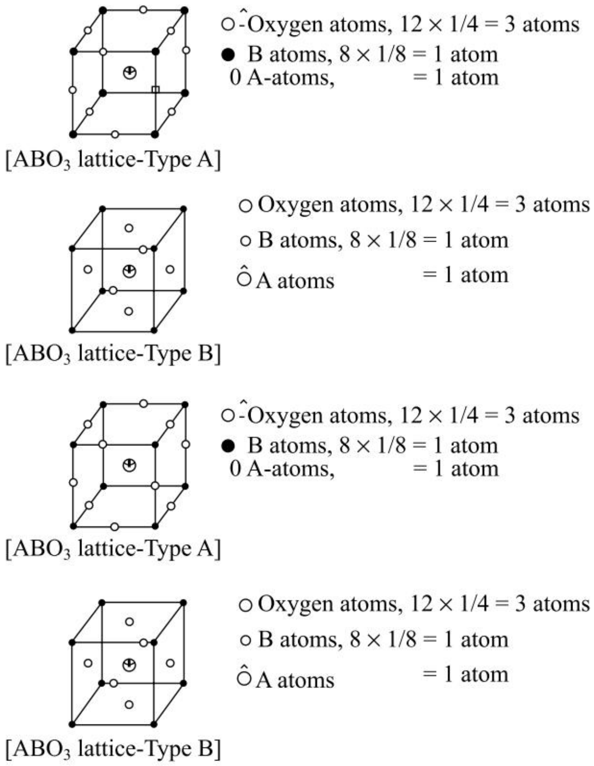
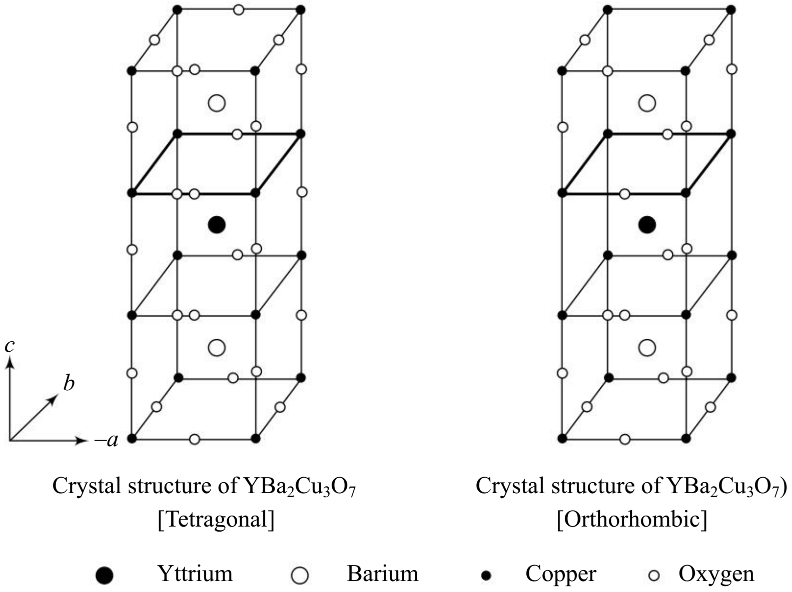


FIGURE 11.20 *Basic structural units of superconductor.*



Crystal structure of $YBa_2Cu_3O_7$
[Tetragonal]

Crystal structure of $YBa_2Cu_3O_7$)
[Orthorhombic]

FIGURE 11.21 *Structure of* $YBa_2Cu_3O_{7-x.}$

If the unit cell of $YBa_2Cu_3O_7$ is stoichiometric, then this unit cell should contain seven oxygen atoms. In order to obtain a nonstoichiometric oxygen-deficient $[O_{7-x}]$ unit cell of 1:2:3 compound, the four oxygen atoms from the vertical edges of the central perovskite cell and two each from the Cu–O horizontal layers edges are removed. These oxygen-deficient vacant sites are believed to be responsible for the distortion of the tetragonal unit cell to orthorhombic structure causing superconductivity.

Review Questions

1. How are the solids classified based on the conductivity?
2. What are conductors, semiconductors and insulators?
3. Describe Band theory of solids.
4. Discuss the conducting properties of solids with overlapping bands with an example.
5. What do you mean by valence band in terms band theory of solids and give its significance?
6. What is conduction band of a solid? Illustrate with a figure in terms of band theory?
7. What is forbidden energy gap in band conduction and its role in conductance of solids?
8. Explain with an example of a conductor in terms of band theory of solid.
9. How does conductivity of a metal and semiconductor vary with temperature?
10. What is an insulator?
11. Describe semiconductors in terms of band theory of solids.
12. How does the conductivity of semiconductors vary with temperature?
13. What are the types of semiconductivity observed in solids?
14. What is intrinsic semiconductivity? Mention few examples.
15. Describe intrinsic semiconductivity of elemental silicon.
16. Explain the semiconducting properties of silicon in terms of band theory.
17. Describe semiconduction of elemental silicon in terms of atomic lattice model.
18. How does the conductivity of an intrinsic semiconductor vary with temperature? Discuss.
19. What do you mean by extrinsic semiconductor? Give two examples.
20. What is doping of a semiconductor? Illustrate its significance.
21. Explain how *n*-type semiconductors are obtained by doping? Illustrate with an example.
22. How *p*-type semiconductors are obtained by doping? Mention the dopant and explain electrical behaviour.
23. Describe *n*-type extrinsic semiconduction of silicon in terms of band theory.
24. Explain the mechanism of *n*-type semiconduction of silicon in terms of atomic lattice model.
25. What is a *p*-type extrinsic semiconductor? How such materials can be obtained by dopants? Explain.
26. Describe *p*-type semiconduction of silicon in terms of Band theory.
27. How *p*-type semiconduction of doped silicon can be explained in terms of atomic lattice model?
28. What are nonelemental semiconductors? Mention few examples.
29. What do you mean by stoichiometric or compound semiconductors? Give examples.
30. Write a note on defects in solid state.
31. What are nonstoichiometric semiconductors?
32. Describe semiconductivity of a nonstoichiometric *n*-type semiconductor with an example.
33. What is an *n*-type nonstoichiometric semiconductor? Explain the mechanism of semiconduction process.
34. Write a note on valence controlled semiconductors.
35. Name chalcogen semiconductors. Give their applications.
36. Describe photocopying process with diagrams.
37. How do you prepare a pure Silicon or germanium semiconductor?

12

Biotechnology

Chapter Outline

Introduction. New trends, scope and importance of biotechnology. Chemical synthesis through biotechnology—fermentation, production of alcohol, production of acetic acid, production of acetone, production of lactic acid, vitamins, preparation of vitamin B_{12}, release of ammonia by bacteria. Industrial enzymes, enzyme processing for textiles. Animal feeds—an account of animal feed enzymes, classification of livestock feeds. Biofuels: types of biofuels, advantages of biofuels, disadvantages of biofuels. Biofuels from energy crops, cellulose as a source of biofuel, biogas. Biomass,—composition of biomass, advantages of biomass, economic benefits, energy benefits, environmental benefits. Biofertilizer, need for biofertilizer, symbiotic nitrogen fertilizer, asymbiotic free nitrogen fixer, azospirillumbiofertilizer, blue green algal biofertilizers, azollabiofertilizer, phosphorus solubilizer, mycorrhiza, advantages of biofertilizers. Biosurfactants. Biosensors, alcohol biosensor, glucose biosensor, enzymatic biosensor, applications of biosensors. Biomembranes, chemical aspects of biomembrane.

12.1 Introduction

Science and technology has developed enormously, and as a result, the demand to solve the various problems has been rigidly confined to the strategies of multidisciplinary approach. The interdisciplinary application involves marshalling of concepts and methodologies from a number of separate disciplines and applying them to a specific problem in another area. Unlike any single discipline, biotechnology can draw upon a wide array of relevant fields such as microbiology, biochemistry, cell biology, immunology, protein engineering, food technology, electronics, enzymology, classified breeding and the full process of bioprocess engineering.

The biotechnology is multidisciplinary, and the fields of interest are vast. In this chapter, a brief discussion of the chemical aspect of biotechnology has been discussed.

12.2 New Trends, Scope and Importance of Biotechnology

Biotechnology is a scientific manipulation of living organisms, especially at the molecular genetic level, to produce useful products.

The application of biotechnology encompasses the following aspects:

1. A technology using biological phenomena for copying and manufacturing various kinds of useful substances.
2. The application of scientific and engineering principles to the processing of materials by biological agents to provide goods and services.
3. Biotechnology is the use of living organisms and their components in agriculture, food and other industrial processes.
4. Industrial microbiology involves the commercial and industrial use of microorganisms to produce organic chemicals, antibiotics, other pharmaceuticals and food supplements.
5. Biopolymers, including cyclodextrins, continue to have new roles in a wide variety of medicinal, industrial and consumer products.
6. Microorganisms are linked with electronics to produce 'biosensors'. Biosensors have many medicinal and industrial applications.

The advantages of the new biotechnological processes are that they function at low temperatures and consume little energy and will rely on inexpensive substrates for biosynthesis. Biotechnology encompasses many traditional processes such as brewing, baking, wine industry, cheese production, the production of oriental foods and sewage treatment, where the use of microorganisms has been developed.

The genetic engineering has had profound impact on all areas of traditional biotechnology and further permitted breakthrough in medicine and agriculture, in particular, those that would be impossible by traditional breeding approaches.

Some of the most exciting advances will be in pharmaceutical drugs and gene therapies to treat previously incurable diseases, to produce healthier foods, safer pesticides, innovative environmental technologies and new energy sources.

There is a substantial impact on wide industrial activities which invariably include human and animal food production, provision of chemical feed stocks to replace petrochemical sources, alternative energy sources, waste cycling, pollution control, agriculture and forestry. Biotechnological industries will be based largely on renewable and recyclable materials and so can be adapted to the needs of the society in which energy is ever increasing expensive and scarce.

12.3 Chemical Synthesis through Biotechnology

In the following topics, we rigidly confine ourselves to a brief discussion of chemical aspects of biotechnology, especially, the production of variety of chemicals through biochemical processes.

The commercial and industrial use of microorganisms is found in 'industrial microbiology', which is used to produce organic chemicals, antibiotics and food supplements. In microbiology and biotechnology, particular microorganisms or microbial communities are frequently used.

12.3.1 Fermentation

Fermentation is a slow anaerobic decomposition of large organic molecules into simpler ones using biochemical catalyst such as enzymes, bacteria and moulds. The process takes place with the continuous evolution of carbon dioxide.

A variety of organic compounds used as organic feedstock's for many chemical industries are produced industrially by the fermentation process. A major attention is given to the microbial production of organic compounds that can be used as a substitute for traditional fuels when petroleum and natural gas become scarce. By evolving a suitable biotechnology, large amounts of biomass can be converted into a variety of useful products.

12.3.2 Production of Alcohol

The following biosynthetic processes have been adopted for the production of ethyl alcohol. They are as follows:

(i) Ethyl alcohol from starch

1. The starch is extracted by crushing potatoes (tubers or roots or grains) and treating with superheated steam at high temperatures and pressure. The resulting product is called as mash.
2. Moist barley is germinated in the dark at 35°C and extracted with ether. This is called malt extract. The solution of starch is heated to 50–60°C with malt extract.
3. An enzyme, diastase, present in the malt extract, hydrolyses starch to maltose, a sugar.

$$\underset{\text{starch}}{2[C_6H_{10}O_5]_n} + nH_2O \xrightarrow{\text{diastase}} \underset{\text{maltose}}{n[C_{12}H_{22}O_{11}]}$$

The liquid is cooled to room temperature and yeast is added.

4. Maltase, an enzyme presents in the yeast, catalyses the hydrolysis of maltose to glucose.

$$\underset{\text{maltose}}{C_{12}H_{22}O_{11}} + H_2O \xrightarrow{\text{maltase}} \underset{\text{Glucose}}{2C_6H_{12}O_6}$$

5. Finally, during the fermentation process, glucose is converted into ethyl alcohol by the enzyme, zymase.

$$C_6H_{22}O_6 \xrightarrow{\text{zymase}} 2C_2H_5OH + 2CO_2$$

(ii) Production from molasses

Molasses is the mother liquor left over after the crystallization of sugar from sugarcane juice. It is a dark coloured liquid containing, 55 percent total sugar. The production process involves the following steps:

The solution is acidified with dilute sulphuric acid (pH of 4–5 is maintained) and ammonium sulphate is added to the syrupy solution.

The solution is fermented in the presence of yeast, at 30°C for 3 days.

The enzyme 'invertase' present in the yeast hydrolyses sucrose into glucose and fructose.

$$\underset{\text{sugar}}{C_{12}H_{22}O_{11}} + H_2O \xrightarrow{\textit{invertase}} \underset{\text{glucose}}{C_6H_{12}O_6} + \underset{\text{fructose}}{C_6H_{12}O_6}$$

Glucose and fructose present in the mixture are then converted into ethyl alcohol by the action of enzyme 'Zymase'.

$$C_6H_{12}O_6 \xrightarrow{\textit{zymase}} 2C_2H_5OH + CO_2$$

(iii) Commercial production of ethyl alcohol

The commercial production of ethyl alcohol involves three steps:

1. Preparation of substrate
2. Fermentation
3. Distillation

For commercial ethyl alcohol production, following three substrates are used as renewable raw materials:

(a) Roots, tubers or grains which are rich in starch.
(b) Molasses or juice, derived from sugarcane palm or sugar beet.
(c) Wood or waste products (rich in cellulose) from processed wood.

Ethyl alcohol is produced mainly by fermentation of molasses, which is produced as a by-product in sugar industries and is quite rich in fermentable sugars (Fig. 12.1).

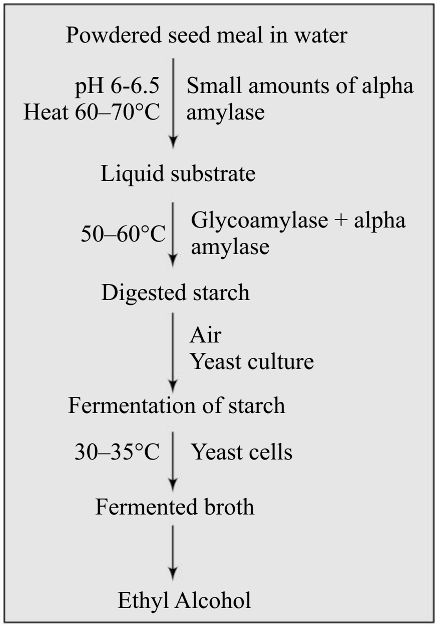


Figure 12.1 *Flow chart of the manufacture of ethyl alcohol.*

12.3.3 Production of Acetic Acid

In this technique of production of acetic acid, an aqueous solution of ethyl alcohol is oxidized by air in the presence of Bacterium aceti (see Fig. 12.2).

A wooden chamber with perforated bottom is packed with beech wood shavings (birch twigs, wood charcoal or fabrics) previously moistened with old vinegar, which contains bacterium aceti. 10 percent ethyl alcohol is allowed to trickle down from the top. The shaving provides a large surface area for oxidation and the added nutrients provide food for bacteria. Air is blown through the inlets at the bottom. Ethyl alcohol undergoes oxidation to acetic acid.

The temperature is maintained at 26–32°C, which is favourable for the growth and activity of the bacteria. The solution is re-circulated till all the alcohol is oxidized to acetic acid. The maximum concentration of acetic acid obtained by this method is 10 percent. The dilute solution of 4–6 percent of acetic acid is called 'vinegar'.

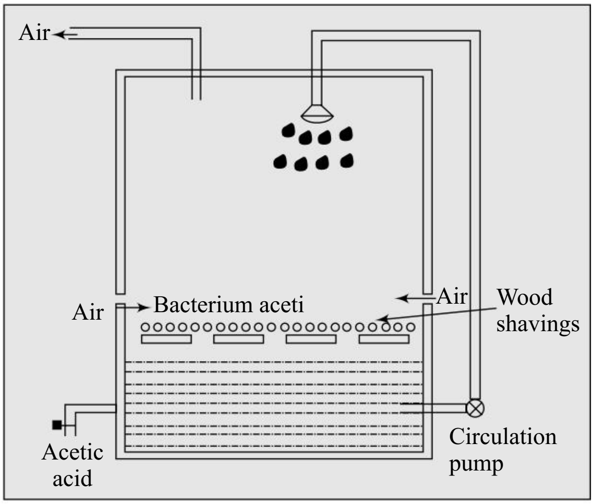


Figure 12.2 *Production of acetic acid.*

$$CH_3CH_2\text{–}OH \xrightarrow{\text{bacterium aceti}} CH_3\text{–}COOH + H_2O$$

Note: At higher concentration of acetic acid, bacteria become inactive.

12.3.4 Production of Acetone

The production of acetone is possible from starch, molasses, wood hydrolysates and pentoses. The fermentation process is carried out using clostridium aceto-butylicium enzyme.

The biosynthesis of acetone derived from acetyl coenzyme-A produced from either glucose or fatty acid dehydration is given in Fig. 12.3.

The substrate like molasses with maize steep liquor is taken in a suitable container in an atmosphere of carbondioxide. The pH of the solution is maintained at 6.0 and the process of fermentation is carried out at 33–35°C. However, the pH drops in the first 18 hours to about 5.2 due to the production of acetic acid and butyric acids, but in the next 18 hours, increase again due to the production of acetone from these acids. Absolute sterilization condition is maintained to avoid bacteria phages in solution during fermentation. Acetone is recovered continuously by fractional distillation.

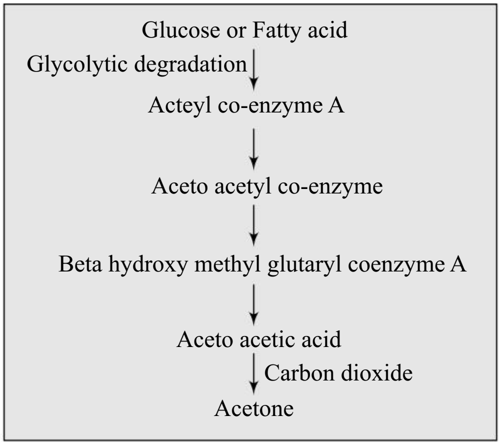
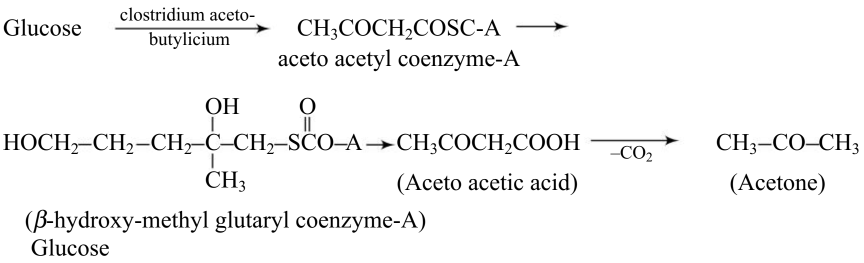


Figure 12.3 *Pathway of biosynthesis of acetone.*



$$\text{Glucose} \xrightarrow[\text{butylicium}]{\text{clostridium aceto-}} \underset{\text{aceto acetyl coenzyme-A}}{CH_3COCH_2COSC\text{-}A} \longrightarrow$$

$$HOCH_2{-}CH_2{-}CH_2{-}\overset{OH}{\underset{CH_3}{C}}{-}CH_2{-}S\overset{O}{\overset{\|}{C}}O{-}A \rightarrow \underset{\text{(Aceto acetic acid)}}{CH_3COCH_2COOH} \xrightarrow[-CO_2]{} \underset{\text{(Acetone)}}{CH_3{-}CO{-}CH_3}$$

(β-hydroxy-methyl glutaryl coenzyme-A)
Glucose

12.3.5 Production of Lactic Acid

A wide variety of organic acids are produced by fermentation process. The lactic acid is produced by the microbial action on sugars such as glucose.

The lactic acid is produced when glucose is used as a substrate in the presence of microorganisms; lactobacillus delbruckii and lacto bacillus leichmannii. During the production, bioreactors are run with the complete depletion of oxygen. The fermentation medium in reactor must contain B-vitamins, in addition to 13 percent glucose and ammonium hydrogen sulphate (0.25 percent). The production is carried out at 45–50°C with the addition of calcium carbonate to maintain the pH between 5.5 and 6.5.

The fermentation duration is about 72 hours. Since, the lactic acid produced is toxic to the microorganisms, the acid has to be removed continuously. +L-lactic acid is produced commercially when the fermentation is completed. The calcium lactate is dissolved by heating the broth. Calcium carbonate is precipitated by treating the broth with dilute sulphuric acid and broth is filtered. After concentration of lactic acid, it is purified.

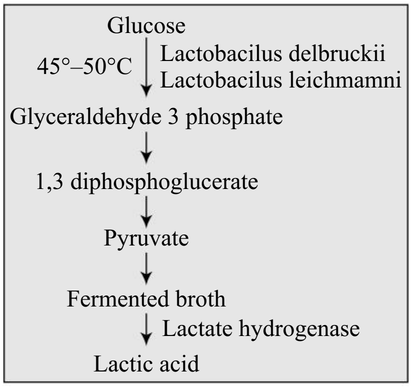


Figure 12.4 *Production process of lactic acid.*

The biosynthesis of lactic acid from glucose proceeds via glyceraldehyde 3-phosphate, 1,3 diphosphate glycerate and pyruvate. The biosynthesis of lactic acid takes the following steps (See Fig. 12.4):

12.4 Vitamins

Vitamins are a class of organic compounds required in small amounts in diets for a stimulatory effect on the growth of microorganisms for healthy nutrition in human and animals.

- Vitamins are not produced in the human body and animals but they have to be introduced into the body with diet either as free vitamins or as a precursor form.
- Deficiency of these vitamins causes different deceases, which can be cured by supply of required vitamins in diet.
- Depending on the solubility of vitamins in water or polar solvents, they are classified as:
  1. Fat soluble vitamins: Vitamins A, D, E and K.
  2. Water soluble vitamins: Vitamins C and B-complex (B_1, B_2, B_6, B_{12}, niacin, Folic acid, etc.)

Vitamins are growth co-factors that are increasingly being introduced as food additives, as medical or therapeutic agents, and as health aids. Although variety of vitamins can be produced by microorganisms, only bacterial production of vitamin B_{12} and ascorbic acid (vitamin C) are of economic significance. Humans obtain vitamin B12 only from food since vitamin B_{12} synthesized by microorganisms in the lower intestinal tract cannot be absorbed. Vitamin B_{12} was first obtained as a by-product of an antibiotic production with streptomycesolivaceus.

For industrial production purposes, propionibacterium freudenreichii, p-shermanii and pseudomonasdenitrificans are used.

12.4.1 Preparation of Vitamin B_{12} (Cobalamin)

Anaerobic fermentation of a mixture carbohydrate, protein, yeast, calcium carbonate (0.5 percent) and cobalt salt (about 10 ppm) using streptomycesolivaceus yields vitamin B_{12} at a given set of conditions.

Commercial method:

The microaerophilic *Propionibacterium* species produce cobalt corrinoids in conventional media (e.g. in molasses or corn steep liquor) supplemented with cobalt in the absence of aeration. As these bacteria can synthesize 5,6-dimethylbenzimidazole under toxic culture conditions, a two stage process was developed.

In the first anaerobic stage (2–4 days), the bacteria grow and produce 5-deoxyadenosyl-cobinamide. Subsequently, a shift to the aerobic phase (3–4 days) promotes the biosynthesis of 5,6-dimethylbenzimidazole so that 5-deoxyadenosyl cobalamin (coenzyme B12) can be produced.

Denitrificans are used in one stage process; bacterial growth parallels vitamin B12 synthesis under toxic conditions when cobalt and 5,6-dimethylbenzimidazole are added as supplements.

A brief detail of sources and deficiency deceases of vitamins is given below:

Vitamin	Source	Deficiency decease's
(1) Vitamin-A	Sweet potatoes, carrots, cabbages, papayas, etc.	Night blindness, xeropthalmia, dermatitis, decrease in resistance, abnormal growth, etc.
(2) Vitamin-D	Irradiation of Sterols/ergosterols found in mould and yeast	Rickets, osteomalacia, etc.
(3) Vitamin-E	Distributed in both plant and animal tissues. cotton seed oil, soya bean oil, rice: livers, heart, kidney, etc.	Decreased erythrocyte life span, hepatic necrosis, anaemia, neurological disorder, muscular dystrophy, etc.
(4) Vitamin-K	Putrefied fish, green leafy vegetables, spinach, cauliflower, soya bean, carrots, etc.	Increased clotting time of blood-haemorrhage conditions, pancreatic dysfunction, biliary disease, etc.

Table 12.1

Fat soluble vitamins

Vitamin	Source	Deficiency decease's
(1) Vitamin-C	Citrus fruits like lemon, oranges, cabbages, guava, grapes, apple, jackfruits, milk, liver, meat and fish.	Scurvy, structural changes in the cartilage bone, anaemia, longer coagulation time and weakness.
(2) Vitamin-B complex	Yeast, liver, rice polishing, cereals, pulses, oil seeds, eggs, fruits, and vegetables.	
(i) Vitamin-B_1 (thiamine)		Beri-beri,
(ii) Vitamin-B_2 (Riboflavin)		Cheilosis, corneal vascularization and inflammation, dermatitis, painful tongue.
(iii) Niacin		Pellagra-lesions of parts of body, diarrhoea, headache, depression, anxiety, insomnia.
(v) Folic acid		Anaemia, infertility, retardation of growth, inadequate lactation in females, weakness.
(vi) Vitamin-B_1		Nervousness, insomnia, weakness, anaemia.
(vii) Biotin		Nausea, anorexia, muscular pain.
(viii) Vitamin-B_{12}		Microcytic anaemia, inflammation of tongue and mouth, diseases of nervous system.

Table 12.2

Water soluble vitamins

12.5 Release of Ammonia by Bacteria

Nitrogen found in compounds like amino acids, purines, pyrimidines, etc. ultimately comes from atmospheric nitrogen. During the process of reduction of nitrogen to ammonia, microorganisms use ATP (adenine tri phosphate) and a powerful reductant. This process of conversion known as nitrogen fixation is carried out by bacteria and cyanobacteria, a blue-green algae.

Rhizobium bacteria, a symbiotic microorganism, attack the roots of the leguminous plants and form root nodules, where nitrogen fixation takes place, supplying both bacteria and plants.

It is well known that the biological process of fixation of nitrogen requires a complex enzyme with multiple redox centres. A two-component enzyme, nitrogenase, catalyses the intricated reaction by the conversion of nitrogen into ammonia.

It is the photosynthesized product reduced ferredoxin found in chloroplasts of plants release the requisite number of electrons to reductase (an iron protein) and thereafter, to nitrogenase component (an Mo-Fe protein) for the intricated bioreduction reaction of nitrogen to ammonia. During the process, ATP gets converted to ADP (adenine diphosphate) by the reductase. The nitrogenase complex is exquisitely sensitive to the inactivation by oxygen.

$$N_2 \xrightarrow{\text{nitrogenase complex}} 2NH_3$$

12.6 Industrial Enzymes

Any living cells contain complex molecules called proteins, which catalyze the desired biochemical reactions. This unique ability of enzyme to perform their specific chemical conversions in isolation has led to an ever-increasing use of enzymes in industrial processes, collectively, termed as enzyme technology. The catalytic activity of an enzyme in a process is highly specific in nature. Very large conversions can be achieved in the enzymatic chemical reactions. It is a fact that a very minute quantity of enzyme can interact with large amounts of substrate. For some enzymes, there is an obligatory need for additional factors, termed 'co-factors', that can be ions of metals, nucleotides, etc. It is known that enzymes can catalyze reactions at relatively low temperatures (0–100°C) and the pH range of 2–14. Enzyme technology will undoubtedly contribute to the solution of some of the most vital problems with which modern-day society is confronted, e.g. food production, energy shortages, preservation and improvement of environment. Applications of industrial enzymes are numerous and are listed briefly in Table 12.3.

Industry	Enzyme used	Applications
1. Baking industry	Fungal-amylase Proteinase	Catalytic breakdown of starch into sugar which can be used by yeast, baking of breads, buns etc. Baking of biscuits
2. Brewing industry	Enzymes produced from barley Amylases, gluconases and proteinases. Proteinase	Degrade starch and proteins to simple sugars, amino acids and peptides. Alcohol production. Split polysaccharides and proteins. Remove cloudiness during storage of beers
3. Starch industry	Amylases, amylo-glycosides, glucoamylases Glucose isomerase	Converts starch into glucose and other syrupy substrates. Converts glucose into fructose.
4. Textile industry	Amylases	Now widely used to remove starch, which is used as an adhesive or size on treads of certain fabrics to prevent damage during washing.
5. Biological detergents	Proteinases Amylases	Used in presoaked conditions and direct liquid applications. detergents for machines-dish washing to remove resistant starch materials
6. Dairy industry	Rennin Lipases Lactases	Manufacture of cheese used to split proteins. Enhance ripening of mould cheeses. Breakdown of lactose to glucose and galactose.

Table 12.3

Applications of enzymes

12.7 Enzyme Processing for Textiles

Nature has provided us with textile fibres such as cotton, wool and silk but there is now the potential to harness biotechnology and produce new or modified fibres as well as improving the production yields of existing fibres. Long-term goals of biotechnology include the modification of fibre quality and properties (e.g. length and strength) leading to the development of high performance cotton.

The enzymatic desizing of starch used for cotton weaving is a traditional biotechnological process. The fibres, yarn or fabric to be finished, the material should be soaked or pre-heated in water to remove natural inhibitors present on the substrate. During weaving, there are ample chances of breakage of yarn due to mechanical friction, and to avoid it, yarn is generally passed through starchy paste to get its surface cemented. This process is sizing and the most important process in weaving. The viscosity of the paste is maintained by breaking starchy macromolecules using the enzyme amylase. They have shown excellent capacity of degrading the starch from all surfaces of yarn even in woven form, and that too at normal conditions, ultimately resulting in further uniform finishing of fabric. Removal of starch from yarn or fabric surface is termed desizing and it is an important step in wet preparatory process. Amylases, lipases or bacterial enzymes are preferred for desizing, since they are able to withstand working temperature up to 110°C.

Cellulase enzyme has entered the detergent market and unlike the other enzymes that degrade particular strains, the cellulases act directly on fabric. A new, cotton fibre consists smooth fabric but with prolonged use and washing, micro fibrils or broken strands of fibre create a fuzz or roughness on the fabric surface. The novel application of cellulase enzymes in textiles is in bio-polishing: the removal of fuzz from the surface of cellulosic fibres, which eliminates pilling and makes the fabrics smoother and cleaner looking.

As starch is costly, tamarind kernel powder (TKP) is successfully replacing starch in sizing. Fabric sized with TKP can be desized with cellulose enzyme of Penicillium funiculosum. Natural fibres like cotton and wool are coated with waxy material which is essentially dewaxed before their chemical processing. These waxes can be degraded using enzymes from microorganisms neisseriacatarrhalis, nocardiaopaca and pseudomonas fluorescents. Preparations of enzymes from the anaerobe clostridium kluyveri can also be useful for removal of fatty coating. Thus, wax eating bacteria can be used along with pectinase enzyme available in aspergillusniger in scouring treatment of natural fibres which otherwise has to be brought out in alkali bath under high pressure and temperature for long time.

The best fibres like linen contain high amount of hemicellulose and pectin substance, in addition to the abundant natural polymer, cellulose. The presence of these substances on fibre surface results in decreasing absorbency, and difficulties in wetting of the fibres during processing arise. Enzymes like hemicellulose and pectinase, both derived from a strain aspergillusniger can selectively cleave respective substrate from fibres under mild conditions, without damaging cellulose chains in the fibres and consequently, less strength losses.

It is a common practice to wash cloths for longer periods under cold conditions. There is an enormous demand for enzymatic materials for textile processing, ecofriendly detergency and for the durability of cloths. Proteases have dominated the detergent market, while amylases and lipases re-widely used for the removal of starch, which is used as an adhesive or size on treads of certain fabrics to prevent damage during washing.

Cellulase has entered the detergent market and unlike the other enzymes that degrade particular strains, the cellulases act directly on fabric. When new cotton consists of smooth fabric but with prolonged use and washing, microfibrils or broken strands of fibre create a fuzz or roughness on the fabric surface. The cellulases remove this and so improve the appearance and feel or smoothness of the fabric.

12.8 Animal Feeds

All living organisms require specific essential nutrients to satisfy the biological processes associated with tissue maintenance and repair, for growth and all other productive activities including reproduction, lactation or work. Unlike green plants, animal cannot capture solar energy and combine this with basic elements to provide nutrients but must rely on finding, ingesting and digesting suitable feeds to satisfy their requirements.

Animals must receive sufficient amount of all essential nutrients (water, energy, amino acids, vitamins, minerals) to remain healthy, to grow and to produce.

'Ration formulation' involves combining the various ingredients so that animal's nutritional requirements are met. One of the major challenges facing the livestock production industry today is providing adequate amounts of a balanced ration at a reasonable cost.

Most potential feeds have complex chemical structures that must be digested (broken down) into simple compounds before they can be taken into (absorbed) and used within the animal body. This process includes the ingestion of feeds, physical and chemical reduction to simple products for absorption from the digestive tract and subsequent elimination of indigestible residues.

12.8.1 An Account of Animal Feed Enzymes

A brief account of enzymes used in animal feeds for growth, protein digestion and absorption hydrolyses are given below:

(a) Amylase

Amylase is widely used in the animal feed industry as a fungal alpha amylase which helps in digestion of starch and it can readily hydrolyse starch molecules into small oligo-carbohydrates and eventually glucose and maltose. Its function in upper digestive tract is to correct incomplete starch digestion.

(b) Protease

Protease enzyme helps in the availability of low grade proteins by releasing readily digestible peptides and facilitates to absorb amino acids. Protease increases digestive capacity of animals and birds and also ensures availability of adequate nutrients supply for better growth and performance.

(c) β-Glucanase

β-Glucanase digests high molecular weight beta-glucans in grain- and cereal-based feeds and can be used in the treatment of endosperm cell walls which contains about 70 percent beta-glucans. Addition of beta-glucanase to grain based feed offers solution to many problems associated with beta-glucans. Glucanase enzyme specifically catalyses the hydrolytic break of glucans (e.g. 1,3–1,4-β-glucanase).

(d) Cellulase

The use of cellulase enzymes in animal feed industry is that it degrades cellulose. Cellulose is probably the most abundant biological compound on earth and is the major component of cereal walls. Cellulase enzyme catalyses the hydrolytic breakdown of cellulose (beta 1,4-glycosidic bonds). Cellulase degrades mainly fibrecomponents in the intestinal tract.

(e) Hemi-cellulase

Hemi-cellulase is widely used in the animal feed industry, due to its ability to degrade various plant fibrous, to facilitate extractions, improve separations, reduce viscosity and modify or completely hydrolyse cellulose.

(f) Lipases

Lipases are useful in the feed industry for the improvement of fat digestion.

12.8.2 Classification of Livestock Feeds

Livestock feeds can be classified as:

1. Roughages or forages

2. Cereals and
3. Supplements:
    - (i) Energy supplements
    - (ii) Protein supplements
    - (iii) Minerals
    - (iv) Vitamins
    - (v) Water

(i) Roughages

The roughage or forage utilization comes from pastures, crop harvested and stored specifically for feeding animals or residues of crop grown for other purposes.

'Roughages' are usually feeds that have a low density, but because of the variety of crops which are considered to be roughages, there is a great deal of variation in composition. Most feeds in this group have a high crude fibre content and a low digestible energy content.

(i) Corn silage has high levels of digestible energy. The protein, mineral and vitamin contents can vary substantially between roughages and also between crops of single roughage. It is very palatable with a fairly high level of digestible energy, but reasonably low levels of protein, calcium, phosphorus, urea, anhydrous ammonia or other sources of nitrogen can be added to increase the level of crude protein.

(ii) Forages form the major part of the most ruminant rations, since they are less expensive than other feeds and the ruminant animal is able to digest most plant components.

All 'pastures' are composed of grass species alone or mixed with legumes and the regional climate is the major consideration in deciding which types or combinations predominate. Hay and green chop are all made from grasses or grass-legume mixtures.

(iii) Straw is the example of crop residues that may be fed to ruminant animals or herbivores. These are the stalks and leaves remaining in the fields after the corn is removed for grain. Waste from sweet corn, peas, bean and other vegetable crops, may be fed effectively.

(ii) Cereals

Cereal grains contain substantial quantities of readily available carbohydrates, sugars, starches, fats or oils. They usually contain little fibre and are highly digestible, so form the major part of diets for animals with simple stomachs. Corn, wheat, oats, Bengal gram and barley are the common cereal grains fed and are generally less variable in nutrient composition than roughages. Cereal grains are all low in Ca and while the P level is quite high. Soya beans can be fed to ruminant animals as both protein and energy source. The protein level is 35–40 percent and although it is deficient in methionine it is not of much importance for ruminants.

(iii) Supplements

Supplements are feeds which are high in one or more of the nutrients required by the animals. They can be classified into (i) protein supplement, (ii) energy supplement, (iii) vitamin and (iv) mineral supplements.

(i) Energy supplement

Many energy-rich feeds are by-products of cereal milling. These by-product feeds usually comes from various layers of the seed covering, which along with higher in fibre, are higher in protein and minerals than the seed itself, which is mainly starch.

(a) Molasses is the major by-product of sugar production, coming mainly from sugarcane but also from sugar beet. It is a good energy source for animals. Molasses is good source of trace minerals, the protein and vitamin level being quite low. It is used to stimulate eating, to reduce dustiness in feeds, as pellet binder, and when fortified with anitrogen source, as a ruminant feed known as a 'liquid protein supplement'.

(b) Beet pulp is another by-product of sugar extraction. It is good feed along with little molasses content. It is very palatable and, despite its relatively high fibre level, is quite digestible. It is often used in lactating cow rations to stimulate appetite.

(c) Surplus fats may be fed as an energy source only, since they supply almost no protein, vitamins or minerals. Fats may be used in feeds to decrease dustiness and increase palatability.

(ii) Protein supplements

A protein supplement is a feed which has more than 20 percent crude protein. Most of the plant protein-rich feeds are by-products of the oil extraction industry. The protein is usually low in sulphur containing amino acids: 'cysteine' and 'methionine' and may be low in 'lysine'. Soya bean is probably the most common protein source (about 48 percent) and is quite palatable, highly digestible and has fairly high energy levels.

(a) Cotton seed meal and linseed meal are the two other potential feedstuffs. Linseed oil cake meal is most popular as a feed for cattle.

(b) Distillers dried grains contain 30 percent protein have relatively high fibre levels and due to the yeast present, a high vitamin-B content. Brewer's dried grains have considerably more fibre (20 percent) and as a result are more suitable as a protein supplement for ruminant animals.

(c) Fish meal protein is very digestible and is very high in lysine, which makes it a good protein source for supplement feed cereal grains. Fish meal is a good source of the vitamin-B, all minerals, and seems to have unidentified growth factors.

(d) Single cell protein: Certain bacteria, yeast and algae can ferment residues from pulp mills, food processing, alcohol production. etc. The protein concentrations range from 48–50 percent and are reasonably well digested. These mostly contain moderate energy and may provide other nutrients such as vitamin-B.

(iii) Minerals

'Minerals' are essential for animal nutrition. Macrominerals are also needed for animal tissue structure and for milk secretion. Feed analysis is necessary to indicate the actual composition so that diets can be fortified as required (Table 12.4).

Macro minerals	Micro minerals
Calcium	Iron
Phosphorus	Copper
Sodium	Zinc
Chlorine	Manganese
Potassium	Cobalt
Sulphur	Iodine
Magnesium	Selenium, etc.

Table 12.4

Fortification of diets for animal nutrition

These minerals are quite essential in requisite amounts in animal feeds for a balanced diet, if otherwise, animals suffer from mineral deficiencies.

(iv) Vitamins

'Vitamins' function as co-factors or enzyme activators in metabolic processes; are also necessary for all general body functions and maintenance of health. All green, growing plants contain carotene which animals can convert into vitamin-A. A supplemental vitamin-A may be necessary to ensure an adequate supply whenever herbivores are not grazing good quality pasture. The rumen microorganisms produce vitamin-B so that these are unlikely to be deficient in ruminants. Animal diets usually contain sufficient vitamins-E and -K but deficiencies can occur. As with all other nutrients, diets should be designed to meet individual requirements and fortified to provide a slight excess of any vitamins that could be deficient under normal conditions.

(v) **Water**

Water is a vital nutrient for all the animals. Requirements vary with age, production and especially with the season of the year. An adequate supply of potable water is essential for all livestock production.

12.9 Biofuels

Biofuel refers to any solid, liquid or gas fuel that has been derived from biomass. It can be produced from any carbon source that is easy to replenish: such as plants. Thus, it could be oils from plants, manure from cows, wood from trees and so on. It is a renewable energy source. Biomass is defined as the biodegradable part of products, waste and residues from agriculture (including vegetal and animal substances), forestry and related industries, as well as the biodegradable part of industrial and municipal waste.

One of the main challenges when producing biofuel is to develop energy that can be used specifically in liquid fuels for transportation. The most common strategies used to achieve this are:

- Grow plants: Plants that naturally produce oils include oil palm, jatropha, soybean and algae. When heated resistance (viscosity) is reduced they can be burned within a diesel engine or they can be processed to form biodiesel.
- Grow sugar crops or starch: These include sugar cane, sugar beet, corn and maize which are then turned into ethanol through the process of yeast fermentation.
- Woods: By-products from woods can be converted into biofuels including methanol, ethanol and wood gas.

(i) Different types of biofuels

- Different biofuels are
  1. E10 fuel: which is actually a mixture of 10 percent ethanol and 90 percent petroleum;
  2. E15 fuel: (15 percent ethanol, 85 percent petroleum);
  3. E20 fuel: (20 percent ethanol, 80 percent petroleum);
  4. E85 fuel: (85 percent ethanol, 15 percent petroleum);
  5. E95 fuel: (95 percent ethanol, 15 percent petroleum) and;
  6. E100 fuel: which is ethanol with up to 4 percent water.
- Biodiesel is the most popular form of biofuel: it can be used in any diesel engine when mixed with mineral diesel. This is produced from oils and fats and is now readily available at many petrol stations.
- This technology has been expanded with the introduction of 'second generation' biofuels: which use biomass to liquid technology. Examples include biohydrogen, biomethanol and mixed alcohols.
- Third-generation biofuels are also known as algae fuels. They have many advantages including have a low input and a high yield level: they produce 30 times more energy per acre than land, and are also biodegradable. As a result, they are relatively harmless to the environment if spilled.

(ii) The advantages of biofuels

The aim of all biofuels is to be carbon neutral. They reduce greenhouse gas emissions when compared to the conventional transport fuels.

In reality, biofuels are not carbon neutral simply because it requires energy to grow the crops and convert them into fuel. The amount of fuel used during this production (to power machinery, to transport crops, etc.) does have a large impact on the overall savings achieved by the biofuels. However, biofuels still prove to be substantially more environmentally friendly than their alternatives.

Another advantage of biofuels is that they save drivers money. With petrol prices on the rise, replacing petroleum with a renewable energy source should also offer significant savings at the pump in the long term, particularly, when biofuels are more readily available.

There are arguments too that biofuels are helping to tackle poverty around the world.

(iii) The disadvantages of biofuels

There are several concerns about biofuels, and particularly including:

- Biodiversity: A fear among environmentalists is that by adapting more land to produce crops for biofuels, more habitats will be lost for animals and wild plants. It is feared, for example, that some Asian countries will sacrifice their rainforests to build more oil plantations.
- Another concern is that if biofuels become lucrative for farmers, they may grow crops for biofuel production instead of food production. Less food production will increase prices and cause a rise in inflation.
- Carbon emissions: Most investigations show that the burning of biofuels substantially reduces greenhouse gas emissions when compared to petroleum and diesel.
- Nonsustainable biofuel production: Many first-generation biofuels are not sustainable. It is necessary to create sustainable biofuel production that does not affect food production, and that does not cause environmental problems.

12.9.1 Fuels from Energy Crop

(a) From hydrocarbon producing plants

In the recent years, large scale growing of hydrocarbon producing plants is encouraged. Hydrocarbon producing plants include a large number of *Euphorbia* and *Asclepias* species. Organic materials are extracted from these latex bearing dry plants. After the water is removed, the latex gives rise to a liquid, which is a mixture of terpenoids and simple sugars (hexoses). The terpenoids can be converted into a gasoline like product and hexoses can be fermented to produce ethanol.

Hydrocarbons are also produced by unicellular alga Botryococcusbraunii, which possess a high hydrocarbon content as high as 75 percent of its dry weight, the highest reported in any kind of biomass. Therefore, cultivation of alga affords a direct and renewable source of solar fuel.

(b) Production of biodiesel

Biodiesel is a renewable, oxygenated fuel made from a variety of agricultural resources such as soybeans or rape seeds.

Biodiesel is a nontoxic, biodegradable replacement for petroleum and diesel. Biodiesel is made from natural, renewable sources such as new and used vegetable (rape seed) oil and animal fats (tallow). Chemically, biodiesel is described as a fatty acid mono alkyl ester.

Through a process called esterification, oils (say, rape seed) and fats are reacted with CH_3OH and NaOH catalyst at $500^{\circ}C$ to produce fatty acid methyl ester along with the co-products: glycerin, glycerin bottoms, soluble potash and soaps.

$$\text{Rape seed oil} + CH_3OH \xrightarrow{\text{NaOH, 500°C}} \text{biodiesel} + \text{co-products.}$$

Using biodiesel in a conventional diesel engine substantially reduces emissions of unburnt hydrocarbons, carbon monoxide, sulphates, polycyclic hydrocarbons and nitrated polycyclic aromatic hydrocarbons.

One of the primary advantages of biodiesel is its renewability. As a renewable, domestic energy source, biodiesel can help reducing dependence on petroleum imports.

Biodiesel is nontoxic, biodegradable and suitable for sensitive environments. Biodiesel contains no petroleum but can be blended at any level with petroleum diesel to create a biodiesel blend. Biodiesel refers to the pure alternative fuel before blending with petroleum-based diesel fuel.

(b) Large scale cultivation of sugar crops

Largescale cultivation of sugar crops: (sugar cane, sugar beet) and starch crops (corn), can be used as a source of biofuels. These crops not only produce sugars which may be used directly or converted into ethyl alcohol (liquid fuel) but also give a valuable by-product, 'bagasse', which can be used as a fuel.

(c) Cellulose as source of biofuel

Biotechnologists have put tremendous effort in developing the following step processes converting cellulose (from trees, weeds shrubs, straw, etc.) into fuel alcohol.

Lignocellulose may be degraded with the help of certain fungi sugars (action of ligase and cellulase), and sugar containing crops are converted into alcohol (yeast fermentation).

Biofuel is obtained from lignocellulose materials. The process involves the conversion of cellulosic and lignocellulosic materials into alcohols (acids and animal feeds) by using microbial enzymes, e.g., cellulase, hemi-cellulase, amylase, pectinase, etc.

Cellulose constitutes the major portion of the plant cell wall, the fundamental unit of which is glucose. A very flourishing industry can be established, if alcohol can be derived from cellulose. Each cellulose molecule consists of few hundred to several thousand glucose units linked together. If cellulose can be degraded into glucose, the glucose can be easily used for the production of fuel alcohol. Several fungi are known to possess the enzyme cellulase and therefore, degrade cellulose into glucose.

There is a growing demand for nonconventional energy sources. Biomass is used as a source for energy generation.

Biomass is defined as living matter or its residues, which is a perpetual or renewable source of energy.

The common examples of biomass are wood, grass, herbage, grains or bagasse. Waste materials including those derived from agriculture, forestry and municipal and domestic wastes.

(d) Growing energy crops involving short rotation forest plantations

A large quantity of biomass materials can be obtained from plants, animals and domestic or municipality wastes.

The waste materials are wood, green-plant matter, paddy straw, rice bran husk, saw dust, etc. Wood can be converted into a liquid fuel, methanol. Green-plant matter and straw are used as animal feeds or converted into an efficient energy source through anaerobic digestion. Rice bran is used as a source of oil, which is converted into methyl ester for use as a fuel. Saw dust is converted into low calorific value producer gas for thermal power generation based on the principle of potential energy source through anaerobic digestion used for biogas production.

Bioconversion is a process involving the conversion of organic materials into energy, fertilizers, food and chemicals through biological agency. In a broad sense, the two bio-conversion steps are as follows:

1. Photosynthesis of biomass and
2. Subsequent conversion into more useful energy forms.

12.9.2 Biogas

Biogas is one of the potential energy sources. Biogases, a mixture of different gases produced when animal, domestic or agricultural wastes are digested with microorganisms under anaerobic condition. It contains methane (50–60 percent) and other gases in relatively low proportions, namely;

CO_2 [25–35 percent], H_2 [1–5 percent], N_2 [2–7 percent] and O_2 [0.01 percent].

The production of biogas involves three steps:

- Hydrolysis: which converts organic polymers into monomers (with the help of hydrolylic bacteria)
- Acid formation: which involves conversion of monomers into simple compounds such as CO_2, NH_3 and H_2, using a group acid forming bacteria (acetogenic bacteria) and
- Methane formation: involving conversion of simple compounds into CH_4 and CO_2, utilizing anaerobic bacteria (methanogenic archae bacteria).

12.9.3 Cellulose as Source of Biofuel

Biotechnologists have put tremendous efforts in developing the following step processes converting cellulose (from trees, weeds shrubs, straw, etc.) into fuel alcohol.

- lignocellulose may be degraded with the help of certain fungi sugars (action of ligninase and cellulase), and
- sugar containing crops are converted into alcohol (yeast fermentation).

Biofuel is obtained from lignocellulose materials. The process involves the conversion of cellulosic and lignocellulosic materials into alcohols (acids and animal feeds) by using microbial enzymes, e.g., cellulase, hemi-cellulase, amylase, pectinase, etc.

Cellulose constitutes the major portion of the plant cell wall, the fundamental unit of which is glucose. A very flourishing industry can be established, if alcohol can be derived from cellulose. Each cellulose molecule consists of few hundred to several thousand glucose units linked together. If cellulose can be degraded into glucose, the later can be easily used for the production of fuel alcohol. Several fungi are known to possess the enzyme cellulase and therefore, degrade cellulose into glucose.

12.9.4 Biomass

Biomass energy is the utilization of energy stored in an organic matter. Examples of biomass include wood, leaves, animal waste, crops, bones and scales. The abundant plant life on our planet is nature's store house of solar energy and chemical resources. Whether cultivated by man, or growing wild, plant matter represents a massive quantity of a renewable resource that we call biomass. Put another way, *biomass is a stored solar energy that can be converted to electricity or fuel. Biomass is a renewable resource.*

Carbon dioxide from the atmosphere and water from the earth are combined in the photosynthetic process to produce carbohydrates or sugars. These sugars form the building blocks of biomass. The solar energy that drives photosynthesis is stored in the chemical bonds of the structural components of biomass. If biomass is burnt efficiently, that is to extract the energy stored in the chemical bonds, oxygen from the atmosphere combines with the carbon in plants to produce carbon dioxide and water. The process is cyclic because the carbon dioxide is then available to produce new biomass. Thus biomass is definitely a renewable energy source. The organic matters are burned directly to produce heat or they are refined to produce fuel-like ethanol or other alcoholic fuels.

(a) The composition of biomass

The chemical composition of biomass varies among different species, but in general, biomass consists of

- 25 percent lignin
- 75 percent carbohydrates or sugars

Within this range of lignin and carbohydrates most species also contain about 5 percent of a third portion of smaller molecular fragments called extractives.

(b) The advantages of biomass

The most important advantage of biomass is that it is everywhere and very easily available. In the agriculture industry, residuals like bagasse (fibres) from sugarcane, straw from rice and wheat, hulls and nutshells, as well as manure lagoons from cattle, poultry and hog farms are usable. Similarly, the timber industry has a lot to offer. Wood wastes such as sawdust, timber slash and mill scrap are considered to be organic materials. Even in the cities, paper and yard wastes are usable. Fully utilized biomass reduces pollution in underground water bodies by off-setting the amount of waste in landfills. Methane and other poisonous gases that form from dead organic matters can be found in landfills and water treatment plants. These can be captured and converted into fuels suitable for generating electricity.

(c) Economic benefits

Rural economies will grow because of the development of a local industry to convert biomass to either electricity or transportation fuel. Because biomass feedstock is bulky and costly to transport, conversion facilities will be located where the crop is grown. That means more people have chances of getting employed. Farmers will see their income rise and will thank to these new markets: for both agricultural wastes and crops that can be grown sustainably on marginal land. As new markets are created, the rural economy will become more diversified.

(d) Energy benefits

Energy producers and consumers will have available a renewable energy option with uniquely desirable characteristics. Biomass has the greatest potential of any renewable energy option for base load electric power production.

(e) Environmental benefits

Agricultural land that might otherwise be converted to residential or industrial use, because a need for fewer and fewer acres to meet the market demand for food, can be used to grow biomass crops that will restore soil carbon, reduce erosion and chemical runoff, and enhance wildlife habitat.

Perennial energy crops can be harvested without damage to the root structure and thus continue to serve as a soil stabilizer and stream buffer and habitat for wildlife. The use of biomass will greatly reduce greenhouse gas emissions.

Fossil fuels remove carbon that is stored underground and transfer it to the atmosphere. In a combustion system, biomass releases carbon dioxide as it burns, but biomass also needs carbon dioxide to grow, thus creating a closed carbon cycle.

(f) The disadvantages of biomass

The burning method of biomass is not clean. It is similar to the burning of fossil fuels and produces large amounts of carbon dioxide. However, it produces much less harmful pollutants (e.g. sulphur), as the main elements found in organic materials are hydrogen, carbon, oxygen and nitrogen. Furthermore, the extra energy crops and other plants can consume the additional carbon dioxide through photosynthesis.

12.10 Biofertilizers

It is a living fertilizer compound of microbial inoculants or groups of microorganisms which are able to fix atmospheric nitrogen or solubilize phosphorus, decompose organic material or oxidize sulphur in the soil. On application, it enhances the growth of plants increase in yield and also improves soil fertility and reduces pollution.

Biofertilizers are organisms that enrich the nutrient quality of the soil. The main sources of biofertilizers are bacteria, fungi and cyanobacteria (blue-green algae). The most striking relationship that these have with plants is symbiosis, in which the partners derive benefits from each other.

Plants have a number of relationships with fungi, bacteria and algae, the most common of which are with mycorrhiza, rhizobium and cyanophyceae. These are known to deliver a number of benefits including plant nutrition, disease resistance and tolerance to adverse soil and climatic conditions. These techniques have proved to be the successful biofertilizers that form a health relationship with the roots.

Biofertilizers will help in solving such problems as increased salinity of the soil and chemical run-offs from the agricultural fields. Thus, biofertilizers are important if we are to ensure a healthy future for the generations to come.

There is a variety of nitrogen fixing microorganisms present in the nature.

Biofertilizers include the following:

1. Symbiotic fixation of nitrogen: rhizobium spp.
2. Asymbiotic nitrogen fixers: azobacter, azospirillum, etc.
3. Algae fertilizers: blue green algae
4. Phosphate solubilizing bacteria
5. Mycorrhizae
6. Organic fertilizers

These microorganisms supply in addition to nitrogen, considerable amount of organic matter enriching structure of soil. Inoculants of these microorganisms have proved their technical feasibility, economic viability and social acceptability. They are therefore called as 'biofertilizers'.

Hence, the term 'biofertilizer' or microbial inoculants is defined as the preparations that contains living or latent cells of efficient strains of nitrogen fixing, phosphorous solubilizing or cellulolytic microorganisms.

12.10.1 Need for Biofertilizers

Biofertilizers have definite advantage over chemical fertilizers. Chemical fertilizers supply over nitrogen whereas biofertilizers provide in addition to nitrogen certain growth promoting substances like hormones, vitamins, amino acids, etc., crops have to be provided with chemical fertilizers repeatedly to replenish the loss of nitrogen utilized for crop growth. On the other hand, biofertilizers supply the nitrogen continuously throughout the entire period of crop growth in the field under favourable conditions. Continuous use of chemical fertilizers adversely affects the soil structure whereas biofertilizers when applied to soil improve the soil structure. The deleterious effects of chemical fertilizers are that they are toxic at higher doses. Biofertilizers, however, have no toxic effects.

Followings are some of the important types of biofertilizers which can be considered for agro-based industries.

12.10.2 Symbiotic Nitrogen Fixers (Rhizobium Biofertilizers)

Rhizobium bacteria in association with leguminous plants fix atmospheric nitrogen in nodules formed on the roots of plants. These nodules are considered as miniature nitrogen production factories in the field. A pure and efficient strain of Rhizobium can be multiplied in the laboratory on a suitable medium by using shake technology or fermentation technology.

12.10.3 Asymbiotic Free Nitrogen Fixer (Azotobacter Biofertilizer)

Azotobacter are free-living microorganisms, and they grow in the rhizosphere and fix atmospheric nitrogen nonsymbiotically and make it available to particular cereals. In addition, these bacteria produce growth promoting substances thereby enhancing the plant growth and finally yield. A highly efficient strain of azotobacteris grown in the laboratory either as shake culture or using fermenter.

12.10.4 Azospirillum Biofertilizers

They are called as associative endosymbiont on roots of grasses and similar types of plants. They are also known to fix atmospheric nitrogen and benefit host plants by supplying growth hormones and vitamins. These bacteria are commonly used for the preparation of commercial inoculants on a large scale using common technique.

12.10.5 Blue Green Algal Biofertilizers

Blue green algae are considered as an important group of micro-organism capable of fixing atmospheric nitrogen nonsymbiotically mostly in rice fields in heterocysts cells specially known as sites of nitrogen fixation. There is a variation between species of blue green algae in fixation of nitrogen and other qualities of biofertilizers. Efficient strains of blue green algae are used for multiplication on a large scale. Production of inoculum in an artificially controlled conditions is defined but more expensive. On the other hand, open air soil culture is most simple, less expensive and easily adaptable by the farmers. This is based on the use of starter culture that is multistrain inoculum of aulosira, tolupothrix, scytonema, nostocand anabaena. This is a paying industry in rural areas and unskilled labour can also undertake the multiplication of blue green algae.

12.10.6 Azolla Biofertilizers

Azolla is a water fern, which grows on the nitrogen fixing blue green algae, anbaena. It contains 2–3 percent nitrogen when wet, and also produces organic matter in the soil. This type of biofertilizer is used all over the world. This can be grown in a cooler region. But there is a need to develop a strain tolerant to high temperature, tolerant to salinity and strains resistant to pests and diseases. Production technology is very easy and can be adopted by rice farmers. The only constraint in azolla is that it has to be kept always growing on water and water becomes limiting factory particularly in summer since viability of azolla seeds is very poor.

12.10.7 Phosphorus-Solubilizer (Biofertilizer)

Phosphorus is one of the important elements required for plant growth and water yield. This element is also essential for nodulation by rhizobium. Phospho-microorganisms are mostly bacteria and fungi are used for commercial purposes. Mycorrhizae have high potential of phosphorus accumulation in plants. There are two types of mycorrhiza, viz., endo- and ecto-trophics. VA mycorrhiza is most popular and used for commercial purpose. In general, it will be seen from above information that biofertilizer manufacturing is a budding industry and need some more years to establish variable.

12.10.8 Mycorrhiza

Mycorrhizae are a group of fungi that include a number of types based on the different structures formed inside or outside the root. These are specific fungi that match with a number of favourable parameters of the host plant on which it grows. This includes soil type, the presence of particular chemicals in the soil types, and other conditions.

These fungi grow on the roots of these plants. In fact, seedlings that have mycorrhizal fungi growing on their roots survive better after transplantation and grow faster. The fungal symbiont gets shelter and food from the plant which, in turn, acquires an array of benefits such as better uptake of phosphorus, salinity and drought tolerance, maintenance of water balance and overall increase in plant growth and development.

While selecting fungi, the right fungi have to be matched with the plant. There are specific fungi for vegetables, fodder crops, flowers, trees, etc.

Mycorrhizal fungi can increase the yield of a plot of land by 30–40 percent. It can absorb phosphorus from the soil and pass it on to the plant. Mycorrhizal plants show higher tolerance to high soil temperatures, various soil- and root-borne pathogens, and heavy metal toxicity.

In a larger sense, the term may be used to include all organic resources (manures) for plant growth which are rendered in an available form for plant associations or interactions. Such biological processes may be as complex as that of nitrogenase mediated reactions in nitrogen fixing microorganisms which reduce nitrogen to ammonia or as simple as the organic acid secretion by phosphate dissolving bacteria. Complex polysaccharides are broken down by certain fungi and actinomycetes and a variety of nitrogen inter-conversions takes place in soil due to microorganisms involved in the nitrogen cycle.

12.10.9 Advantages of Biofertilizers

The benefits of biofertilizers are as follows:

1. It is a low-cost fertilizer and can be used by small and marginal farmers.
2. It increases the soil fertility and is eco-friendly.
3. Cyanobacteria present in the biofertilizer provides the necessary growth promoting substances such as amino acids, proteins, vitamins, etc., to the plants. They add adequate amounts of organic matter in soil.
4. Cyanobacteria can grow and multiply under wide range of pH of 6–8.5. Therefore, they can be used as the possible tool to reclaim saline or alkaline soil because of the ameliorating effect on the physico-chemical properties of soil.
5. Azobacter and azospirillum, besides supplying nitrogen to soil, secrete antibiotics which act as pesticides.
6. Biofertilizers increase physico-chemical properties of soils such as soil texture, water-holding capacity and pH by providing several nutrients and sufficient organic matter.

12.11 Biosurfactants

The commercially available surfactants are the products of synthetic chemistry (Table 12.5). In recent years, there is growing demand for the use of ‘biosurfactant’.

Biosurfactants are used for emulsification, increased detergency, wetting and phase dispersion as well for solubilization.

Hydrocarbons, which include petroleum products, oil products and halogenated compounds, form an important class of pollutants on a global scale. The presence of these hydrocarbons in the environment is of considerable public health and ecological concern, because of their persistence, toxicity and ability to bioaccumulate. More recently, one technology that is receiving increasing attention is ‘bioremediation’. This is natural process by which microbes modify and breakdown contaminants into other less toxic compounds, and ultimately carbon dioxide and water.

One of the problems associated with the bioremediation of hydrophobic compounds, which include petroleum hydrocarbons, is that they bind to soil particles and have limited solubility in water, resulting in limited availability to soil microorganisms, which in turn can retard and or stop the degradation process. It is generally assumed that by increasing the availability of the hydrophobic compounds to the microorganisms, the rate of bioremediation can be increased.

One method which has been investigated to address the problem is the use of 'surfactants' and 'emulsifiers'. These chemicals increase the solubility and dispersion of hydrophobic compounds. A class of surfactants which has received attention are those produced naturally by microorganisms: they are called 'biosurfactants'

The biosurfactants may be more ecologically acceptable and, unlike many synthetic surfactants they are biodegradable.

The widely used surfactants are 'glycolipids'. An example of a biosurfactant is rhamnolipid, as shown in Fig. 12.5.

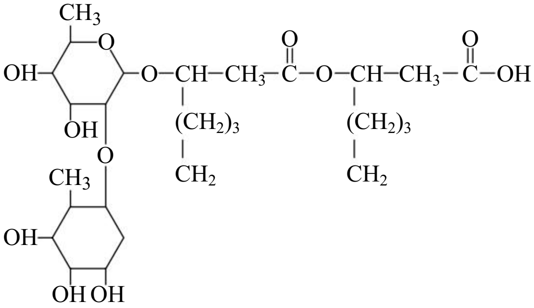


FIGURE 12.5 *Rhamnolipid R, a microbially produced biosurfactant.*

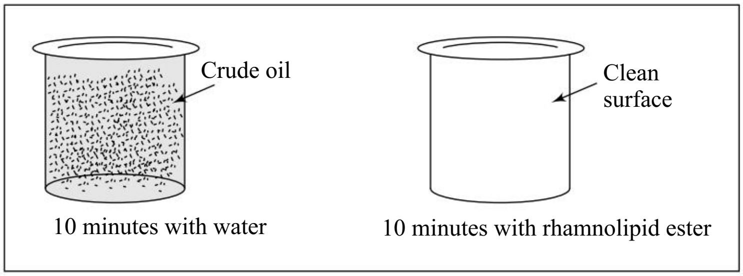


FIGURE 12.6 *Microbial products clean up oil spills. A biosurfactant glycolipid emulsifier improves removal of crude oil from the surface of the container.*

These compounds have distinct 'hydrophilic' and 'hydrophobic' regions and the final compound structures and characteristics depend on the structure growth conditions and carbon source used. Good yields are often obtained with insoluble substrates. These biosurfactants are excellent dispersing agents as shown in Fig. 12.6.

Microorganisms	**Biosurfactants**
Bacillus subtilis	Surfactin
Bacilluspumilus A1	Surfactin
Bacillus *sp*. C-14	Hydrocarbon –lipid-protein
Bacillus *sp*. AB-2	Ramnolipids
Coryne bacterium insidiosum	Phospholipids
Pseudomonas aeruginosa	Rhamnolipid

Table 12.5

A list of various biosurfactants produced by microorganisms

12.12 Biosensors

A biosensor is the term used for a whole class of sensors that utilize a biochemical reaction to determine a specific compound. A biosensor is generally an immobilized enzyme column or cell that is combined with a transducer to monitor a specific change in the micro-environment. The probe tip is immersed in the liquid phase and is in contact with the process either directly or through a membrane.

In analytical and clinical chemistry, the enzyme polymer conjugates are used extensively. Enzyme electrodes are a new type of detector or biosensor designed for the potentiometric or amperometric assay of substrates such as urea, amino acids, glucose, alcohol and lactic acid. Biosensors are finding increasing applications in medicine, industrial microbiology and environmental monitoring.

Microbial sensors are suitable for the industrial processes because they are stable for along time. Two different types of microbial sensors were developed for the measurement of organic compounds.

Microbial sensors consisting of immobilized whole cells and an oxygen probe when used for determination of substrates and products. The concentration of compounds was determined from the microbial respiration activity which could be directly measured by an oxygen probe.

Microbial sensors consisting of immobilized microorganisms and an electrode was used for the determination of organic compounds. The concentration of compounds was indirectly determined from electroactive metabolites such as proton, carbon dioxide, hydrogen, formic acid and reduced co-factors which can be measured by the electrode.

In a biosensor, a biomolecule or whole microorganism carries out a biological reaction and the reaction products are used to produce an electrical signal.

The biosensor design process is schematically illustrated in Fig. 12.7.

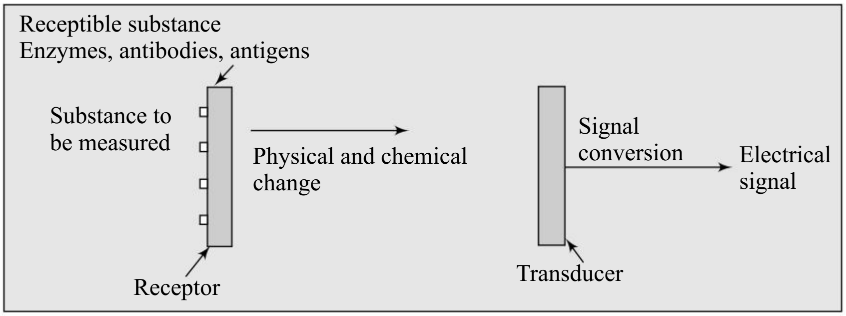


FIGURE 12.7 *A biosensor design.*

In a typical biosensor, the electrode is in close contact with a thin permeable enzyme membrane, which is capable of reacting specifically with the given substrates (Fig. 12.8).

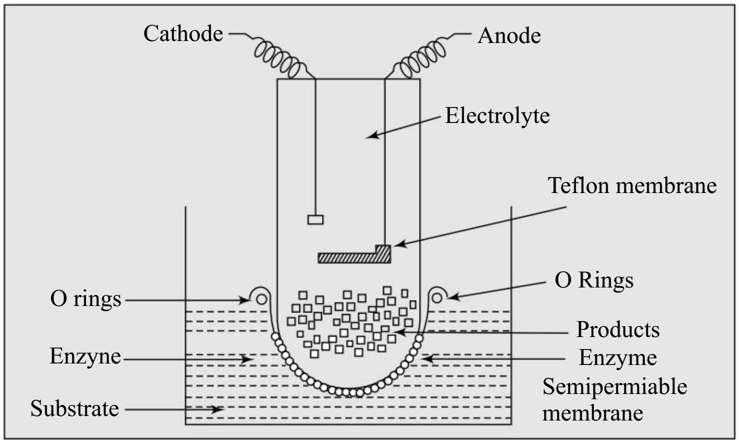


FIGURE 12.8 *A typical biosensor.*

The embedded enzymes in the membranes produce oxygen, hydrogen ions, ammonium ions, carbon dioxide or other small molecules depending on the enzymatic reactions occurring, which are readily detected by the specific sensor, the magnitude of the response determine the concentration of the substrate.

Although, biological component in a biosensor may be more often to be an enzyme or multi-enzyme system or a microbial cell, it is possible to measure the concentration of substances from many different environments.

A brief discussion of the following biosensors is given as:

12.12.1 Alcohol Biosensor

Continual measurement of ethyl alcohol concentration in culture broth is required for fermentation industries. Furthermore, in the cultivation of yeasts using sugar as a carbon source, it is well known that ethyl alcohol as a by-product decreases the sugar basis yield of whole cells.

It is well known that many micro-organisms utilize alcohols as carbon source. As similation of alcohols by micro-organism can be determined from the respiratory activity of micro-organisms. The respiration activity is directly measured with an oxygen electrode. Therefore, it is possible to construct a microbial sensor for alcohols using immobilized micro-organisms and an oxygen electrode.

- An alcohol biosensor system consists of the following components
- Immobilized yeast or bacteria (trichosporonbrassicae)
- A gas permeable membrane (teflon)
- An oxygen electrode

12.12.2 The Glucose Sensor

The determination of glucose is important for the process control. Assimilation of glucose by microorganisms can be determined by an oxygen electrode because respiration activity increases after assimilation of organic compounds. Therefore, it is possible to construct a microbial electrode sensor for glucose using immobilized whole cells which utilize mainly glucose and an oxygen electrode.

A microbial electrode consisting of immobilized whole cells of pseudomonasfluorescens and an oxygen electrode was developed for the determination of glucose. The microbial sensor for glucose was applied to molasses. The concentration of glucose was determined both by the microbial sensor and by the enzymatic method.

12.12.3 Enzymatic Biosensors

Another important enzyme biosensor, very useful in food industry is monoamine oxidase electrode sensor. It is used in freshness testing of meat. Various compounds such as amines, carboxylates, aldehydes, ammonia and carbon dioxide are produced in the meat putrefaction process.

Monoamine oxidase (monooxydoreductase from aspergillius niger) has been immobilized in a collagen membrane. The biosensor, consisting of a monoamine oxidase-collagen membrane and an oxygen electrode, could be developed and used to determine monoamines in meat and meat products by determining oxidation of monoamines with dissolved oxygen in the presence of monoamine oxidase.

12.12.4 Applications of Biosensors

Biosensors find applications in the following:

1. Clinical diagnosis and biomedical monitoring.
2. Agricultural, horticultural and veterinary analysis.
3. Detection of pollution and microbial contamination of water.

4. Fermentation analysis and control.
5. Monitoring industrial gases and liquids.
6. Measuring toxic gases in mining industry.
7. Direct biological measurement of flavours, essences and pheromones.

12.13 Biomembranes

Every cell is surrounded by plasma membrane, which creates a compartment where functions of life can proceed in relative isolation from the world. The plasma membrane keeps proteins and other essential materials inside the cell. But the plasma membrane is not simply an inert barrier. Proteins embedded in it work to bring nutrients into cell and extrude waste products. Other membrane bound proteins may sense the cell surroundings, communicate with other cells or act to move the cell to a new location. In aerobic bacteria, plasma membranes house the electron-transfer reactions that provide the cell with energy.

Although membranes from different organelles or cells have different activities, they share the following properties:

1. Biological membranes are impermeable to polar molecules or ions.
2. Membranes are not rigid but have adapted flexibility to changes in cellular or organellar shape and size.
3. They are durable.
4. Membranes contain proteins that are not simply structural in nature, but have a variety of enzymatic activities.

Biological membranes consist primarily of proteins and lipids. The relative amounts of these materials vary considerably, depending on the source of the membrane. At one extreme, the inner mitochondrial membrane is about 80 percent protein and 20 percent lipid by weight; at the other, mycelin sheath membrane of human erythrocytes contain about equal amounts of protein and lipid. Many membranes also contain small amounts of carbohydrates.

Biological membranes provide another example of an important biological structure stabilized by multiple noncovalent bonds. All biological membranes have a hydrophobic core, and they all separate two aqueous solutions. The plasma membrane, for example, separates the interior of the cell from its surroundings. Similarly, that surrounds the organelles of eukaryotic cells separate one aqueous phase— the cell cytosol—from one another—the interior of the organelle. Although biological membranes contain proteins that give each type of membrane its unique properties, all biomembranes are built of the same kinds of phospholipid molecules linked together by multiple noncovalent interactions.

12.13.1 Chemical Aspects of Biomembrane

The chief constituents of biomembrane are phospholipids. The fatty acid molecules are the key components containing a long hydrocarbon chain attached to a carbonyl group [–COOH]. The fatty acids differ in length and in the extent and position of their double bonds. Most fatty acids in cells have 16, 18 or 20 carbon atoms and 0–3 double bonds (unsaturated).

These fatty acids are insoluble in water and salt solutions. The lipid in membranes contains long chain fatty acyl groups, but these are linked (usually by an ester bond) to small highly hydrophilic groups. Consequently, membrane lipids, unlike tristearin, do not clump together in droplets but orient themselves in sheets to expose their hydrophilic ends to the aqueous environment. Molecules with water and other end (the tail) is hydrophobic are said to be amphipathic.

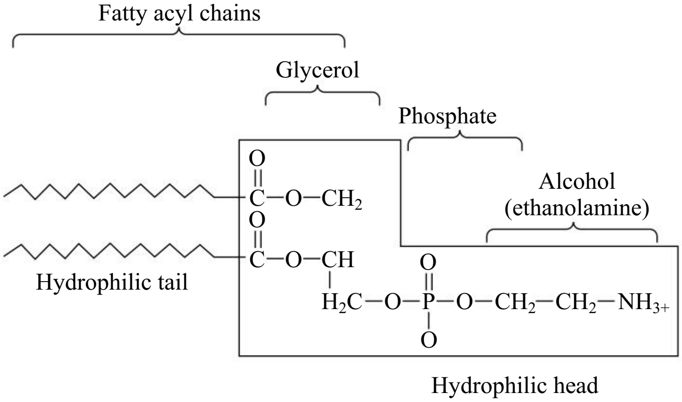


Figure 12.9 *Chemical aspects of biomembrane.*

The tendency of amphipathic molecules to form organized structures spontaneously in water is the key to the structure of cell membranes which typically contain a large proportion of amphipathic liquids.

The phospholipids are the most abundant amphipathic lipids. Phosphoglycerides, a principle class of phospholipids, contain fatty acyl side chains esterified to phosphate. The simplest phospholipid, phosphatidic acid contains only these components. In most phospholipids, however, the phosphate group on a hydrophilic compound such as ethanolamine HO–CH_2–CH_2–NH_2 or serine or choline or glycerol so that the hydrophilic head group is somewhat larger (see Fig. 12.9). Both of the fatty acyl side chains may be saturated or unsaturated. The negative charges on the phosphate as well as the charged groups or hydroxyl groups on the alcohol esterified to it interact strongly with water. Phospholipids assume different forms in aqueous solutions: micelles, bilayer sheets and liposomes (Fig. 12.10).

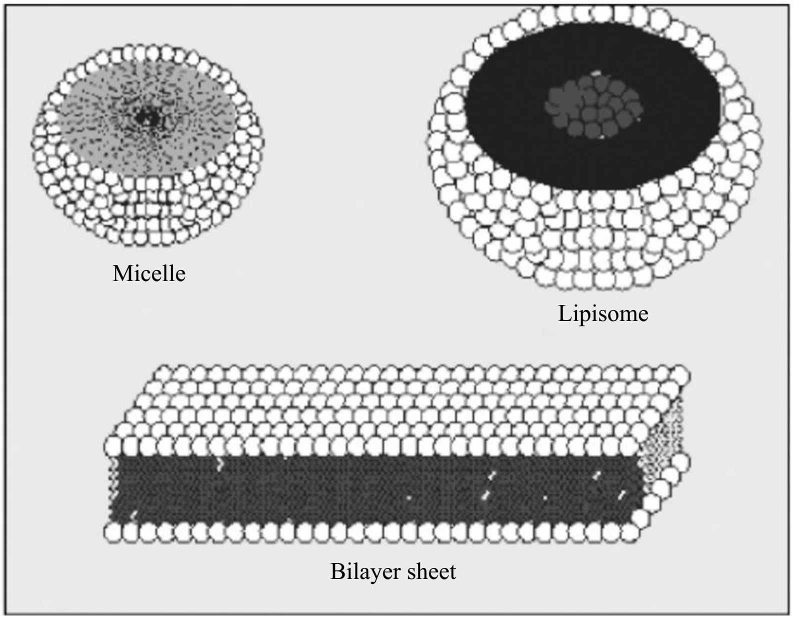


Figure 12.10 *Different forms of phospholipids configurations in aqueous solution.*

Review Questions

1. What is biotechnology?
2. Enumerate the new trends, scope and importance of biotechnology.
3. What is fermentation process?
4. Describe alcoholic fermentation through Embden–Meyerhof–Parnas pathway.
5. How ethyl alcohol is biosynthesized from starch?
6. Describe the processes of fermentation to produce ethyl alcohol?
7. What are the renewable source materials used for the commercial production of ethyl alcohol? Give a flow chart showing the various steps involved in the manufacture of alcohol.
8. Describe with diagram how acetic acid is produced biosynthetically.
9. How acetone is obtained by a bio process from glucose or fatty acid?
10. Describe the pathway adopted for the production of lactic acid.
11. Write a note on release of ammonia by bacteria.
12. What is an enzyme? What are their salient features?
13. Give consolidated applications of industrial enzymes.
14. Trace the importance of few enzymes used in starch processing for textiles.
15. What are food beverages? Why are fermented foods and beverages preferred?
16. How are beverages classified? Give examples in each case.
17. What are alcoholic beverages?
18. Describe how wine is produced by fermentation process.
19. Describe how is beer produced from starchy cereals.
20. Write note on nonalcoholic beverages.
21. Write a note on malt and juice beverages.
22. Write a note on milk and other dairy beverage products.
23. Write a brief note on tea and coffee beverages.
24. What is an animal feed? How they are classified?
25. Give an account of animal feed enzymes.
26. Discuss animal feeds.
27. What is a biofuel?
28. What is a biomass? Enumerate the advantages and disadvantages.
29. What is bioconversion?
30. What are the various forms of biofuels?
31. Write a note on biogas.
32. How biodiesel is obtained from rapeseed oil?
33. Describe how biofuel can be obtained from energy crops?
34. Write note on cellulose as a source of biofuel.
35. What is a biofertilizer?
36. Write a note on biofertilizers?
37. What is the need for biofertilizers?
38. Mention the advantages of biofertilizers.
39. How are biofertilizers classified?
40. Write a note on symbiotic nitrogen fixers.
41. Write a note on asymbiotic nitrogen fixers.
42. Write a note on soluble phosphate biofertilizers.
43. Write notes on (i) azosporillum (ii) azolla biofertilizers.
44. What is a biosurfactant? Illustrate with an example, how it works.
45. Mention a few biosurfactants.
46. What is a biosensor? Explain a biosensor design.
47. Construct and explain the working principle of a biosensor.
48. Give the applications biosensors.
49. Write a note on enzyme biosensor.
50. Describe the construction and working of glucose biosensor.
51. What is an alcohol biosensor? Explain.

Environmental Chemistry

13

Chapter Outline

Introduction. Environmental pollution, causes of pollution. Types of pollution, air pollution, source of air pollution. Nature of pollution sources, impacts of air pollution. Classification of air pollutants, sources, ill effects and remedies of air pollutants, sulphur dioxide, carbon monoxide, organic vapours, oxides of nitrogen. Particulate matter, control of particulate matter. Acid rain, harmful effects of acid rain, conservation of resources to minimize acid rain. Ozone depletion, causes of ozone depletion, protecting the stratospheric ozone layer. Global warming, greenhouse effect, greenhouse gases, global climate trend, control measures. Water pollution—classification of water pollution, sources of water pollution, ill effects of heavy metals on water. Sewage and its characteristics, sewage treatment. Land pollution, sources and methods. Waste disposal. Toxic chemicals—List of dangerous chemicals. Thermal pollution, sources and methods. Nuclear pollution, sources and methods. Noise pollution, sources and methods.

13.1 Introduction

Environment is the surrounding in which we live. Environmental chemistry is a science that deals with the chemical constitution and composition of the environment, the inter and intra interactions between different species in the environment and the cause and effect of changes in the environment on the life cycle of the species.

Our environment is a dynamic wonderful life-sustaining system. Various forms of life, which are independent and interdependent and the large numbers of natural processes, are in equilibrium. Many of the needs of man are met by nature. Nature has provided him with good drinking water, pure air to breath and ozone layer to protect him from harmful UV and cosmic rays. The green vegetation around him absorbs the carbon dioxide he breathes out and converts it to oxygen required to sustain life.

Man is always greedy. In his attempt to keep his comfort he has exploited nature too much, disobeyed natural laws and is in the act of destroying the wonderful environmental equilibrium. Rapid increase in population and industrialization are the two main reasons of environmental pollution. Man has overtaken natural causes of pollution like volcanic eruption, forest fire, etc. His polluting capacity has reached a level of more than a thousand times that of natural causes. Maximum danger is produced by highly industrialized countries. Automobiles which use fossil fuel are another highly polluting device man has invented. Automobiles have become automatic mobile polluting stations. Environmental Pollution is multidisciplinary subject involving chemistry, physics, life science, medical science, engineering and agricultural science.

Environmental pollution causes various problems to the living systems. Some of the water pollutants endanger fauna and flora and also aquatic life. Atmospheric and water pollution produces lung problems, reduces visibility, impair health and even causes death. Pollution also destroys historical monuments and other manmade and natural materials. Hence, it has become mandatory on each individual human being to help adopting pollution control methods. It is already late but not too late to repent.

Vertical section of air may be divided into three layers: (1) troposphere which extends up to 18 km, (2) stratosphere extending from 18 to 150 km and (3) ionosphere—above 150 km. Air pollutants concern mainly with troposphere. The densest part of air is troposphere and it contains approximately 78.09 percent N_2, 20.94 percent oxygen, 0.93 percent noble gases, CO_2, CH_4, NO, CO, ozone, water vapour and dust as minor components.

13.2 Environmental Pollution

Environmental pollution is defined as *the excessive discharge or addition of undesirable foreign substances into the environment, which adversely affect the natural quality of the environment causing damages to the living conditions of human beings, animals and plants.*

13.2.1 Causes of Pollution

Before industrialization started, most of the pollutions were from natural causes like volcanic eruption, forest fire, dust storm, etc. Pollution was caused by human beings, and other living organisms through their excreta, death and decay. Most of these pollutants are biodegradable. Now desire of man to have comforts in life has increased the extent of pollution by leaps and bounds. Most of the plastics commonly used in various forms, automobile exhaust, industrial waste, pesticides and herbicides from farm land are some of the main sources of pollution as shown in Fig. 13.1. The causes of pollution may be listed as follows.

1. Rapid growth in population
2. Rapid industrialization involving pollution of earth, water and air
3. Rapid urbanization
4. Misuse of natural resources
5. Natural phenomena like volcanic eruption
6. All transportation mechanisms that use fossil fuels

13.3 Types of Pollutions

The major types of pollutions are as follows:

- Air pollution
- Water pollution

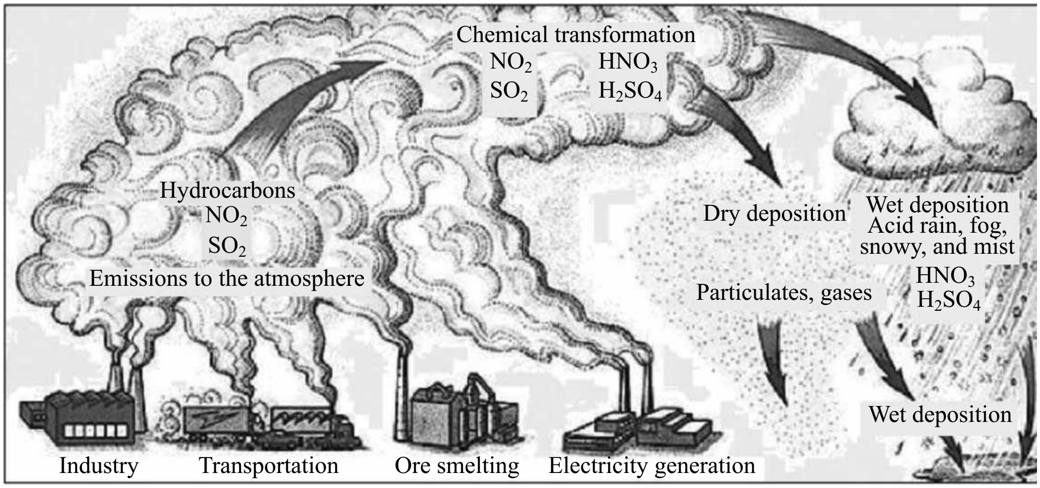


FIGURE 13.1 *An illustration of polluting environment of various forms.*

- Land/soil pollution
- Nuclear pollution
- Thermal pollution
- Noise pollution

13.3.1 Air Pollution

Air pollution can be defined as *the extensive discharge of undesirable foreign substances into air that adversely affect living conditions of human beings, animals and plants and also cause damages to property*.

Air pollution is an atmospheric phenomenon that has no natural cause. The largest sources of pollution are found in the industrialized countries, because the polluting substances can spread themselves easily in the atmosphere. The industry consists of a large number of processes where several substances are released that causes air pollution as shown in Fig. 13.1.

Air pollution is an indication of disturbances to the composition of compounds in the atmosphere and it may be summarized as follows:

- Excess emission of gases/vapours into the atmosphere
- Saturation of chemical compounds/particulates
- Emergence of new chemical reactions of reactive and nonbiodegradable compounds
- Global warming, acid rain, smog and ozone depletion are some effects of air pollution

13.3.2 The Sources of Air Pollution

Air pollution may be personal, occupational or community pollution.

- *Personal pollution* is due to individuals exposing themselves to pollution, e.g. cigarette smoking (Fig. 13.2).

FIGURE 13.2 *Smoking of cigarette by humans.*

- *Occupational pollution* is due to the pollution in the work place, e.g. coal dust and cement dust in cement factories accumulates as polluted cloud as shown in Fig. 13.3.
- Community pollution involves a variety of pollutants from the chimney exhausts from factory gases, effluents from automobile exhausts, etc. (Fig. 13.4).

FIGURE 13.3 *Cloud of smoke and dust.*

FIGURE 13.4 *Chimney exhausts.*

The major sources to air pollution are due to the following categories:

- Motor vehicle exhaust
- Heat and power generation facilities
- Industrial processes
- Auto manufacturing
- Fertilizer plants
- Building demolition
- Solid waste disposal
- Volcanic eruption
- Fuel production
- Roadway construction
- Electrical components manufacturing
- Solvent evaporation
- Forest fires

Air pollution is the accumulation in the atmosphere of substances that, in sufficient concentrations, endanger human health or produce other measured effects on living matter and other materials. The six major common types of pollutants are carbon monoxide, hydrocarbons, nitrogen oxides, particulates, sulphur dioxide, and photochemical oxidants. Among the major sources of pollution are power and heat generation, the burning of solid wastes, industrial processes, and, especially, transportation, etc.

Air pollution is the presence of substances in air that on itself or with other substances has disadvantageous effects on the health of people, animals or plants. If the change of composition of air damages the environment, it is called air pollution. Because of a large mixture of processes in the industry, therefore a large number of air pollutants are released in the atmosphere as shown in Fig. 13.5. Air pollution caused by air pollutants has bad consequences on the health and the environment. Some polluting substances having similar functioning (NO_x and SO_2) contributes to acidification. The small number of polluted substances are carbon dioxide (CO_2), nitrous-oxide (N_2O), carbon monoxide (CO), sulphur dioxide (SO_2), methane (CH_4), volatile organic compounds (VOC), hydrogen disulphide (H_2S) and ammonia (NH_3).

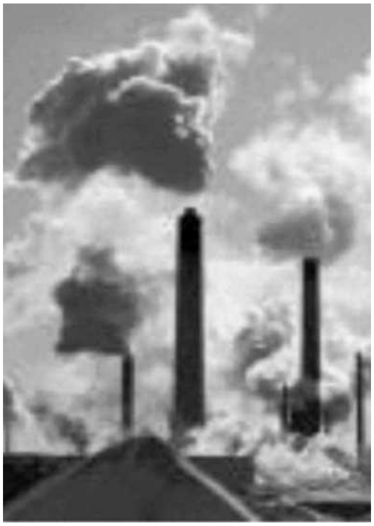


FIGURE 13.5 *Industrial gaseous waste.*

The chemicals that cause air pollution and are bad for the environment and people are as follows:

- *Ozone*: Ozone is produced when other pollutant chemicals combine. It is the basic element of smog. It causes many different kinds of health issues dealing with the lungs. It can damage plants and limit sight. It can also cause a lot of property damage.
- *VOCs* (volatile organic compounds, smog formers): VOCs are let into the air when fuel is burned. This chemical can cause cancer. It can also harm plants.

- *NO_x* (nitrogen dioxide): This chemical forms smog. It is also formed by burning sources of energy, such as gas, coal and oil, and by cars. This chemical causes problems in the respiratory system (including the lungs). It causes acid rain, and it can damage trees. This chemical can eat away buildings and statues.
- *CO* (carbon monoxide): The source of this chemical is burning sources of energy. It causes blood vessel problems and respiratory failures.
- *PM-10* (particulate matter): The source of this chemical is ploughing and burning down of fields. It can cause death and lung damage. It can make it hard for people to breathe. The smoke, soot, ash and dust formed by this chemical can make many cities dirty.
- *Sulphur dioxide*: This chemical is produced by making paper and metals. This chemical can cause permanent lung damage. It can cause acid rain which kills trees and damages building and statues.
- *Lead*: This chemical is in paint, leaded gasoline, smelters and in lead storage batteries. It can cause many brain and nerve damages and digestive problems.

13.3.3 Nature of Pollution Sources

Air pollution can be manmade or naturally occurring. The main sources of air pollution are as follows:

- Stationary and area sources
- Mobile sources
- Agricultural sources
- Natural sources

(i) Stationary and area sources

A stationary source of air pollution refers to an emission source that does not move (i.e. utilities, chemical and manufacturing industries). Often stationary sources are defined as large emitters who release relatively consistent qualities and quantities of pollutants (Fig. 13.6). The term area source is used to describe the many smaller stationary sources located together whose individual emissions may be low but whose collective emissions can be significant.

Figure 13.6 *Accumulation of air pollutant as cloud.*

(ii) Mobile sources

A mobile source of air pollution refers to a source that is capable of moving under its own power. In general, mobile sources imply to on-road transportation (Fig. 13.7). In addition, there is also a non road or off-road category that includes gas-powered lawn tools and mowers, farm and construction equipment, recreational vehicles, boats, planes and trains.

(iii) Agricultural sources

A wide range of contaminants can reach the river either via groundwater or through drainage ditches, including artificial fertilizer residues, insecticides, herbicides, pesticides and farmyard waste, all of which are potentially very harmful. Accidental milk spillage from dairies is a serious contaminant.

Undiluted animal manure (slurry) is one hundred times more concentrated than domestic sewage, and can carry a parasite, cryptosporidium, which is difficult to detect. Silage liquor (from fermented wet grass) is even stronger than slurry, with a low

Figure 13.7 *Atmospheric pollution from moving vehicles.*

pH and very high BOD (biological oxygen demand). With a low pH, silage liquor can be highly corrosive; it can attack synthetic materials, causing damage to storage equipment and leading to accidental spillage.

Agricultural operations, those that raise animals and grow crops, can generate emissions of gases, particulate matter and chemical compounds (Fig. 13.8). For example, animals confined to a barn or area (rather than field grazing), produce large amounts of manure. Manure emits various gases, particularly ammonia into the air. This ammonia can be emitted from the animal houses, manure storage areas or from the land after the manure is applied. In crop production, the misapplication of fertilizers, herbicides and pesticides can potentially result in aerial drift of these materials.

Figure 13.8 *Spraying insecticide, pesticide or fertilizer during agricultural operations.*

(iv) Natural sources

Natural sources of air pollution are sources not caused by people or their activities. An erupting volcano (Fig. 13.9) emits particulate matter and forest fires (Fig. 13.10) can emit large quantities of pollutants, plants and trees emit hydrocarbons and dust storms can create large amounts of particulate matter. Wild animals in their natural habitat are also considered natural sources of pollution given that there is a certain amount of natural pollution, it is very important to control the 'excess' pollution caused by man's activities.

Figure 13.9 *Volcano pollution.*

Figure 13.10 *Forest fire pollution.*

13.3.4 Impacts of Air Pollution

Pollutions are due to atmospheric deposition of nitrogen (NO_x) and other chemical contaminants. These pollutants dramatically impact the watershed as shown in Fig. 13.1.

The effects of nitrogen can be seen in the following:

(i) *Acid rain*: Nitrogen oxide (NO_x) is one of the key air pollutants which causes acid deposition, and results in adverse effects on aquatic and terrestrial ecosystems (Fig. 13.11). Acid deposition increases the acidity of water and soils. Increases in water acidity can impair the ability of certain fish and aquatic life to grow, reproduce and survive. Increases in soil acidity can impair the ability of some types of trees to grow and resist disease.

(ii) *Smog*: It is a collection of pollutants. It is formed by NO_x, particulate matter and humidity, all mixed together. Smog reduces how far and how clearly we can see through the air, an effect called visibility reduction or regional haze. General atmospheric conditions are that haze and ozone occur at the same time and the mix of the two is called smog (Fig. 13.12).

Figure 13.11 *Formation of acid rain cloud due to the presence of CO_2, SO_2, NO_x,etc. with moisture.*

Figure 13.12 *Smog formation in the atmosphere.*

(iii) *Eutrophication*: Reduced levels of dissolved oxygen in water due to increased mineral and organic nutrient deposits produce algae and other water plants that choke other forms of life in the oxygen competition. Soil erosion, phosphorous and direct runoff from feedlot operations and intensive agriculture are the main cause.

When excess nitrogen causes accelerated growth of algae, the algae blocks sunlight, needed for submerged aquatic vegetation to grow, when the algae dies it sinks to the bottom and decomposes in a process which depletes the water of oxygen.

(iv) *Accumulation*: Nitrogen compounds percolate through soil and reach drinking water sources, and these nitrate contaminants pollute the water.

The effects of chemical contaminants can be seen in the following:

(v) *Bioaccumulation*: Chemical contaminants increase with concentration as it moves through the food chain. Example: an invertebrate eats contaminated algae, a small fish eats many contaminate invertebrates, a large fish eats many small fish, and humans eat the big fish which is now loaded with the chemical. Chemical contaminant can accumulate and bind to the sediments they deposit on. When this happens, the chemical contaminants are moved wherever the sediments are moved. Chemical contaminants can change a plant species composition, and make species more susceptible to disease, weather and insect damage. Changes such as these challenge a species ability to reproduce and develop.

(vi) *Persistent*: Chemical contaminants do not break down or diminish over time.

13.3.5 Classification of Air Pollutants

Air pollutants are classified into two types according to their origin.

(i) *Primary pollutants* which are directly injected into air, e.g. SO_2, CO, oxides of nitrogen, dust particles, etc.

(ii) *Secondary pollutants* which are not directly introduced into air but are found in the atmosphere by chemical or photochemical reactions among the constituent of air, e.g. ozone, PAN, etc.

Another method of classification of air pollutants is based on their physical state.

(i) *Gases*, e.g. CO, NO, hydrocarbon vapours, SO_2, etc.

(ii) *Aerosol particulates* which involve solids such as smoke and dust and liquids like fog, sprays, etc.

A third type of classification is based on the chemical composition of the pollutant.

(i) *Inorganic pollutants:* Oxides of N_2, CO, SO_2, H_2S, etc.

(ii) *Organic pollutants:* Hydrocarbons, aldehydes, ketones, etc.

13.3.6 Sources, Ill Effects and Remedies of Air Pollutants

There are six primary pollutants of air. They are (a) SO_2, (b) CO, (c) oxides of nitrogen, (d) volatile organic vapours, (e) particulate matter and (f) ozone.

Their natural and anthropogenic (manmade) sources, ill effects and remedial measures are given below.

13.3.6.1 Sulphur Dioxide

(a) *Natural source*: Volcanic eruptions produce gases containing SO_2. When plants decay they produce H_2S, which gets oxidised in air by ozone to form SO_2.

$$H_2S + O_3 \rightarrow H_2O + SO_4^{2-}$$

(b) *Anthropogenic sources:* Combustion of sulphur containing fossil fuels is the main manmade source. Roasting of sulphide ores of metals also produce SO_2.

$$2FeS + 3O_2 \rightarrow 2FeO + 2SO_2$$

The concentration of SO_2 in air always remained below the permissible level. The concentration of SO_2 is high only in low altitude regions due to anthropogenic forces.

(c) *Ill effects of SO_2*:

(i) *On human beings:* Concentration of SO_2 to about 20 ppm may cause eye irritation and affect digestive systems. Concentration above 400 ppm may be fatal.

(ii) *On plants:* Plants are damaged even at 1 ppm and for some plants yellowing of leaves takes place due to the presence of sulphur dioxide content in air.

(iii) *On materials:* Yellowing of paper, loss of strength due to corrosion of metals and loss of strength of buildings may be caused by SO_2. Acid rain is also caused by SO_2.

(d) *Control of SO_2 pollution*: The main source of SO_2 is the burning of coal. Control of SO_2 pollution may be done by passing the emitted gases through a slurry of limestone when SO_2 is absorbed. Air is blown through the slurry to convert calcium sulphite to calcium sulphate.

$$CaCO_3 + SO_2 \rightarrow CaSO_3 + CO_2$$

$$CaSO_3 + ½\, O_2 \rightarrow CaSO_4$$

In fluidized bed combustion finely powdered coal and limestone is fed into the chamber and fluidized by blowing air at about 1000°C. All the above changes take place in the chamber.

13.3.6.2 Carbon Monoxide

(a) *Natural source* by the oxidation of methane from marshy places by O_2.

$$CH_4 + \frac{1}{2}O_2 \rightarrow CO + 2H_2O$$

(b) *Anthropogenic source:*

(i) Incomplete combustion of carbon and carbon-containing fuels. CO is one of the major pollutants present in automobile exhaust gases.

(ii) Reaction of CO_2 with carbon at high temperature.

$$CO_2 + C \rightarrow 2CO$$

(c) *Ill effects on human beings:* 10 ppm is the admissible level of CO in air. Concentration above 110 ppm affects the respiratory system. At higher concentration it leads to headache, fatigue, unconsciousness and finally to death. This is caused because CO forms a stable complex carboxyhaemoglobin with haemoglobin of the blood. Haemoglobin loses its capacity to carry oxygen to the cells.

(d) *Control of CO Pollution:* The main source of CO pollution is automobile exhaust. These emissions are controlled by using a catalytic converter where CO is oxidised to CO_2.

$$CO + \frac{1}{2}O_2 \rightarrow CO_2$$

13.3.6.3 Organic Vapours

(a) *Natural source:* From marsh gas and from gases from petroleum wells.

(b) *Anthropogenic source:* The major source is from the unburnt gases of automobile exhaust from various industrial processes, perfumes we use in day-to-day life.

(c) *Ill effects:* On their own hydrocarbons do not produce any pollution. They are however converted to photochemical oxidants called PAN (peroxyacyl nitrate) by a series of reactions. PAN is the cause of photochemical smog.

(d) *Control of hydrocarbon pollution*: The main source of anthropogenic hydrocarbon pollution is exhaust gases of automobiles. Control can be done by having a catalytic converter which helps in the oxidation of hydrocarbons to CO_2 and water.

$$C_x\frac{H}{y} + \left(x + \frac{y}{4}\right)O_2 \rightarrow x\,CO_2 + \frac{y}{2}H_2O$$

13.3.6.4 Oxides of Nitrogen

The three oxides of nitrogen found in the atmosphere are (N_2O) nitrous oxide, (NO) nitric oxide and (NO_2) nitrogen dioxide.

(a) *Natural source:*

  (i) By biological oxidation of nitrogenous compounds in the soil N_2O and NO are formed. This easily gets oxidized to NO_2, in air.

  (ii) During lightning N_2 and O_2 of air combine to form NO which then gets oxidized to NO_2.

(b) *Anthropogenic source:* Internal combustion engines operate at high temperatures. N_2 and O_2 of the air in the engine form NO and then NO_2. Thus, the exhaust gases from IC engine contain NO and NO_2.

(c) *Ill effects of oxides of nitrogen*

On human beings: Causes eye irritation and respiratory problems, produces acid-induced irritation and lung cancer.

On plants: Nil

On materials: (i) Causes acid rain which damages old monuments and buildings. (ii) Causes formation of photochemical smog which leads to poor visibility of roads and landing difficulties to aircrafts.

(d) *Control of the oxides of nitrogen:* In the IC engines catalytic converter is used for the decomposition of the oxides of nitrogen to N_2 and O_2.

$$\underset{\text{(exhaust gas)}}{2NO_2} \xrightarrow{\text{Pt/Rh. catalyst}} N_2 + O_2$$

13.3.6.5 Particulate Matter

Particles (both solids and liquids) with a size around 10 μm form particulate pollution in air. They are also called aerosol. Dust and soot are solid sols while fog and mist form liquid aerosol.

(a) *Natural source* by dust storms, forest fire, etc.

(b) *Anthropogenic source:*

  (i) Soot is produced during the combustion of coal, wood, fuel oil, house and municipal garbage and tobacco smoking.

  (ii) Dust is produced during material handling and crushing, grinding of ores, etc. Dust is also produced during mixing and packaging of powdered substances like chemicals, flours, starches, etc. Construction of houses, roads, dams and mining also cause formation of dust. House cleaning, body sprays, spraying of crops, engine exhaust also produce aerosols. Flyash from thermal power plants and cement industries is another major source.

(c) *Ill effects of particulate pollution:* Smoke causes respiratory problems and may lead to TB. Silica dust causes asbestosis. Lead dust produces lead poisoning and mercury dust produces kidney problems.

13.3.6.5.1 Control of Particulate Matter

Various methods for the control of particulate matter in air are available depending on the size of particles, their physical nature and the economy involved.

(a) Fabric filter

Dust particles from dry gases can be removed using fabric filters. It consists of fibrous materials with fine pores suspended in a chamber (Fig. 13.13). The gas is allowed to pass through the filters when particulate matter is filtered at the bags. When the mechanical shaker is operated the collected dust particles fall down. Blowing air in the opposite direction can also be done to remove dust particles from the bags.

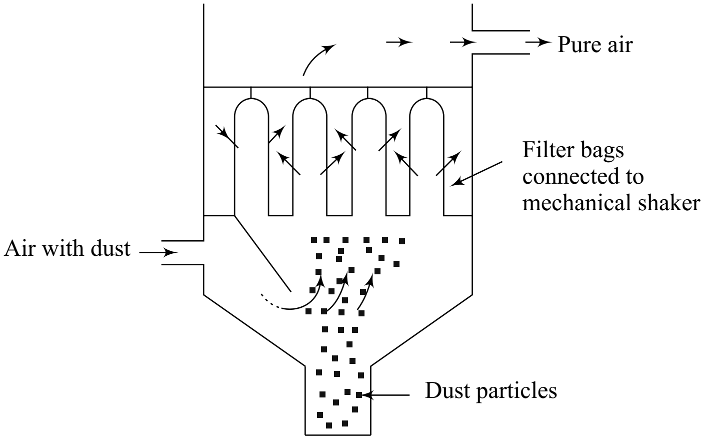


Figure 13.13 *Fabric filter.*

Fabric filters are simple to operate, and have high collection efficiency. But it can be used only for dry gases, requires maintenance and replacement of costly fabrics.

(b) Gravity separation

Air is allowed to flow through dust-settling tanks slowly in one direction. The suspended heavier particles settle down and collect at the conical portions of the tank. Purified air flows out (Fig. 13.14).

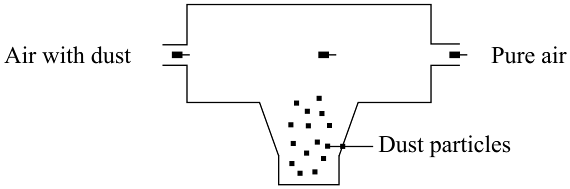


Figure 13.14 *Gravity Separator.*

(c) Cyclone separator (centrifugal separator)

The principle of operation of a cyclone separator is based on centrifugal force to separate dust particles from gas streams. The principle is shown in Fig. 13.15.

A cyclone separator consists of a cylindrical centrifugal chamber, inner tube, conical base and dust collector. The impure air containing dust particles is admitted into the chamber tangent to the surface. It is made to move in a spherical manner down the chamber and pass up in smaller spirals through the inner tube. Under the influence of the centrifugal force dust particles get separated and collect in the conical base. Purified air leaves the chamber through the top of the inner tube.

This method is simple, maintenance free and is economical. But it has low efficiency to smaller particles.

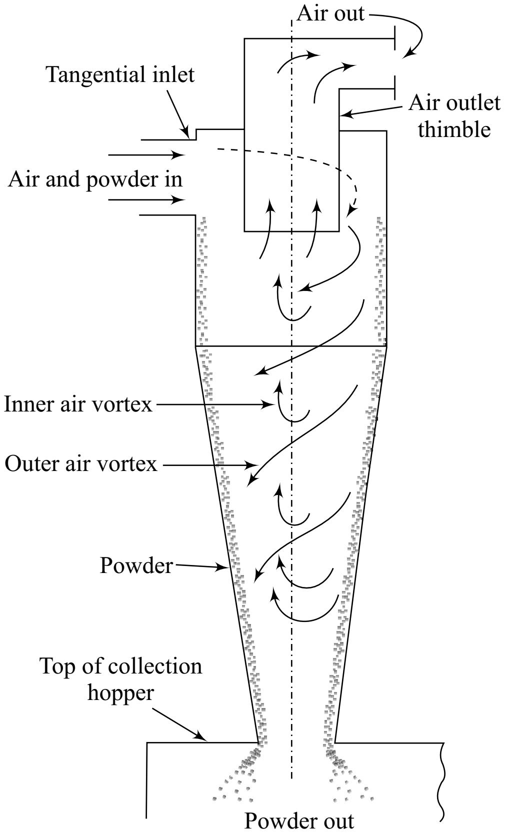


Figure 13.15 *A typical cyclone separator.*

(d) Wet scrubber

Impure air is allowed to ascend a tower and water is sprayed from the top (Fig. 13.16).

Descending water wets the dust particles and settles the particle at the bottom of the chamber as sludge. Scrubber also removes water-soluble gaseous impurities.

This method is also simple and removes both dust particles and water-soluble gases. It is highly efficient. But disposal of sludge becomes costly, large quantities of water is required and air pollution now becomes water pollution.

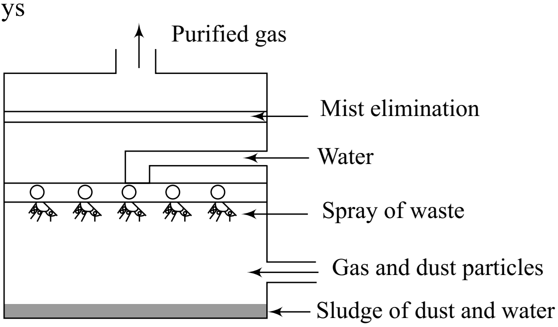


Figure 13.16 *Wet scrubber.*

(e) Electrostatic precipitator

Cottrell electrostatic precipitators are widely used in industries to remove particulate matter from gaseous effluents. They are connected to the exhaust outlets of chimneys Fig. 13.17. The exhaust gas stream of industrial plant is allowed to pass between metal electrodes maintained at a high potential of 50,000 volts.

The pointed electrodes develop high intensity current and produce coronas in which the gas molecules get ionized releasing electrons from the gases. The electrons so generated collide with the suspended particles present in the gas and the particles become negatively charged. These negative particles are attracted by the collection electrode. The discharged

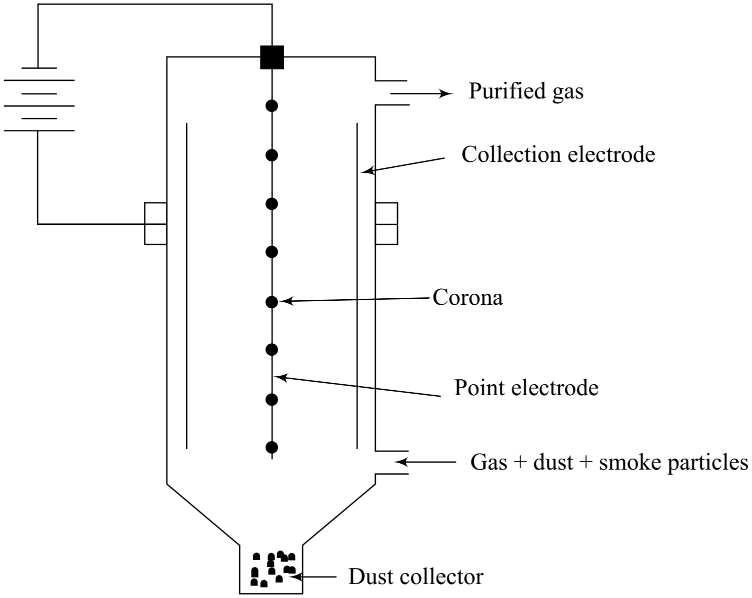


FIGURE 13.17 *An electrolytic precipitator.*

particles are removed by tapping or vibrations of the electrode. The particles collect in the conical collector. Dust-free gas flows out at the top.

The method has low maintenance cost with high collection efficiency, large quantities of gas can be cleaned. Initial cost is very high. It can be used only for particulate impurities, and gaseous impurities cannot be removed.

13.3.7 Acid Rain

Acid rain is a result of air pollution. When any type of fuel is burnt, a large number of different chemicals are produced. The smoke that comes from a fire or the fumes that come out of a car exhaust do not just contain the sooty grey particles that you can see—they also containa lot of invisible gases that can be even more harmful to our environment.

Power stations, factories and cars all burn fuels and therefore they all produce polluting gases. Some of these gases (especially nitrogen oxides and sulphur dioxide) react with the tiny droplets of water in clouds to form sulphuric and nitric acids. The rain from these clouds then falls as very weak acid—which is why it is known as 'acid rain'.

Rainwater is slightly acidic since it dissolves varying amounts of CO_2 from air. The lowest pH of rainwater is 5.6.

$$H_2O + CO_2 \rightarrow \underset{\text{carbonic acid}}{H_2CO_3}$$

In some places in western countries sometimes the pH of rainwater comes down to 4.5 and even up to 2.4. Such a rain is called acid rain (Fig. 13.18). Two main sources of acidity in rain are sulphur dioxide and oxides of nitrogen. When sulphur containing fossil fuels are burnt they produce SO_2 which in presence of rainwater and air forms H_2SO_4.

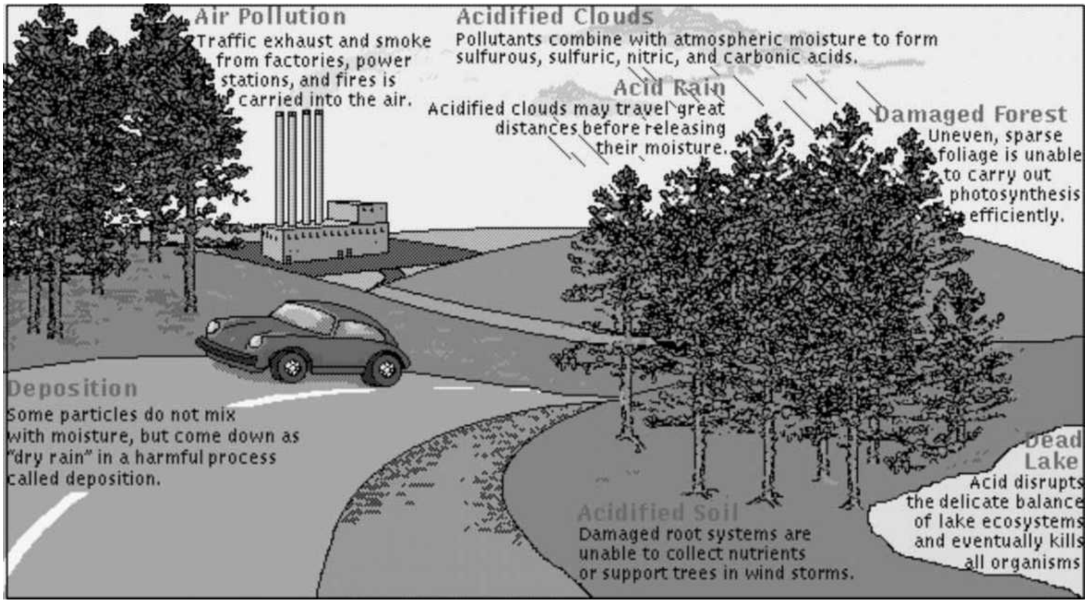


FIGURE 13.18 *An illustration of formation of acid rain and its ill effects.*

$$S + O_2 \rightarrow SO_2$$

$$H_2O + SO_2 + \frac{1}{2}O_2 \rightarrow H_2SO_4$$

Automobile exhaust contains nitric oxide. In contact with air it forms nitrogen dioxide which dissolves in water to form nitrous and nitric acids.

$$NO + \frac{1}{2}O_2 \rightarrow NO_2$$

$$2NO_2 + H_2O \rightarrow HNO_2 + HNO_3$$

1. Acid rain results in acidifications of streams and lakes. This endangers the life of aquatic animals and plants. Fishes are killed and their reproductivity is severely reduced. Growth of green algae is impaired. Many useful bacteria are killed and hence decomposition of organic matter in water is reduced making the water useless. Acid dissolves metals like aluminium, lead, zinc, iron, etc. increasing metal ion concentrations in lakes and stream. This again affects life span of aquatic living beings.
2. With decrease in pH, the nutrients like potassium, iron, magnesium, etc., are washed away from the soil. This affects the growth of plants (Fig. 13.19). The nitrifying bacteria present in the roots of leguminous plants are destroyed causing decrease in fertility of the soil.
3. Acid rain causes extensive damage and destruction to buildings as it leaches away the reinforcement.

FIGURE 13.19 *The growth of trees are affected by acid rain.*

4. Acid rain damages building materials made of marble, limestone, slate, metals, etc. The action of acid on marble or lime-stone makes the surfaces look pitted. This is called *stone leprosy*. The reaction is

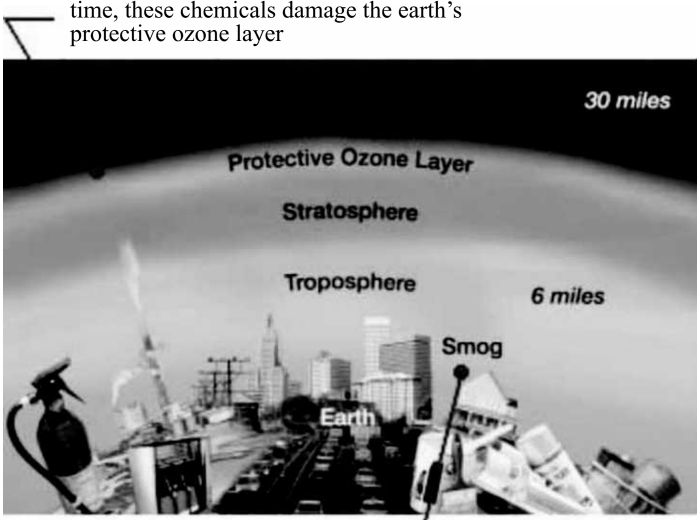
$$CaCO_3 + 2H^+ \rightarrow Ca^{+2} + H_2O + CO_2$$

13.3.7.1 Conserving Resources to Minimize Acid Rain

- Greater subsidies of public transport by the government to encourage people to use public transport rather than always travelling by car.
- Every individual can make an effort to save energy by switching off lights when they are not being used and using energy-saving appliances—when less electricity is being used, pollution from power plants decreases.
- Walking, cycling and sharing cars all reduce the pollution from vehicles.

13.3.8 Ozone Depletion

Ozone is a gas that occurs both in the earth's upper atmosphere and at ground level (Fig. 13.20). Ozone can be 'good' or 'bad' for your health and the environment, depending on its location in the atmosphere. Ozone occurs in two layers of the atmosphere. The layer closest to the earth's surface is the troposphere. Here, ground-level or 'bad' ozone is an air pollutant that is harmful to breathe and it damages crops, trees and other vegetation. It is a main ingredient of urban smog. The troposphere generally extends to a level about 6 miles up, where it meets the second layer, the stratosphere. The stratosphere or 'good' ozone layer extends upwards from about 6 to 30 miles and protects life on Earth from the sun's harmful ultraviolet (UV) rays.



Too little there…Many popular consumer products like air conditioners and refrigerators involve CFCs or halons during either manufacture or use. Over time, these chemicals damage the earth's protective ozone layer

Too much here… Cars, trucks, power plants and factories all emit air pollution that forms ground-level ozone, a primary component of smog.

FIGURE 13.20 *An illustration of ozone layers in different strata of atmosphere.*

In the stratosphere, ozone is formed from oxygen by photochemical means. Oxygen molecule absorbs UV light and dissociates to form oxygen atoms which combine with O_2 molecules to form ozone.

$$O_2 \rightarrow 2(O)$$

$$O_2 + (O) \rightarrow O_3$$

Ozone in turn absorbs short wave UV and decomposes to O_2 and O.

$$O_3 \rightarrow O_2 + (O)$$

Thus, a dynamic equilibrium exists between O_2 and O_3 and a constant concentration of ozone is maintained in the stratosphere. This ozone cover prevents harmful radiations to reach the earth's surface. The thin layer of ozone gas in the stratosphere (ozone layer) is shielding life on earth from the harmful UV light coming from the sun ('good ozone') (Fig. 13.21).

Ozone is harmful at the earth level, being very reactive and irritant to the human eyes (the so-called bad ozone).

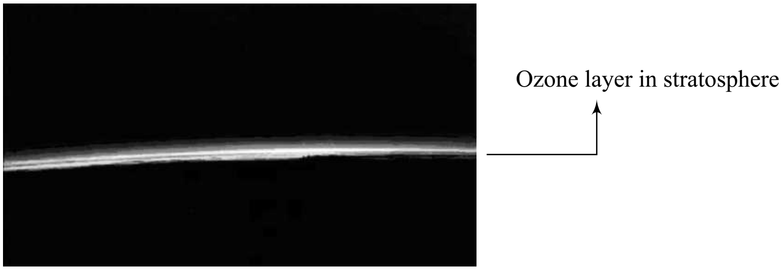


FIGURE 13.21 *Ozone layer shielding the earth at stratosphere.*

Ozone is produced naturally in the stratosphere. But this 'good' ozone is gradually being destroyed by man-made chemicals referred to as *ozone-depleting substances* (ODS), including chlorofluorocarbons (CFCs), hydro-chlorofluorocarbons (HCFCs), carbon tetrachloride and methyl chloroform. These substances were formerly used and sometimes still are used in coolants, foaming agents, fire extinguishers, solvents, pesticides, aerosol propellants, refrigerators and blowing agents for foams; the 'Halons', used for firefighting; methyl bromide, used in agriculture. Once released into the air these ozone-depleting substances degrade very slowly. In fact, they can remain intact for years as they move through the troposphere until they reach the stratosphere. There they are broken down by the intensity of the sun's UV rays and release chlorine and bromine molecules, which destroy the 'good' ozone. The overall amount of ozone is essentially stable in a natural cycle. This has been true for millions of years.

The most important disturbing chemicals are chlorofluorocarbons (CFC) which are mainly used as refrigerants, fire extinguishers, cleaning solvents. CFCs are inert, stable and nonharmful in the atmosphere. CFCs do not dissolve in rain: after several years, carried by the winds, they reach the stratosphere without being modified. Because of their inertness they do not have a natural sink and hence they slowly rise up to the stratosphere. Here their molecules are broken down by the intense UV light, and free chlorine atoms are created by this degradation. Each chlorine atom can destroy several thousands of ozone molecules before being removed from the atmosphere: chlorine is a catalyst for the ozone depletion. Bromine (e.g. from methyl bromide, used by farmers as a fumigant) is even more effective than chlorine.

$$CCl_2F_2 \rightarrow CClF^{\bullet}_2 + Cl$$

The Cl atoms catalyse the dissociation of ozone

$$Cl^{\bullet} + O_3 \rightarrow ClO^{\bullet} + O_2$$

$$ClO^{\bullet} + O \rightarrow Cl^{\bullet} + O_2$$

The overall reaction is

$$O_3 + O \rightarrow 2O_2$$

The regenerated chlorine atom causes further dissociation of ozone. Thus, ozone depletion (ozone hole) is produced in the stratosphere.

13.3.8.1 Causes of Ozone Depletion

Ozone depletion can cause increased amounts of UV radiation to reach the earth which can lead to more cases of skin cancer, cataracts and impaired immune systems. Overexposure to UV is believed to be contributing to the increase in melanoma, the most fatal of all skin cancers. UV light kills microorganisms and cause biological imbalance. UV can also damage sensitive crops, such as soybeans, and reduce crop yields. Some scientists suggest that marine phytoplankton, which are the base of the ocean food chain, are already under stress from UV radiation. This stress could have adverse consequences for human food supplies from the oceans.

According to scientists, certain manmade chemicals are major contributors to the problem. These chemicals are called *ozone-depleting substances* (ODS) and include many gases containing chlorine and bromine, such as: chlorofluorocarbons (CFCs, substances containing chlorine, fluorine and carbon) used in refrigerators and blowing agents for foams; the 'Halons', used for fire fighting; methyl bromide, used in agriculture.

Scientific community has woken up to the harmful effects of ozone hole. They have started replacing chlorofluorocarbons with hydrofluorochlorocarbon (HCFC) and hydrofluoroalkane (HFA). These compounds with one or more hydrogen atoms are biodegradable. But these substances are greenhouse gases. Hence scientists are returning to hydrocarbon gases like propane, butane, etc., as refrigerants.

13.3.8.2 Protecting the Stratospheric Ozone Layer

Ozone-destroying chemicals escape into the air and reach the stratosphere. Over time they reduce the layer of stratospheric ozone that protects us. When the protective ozone layer is damaged, there is an increase in harmful rays from the sun reaching the Earth. These rays can harm both health and the environment. Service stations must have special equipment that prevents release of refrigerant chemicals to the air when they are recharging car air conditioning systems.

Ozone can be good or bad depending on where it is located. Close to the earth's surface, ground-level ozone is a harmful air pollutant. Ozone in the stratosphere, high above the Earth, protects human health and the environment from the sun's harmful ultraviolet radiation. This natural shield has been gradually depleted by manmade chemicals.

Ozone in the stratosphere, a layer of the atmosphere located 10–30 miles above the Earth, serves as a shield, protecting people and the environment from the sun's harmful ultraviolet radiation. The stratospheric ozone layer filters out harmful sun rays, including a type of sunlight called *ultraviolet B*. Exposure to ultraviolet B (UVB) has been linked to cataracts (eye damage) and skin cancer. Scientists have also linked increased UVB exposures to crop injury and damage to ocean plant life.

In the mid-1970s, scientists became concerned that chlorofluorocarbons (CFCs) could destroy stratospheric ozone. At that time, CFCs were widely used as aerosol propellants in consumer products such as hairsprays and deodorants, and as coolants in refrigerators and air conditioners.

There is an urgent need to replace the CFCs and other ozone depleting substances with environmentally safe substances. Research studies are conducted for identifying the best alternative substances; presently HCFCs (hydrochlorofluorocarbons, substances containing hydrogen, chlorine, fluorine, carbon) are replacing CFCs, being much less harmful for the ozone layer. In the future, HCFCs will be phased out too.

The Clean Air Act includes other steps to protect the ozone layer. The act encourages the development of 'ozone-friendly' substitutes for ozone-destroying chemicals. Many products and processes have been reformulated to be more 'ozone-friendly'. For instance, refrigerators no longer use CFCs.

Sometimes it is not easy to phase out an ozone-destroying chemical. For instance, substitutes have not been found for CFCs used in certain medical applications. The limit on the production of methyl bromide, a pesticide, was extended because farmers did not yet have an effective alternative. Despite the inevitable delays because of technical and economic concerns, ozone-destroying chemicals are being phased out, and, with continued work, over time the protective ozone layer will be repaired.

13.3.9 Global Warming

Global warming is the increase of average world temperatures as a result of what is known as the greenhouse effect. It happens when greenhouse gases (carbon dioxide, water vapour, nitrous oxide and methane) trap heat and light from the sun in the earth's atmosphere, which increases the temperature. As the greenhouse gases build up in the atmosphere the earth gets hotter (Fig. 13.22). This hurts people, animals and plants. Many cannot take the change, so they die.

Figure 13.22 *An imaginative picture depicting global warming.*

Sun's radiations reaching the earth are made up of mostly visible radiation and small quantities of UV and infrared radiation. Earth's surface absorbs most of these radiations and reemits a part of this radiation. Thus, equilibrium exists between the radiations absorbed and those reflected back. This is the reason for the constancy of earth's temperature, conducive for the existence of living things.

The radiations reflected back by the earth is in the form of heat and its wavelength range is around 4000–5000 mm. This corresponds to IR radiation and hence it is called thermal IR. Some of the gases like CO_2, water vapour, methane and ozone absorb these thermal IR and reemit it to the earth's surface (like glass in a greenhouse) making the earth's surface warm. This phenomenon is called greenhouse effect or global warming. Only because of greenhouse effect the average temperature of the earth's surface is +15° and in the absence of the effect the temperature would have been –15°.

Major contributing factor for greenhouse effect is CO_2. A lot of CO_2 gets released into the atmosphere by combustion of fuels in various fields like thermal power plants, automobiles, etc. The increase in concentration of CO_2 is expected to increase the temperature of the earth by 0.4 percent every year. To aggregate the situation the greeneries of the earth which act as CO_2 links are being destroyed by man.

13.3.9.1 Greenhouse Effect

1. Due to increase in temperature of the earth's evaporation of surface water will increase water vapour content in the atmosphere and in turn increase the atmospheric temperature. This leads to the melting of ice in Arctic and Antarctic regions, increase in sea water levels and many cities on the sea shore may get submerged.
2. Global warming may result in warmer tropical oceans causing the formation of cyclones and hurricanes.

13.3.9.2 Greenhouse Gases

The main greenhouse gases are as follows: carbon dioxide (CO_2), methane, nitrous oxide (N_2O), chlorofluorocarbons (CFCs); also water vapour (H_2O) is producing the greenhouse effect.

Some gases in the atmosphere produce the 'greenhouse effect', trapping the heat of Earth without allowing it to escape in the outer space. The greenhouse effect is normally natural and beneficial: without it Earth would be at least 15°C colder; it is becoming more and more important due to the increasing concentration of these gases ('greenhouse gases') in the atmosphere due to human activity.

(1) Carbon dioxide

The concentration of the most important greenhouse gas, carbon dioxide, has increased in the atmosphere from 290 ppmv (parts per millions by volume) in 1880 to about 380 ppm in 2006, and is going to increase in the near future, because carbon dioxide, with water, is the final product of the combustion of fossil fuels (oil and derivatives, methane and hydrocarbons, coal), and of living and dead vegetation (biomass burning). The fossil fuels can be considered reservoirs of carbon, made ages ago; their combustion lets carbon return (as dioxide) into the atmosphere, increasing the greenhouse effect.

Carbon dioxide is easily soluble in water: the oceans contain enormous amounts of it, but the temperature increase (due to the greenhouse effect) reduces its water solubility, releasing new gas into the atmosphere, and accelerating the greenhouse effect.

(2) Methane

Methane is released during coal-mining activities, oil exploration and when vegetation is burnt during land clearance. The main source of methane though is agricultural activity. It is released from wetlands such as rice paddies and from animals, particularly cud-chewing species like cows. The problem with methane is that as the world population increases, agricultural activity must increase and so emissions of methane will also increase. Since the 1960s the amount of methane in the air has increased by 1 percent per year—twice as fast as the build-up of CO_2.

(3) Nitrous oxide

Nitrous oxide comes from both natural and manmade processes. Man-influenced sources, which represent about 45 percent of output to the atmosphere, are mainly fossil fuel combustion, as in power stations; use of nitrogenous fertilizers; burning rain forests and human and animal waste. N_2O contributes about 6 percent to the greenhouse effect at the moment.

(4) CFCs

CFCs found in fridges, air conditioners, aerosols, etc. are extremely effective greenhouse gases. Although there are lower concentrations of CFCs in the atmosphere than CO_2 they trap more heat. A CFC molecule is 10,000 times more effective in trapping heat than a CO_2 molecule, methane is about 30 times more effective. Methane molecules survive for 10 years in the atmosphere and CFCs for 110 years. It is this that causes people to want to ban them completely.

13.3.9.3 Global Climatic Trend

With the greatly improved methods compared to the past, there is a big increase in atmospheric CO_2 concentration due to human activity, with significant climatic consequences.

The globally averaged surface temperature is projected to increase by 1.4–5.8°C over the period of 1990–2100: the rate of warming, too, should increase compared to last century.

1. Heat waves, droughts, drier soils.
2. Increased evaporation and precipitation due to higher energy in the atmosphere: more frequent and extreme weather events (storms, tornados, hurricanes). Evidence is building up at an alarming rate.

3. Sea levels will go on raising: erosion of sandy beaches or flooding of coastal sea levels are already rising at a rate of 1–2 mm each year due to expansion of the top layer of the oceans as they warm and the melting of the polar ice caps. The predicted rise by 2050 is between 20 and 50 cm. This will cause increased flooding in coastal areas and river estuaries.
4. Easier transmission of some infectious diseases, including malaria and yellow fever.

The emissions of the most persistent greenhouse gases (carbon dioxide, nitrogen dioxide, perfluorocarbons) have a lasting effect on the climate, e.g. about a quarter of carbon dioxide persists in the atmosphere several centuries after the emission. Even if greenhouse gases concentration could be stabilized, the average surface temperatures and sea level would go on rising for centuries, due to the fact that deep oceans follow climatic variations with big delay.

13.3.9.4 Control Measures

Although individual people can help to combat air pollution in their own immediate environment, efficient control can be best achieved by legislation. Some commonly enforced control measures include the establishment of more smokeless zones; control over the kinds of fuel used in cars, aeroplanes, power stations, etc. The best way to fight the air pollution is to ensure that is no production of air-polluting materials during the process. One can do this to use other fuels or to change the process. Besides these steps one can ensure that if air pollutants are formed, these can be removed and caught before they come into the environment. Preventing the consequences for the environment directives have been established for the emission of air-polluted substances.

It is important to slow the warming as much as possible. This means using less fossil fuel, eliminating CFCs altogether and slowing down deforestation.

This can be achieved best through

- energy conservation, including better use of public transport;
- cleaner, more efficient cars;
- energy efficiency by greater use of gas which produces less CO_2 than coal and oil, and through pollution-free renewable energy such as solar power, wave and wind energy;
- we need to stop destroying rain forests (deforestation) and start replanting trees (afforestation) to soak up carbon dioxide.

13.4 Water Pollution

Water pollution can be defined as *the discharge of foreign matter into water that affects the physical, chemical and biological properties of water, which in turn decreases the utility of water.*

Examples of water pollution are as follows:

(1) Industrial effluents

Water is discharged from after having been used in production processes. This waste water may contain acids, alkalis, salts, poisons, oils and in some cases harmful bacteria.

(2) Mining and agricultural wastes

Mines, especially gold and coal mines, are responsible for large quantities of acid water. Agricultural pesticides, fertilizers and herbicides may wash into rivers and stagnant water bodies.

(3) Sewage disposal and domestic wastes

Sewage as well as domestic and farm wastes were often allowed to pollute rivers and dams.

Water is an important substance required for the existence of human beings, animals and plants. Contamination of water has led to the increase in instances of waterborne diseases in underdeveloped countries.

The sale of safety mineral water for drinking in our country has become a million-dollar business indicating the gravity of unavailability of pure drinking water in India. A minimum of 50 percent diseases can be prevented if pure drinking water is made available for all people.

Water pollution is contamination of water by foreign matter that deteriorates the quality of the water (Fig. 13.23). Water pollution covers pollutions in liquid forms like ocean pollution and river pollution. As the term applies, liquid pollution occurs in the oceans, lakes, streams, rivers, underground water and bays, in short liquid-containing areas. It involves the release of toxic substances, pathogenic germs, substances that require much oxygen to decompose, easy-soluble substances, radioactivity, etc. that becomes deposited upon the bottom and their accumulations will interfere with the condition of aquatic ecosystems. For example, the eutrophication: lack of oxygen in a water body caused by excessive algae growths because of enrichment of pollutants. This process ranges from simple addition of dissolved or suspended solids to discharge of the most insidious and persistent toxic pollutants (such as pesticides, heavy metals, and nondegradable, bioaccumulative, chemical compounds).

Figure 13.23 *Pollutants in water.*

13.4.1 Classification of Water Pollution

The major sources of water pollution can be classified as follows:

1. Municipal
2. Industrial
3. Agricultural

1. Municipal water pollution consists of wastewater from homes and commercial establishments. For many years, the main goal of treating municipal wastewater was simply to reduce its content of suspended solids, oxygen-demanding materials, dissolved inorganic compounds and harmful bacteria. In recent years, however, more stress has been placed on improving means of disposal of the solid residues from the municipal treatment processes. The basic methods of treating municipal wastewater fall into three stages: primary treatment, including grit removal, screening, grinding and sedimentation; secondary treatment, which entails oxidation of dissolved organic matter by means of using biologically active sludge, which is then filtered off; and tertiary treatment, in which advanced biological methods of nitrogen removal and chemical and physical methods such as granular filtration and activated carbon absorption are employed. The handling and disposal of solid residues can account for 25–50 percent of the capital and operational costs of a treatment plant.
2. The characteristics of industrial wastewaters can differ considerably both within and among industries. The impact of industrial discharges depends not only on their collective characteristics, such as biochemical oxygen demand and the amount of suspended solids, but also on their content of specific inorganic and organic substances. Three options are available in controlling industrial wastewater. Control can take place at the point of generation in the plant; wastewater can be pretreated for discharge to municipal treatment sources; or wastewater can be treated completely at the plant and either reused or discharged directly into receiving waters.

13.4.2 Sources of Water Pollution

Pollution of water arises from various sources. Some of the important polluting sources are as follows:

(i) *Domestic sewage:* Domestic sewage contains human excreta, kitchen waste, soaps, detergents, wastes from streets, organic wastes from trees and plants. The release of these wastes through drainage into canals, lakes and rivers cause major water pollution.

(ii) *Industrial sewage:* Industrial sewage contains a variety of pollutants like acids, alkalis, soaps, detergents, grease, pesticides, fungicides, insecticides, metal salts, animal body wastes from butcheries, heavy metal salts from tannery, organic pollutants from breweries and distilleries, oil refineries, chemical and pharmaceutical industries, etc.

(iii) *Farm sewage:* Farm sewage includes excreta from animals, fertilizers, pesticides, insecticides fungicides, silt, organic wastes from trees and plants, etc.

(iv) *Other polluting agents:* Oil from oil spills, washing of automobiles is another major pollutant. Radioactive pollutants arise from wastes of atomic power plants, radioactive mineral processing units, research laboratories and hospitals minerals, fine soil particles, clay, hazardous wastes, infectious organisms and organic wastes from forests through which the river flows make another polluting system.

13.4.3 Ill Effects of Heavy Metals on Water

Various heavy metals get dissolved in water and cause ill effects. The important heavy metals that cause maximum damage are as follows:

(1) Cadmium

Source: From Ni–Cd battery and gold ornaments industries.

Ill effects: Gets absorbed on suspended matter in the water which when ingested, cause 'itai-itai' disease (increased salivation, acute gastritis, liver and kidney damage).

(2) Lead

Source: From lead–acid battery industries, galvanized iron (GI) pipe connections and plastic industries.

Ill effects: Lead is cumulative poison, causing constipation, loss of appetite, abdominal pain and anaemia, gradual paralysis of muscles, mental retardation, a blue line at the junction of teeth and gums, and finally nervous and skeletal disorders and brain damage.

(3) Mercury

Source: From alkali, thermometer industries and certain pharmaceutical industries.

Ill effects: When present in water in amounts exceeding 1–10 mg/l, are converted into the highly toxic methyl mercury. It gets concentrated in fish and shell-fish in the sea, which when ingested cause nerve and brain damage and paralysis followed by death in human beings and other animals.

Ill effects of inorganic Species:

(1) Cyanide

Source: Many metallurgical operations, like gold and silver refinery, heat-treatment industries, etc.

Ill effects: Even very small dosages of cyanide (very few parts per million) cause formation of cyanohaemoglobin in blood, which destroys the oxygen carrying capacity of blood and lead to death. As you are aware that a few micrograms of cyanide are good enough to cause instant death.

(2) Ammonia

Source: Many fertilizer, chemical industries and biological degradation of vegetable matter add ammonia to water.

Ill effects: Ammonia in water increases pH value, and when consumed produces high alkalinity, causing stomach disorder. In ponds, increase in ammonia produces *algae bloom*, which in turn reduces dissolved oxygen, causing damage to aquatic life.

(3) Hydrogen sulphide

Source: Detergent industries, decay of vegetable matter and decomposition sulphide containing materials, lead to addition of hydrogen sulphide into water.

Ill effects: It is toxic in nature and causes harmful effects to human beings. It produces bad odour (rotten-egg odour), it is highly corrosive to metals.

(4) Ill effects of organic matters

(i) Pesticides

Source: From pesticide industries and farm drains, etc.

Ill effects: They are toxic to human beings and animals. They cause biological disorders and damage to human body. They may also lead to cancer.

(ii) Herbicides

Source: Mostly are from herbicide industries and farms.

Ill effects: They are toxic to human beings and animals. They cause biological disorders and damage to human body. They may also lead to cancer.

(iii) Insecticide

Source: From insecticide, industry wastes, various insecticides from the farm, poultry and domestic uses.

Ill effects: They are toxic to human beings and animals. They cause biological disorders and damage to human body. They may lead to cancer, also.

13.4.4 Sewage and Its Characteristics

Any fresh sewage is usually green or yellowish green. With aging, it becomes dark brown. Fresh sewage is odourless but when it becomes stale it produces obnoxious smell due to the release of gases like H_2S, phosphine, etc. Sewage water is normally turbid and has a temperature slightly higher than ordinary water. It contains dissolved and suspended impurities and also colloidal dispersion. Sewage contains both organic and inorganic impurities. The pH of fresh sewage is higher than seven (alkaline) but after a few days it becomes less than seven (acidic). Sewage also contains pathogenic (disease producing) bacteria and nonpathogenic bacteria. There are two types of nonpathogenic bacteria.

(i) *Aerobic bacteria* which live and grow on free oxygen and dissolved oxygen. These bacteria act as catalysts for the oxidation of fresh sewage. The products of oxidation are nitrites, nitrates, sulphates, etc. These do not produce any offensive smell.

(ii) *Anaerobic bacteria* which live and grow in the absence of free oxygen. When the dissolved oxygen content in water decreases below a particular level anaerobic bacteria try to decompose the organic compounds in water and produces methane, H_2S, phosphine, etc. responsible for the offensive smell. The water is said to be stale and the process of oxidation is anaerobic oxidation.

13.4.4.1 Sewage Treatment

Domestic sewage contains colour, bad odour, organic and inorganic impurities, pathogenic bacteria, etc. If this sewage is discharged into rivers, lakes and sea, the water in them gets polluted and causes several harmful effects (Fig. 13.24)

- It colours and produces bad odour affecting quality of water
- It causes reduction in dissolved oxygen. This endangers the life of fish and aquatic life
- Pathogenic bacteria lead to waterborne diseases

Figure 13.24 *Discharge of sewage through an open pipe to a drain.*

Domestic sewage therefore needs proper treatment before it is run into natural water bodies. Treatment of sewage involves removal of organic impurities, suspended and floating materials, inorganic salts, pathogenic bacteria, etc.

Sewage treatment is carried out in three stages (a) primary treatment, (b) secondary treatment, and (c) tertiary treatment (Fig. 13.25).

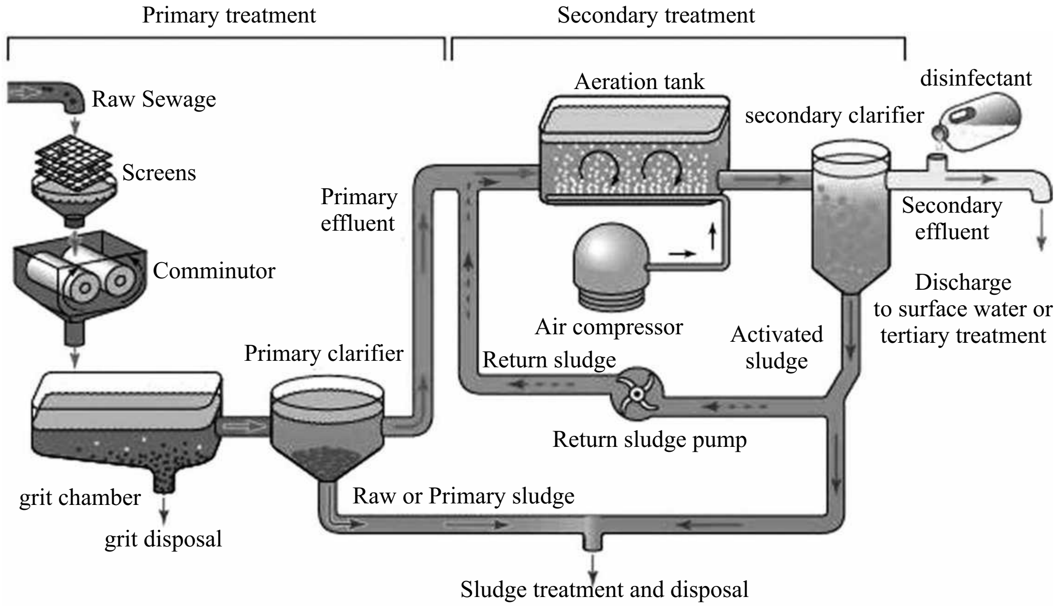


Figure 13.25 *Schematic view of sewage treatment process.*

(1) Primary treatment

Primary treatment is used to remove suspended and floating solids from waste water by physical and chemical methods. It involves the following steps.

(i) *Screening* using bar screens and mesh screens floating, suspended and coarse particles are removed by passing sewage water through it.

(ii) *Slit and grit removal:* Sand, powdered glass, etc. called grit are removed by slowly passing sewage water through grit chambers. Heavier sand and broken glasses settle down by gravity.

(iii) *Removal of oil and grease:* Sewage water is kept in a skimming tank and compressed air is blown through it. Oil and grease form froth and float on the surface. It is skimmed off.

Sedimentation process:

In this process fine suspended particles which do not settle down by gravity are coagulated by the addition of coagulating agents like alum, $FeSO_4$, etc. The sedimented particles are filtered off.

(2) Secondary treatment (biological treatment)

Activated sludge method:

Waste water after sedimentation is mixed with required quantity of activated sludge (containing microorganisms and aerobic bacteria) in aeration tank. The mixture is aerated by blowing air through it for several hours. Because of the aerobic conditions the organic matter in the sewage is fully oxidized. The purified water + sludge is sent to a tank where sludge settles down. A part of the sludge is used for purification of fresh batch of sewage while the rest is pumped to sludge disposal tank. Purified water is pumped out and collected separately.

(3) Tertiary treatment

Tertiary treatment involves further purification. It involves the following steps of removing dissolved salts, colloidal particles, gases, etc.

(i) *Removal of phosphate* is done by adding lime. A precipitate of calcium phosphate is formed at a pH of 10. Ammonium salts are converted to ammonia.

$$3Ca(OH)_2 + 2PO_4^{-3} \rightarrow Ca_3(PO_4)_2 + 6OH^-$$

$$NH_4^+ + OH^- \rightarrow NH_3 + H_2O$$

(ii) *Coagulation and sedimentation*: Colloidally disposed fine particles are coagulated in a sedimentation tank by the addition of alum or $FeSO_4$. The highly charged Al^{3+} and Fe^{+2} ions neutralize the charge on the colloidal particles. $Al(OH)_3$ and $Fe(OH)_2$ formed as flocculant precipitates bring down the coagulating colloids.

(iii) *Filtration:* The water is passed through conventional filter beds filled with gravels, coarse sand and fine sand to remove last traces of suspended matter.

(iv) *Degasification:* Dissolved gases are moved by trickling the water through tall tower filled with a number of perforated hot plates. Large surface area and heat liberates the dissolved gases like NH_3, CO_2, H_2S, etc.

(v) *Disinfection:* Harmful pathogenic bacteria are removed either by passing Cl_2 gas or ozone or UV light through the water.

13.5 Land Pollution

Land pollution is the degradation of the earth's land through human misuse of the soil. Human influences such as poor agricultural practices, the digging up of important resources, industrial waste dumping and careless disposal of trash continue to pollute the land rapidly.

Land pollution involves the following mechanism:

- Deposition of solid waste
- Accumulation of nonbiodegradable materials
- Toxification of chemicals into poisons
- Alteration of soil chemical composition.

13.5.1 Sources and Methods

We can classify major sources that lead to land pollution to the following categories:

- Agriculture
- Mining and quarrying
- Sewage sludge
- Dredged spoils
- Household
- Demolitions and constructions
- Industrial

Human impact on the land plays an important role in the cleanliness and prosperity of our earth's future.

(i) Soil misuse

Soil erosion is when the top soil is stripped of essential nutrients needed for plant and vegetative growth. When the soil erodes, the rich top layer called humus is removed.

(ii) Solutions

Solutions to the problems of land pollution have become increasingly recognized over the years. The most common and convenient method of waste disposal is by the sanitary landfill. The sanitary landfill is a legal replacement for the open dumps that once were present. Sanitary landfills are the most popular for cleanliness; however, often times industrial waste of unknown contents is commingled with domestic wastes.

(iii) Recycling

Recycling is practical to some extent for much municipal and some industrial wastes, and a small but growing proportion of solid waste are being recycled. When wastes are commingled, however, recovery becomes difficult and expensive. New processes of sorting ferrous and nonferrous metals, paper, glass and plastics have been developed, and many communities with recycling programmes now require reuse separation. Crucial issues in recycling are devising better processing methods, inventing new products for the recycling materials and finding new materials for them.

(iv) What happens after soil erosion?

The result of soil erosion over a long period of time includes the decaying of vegetation and negative microbial effects. This manmade disaster leaves large parcels of land destroyed and unfit for regrowth for many years, and thus, thousands of acres of land are wasted each year. Furthermore, the increase of the earth's population also increases land pollution covering good humus with more apartment buildings and businesses. Humans must find a way to work around nature and its living creatures. Deforestation, especially of the rain forest, can kill off species of plants, animals and bacteria never known to man. Disease cures and medicines could be wiped out without ever being introduced to humans.

(v) Incineration

This is another method of disposing solid wastes. Wastes must be burned at very high temperatures and incinerator exhausts must be equipped with sophisticated scrubbers and other devices for removing dioxin and other toxic pollutants. Problems remain, however: incinerator ash contains high ratios of heavy metals becoming a hazardous waste in itself.

(vi) Composting

This is using degradable foods from your household as manure/fertilizer for the earth. Again, this process alters the environment of the animals and their habitats in the area, leading to a change in the ecosystem.

13.5.2 Waste Disposal

A disposal of normally solid or semisolid materials, resulting from human and animal activities, that are useless, unwanted or hazardous (Fig. 13.26). Solid wastes typically may be classified as follows:

FIGURE 13.26 *Land pollution by dumping solid waste.*

1. *Garbage*: decomposable wastes from food.
2. *Rubbish*: nondecomposable wastes, either combustible (such as paper, wood, and cloth) or noncombustible (such as metal, glass and ceramics).
3. *Ashes*: residues of the combustion of solid fuels.
4. *Large wastes*: demolition and construction debris and trees.
5. Dead animals.
6. Sewage-treatment solids: material retained on sewage-treatment screens, settled solids and biomass sludge.
7. *Industrial wastes*: such materials as chemicals, paints and sand.
8. *Mining wastes:* slag heaps and coal refuse piles.
9. *Agricultural wastes*: farm animal manure and crop residues.

13.5.3 Waste Management and Recycling

If you cannot *reduce* or *reuse* an item, can it be recycled or, if it is organic material like grass cuttings or food scrapings, composted material. *Recycling means using valuable raw materials like glass, paper, tin and aluminium over and over again so that they need not be produced again from scratch.*

Glass jars, bottles, aluminium cans, newspapers, plastic bottles, tin cans and clothes, etc. can be recycled. Benefits of recycling include the following:

- Cutting down the amount of material
- Reducing transport movements and air pollution
- Waste landfilling is less
- Reducing energy consumption from making new materials
- Creating new jobs from recycling initiatives
- Encouraging people and businesses to take more responsibility for the waste they produce
- Improving competitiveness as businesses use their materials more efficiently

13.5.4 Toxic and Dangerous Chemicals

Pollution by toxic chemicals may be released from the following sources:

- Industrial wastes (e.g. chemical industry)
- Transportation (e.g. oil tankers)
- Agriculture (fertilizers, herbicides, pesticides)
- Consumption (e.g. paints, PVC bottles)
- Civil waste (e.g. detergents)

Some toxic chemicals are resistant to degradation and, due to their environmental persistence, may move long distance and concentrate very far from the emission sites, or, being fat-soluble, bioaccumulate in the tissues of humans or animals (at the top of food chain).

The most common among these substances, DDT and PCB (poly chlorinated biphenyls), are widely contaminating the oceans: they have been found in whales and other marine mammals.

PCBs persist in the environment for years and may bioaccumulate up to 70,000 times. It is reasonable to believe that many other *persistent organic pollutants* (POP), contaminate the Oceans. Toxic chemicals are chlorinated pesticides (dieldrin, DDT, toxaphene, hexachlorobenzene, endrin, chlordane), and industrial chemicals (PCBs), etc.:

The other harmful chemicals are the poly aromatic hydrocarbons (PAHs), the short chained chlorinated paraffins, mercury and organomercury compounds, cadmium, lead and organolead compounds, organotin compounds (including TBT (tributyltin), used in naval paints), and some phthalates (dibutylphthalate, di-etyl-hexyl-phthalate).

13.6 Thermal Pollution

This has become an increasing and the most current pollution, owing to the increasing call of globalization everywhere. Heat produced from industries is a major contribution to the pollution, much to the operation of the heavy industries which produces high amount of heat energy (Fig. 13.27).

Figure 13.27 *Pollution from a thermal power plant.*

Pollution of environment may be from:

- Raw materials for productivity (organic and inorganic products)
- Undergo different chemical reactions with several process
- Excess heat energy is produced as a waste product
- Heat is released through into atmosphere (vapour)
- Increase of temperature of environmental system

In view of the pollution, global temperature had increased significantly.

13.6.1 Sources and Methods

We can classify major sources that lead to thermal pollution into the following categories:

- Power plants creating electricity from fossil fuel
- Water as a cooling agent in industrial facilities
- Deforestation of the shoreline
- Soil erosion

13.7 Nuclear Pollution

Nuclear energy is a form of energy that is released by the splitting of atoms. Since scientists have found a way to make use of the energy, it has also been used to generate electricity.

Nuclear energy has been recognized as a clean energy because it does not release pollutants such as CO_2 to the atmosphere after its reaction that could damage our environment. It is also known that nuclear energy has reduced the amount of greenhouse gas emission, reducing emissions of CO_2 for about 500 million metric tons of carbon.

Despite the advantage of nuclear as a clean energy, the big concern is the waste resulted from nuclear reaction, which is a form of pollution, called radioactivity. Radioactivity is a form of radiation. Some elements in this world are naturally radioactive while some others are made to be. Radioactivity is emitted when a radioactive element become unstable and begin to decay in the attempt to regain their molecular stability. When an element decays, it emits energy and small particles. If it is still radioactive, it will repeat the process, until it finally regains its molecular stability and stop decaying. The time that it takes for half way of decaying process is called half-life, and this differs for each radioactive element. It possibly takes up to 4.5 billion years (Uranium 238) and as short as 8 days (Iodine 131). This process constantly remains, not considering external factors such as pressure or temperature (Fig. 13.27).

There are commonly three types of radiation, namely:

- Alpha particles can be blocked by a piece of paper and human skin.
- Beta particles can penetrate through skin, while can be blocked by some pieces of glass and metal.
- Gamma rays can penetrate easily to human skin and damage cells on its way through, reaching far, and can only be blocked by a very thick, strong, massive piece of concrete.

13.7.1 Sources and Methods

We can classify major sources that lead to radioactive pollution into the following categories:

- Nuclear power plants
- Nuclear weapon
- Transportation
- Disposal of nuclear waste
- Uranium mining

13.8 Noise Pollution

This particular pollution is ever increasing with due to the rise in the utilization of heavy duty machineries of industrial facilities and vehicles, synonymous to the increase in the standard of living in most countries (Fig. 13.28). We make sounds practically every second of our day, but to the extent it has reached an unfavourable high intensity, which has caused many disturbances and irritation to others emotionally and has adverse effects also on our daily activities.

These are a few examples of threshold decibels of noises made:

Figure 13.28 *Noise pollution of an aircraft.*

Threshold of hearing	0 dB	Motor cycle (30 feet)	88 dB
Rustling leaves	20 dB	Food blender (3 feet)	90 dB
Quiet whisper (3 feet)	30 dB	Subway (inside)	94 dB
Quiet home	40 dB	Diesel truck (30 feet)	100 dB
Quiet street	50 dB	Power mower (3 feet)	107 dB
Normal conversation	60 dB	Pneumatic riveter (3 feet)	115 dB
Inside car	70 dB	Chainsaw (3 feet)	117 dB
Loud singing (3 feet)	75 dB	Amplified Rock and Roll (6 feet)	120 dB
Automobile (25 feet)	80 dB	Jet plane (100 feet)	130 dB

13.8.1 Sources and Methods

We can classify major sources that lead to noise pollution into the following categories:

- Road traffic noise
- Air traffic
- Rail traffic
- Neighbourhood and domestic noise
- Incompatible land use
- Industrial noises
- Burning of crackers and fireworks

Review Questions

1. Define the term environmental pollution. List out the causes for pollution.
2. Define air pollution. Explain the various natural and anthropogenic causes of air pollution.
3. Mention the sources, ill effects and control of the following atmospheric pollution, (i) SO_2, (ii) oxides of nitrogen, (iii) CO, and (iv) hydrocarbons. Explain their all effects.
4. Mention the various sources of particulate matter. Describe two methods of removing particulate matter from air. [VTU: March 1999]
5. Give an account of acid rain. [VTU: Aug. 1999]
6. Write a note on Cottrell electrostatic precipitator. [VTU: Aug.1999, Feb.2000]
7. Write a short note on greenhouse effect. [VTU: Aug.1999, Feb.2000, March 1999]
8. Discuss to harmful effects and prevention of NO_x pollutants. [VTU: Feb. 2000]
9. What is photochemical smog? Explain the mechanism of its formation. [VTU: Aug. 2000, Aug. 2001]
10. Mention the sources of water pollution. Explain the terms BOD and COD. [VTU: Aug. 2000, Aug. 2001]
11. What are the sources of air pollution? Give the effect and control of SO_2 pollution. [VTU: Aug. 2001]
12. Describe the primary and secondary methods of treatment of sewage.
13. Write a short note on Lake Eutrophication.
14. Give the ill effects and prevention of ozone depletion and greenhouse effect. [VTU: Aug. 2001]
15. Write a note on centrifugal separation.

16. Explain the following.
    (i) SO_2 pollution is a localised pollution.
    (ii) Ozone depletion is caused by CFC.
    (iii) Waste water produces bad smell after a few days.
17. Write a note on nature of pollution sources.
18. What is agriculture pollution?
19. Write a note on impacts air pollution.
20. How are pollutants classified?
21. Discuss the natural sources, ill effects and control of SO_2 pollutant.
22. Discuss the natural sources, ill effects and control of CO pollutant.
23. Discuss the natural sources, ill effects and control of NO_x pollutant.
24. Write a note on particulate matter in pollution of air.
25. Give brief description of fabric filter for controlling particulate matter.
26. Write a note on cyclone separator.
27. Write a note on electrostatic precipitator.
28. What is acid rain? Explain what are the harmful effects of acid rain?
29. Write a note on ozone depletion.
30. How do you control ozone depletion? What are the measures undertaken to stop ozone depletion?
31. What do you mean by global warming? Give reasons.
32. Describe greenhouse effect.
33. Write a note on greenhouse gases. Explain.
34. What is water pollution? What are the sources of water pollution?
35. Give the classification of water pollution.
36. Write a note on the ill effects of heavy metals on water.
37. Mention the ill effects of inorganic species.
38. Give the ill effects of organic matters.
39. What is a sewage? Give its characteristics.
40. Discuss sewage treatment methods.
41. Write a note on land pollution.
42. Mention the sources of land pollution and discuss its impact on environment.
43. Write a note on waste disposal.
44. Write a note on toxic chemicals and environment.
45. What is thermal pollution? Mention the sources of thermal pollution.
46. Write a note on nuclear pollution.
47. Classify the major sources of nuclear pollution.
48. Write a note on noise pollution.

14

Phase Rule

Chapter Outline

Introduction. The phase rule. The terms involved—phase, components, degree of freedom or variance. Application of phase rule to one-component water system. Application of phase rule to two-component systems. Reduced phase rule—two-component systems with solid and liquid phases in equilibrium. A simple eutectic system—lead–silver system, desilverization of lead. Two-component system—congruent melting compound—tin–magnesium system. Two-component system—incongruent melting compound—phase diagram of Mg_2SiO_4–SiO_2 system. Two-component system—formation of solid solution systems—phase diagram of $NaAlSi_3O_8$–$CaAl_2Si_2O_8$ system, phase diagram of silver–gold system, system forming partially miscible solid solutions—iron–carbon system, phase diagram of Lead—tin system.

14.1 Introduction

In this chapter, a discussion of the application of the phase rule is restricted to few systems where different phases exist in a state of equilibrium. A phase diagram is a graphical representation of chemical equilibrium. Since chemical equilibrium is dependent on the composition of the system, the pressure, and the temperature, a phase diagram should be able to tell us what phases are in equilibrium for any composition at any temperature and pressure of the system. The phase rule also predicts the conditions like temperature, pressure, composition, etc., to be specified for a system to exhibit equilibrium.

14.2 The Phase Rule

The phase rule is an expression of the number of variables and equations that can be used to describe a system in equilibrium. In simple terms, the numbers of variables are the number of chemical components in the system plus the extensive variables, temperature and pressure. The number of phases present will depend on the variance or degrees of freedom of the system. The general form of the phase rule is stated as follows:

$$P + F = C + 2$$

where P = number of phases involved in equilibrium
C = number of components in a system and
F = the degree of freedom of a state of a system.

The phase rule is defined as follows:

'For a heterogeneous system in equilibrium, the number of phases plus the number of degree of freedom equal the number of components plus 2.'

14.3 Terms Involved in the Phase Rule

14.3.1 Phase

In a heterogeneous system, a number of phases are involved.

'A phase is defined as any homogeneous, physically distinct and mechanically separable portion of a system, which is separated from other parts of the system by definite boundary surface.'

Figure 14.1 shows the existence of the possible phase/phases of a system.

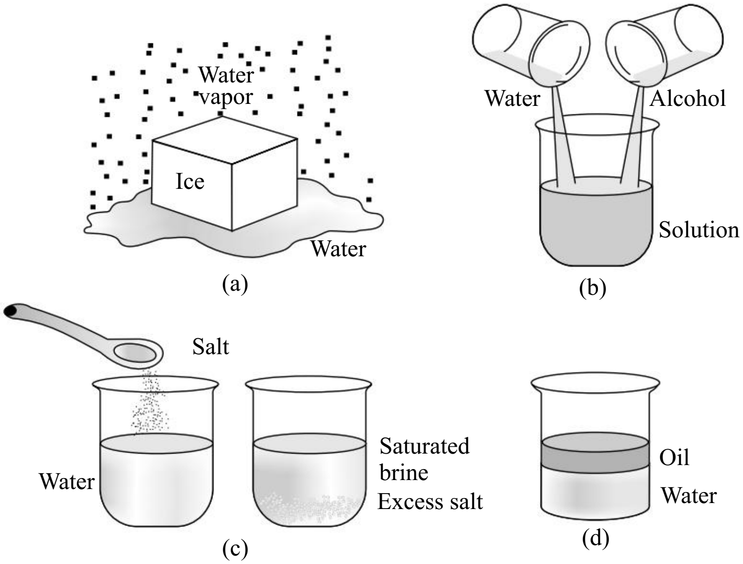


FIGURE 14.1 *Number of possible phases of a system: (a) ice, water and water vapour, (b) water and alcohol, (c) water and excess salt and (d) oil and water.*

Examples

(a) A completely miscible homogeneous mixture of gases (O_2 and N_2) and liquids (water and alcohol) gives a single phase ($P = 1$).
(b) Any pair of immiscible liquids (water and kerosene) gives rise to two phases, i.e. $P = 2$.
(c) At a specified condition, all the three phases of water coexists
Ice water vapour ($P = 3$).

(d) Structurally different solids (e.g. S_{Rh} and S_m) give rise to a multiphase system.

(e) A heterogeneous system of equilibrium, such as, $CaCO_3 \leftrightarrow CaO + CO_2$.

There are two solid phases and one gaseous phase in a state of equilibrium.

14.3.2 Components

The components of a system is defined as 'the smallest number of independently variable constituents taking part in a state of equilibrium by means of which the composition of each phase can be expressed in the form of chemical equations':

Illustrated Examples

(i) In the water system, the composition of all the three phases can be defined in terms of only H_2O. Therefore, it is a one-component system, $c = 1$.

(ii) Lead and silver are miscible in all proportions in the molten state and gives a homogeneous solution of two constituents.

Therefore, to define the composition of this homogeneous molten single phase, we need both the constituents. Therefore, it is a two-component system, $c = 2$

(iii) Consider a heterogeneous chemical system in equilibrium.

$$CaCO_3 \leftrightarrow CaO + 1\ CO_2$$

The composition of any of the three phases can be expressed in the form of an equation in terms of at least any two of the independently variable constituents such as CaO and CO_2, $CaCO_3$ and CO_2 or $CaCO_3$ and CaO. Therefore, it is a two-component system.

That is,

$$CaCO \leftrightarrow 1CaO + 1CO_2$$
$$CaO \leftrightarrow 1CaO + 0CO_2$$
$$CO_2 \leftrightarrow 0CaO + 1CO_2$$

(iv) Consider the dissociation of NH_4Cl in a closed vessel.

$$\underset{(s)}{NH_4Cl} \leftrightarrow \underset{(g)}{[NH_3 + HCl]} \leftrightarrow \underset{(g)}{NH_3} + \underset{(g)}{HCl}$$

(a) If equivalent amounts [say, a g mole each] of NH_3 and HCl(g) are formed from a g mole molecule of NH_4Cl (s) and this also would be equivalent to 1 g mole molecule of NH_4Cl(g).

Therefore, the composition of all the phases involved can be expressed by considering only one NH_4Cl. Therefore, $C = 1$, i.e. it is a one-component system.

(b) However, if NH_4Cl (s) is in excess to the other constituents in the system, then it is a two-component system, $C = 2$.

14.3.3 Degrees of Freedom or Variance

The 'state of a system' in equilibrium can be defined in terms of the thermodynamic variables such as the pressures (P), temperature (T), composition (C), etc.

Therefore, *'the degree of freedom of a system is defined as the number of independently variable, such as pressure temperature and composition which must be specified in order to define the state of a system'.*

Illustrated Examples

Consider any single phase of a one-component water system.

H_2O (s), or H_2O (l), or H_2O (g), The state of any single phase in the water system can be defined by specifying two variables like pressure and temperature. In other words, any single phase of the water system has two degrees of freedom. Hence, $F = 2$ or $f(P$ and $T)$.

(i) Consider any two phases in equilibrium for a one-component water system

$$H_2O\ (s) \leftrightarrow H_2O\ (l), \text{ or } H_2O\ (l) \leftrightarrow H_2O\ (g), \text{ or } H_2O\ (g) \leftrightarrow H_2O\ (s),$$

Thus, to describe the state of such two-phase equilibria in the water system, we need to fix only one variable, that is, either pressure (P) or temperature (T). Therefore, $F = 1$ or $f(P$ or $T)$.

(ii) The coexistence of all the three phases in a one-component water system.

All the three phases coexist only at a particular pressure and temperature. Therefore, the above system has no degrees of freedom.

$F = 0$, the system is invariant.

14.4 Application of the Phase Rule to a One-Component Water System

Water can exist in three possible phases, viz. $H_2O(s)$, $H_2O(l)$, and $H_2O(g)$, and depending on the thermodynamic variables, pressure and temperature. The composition of all the three phases of water is only H_2O and hence, it is a one-component system, that is, $C = 1$.

In the water system, the following systems exist in equilibrium, depending on their pressure and temperature.

1. $H_2O(s) \leftrightarrow H_2O(l)$
2. $H_2O(l) \leftrightarrow H_2O(g)$,
3. $H_2O(g) \leftrightarrow H_2O(s)$,

In the phase diagram of the water system (Fig. 14.2), the application of the phase rule predicts the conditions (pressure and temperature) that must be fixed to define states of system in equilibrium.

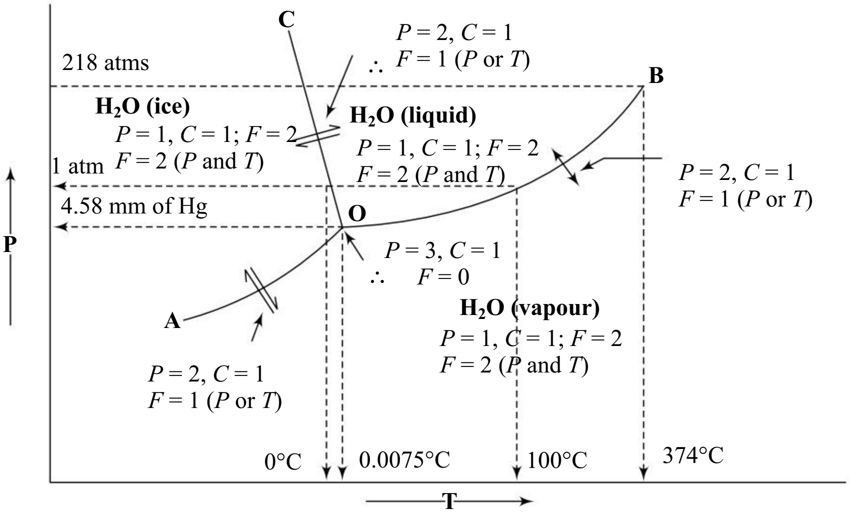


FIGURE 14.2 *Phase diagram of water.*

One can infer from the application of the phase rule that the degrees of freedom of a one-component water system depends on the number of phases present in the equilibrium.

Applying the phase rule, we have

$$F = C + 2 - P$$

If (i) $P = 1, C = 1, F = 2$ (bivariant), $f(P$ and $T)$
(ii) $P = 2, C = 1, F = 1$ (univariant), $f(P$ or $T)$
(iii) $P = 3, C = 1, F = 0$ none

Areas: Single-phase system

AOC, *BOC* and *AOB* are the areas in the phase diagram of water that represent the state of existence of water as (ice), (liquid) and (vapour) phases, respectively.

The existence of any single phase in a one-component water system can be described by ascertaining its degree of freedom.

By applying the phase rule,

$$\text{If } P = 1, C = 1, \text{ therefore } F = 1 + 2 - 1 = 2\ (P \text{ and } T)$$

Thus, the phase rule predicts that both pressure and temperature have to be specified to define any single phase, $H_2O(s)$, or $H_2O(l)$, or $H_2O(g)$; and hence, the state of single phase (*AOB*, *BOC* and *AOC*) in equilibrium is bivariant (*P* and *T*).

Curves *OA*, *OB* and *OC*: two-phase system

In the phase diagram, the boundary lines of two phases, *OA*, *OB* and *OC* represent the coexistence of two phases of water in equilibrium for a one-component system. The existence of the state of equilibrium between these two phases can be explained by the application of the phase rule.

For any curve *OA*, *OB* or *OC*, we have $P = 2$, $C = 1$, therefore $F = 1$ (*P* or *T*)

The degree of freedom of such a two-phase system is ‘univariant’. In other words, it predicts the condition that one of the variables (either pressure or temperature) can be varied without disturbing the existence of two phases in equilibrium.

(i) The curve *OA* is the ‘sublimation curve of ice’. It gives the condition under which ice ($H_2O(s)$) is in equilibrium with vapour ($H_2O(g)$)
Applying the phase rule,

$$P = 2, C = 1, \text{ therefore } F = 1\ (P \text{ or } T)$$

The curve *OA* is univariant.

(ii) The curve *OB* represents the equilibrium between H_2O (l) and H_2O (g),.
This curve is known as ‘vaporization curve’ of water. The curve *OB* is also univariant. The curve *OB* has a natural upper limit at 374°C, which is the critical point, beyond which the liquid phase merges into vapour.

(iii) The curve *OC* represents the equilibrium between $H_2O(s)$ and $H_2O(l)$.
Applying the phase rule,

$$P = 2, C = 1, \text{ therefore } F = 1\ (P \text{ or } T)$$

The curve *OC* is univariant and is $f(P$ or $T)$. The curve *OC* is called the ‘fusion curve’ of ice which indicates temperature (or pressure) at which ice and water can coexist in equilibrium. Also, the curve/line *OC* indicates how the freezing temperature of water varies with pressure. Interestingly, ‘the slope *OC* towards the pressure axis shows that the melting point of ice decreases with corresponding pressure increase’.

Triple Point: three-phase system

'Triple point' in a one-component water system denotes a point at which all the three phases, $H_2O(s)$, $H_2O(l)$, and $H_2O(g)$ of water coexists in equilibrium.

Applying the phase rule, we have

$$P = 3, C = 1, \text{ therefore } F = 1 + 2 - 3 = 0$$

Since, all the three phases of water coexists at a triple point *O*, the system is 'invariant'. Triple point *O* for the water system exists at 4.58 mm of Hg and 0.0075°C.

Metastable Equilibrium Curve *OA′*

It is possible to super cool water several degrees below the freezing temperature without solidification along the curve *OA′*. This super cooled liquid vapour system along the curve *OA* is said to be in 'metastable equilibrium'. It may be noted that the vapour pressure of metastable super cooled liquid is higher than curve *OA* and hence, any disturbance induced leads to crystallization of ice.

14.5 Application of the Phase Rule to a Two-Component System

14.5.1 Reduced Phase Rule

The degrees of freedom of any two-component system can be ascertained by the application of the phase rule,

$$F = C + 2 - P = 2 + 2 - P = 4 - P$$

Since, the minimum number of phase is 1 in any two-component system, the maximum number of degrees of freedom is 3.

Thus, all the three thermodynamic variables (pressure, temperature and composition) are necessary to define any single phase in a two-component system.

In practice, this approach is cumbersome and not generally adopted.

For the sake of convenience, all two-component systems in equilibrium are described by a 'reduced phase rule', since the effect of pressure on such solids and liquids system is negligible. The degree of freedom of the system in equilibrium is reduced by 1. In such a case, the phase rule reduces to

$$P + F = C + 1$$

This is known as the 'condensed' or the 'reduced phase rule'.

Therefore, 'to define any state of a system in equilibrium, the reduced phase rule demands only variables such as temperature and composition'.

14.5.2 Two-Component Systems with Solid and Liquid Phases in Equilibrium

Depending on the mutual solubility of the components in the liquid phase, and on the nature of the solid that separates out from the liquid mixture, the systems are classified as given:

- A simple eutectic formation—pure solid *A* or *B* separates out from the liquid solution.
- A new solid compound having congruent melting point separates out from the solution.
- A new solid with incongruent melting point crystallizes out from solution
- A completely miscible solid solution crystallizes out from the solution.

The two solids crystallize out of the solution and are partially miscible among themselves.

14.5.2.1 A Simple Eutectic System

Consider the case of binary solids A and B that are completely miscible in all proportions in the molten state and also they do not form any compound (Fig. 14.3).

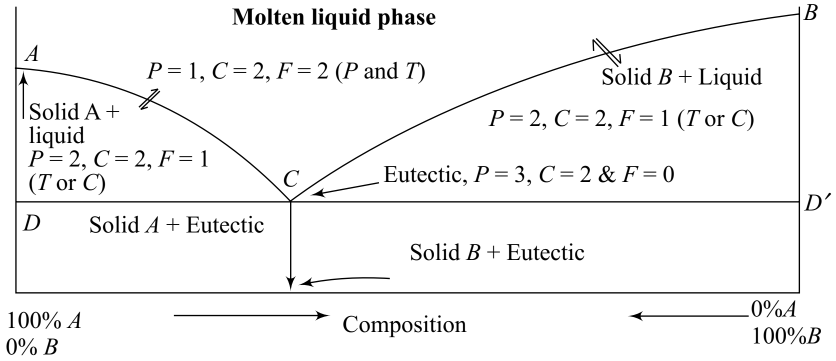


Figure 14.3 *Phase diagram of a simple eutectic of a two-component system.*

(i) Area of the molten phase (ACB)

Within this area ACB, both solids A and B are completely miscible in all proportions in the molten state. This is a two-component system.

Applying the reduced phase rule, $P + F = C + 1$, to this homogeneous single phase, we have

$$P = 1, C = 2, \text{ therefore } F = 2 + 1 - 2 = 2 \ (T \text{ and } C)$$

Therefore, the molten homogeneous phase of two components is 'bivariant'.

This means, 'to define any point in the area ACB, it is necessary to specify both temperature and composition'.

(ii) The curves AC and BC

In the phase diagram, A and B denotes the melting points of pure A and B.

The curves AC and BC represent the melting point curves of both A in B and B in A, respectively.

It is observed that the melting point of A decreases along the curve AC with increasing amounts of B. Similarly, as increasing quantities of A are added to B, the melting point of B falls along the curve BC.

The states of equilibrium, curves AC and BC, represent two-phase equilibrium.

1. Solid $A \leftrightarrow$ molten liquid
2. Solid $B \leftrightarrow$ molten liquid

Applying the reduced phase rule to such two-phase equilibrium, we have

$$P = 2, C = 2, \text{ therefore } F = 1 \ (T \text{ or } C)$$

Hence, the curves AC and BC are 'invariant'. The composition of the binary system varies with temperature along AC or BC.

(iii) Eutectic point

The two melting point curves AC and BC intersect at a 'lowest temperature' C. At this point C, Solid A, Solid B and molten liquid having both the components coexist at a 'definite composition' as shown in Fig. 14.3.

Therefore, all the three phases coexist at eutectic point C.

Applying the reduced phase rule, we have

$$P = 2, C = 2, \text{ therefore } F = 3 - 3 = 0$$

Eutectic point: The point on a phase diagram where the maximum number of allowable phases is in equilibrium. When this point is reached, the temperature must remain constant until one of the phases disappears. An eutectic is an invariant point.

'The 'eutectic' is a state of two components in a homogeneous solid of definite composition which melts at the lowest temperature at point C.'

14.5.3 Lead–Silver System

Two elements Pb and Ag are completely miscible in all proportions in their molten state and gives rise to a single homogeneous phase containing two components Pb and Ag.

The coexistence of different phases in equilibrium at different set of conditions is shown in the lead–silver phase diagram (Fig. 14.4).

(a) Area ACB

Within the area *ACB* in the phase diagram, a homogeneous single phase of molten liquid of two components (Pb and Ag) in all proportions exists.

Applying the reduced phase rule, $P + F = C + 1$, we have

$$P = 1, C = 2, \text{ therefore } F = 2 + 1 - 1 = 2 \ (T \text{ and } C).$$

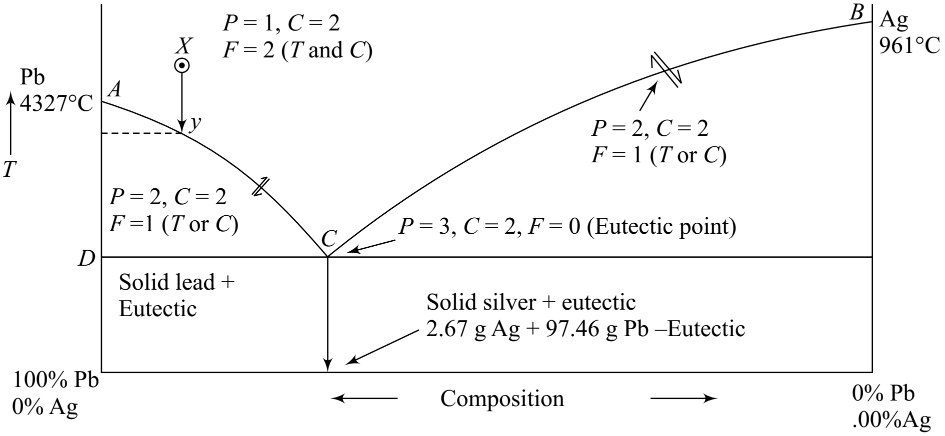


FIGURE 14.4 *Phase diagram of the lead–silver system.*

The homogeneous molten phase of Pb and Ag within area *ACB* is invariant and is a *f*(*T* and *C*). Therefore, 'to describe any point in the area *ACB*, both temperature and composition of the system has to be specified'.

(b) Melting point curves AC and BC

In the phase diagram of lead–silver, *A* and *B* denotes the melting points of pure lead (327°C) and silver, (961°C) respectively. The curve *AC* denotes the variation of melting temperatures of silver in Pb, while the curve *BC* indicates the decrease of melting temperatures of silver with the addition of lead.

The states of equilibrium, curves *AC* and *BC*, represent two-phase equilibrium.

(1) Solid lead ↔ molten liquid

(2) Solid silver ↔ molten liquid

Applying the reduced phase rule to such two-phase equilibrium, we have

$$P = 2,\ C = 2, \text{ therefore } F = 2 + 1 - 2 = 1\ (T \text{ or } C)$$

Hence, the curves AC and BC in a lead–silver system are 'univariant', that is, the composition changes with temperature without disturbing the two-phase equilibrium.

(c) Eutectic of lead and silver

In the phase diagram of lead–silver, point 'C' denotes an 'eutectic point'. It is evident that all the three phases, solid lead, solid silver and molten liquid coexists at a definite composition and temperature.

Applying the reduced phase rule,

$$P = 3,\ C = 2, \text{ therefore } F = 2 + 1 - 3 = 0$$

Therefore, the eutectic point, C, of a lead–silver system is 'invariant'.

'Eutectic point C is a state of two components, Pb and Ag, a homogeneous solid of definite composition which melts at the lowest temperature (303°C)'.

'In other words, a homogeneous solid mixture of two components Pb and Ag of constant composition (2.6 percent Ag and 97.4 percent Pb) melts at definite lowest temperature of 303°C. This is known as the 'eutectic' of lead and silver'.

It can be seen in the phase diagram that the temperatures below the tie line DD' corresponding to the lowest temperature (303°C), only solid phases coexist. For composition between A and C, only solid Pb and solid eutectic crystallizes out and for composition between BC, only solid Ag and solid eutectic exists. For composition exactly at C, only a solid of eutectic (2.6 percent Ag + 97.4 percent Pb) composition crystallizes out. This eutectic of Pb and Ag melts exactly at 303°C.

14.5.4 Desilverization of Lead

Pattinson's process

A knowledge of the phase diagram of lead and silver system (14.5) is useful to understand the concept of Pattinson's process of desilverization of lead and the enrichment of silver content in argentiferrous lead.

In the phase diagram of the Pb and Ag system, the curve AC represents the decrease of m.p. of Pb with different amounts of silver, the curve AC is, thus, univariant (Fig. 14.5).

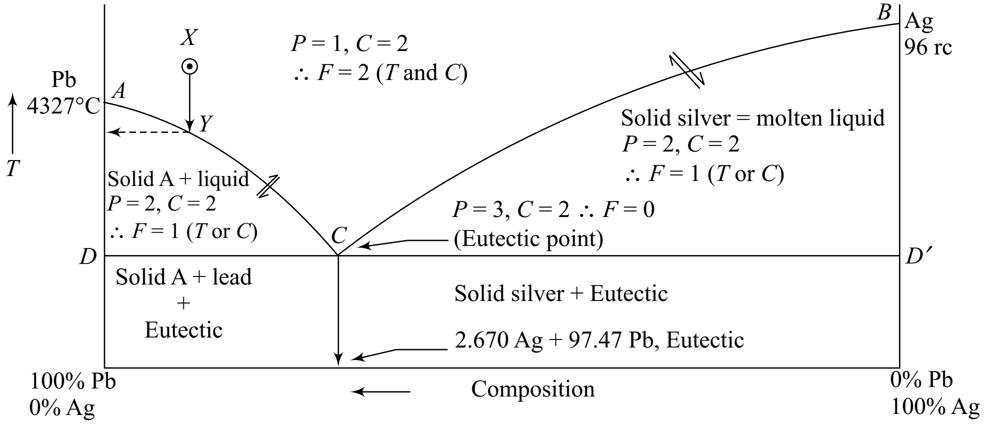


FIGURE 14.5 *Phase diagram of the Ag–Pb system and desilverization of lead.*

'In other words, lowering of the temperature of the molten liquid results in the change in the composition of the two components till the constant composition, eutectic *C*'. This principle is used in the Pattinson's process.

If a molten liquid of argentiferrous lead of composition 'X' (having 0.1 percent Ag) is allowed to cool gradually; solid Pb crystallizes out at the point 'Y' (at *T*′) on the AC curve and the composition of the argentiferrous lead changes along 'YC' with the enrichment of Ag. The crystallized Pb is constantly removed by wooden ladles. The melt becomes continuously richer and richer in the silver content until the eutectic point *C* is reached where the content of Ag is 2.6 percent.

The process of increasing the relative proportion of silver in argentiferrous lead is known as 'Pattinson's process'.

14.6 Two-Component Congruent Melting Compound

14.6.1 Tin–Magnesium System

An example of congruent melting compound is the tin–magnesium system (Fig. 14.6). A compound Mg_2Sn with congruent melting point is formed when Mg is added to tin or vice versa. The melting point of pure tin (P) is 232°C and of pure Mg (W) is 651°C.

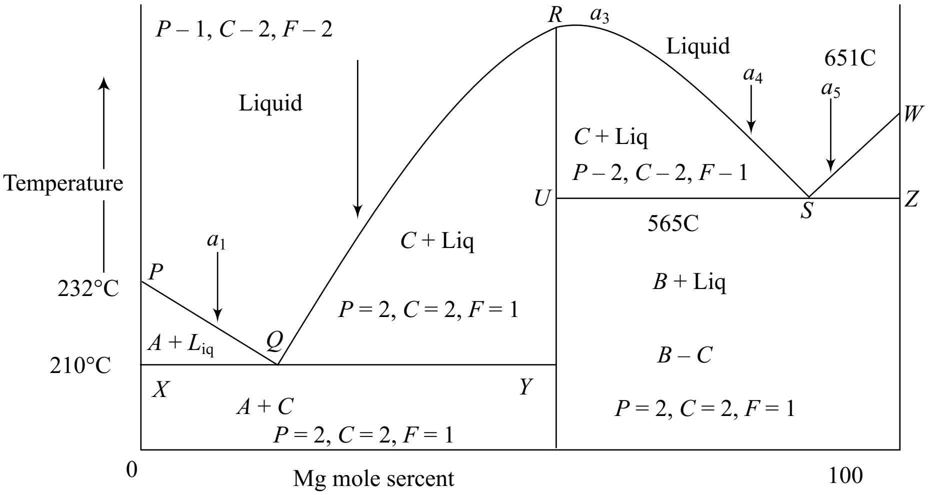


Figure 14.6 *Phase diagram of the magnesium–tin congruent melting system.*

As magnesium is gradually added to tin the melting point of the latter decreases to a minimum of 210°C (point *Q*), and thereafter increases, ultimately reaches a maximum of 783°C (point *R*), then decreases up to point *S*, finally increasing to W. The whole liquidus curve in the phase diagram of Tin and magnesium is, therefore, PQRSW. Above this liquidus curve, the system is completely homogenous liquid of two ($P = 1$). i.e. $F = 2 + 1 – 1 = 2$; it is bivariant and therefore, $f(T$ and $C)$.

A continued addition of Mg to tin or vice versa, a congruent melting compound is formed at point '*R*' having the composition Mg_2Sn, which melts at 783°C. It may be observed from the diagram that the line RY has divided the whole diagram into the diagrams of two simple eutectic systems, namely,

- Sn and Mg_2Sn, and
- Mg_2Sn and Mg.

The two systems are characterized by their own liquidus and solidus curves, as also by their own eutectic compositions. Thus, there are two solidus curves XY and UZ with two eutectic points Q(210°C) and S(565°C) belonging to two systems, respectively. If A = pure Sn, B = pure Mg and $C = Mg_2Sn$, then the regions:

1. PQX A + liquid, $P = 2, C = 2,$ therefore, $F = 2 + 1 - 2 = 1$. $f(T$ or $C)$.
2. RQY C + liquid, $P = 2, C = 2,$ therefore, $F = 2 + 1 - 2 = 1$. $f(T$ or $C)$.
3. RSU C + liquid, $P = 2, C = 2,$ therefore, $F = 2 + 1 - 2 = 1$. $f(T$ or $C)$.
4. WSZ B + liquid, $P = 2, C = 2,$ therefore, $F = 2 + 1 - 2 = 1$. $f(T$ or $C)$.

Below XY, we get a mixture of two solids A and C ($P = 2$) and below UZ a mixture of B and C ($P = 2$). Thus, no eutectic is formed between A and B. If one takes a mixture of two metals and tries to separate them by fractional crystallization of the molten mixture, one would not be able to obtain the pure metals A and B; instead either A and C, or B and C shall be obtained. In any case, C separates out on cooling the mixture on either side of the line RY.

Upon cooling a molten mixture of the two liquids on either side of point 'R', we get the separation of one of the components (either tin or magnesium) as solid. Which component is separated is determined by the composition of the initial mixture. For example, from the melts corresponding to a_1, a_2, a_4 and a_5, the components that separate out as solid are A (tin), C (Mg_2Sn), C(Mg_2Sn) and B(Mg), respectively. At a_3 the pure compound Mg_2Sn is obtained. The eutectic points Q and S represent A(Tin) + C(Mg_2Sn), and B(Mg) + C(Mg_2Sn), respectively. The number of degrees of freedom (F) for different parts of the diagram may be determined as follows.

Above the liquidus, the system has one phase, and $F = C - P + 1 = 3 - 1 = 2$ so the system is bivariant, i.e. $f(T$ and $C)$, and within the regions the system has two phase and is univariant, $f(T$ or $C)$, and so $F = 1$. At the eutectic, the number of phases ($P = 3$) is three and so $F = 0$, i.e. it is invariant.

The special feature of the diagram is the hump in the intermediate region. Any hump which divides the diagram into two parts corresponds to the formation of a definite compound with a definite melting point. If, however, the hump is not well formed then it indicates the formation of a compound with an incongruent melting point. Such a compound breaks up before reaching its normal melting point.

14.7 Two-Component Incongruent Melting Compound Formation

14.7.1 Mg_2SiO_4–SiO_2 System

For incongruent melting system, consider the system Mg_2SiO_4 (forsterite)–SiO_2 (silica), which has an intermediate compound, $MgSiO_3$ (enstatite).

Incongruent melting*—melting wherein a phase melts to a liquid with a composition different from the solid and produces a solid of different composition to the original solid.*

A simplified version of the system Mg_2SiO_4–SiO_2 with its intermediate compound (a phase that has a composition intermediate between two other phases) $MgSiO_3$ is shown in Fig. 14.7.

The crystallization histories for compositions X, Y and Z will be documented in the following discussion. Each of these compositions behaves in a slightly different manner.

Areas in the phase rule diagram

1. Area APEB, $P = 1, C = 2,$ and $F = 2$. The single phase is bivariant. $f(T$ and $C)$.
2. Area APC, $P = 2, C = 2,$ and $F = 1$. The phase is monovariant. $f(T$ or $C)$.
3. Area BEF, $P = 2, C = 2,$ and $F = 1$. The phase is monovariant. $f(T$ or $C)$.
4. Area PED, $P = 2, C = 2,$ and $F = 1$. The phase is monovariant. $f(T$ or $C)$.
5. Area CJKO, $P = 2, C = 2,$ and $F = 1$. The phase is monovariant. $f(T$ or $C)$.
6. Area KDF, $P = 2, C = 2,$ and $F = 1$. The phase is monovariant. $f(T$ or $C)$.

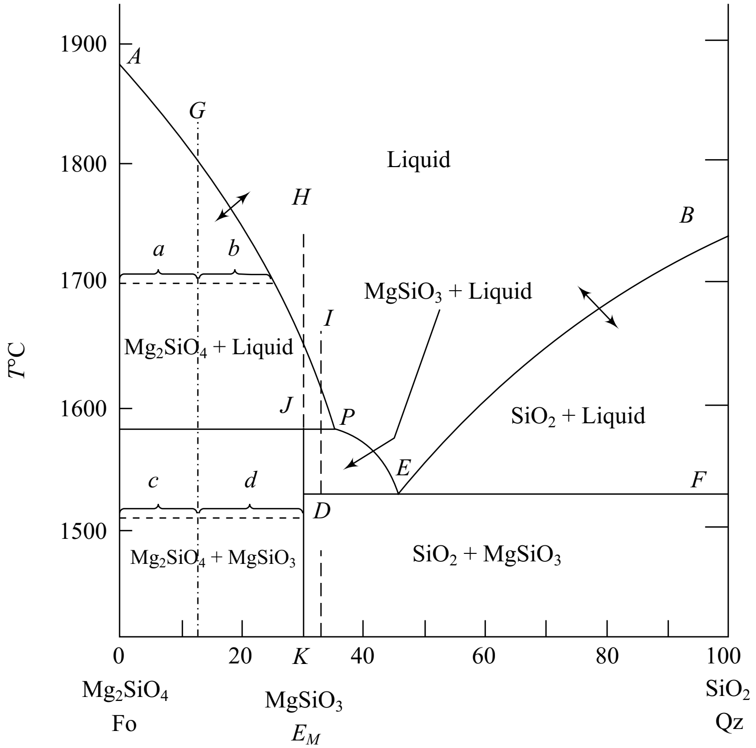


FIGURE 14.7 *Phase diagram of Mg_2SiO_4–SiO_2.*

Curves AP, PE and EB

1. Curve AP:
   $P = 2$, $C = 2$ and $F = 1$: The phase is monovariant. $f(T$ or $C)$.
   Solid Mg_2SiO_4 is in equilibrium with liquid phase.
2. Curve PE:
   $P = 2$, $C = 2$ and $F = 1$: The phase is monovariant. $f(T$ or $C)$
   Solid $MgSiO_3$ is in equilibrium with liquid phase.
3. Curve EB:
   $P = 2$, $C = 2$ and $F = 1$: The phase is monovariant. $f(T$ or $C)$.
   Solid SiO_2 is in equilibrium with liquid phase.

Points *P* and *E*

All three phases are in equilibrium at points *P* and *E*.
$F = 2 + 1 - 3 = 0$ and $F = 0$: Peritectic is an invariant point.

- Point '*P*' = peritectic point, Mg_2SiO_4 (Fo), liquid and $MgSiO_3$ (En).
- Point '*E*' = eutectic point, liquid, SiO_2 and $MgSiO_3$.

Crystallization of Composition

Composition 'G' is a mixture of 13 wt percent SiO_2 and 87 wt percent Mg_2SiO_4. Because this composition falls between the compositions of pure Mg_2SiO_4 and pure $MgSiO_3$, it must end its crystallization history containing only crystals of Mg_2SiO_4(Fo) and $MgSiO_3$(En). That is, no quartz will occur in the final crystalline mixture.

(i) Areas in the phase diagram

If a mixture such as composition 'G' is taken to a temperature above its liquidus (i.e. above 1800°C in Fig. 14.7) it will be in an all liquid state.

$$P = 3,\ C = 2,\ F = 2 + 1 - 3 = 0,\ P = 1,\ C = 1 \text{ and therefore, } F = 2 + 1 - 1 = 2;$$

The single phase area above APEB is bivariant, i.e. $f(T$ and $C)$.

(ii) Curves

Curve AP,	$Mg_2SiO_4 \leftrightarrow$ Liquid,	$P = 2, C = 2, F = 2 + 1 - 2 = 1,$	$f(T$ or $C)$.	Monovariant
Curve PE,	$MgSiO_3 \leftrightarrow$ Liquid,	$P = 2, C = 2, F = 2 + 1 - 2 = 1,$	$f(T$ or $C)$.	Monovariant
Curve EB,	$SiO_2 \leftrightarrow$ Liquid,	$P = 2, C = 2, F = 2 + 1 - 2 = 1,$	$f(T$ or $C)$.	Monovariant

(iii) Points

- Point P— Peritectic melting point. $P = 3, C = 2, F = 2 + 1 - 3 = 0,$ invariant point.
- Point E— eutectic melting point $P = 3, C = 2, F = 2 + 1 - 3 = 0,$ invariant point.

We now trace the cooling history of composition 'G'. As a liquid of composition G is cooled, nothing will happen until the temperature is equal to the liquidus temperature at 1800°C. At this point and onwards along curve AP, Mg_2SiO_4 (F_o) begin to crystallize out of the liquid ($P = 2$). As the temperature is further lowered, the composition of the liquid will change along the liquidus towards the peritectic (P), and the crystals forming from the liquid will always be pure Mg_2SiO_4 (F_o) until P is reached.

At peritectic point (P) all the three phases coexist.

Peritectic point (P): *The point on a phase diagram where a reaction takes place between a previously precipitated phase and the liquid to produce a new solid phase. When this point is reached, the temperature must remain constant until the reaction has run to completion. A peritectic is also an invariant point.*

At the temperature of the peritectic (P), about 1580°, all the three phases are in equilibrium, Liquid $\leftrightarrow$ Mg_2SiO_4 (F_o) $\leftrightarrow$ $MgSiO_3$ (E_n).

At this point, some of the crystals of Mg_2SiO_4 (F_o) react with the liquid to produce crystals of $MgSiO_3$ (E_n). The reaction that takes place can be written as follows:

$$Mg_2SiO_4\ (F_o) + SiO_2 \rightarrow 2MgSiO_3\ (E_n)$$

After all of the liquid is consumed by this reaction, only crystals of Mg_2SiO_4 (F_o) and $MgSiO_3$ (E_n) will remain. The proportions of Fo and En in the final crystalline product can be found by applying the lever rule.

- Percentage of Mg_2SiO_4crystals = $[d/(c + d)] \times 100$
- Percentage of $MgSiO_3$ crystals = $[c/(c + d)] \times 100$

At any intermediate stage in the process, such as at 1700° the proportion of all phases present (Fo and liquid in this case) can similarly be found by applying the lever rule.

At 1700°C

- Percentage of Mg_2SiO_4 crystals = $[b/(a + b)] \times 100$
- Percentage of liquid = $[a/(a + b)] \times 100$

Note that melting of composition G is exactly the reverse of crystallization. Mixture X will begin to melt at the peritectic temperature. At this point $MgSiO_3$ will melt to crystals of Mg_2SiO_4 (Fo) plus liquid (incongruent melting). As soon as all of the $MgSiO_3$ crystals have been consumed by this reaction, the temperature can be increased until it reaches 1800° at which point all of the Mg_2SiO_4(Fo) crystals will have been consumed and the only phase left will be liquid with a composition of the starting material.

(A) Crystallization of composition H

Composition 'H' is equivalent to pure $MgSiO_3$ (E_n). Thus, only E_n may appear in the final crystalline product if perfect equilibrium is maintained.

If composition 'H' is cooled from an all liquid state it first begins to crystallize at about 1650°. At 1650° crystals of Mg_2SiO_4 will begin to precipitate from the liquid. This will continue with further cooling until the temperature of the peritectic (P) is reached. In this interval, the composition of the liquid must become more enriched in SiO_2 and will thus change along the liquidus until it has the composition of the peritectic, P. At the peritectic temperature (1580°) all of the remaining liquid will react with all of the previously precipitated Mg_2SiO_4 to produce crystals of $MgSiO_3$. The temperature will remain constant until this reaction has gone to completion, after which the only phase present will be pure $MgSiO_3$.

Thus, it can be seen that $MgSiO_3$ melts incongruently. If pure $MgSiO_3$ is heated to a temperature of 1580° it melts to Mg_2SiO_4 plus liquid.

(B) Crystallization of composition I

Since composition 'I' lies between $MgSiO_3$ and SiO_2, it must end up with crystals of $MgSiO_3$ and SiO_2 (Quartz). If such a composition were cooled from some high temperature where it is in the all liquid state, it would remain all liquid until it reached the liquidus temperature at about 1600°. At this temperature crystals of Mg_2SiO_4 would begin to precipitate and the composition of the liquid would begin to change along the liquidus towards the peritectic, *P*. At *P*, all of the Fo previously precipitated would react with the liquid to produce crystals of $MgSiO_3$ (E_n). After this reaction has run to completion, and all of the previously precipitated Mg_2SiO_4 is consumed, there would still remain some liquid. With decreasing temperature, more crystals of En would form, and the liquid composition would change along the liquidus toward the eutectic, *E*. At *E* crystals of SiO_2 (Qz) would begin to form, the temperature would remain constant until all of the liquid was used up, begin to form, the temperature would remain constant until all of the liquid was used up, leaving crystals of SiO_2 and $MgSiO_3$ as the final solid. Note that because composition Z lies very close to the composition of pure $MgSiO_3$, the final crystalline product would consist mostly of $MgSiO_3$ with a very small amount of SiO_2.

For all compositions between *P* and 100 percent SiO_2 the system would behave in an identical fashion to the simple eutectic system discussed previously.

14.8 Two-Component Formation of Solid Solution Systems

In the systems we have discussed so far, all solid phases have been pure phases, i.e. they have one and only one possible composition (Fig. 14.8). This is not usually the case in nature, *since substitution of one element for another often occurs due to the fact that some elements behave in a chemically similar fashion to other elements. When such substitutions occur, the phase can have a range of possible compositions, depending on the amount of substitution that takes place. Such solids that can have various amounts of elemental substitution are called solid solutions.*

A good example of a solid solution mineral is the mineral *olivine.* The general chemical formula for olivine is $(Mg, Fe)_2SiO_4$. Since Mg^{+2} and Fe^{+2} are about the same size and have the same charge, they may substitute for one another in the crystal structure of olivine. Thus, olivine may have a composition anywhere between the pure Mg end member, forsterite (Mg_2SiO_4) and the pure Fe end member, fayalite (Fe_2SiO_4). When all compositions between two end members are possible, the solid solution is said to be a complete solid solution.

Another good example of a complete solid solution is displayed in the *plagioclase feldspars*. In this case the solid solution is between the end members $NaAlSi_3O_8$ (albite) and $CaAl_2Si_2O_8$ (anorthite). In order to maintain charge balance we cannot simply substitute Na^+ for Ca^{+2}, so this solid solution is what is called a

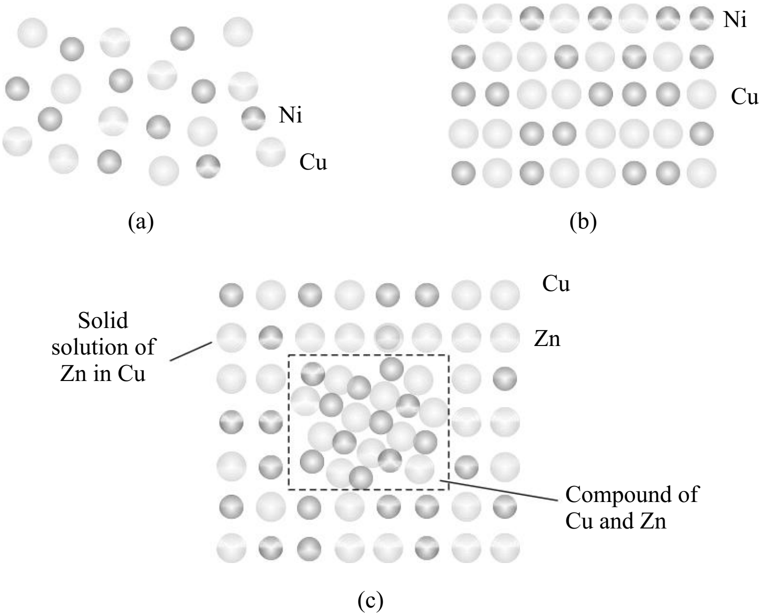


FIGURE 14.8 *Examples of solid solutions of two components.*

coupled solid solution. In this case, Na^+Si^{+4} is substituted for $Ca^{+2}Al^{+3}$ in the plagioclase structure to produce intermediate compositions of plagioclase.

14.8.1 $NaAlSi_3O_8$–$CaAl_2Si_2O_8$ System

Because the elements that substitute are not exactly the same size (they are similar in size) the amount of substitution is dependent on temperature and pressure and the solid solutions behave in a somewhat orderly fashion as illustrated below.

Since plagioclase is one of the most common minerals in the earth's crust, we will discuss the phase diagram for the plagioclase system. The phase relations in the plagioclase system are shown in Fig. 14.9 at constant pressure equal to that of the atmosphere (atmospheric pressure is 1 bar). In Fig. 14.9, the upper curve is called the liquidus and the lower curve is called the solidus. At temperatures above the liquidus everything is liquid, below the solidus everything is solid (crystals of plagioclase solid solution). At temperatures between the solidus and liquidus crystals of plagioclase solid solution coexist in equilibrium with liquid.

Pure $NaAlSi_3O_8$ (albite) melts (or crystallizes) at 1118°C, and pure $CaAl_2Si_2O_8$ (anorthite) melts (or crystallizes) at 1500°C. Note that any composition of plagioclase between the two end members melts or crystallizes over a range of temperatures unlike the pure end members which have only one melting point. Thus, we can read from the diagram that a solid solution containing 50 percent albite and 50 percent anorthite ($Ab_{50}An_{50}$) begins to melt at 1220°, point *F*, and the melting is complete at 1410°, point A. Inversely, if a melt of composition $Ab_{50}An_{50}$ is cooled it will begin to crystallize at 1410° and will be completely crystalline at 1220°.

We will now trace the crystallization history of composition X, which is $Ab_{50}An_{50}$. Composition X is completely liquid above the liquidus curve XECY (above 1410°).
i.e. $P = 1$, $C = 2$ and therefore, $F = 2 + 1 - 1 = 2$. Liquid single phase is bivariant, $f(T \text{ and } C)$.

When this liquid phase is cooled to liquidus at point *A*, a solid solution of An in Ab crystallizes out. The liquidus curve XECY defines the equilibrium of two phases, solid solution crystals and liquid of corresponding

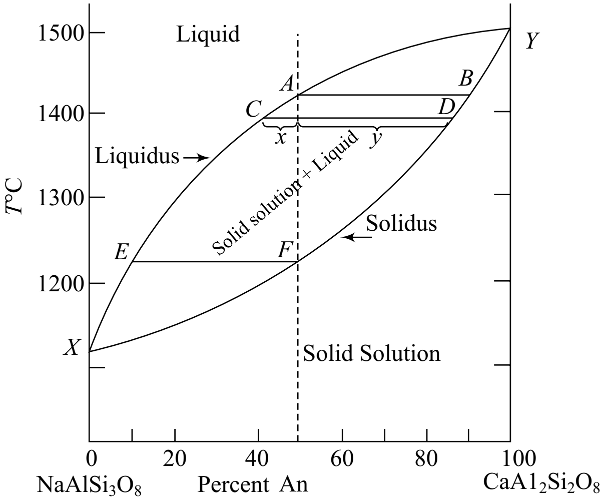


Figure 14.9 *Phase diagram of the $NaAlSi_3O_8$–$CaAl_2Si_2O_8$ system.*

compositions. For these equilibria, $P = 2$, $C = 2$ and $F = 2 + 1 - 2 = 1$. These crystals that are in equilibrium with liquid phase in the system is univariant, $f(T \text{ or } C)$.

The composition of these solid solution crystals can be found by drawing an isotherm (line of constant temperature, a horizontal line in this diagram) through the temperature 1410°. Where this isotherm intersects the solidus (at point *B*), the composition of the solid can be found by drawing a vertical line to the base of the diagram. Thus, it is seen that the first crystals precipitated from composition X will have the composition $Ab_{10}An_{90}$. Note that in this diagram crystals that are in equilibrium with liquid will always be enriched in anorthite component relative to the liquid. As crystallization continues with lowering of temperature the composition of the plagioclase will change along the solidus, continually reacting with the liquid to produce crystals more enriched in the Ab component. Meanwhile, the composition of the liquid will change along the liquidus, thus also becoming more enriched in the Ab component. At a temperature of 1395° the liquid composition will be at point *C*, while the solid composition will be at point *D*. Crystallization proceeds until a temperature of about 1220°, at which point the last remaining liquid will have a composition at *E*, and the solid will have a composition equal to the original starting composition at point *F*.

At this point all of the liquid will be consumed and the final crystalline product will have the composition $Ab_{50}An_{50}$.

During crystallization, the proportion of the solid continually increases while that of the liquid continually decreases. Thus, as the composition of the liquid becomes more sodic, approaching *E*, its volume steadily decreases. Thus, it can be seen that the amount of liquid in equilibrium with the solid of composition *F* will be extremely small.

If at any point during the crystallization, we wish to determine the amount of solid and liquid, we can apply the lever rule. As an example, we will determine the proportions of liquid and solid in the system at a temperature of 1395°. At this point, we measure the distances *OC*, *OD* and *CD*. The percentages of liquid and solid are then given as follows:

- percent solid (with composition *D*) = $[x/(x + y)] \times 100$
- percent liquid (with composition *C*) = $[y/(x + y)] \times 100$.

14.8.2 Silver–Gold System

The silver and gold have similar crystallographic characteristics and ionic radii, since they belong to the same group in the periodic table. They are completely miscible and form solid solutions of any composition. The phase diagram of the system is as shown in Fig. 14.10.

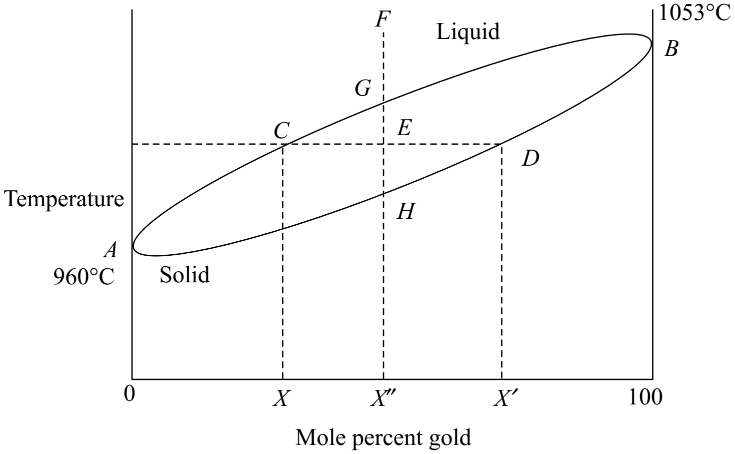


FIGURE 14.10 *Phase diagram of the silver–gold system.*

Pure silver and gold melts at 960.5°C (point *A*) and 1053°C (point *B*), respectively. In the phase diagram, curve AHDB is solidus, while ACGB is liquidus curve.

ACGB is a liquidus curve in the phase diagram, which is a single phase ($P = 1$) and hence, only liquid phase can exist, if $F = C + 1 - 1 = 2$, i.e. $f(T$ and $C)$. Again, AHDB is a solid phase curve which exists as a single phase, if $F = 2 + 1 - 1 = 2$, i.e. $f(T$ and $C)$. Any point inside ACBDA represents equilibrium between the solid crystals and the liquid, the relative amounts of solid and liquid being given by solid : liquid = CE : ED. If we take a liquid solution of composition at a temperature given by *F* and go on cooling the same, solid phase appears at *G*. Further fall of temperature results in a change in composition of the liquid phase, the change takes place along *GA*. The solution becomes richer in Ag. If gold is added to liquid silver at particular temperature, the system remains liquid until the composition reaches '*x*' at point *C*. Further addition of gold will create a solid phase in addition to the liquid phase. Increase of concentration of gold increases the amount of solid phase and the system becomes completely solid after crossing *D*. Fractional crystallization may be employed to separate the pure components from the mixture.

14.9 System Forming Partially Miscible Solid Solutions

14.9.1 Iron–Carbon System

A study of the constitution and structure of all steels and irons must first start with the iron–carbon equilibrium diagram. Many of the basic features of this system (Fig 14.11) influence the behaviour of even the most complex alloy steels. For example, the phases found in the simple binary Fe–C system persist in complex steels, but it is necessary to examine the effects alloying elements have on the formation and properties of these phases. The iron–carbon diagram provides a valuable foundation on which to build knowledge of both plain carbon and alloy steels in their immense variety.

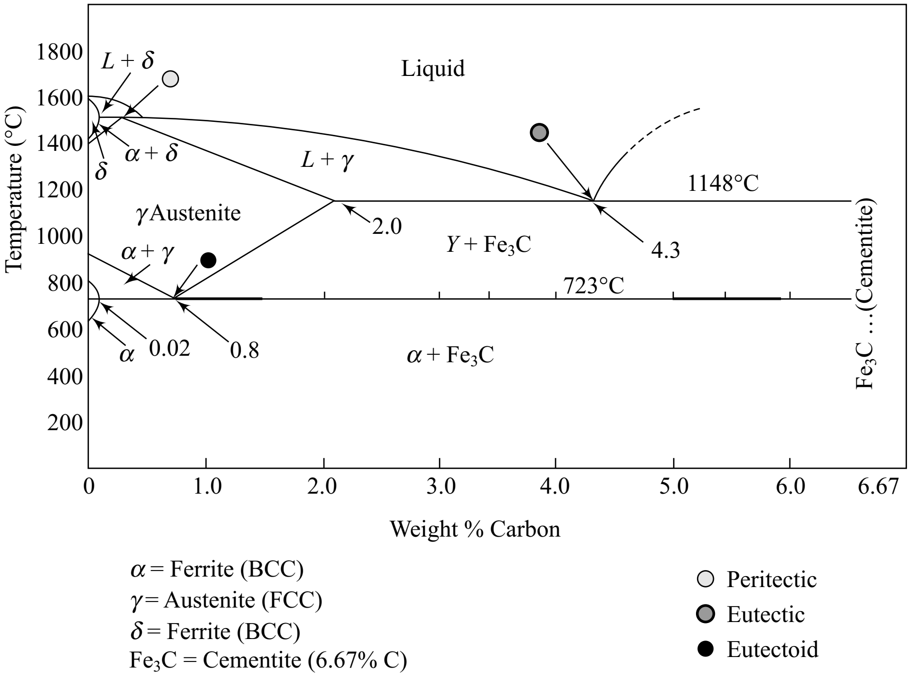


Figure 14.11 *Phase diagram of iron–carbon indicating eutectic and peritectic points.*

The left side of the diagram represents pure iron and the right hand of the diagram represents an alloy (cementite) with 6.67 percent *C* which is obtained on cooling. This is an intermetallic compound (iron carbide (Fe_3C)) which although not 100 percent stable but is to all practical purposes a stable phase. The phase diagram shown is therefore a metastable phase.

(i) Iron can exist in three forms

(a) BCC crystal exists at temperatures up to 910°C
(b) FCC crystal exists at temperature range 910°C–1403°C
(c) BCC crystal exists at temperature range 1403°C–1535°C

(ii) Classification of ferrous alloys

Iron: less than 0.008 wt percent *C* in α-ferrite at room *T*.
Steels: 0.008–2.14 wt percent *C* (usually < 1 wt percent), α-ferrite + Fe_3C at room *T*.
Cast iron: 2.14–6.7 wt percent (usually < 4.5 wt percent).

(iii) Different phases of the iron–carbon system

The phased diagram includes four solid phases and one homogeneous liquid phase.

1. α-iron or ferrite

α-ferrite –

- A solid solution of *C* in BCC Fe.
- Stable form of iron at room temperature.
- The maximum solubility of *C* is 0.022 wt percent at 723°C. At 0°C temperature the solubility falls to 0.008 percent. The carbon atoms are located in the crystal interstices.

- Transforms to FCC γ- austenite at 912°C.
- Magnetic properties: α-ferrite is magnetic below 768°C.

2. γ- *iron or austenite*

- The solid solution of carbon in γ iron is called austenite.
- This has an FCC crystal structure with a high solubility for carbon as compared to α-ferrite.
- The solubility reaches a maximum of 2.08 percent at 1148°C. The solubility decreases to 0.8 percent at 723°C The carbon atoms are dissolved interstitially.
- Transforms to BCC δ-ferrite at 1395°C.
- Is not stable below the eutectic temperature.
- The difference in solubility between the austenite and α-ferrite is the basis for the hardening of steels
- Austenite is nonmagnetic.

3. δ-*iron or ferrite*

- This is a solid solution of carbon in iron and has a BCC crystal structure.
- The same structure as α-ferrite.
- The maximum solubility or *C* in Fe is 0.09 percent at 1495°C.
- Stable only at high *T*, above 1394°C.
- Melts at 1538°C.

4. *Cementite, Fe_3C*

This is an intermetallic compound which contains 6.67 percent *C* and 93.3 percent Fe and is metastable, it remains as a compound indefinitely at room *T*, but decomposes (very slowly, within several years) into α-Fe and *C* (graphite) at 650–700°C. Cementite is a hard brittle compound having orthorhombic crystal structure. Each unit cell has 12 Fe atoms and 4 C atoms.

Mechanical properties: Cementite is very hard and brittle, which can strengthen steels. Mechanical properties also depend on the microstructure, that is, how ferrite and cementite are mixed.

(iv) The phase diagram

Therefore, from the reduced phase rule: $F = C + 1 - P$

Single phases

1. Area AGA1 = α-ferrite; $P = 1$, $C = 2$, therefore $F = 2 + 1 - 1 = 2$ or $f(T \text{ and } C)$.
2. Area A1DBCK = γ- austensite; $P = 1$, $C = 2$, therefore $F = 2 + 1 - 1 = 2$ or $f(T \text{ and } C)$.
3. Area IKJ = ð-ferrite; $P = 1$, $C = 2$, therefore $F = 2 + 1 - 1 = 2$ or $f(T \text{ and } C)$.

The above-mentioned areas are, therefore, bivariant and denoted as $f(T \text{ and } C)$.

Two phases

1. Area ADG = solid $\alpha + \gamma$- austensite; $P = 2$, $C = 2$, therefore $F = 2 + 1 - 2 = 1$ or $f(T \text{ or } C)$.
2. Area IKCBE = γ- austensite + liquid; $P = 2$, $C = 2$, therefore $F = 2 + 1 - 2 = 1$ or $f(T \text{ or } C)$.
3. Area BDHF = γ- austensite + Fe_3C; $P = 2$, $C = 2$, therefore $F = 2 + 1 - 2 = 1$ or $f(T \text{ or } C)$.
4. Area EFL = Fe_3C + liquid; $P = 2$, $C = 2$, therefore $F = 2 + 1 - 2 = 1$ or $f(T \text{ or } C)$.
5. Area AGHM = solid α-ferrite + Fe_3C; $P = 2$, $C = 2$, therefore $F = 2 + 1 - 2 = 1$ or $f(T \text{ or } C)$.

All the lines or curves observed in the phase diagram are monovariant and therefore dependent on either temperature (*T*) or composition (*C*).

i.e. $F = 2 + 1 - 2 = 1$, therefore $f(T \text{ and } C)$.

All *points* in the phase diagram denote invariant points since all the three phases coexist.

i.e. $F = 2 + 1 - 3 = 0$, therefore $f = 0$, i.e. invariant.

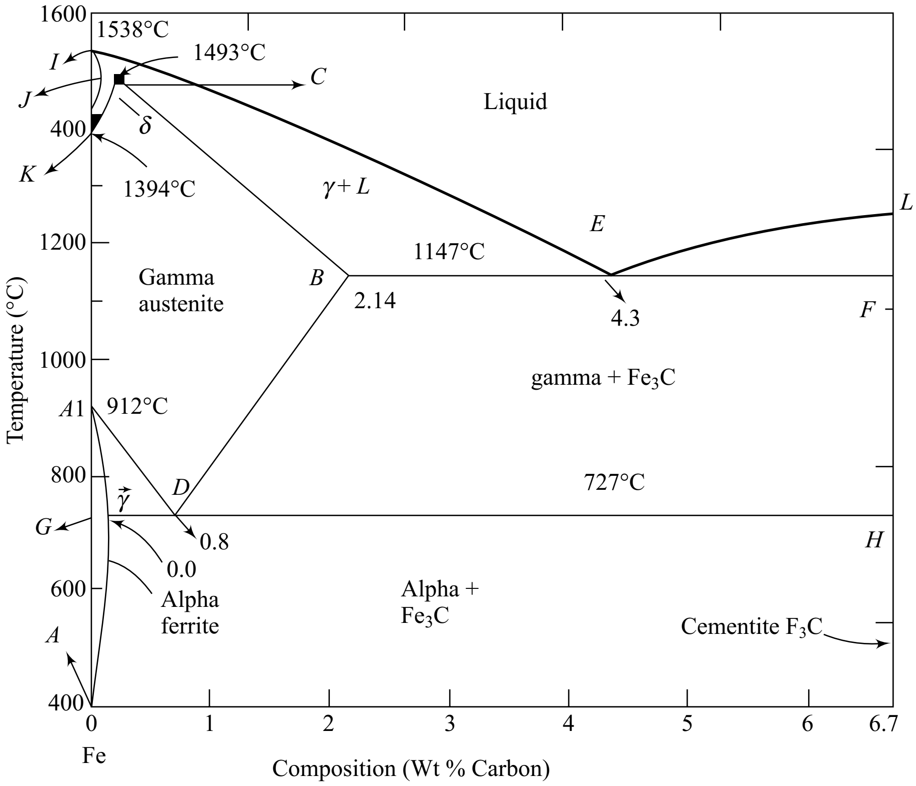


FIGURE 14.12 *(a) Iron–carbon phase diagram—areas indicating all the possible phases.*

Points

At C = peritectic point
D = eutectic point– pearlite phase
E = eutectic point – ledeburite phase and
M = cementite phase at 6.7 percent of carbon.

Sometimes the constituents of a mixture may react to form a new substance, which has significant effects on the phase diagram. When carbon is dissolved in iron, the hard, brittle, white compound *cementite*, Fe_3C, may be formed. On the phase diagram, cementite corresponds to a vertical line at 6.7 percent C, and divides the phase diagrams into two independent sections, one with less carbon and the other with more.

Pure iron has a body-centred cubic crystal structure at room temperature, called α-iron or *ferrite*. At 910°C the structure changes to face-centred cubic, and this material is called γ-iron or *austenite*. At 1403°C, the structure reverts to bcc, and the material is called δ-iron, though it is really the same as α-iron. At 1535°C, the iron melts. Austenite is much softer and more easily worked than ferrite, which explains the blacksmith's heating of iron to red heat for working it.

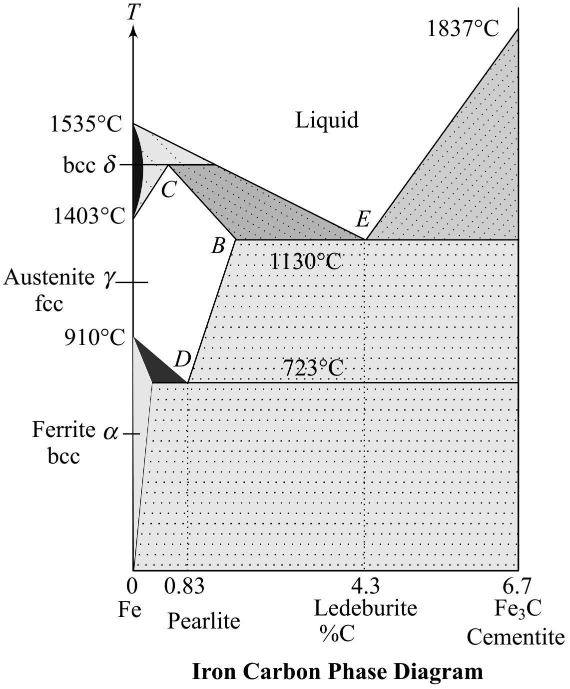
Two very important phase changes take place at 0.83 percent C and at 4.3 percent C.

At 0.83 percent C, the transformation is eutectoid at D, called *pearlite*.

$$\gamma \text{ austenite} \rightarrow \text{alpha} + Fe_3C \ (\textit{cementite})$$

At 4.3 percent C, the transformation is eutectic at E, called *ledeburite*.

$$\text{L(liquid)} \rightarrow \gamma\text{-austenite} + Fe_3C \ (\textit{cementite})$$



Iron Carbon Phase Diagram

Figure 14.12 *(b) Iron–carbon phase diagram.*

In Fig. 14.12(b), Point *E* is easily recognized as an *eutectic.* Its temperature is the lowest fusing temperature of an iron–carbon mixture, 1130°C. An eutectic liquid will solidify to ledeburite, a typical layered *eutectic solid of austenite and cementite.* A liquid to the left of *E* will solidify to austenite + ledeburite, while one to the right will solidify to cementite + ledeburite. At 723°C, there will be a further change to ferrite + cementite. These changes in the solid state are not rapid. Any material whose composition is to the right of point *B* (greater than 1.7 percent carbon) is called *cast iron*. When iron is reduced in the liquid state with carbon, cast iron is the usual result. In early times, when iron men were trying to get a low-carbon iron by directly reducing the iron ore with charcoal, their work would be ruined if the furnace temperature rose above 1130°C, and the iron absorbed carbon and melted. The cast iron could not be hammered or worked; it was just a hard, disappointing, inedible, useless lump. Using cast iron to make castings was a much later development, since for several reasons iron is not as easy to cast as bronze is.

Any material with a composition to the left of *B* is called *steel*. For such a material, heating will cause the steel to become a uniform austenite phase. This is a lucky circumstance that is responsible for the ability to heat-treat steel to suit its application. A steel of composition corresponding to point *D* will become a layered mixture of ferrite and cementite of average composition 0.83 percent carbon, called *pearlite*. It is so-called because it can have a pearly lustre in the microscope, a play of colours produced by the interference of light in its regular layers, which acts as a diffraction grating. This part of the phase diagram looks just like an eutectic, except that the upper material is solid austenite, not a liquid. It is called a *eutectoid*, 'like an eutectic' though there is no melting involved. A steel that at room temperature is a structural mess of different phases and crystals, but of the average composition of the eutectoid, will, on heating above 723°C, become a uniform solid solution after a while, that will cool to a nice pearlite structure throughout. This *normalization* can be applied to any steel, and does not require melting. The steel must be heated to a high enough temperature that it all becomes austenite, however, a temperature that is certainly below 1130°C.

There is a funny bit at high temperatures near the δ-iron region. This looks like an inverted eutectoid, but it is not eutectoid. If we consider a melt with a composition just to the left of point *C*, δ-iron will freeze out and the liquid will become richer in carbon. When the temperature of *C* is reached, the carbon-rich liquid will dissolve some of the carbon-poor δ-iron until all the liquid freezes to a mixture of δ-iron and austenite of the composition of point *C*. Further cooling will decrease the amount of δ-iron and increase the amount of austenite until the solid phase is uniform austenite. What happens at point *C* is called a *peritectic reaction*, as δ-iron is dissolved and austenite formed. The carbon–cementite phase diagram illustrates most of the interesting things that can happen!

Austenite cooled quickly does not become a tough ferrite–pearlite mixture, but a difficult structure called *martensite* that is very hard and brittle. It is as if the carbon, trying to escape as the bcc crystals of austenite shift to fcc, are trapped in the attempt, and hinder the structure change by wedging themselves in and preventing the formation of a stress-free crystal. If you are making a sword or a knife, this may be just what you want. The steel is first cooled slowly, so ferrite and pearlite are formed, and the item can be shaped and worked. Then it is heated above the transition temperature to austenite, and finally *quenched* by plunging it into water, or into the body of a slave on the advice of local priests. The temper can be *drawn* by gentle heating to allow some of the martensite to transform into ferrite, and toughen the object. In most cases, the hard temper is only on the surface anyway, since the interior cannot be quickly quenched. The possibility of heat treatment makes steel a far superior metal for tools and weapons than any other available alloy.

14.9.2 Lead–Tin System

It is more common to have systems of two metals that are miscible in the liquid state, but crystallize in different systems in the solid state, or have ions significantly different in size, so they cannot occupy the same lattice sites without considerable strain. Usually each crystal lattice can accommodate a few foreign ions and make a homogeneous solid solution phase, but when the concentration of the foreign ions reaches a certain limit that depends on the temperature, the phase may refuse to accept any more. Then we have a liquid state that is one phase, while the solid state is a two-phase mixture of the *terminal* phases. This happens in the tin–lead system. The terminal tin phase, α, can hold less than 3 percent lead, while the terminal lead phase, β, can hold no more than 20 percent, and that at an elevated temperature. The solid must, in general, be a mixture of the two phases, nearly pure tin, and nearly pure lead. After a while, they may diffuse into one another to approach homogeneity.

The phase diagram for the tin–lead system is shown in Fig. 14.13.

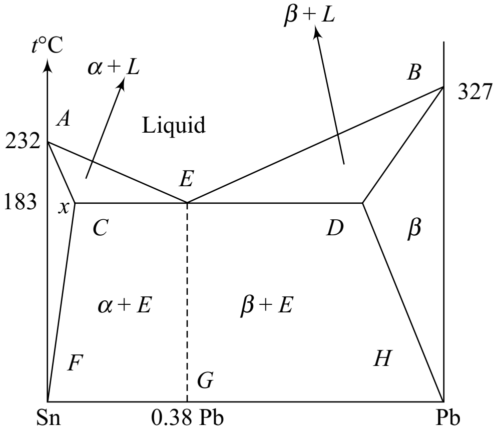


Figure 14.13 *Tin–lead phase diagram.*

1. Single phase

Area above *AEB*, single liquid phase, $P = 1$, $C = 2$ and $F = 2$ or $f(T$ and $C)$,
Therefore, it is bivariant.

Area *ACF* and *BDH* are single phases (α and β) of the solid solutions of tin in lead and vice versa in the phase diagram.

$P = 1$, $C = 2$ and $F = 2$ or $f(T$ and $C)$.

2. Lines *AE* and *EB*

They are the variation of melting point curves or lines of the respective pure phases with the addition of the other. *AE* represents the decrease of melting temperature of tin with the addition of lead whereas curve *EB* indicates the variation of the melting temperature of lead with the addition of tin.

Liquid ↔ α phase and β ↔ Liquid.

$P = 2$, $C = 2$; and $F = 1$ or $f(T$ or $C)$. Therefore, both the curves *AE* and *EB* in the phase diagram are univariant. Similarly, lines *AC*, *CF*, *BD* and *DH* are univariant.

3. Areas FCEG and GEDH phases

Within the area FCEG, only solid α + eutectic phases coexist and β + eutectic solids are found within area GEDH.

$P = 2$, $C = 2$; $F = 1$ or $f(T$ or $C)$ The above areas are univariant.

4. Eutectic point *E*

There are three possible phases, α, β and the liquid. The phase rule says that $f = 2 - 3 + 1 = 0$, so they can all three be present only at one point, which is point *E*. This point is at 183°C and a lead concentration of 38 percent by weight. It is called an *eutectic*, and represents the minimum melting point of the system. Three large regions correspond to areas with two phases: liquid + α, liquid + β and $\alpha + \beta$. The liquidus is composed of the two upper lines meeting at *E*. The solidus is represented by the horizontal line through *E*, i.e. terminated by triple points at each end, together with the lower lines from the melting points. The final two lines represent the limits of stability of the terminal phases.

Pure lead and tin melts at 327 and 232°C. Adding one to the other *lowers* the melting point in either case (lines *AE* and *BD*). Suppose a tin-rich liquid solidifies rather quickly. The first metal to solidify is nearly pure tin α phase. This makes the liquid richer in lead, so its concentration moves to the right. This continues until 183°C is reached. Then we have α phase solid, and a liquid of the eutectic composition. Now this liquid solidifies, depositing characteristic thin layers of alternating α and β phases as it departs from equilibrium one way and then the other, correcting itself every time.

Review Questions

1. What does the phase rule predict?
2. State the phase rule.
3. Define ‘phase’ of a system. Give examples.
4. Find the number phases involved in the equilibrium of $CaCO_3 \leftrightarrow CO_2 + CaO$.
5. Define the components of a system. Give examples.
6. Find the number of components in the decomposition of $CaCO_3$.
7. Give the number of components in the decomposition of $NH_4Cl(s)$.
8. Define the degrees of freedom of a system. Give examples.
9. How do you define the state of a system? Mention a few examples.
10. Discuss the application of the phase rule to a one-component water system.
11. Explain the significance of the curve *OB* in the water system.
12. Explain the significance of the curve *OA* in the water system.
13. Explain the significance of the curve *OC* in the water system.
14. What do you mean by a ‘triple point’ in the water system?
15. Describe the metastable equilibrium state in the water system.
16. What is the degree of freedom in a two-component system?
17. What do you mean by the reduced phase rule?
18. Describe the application of the reduced phase rule to a two-component system.
19. What is an ‘eutectic’ in a two-component system?
20. Define eutectic of lead–silver system.
21. Discuss the phase diagram of the lead–silver system.
22. Describe desilverization of lead or Pattinson’s process.
23. What is the principle involved in the silverization of lead.
24. Composition of the eutectic of Pb and Ag is 2.6 percent Ag and 97.4 percent Pb. Mention the number of phases involved at 303°C.
25. What is congruent melting point of a two-component system?
26. Discuss the phase diagram of the magnesium–tin congruent melting system.
27. What do you mean by incongruent melting?
28. Describe the phase diagram of incongruent melting compound formation with any example.
29. Discuss the phase diagram of the Mg_2SO_4–SiO_2 incongruent melting system.

30. What are the areas, lines and points describe the phase system of Mg_2SO_4–SiO_2 signify? Justify your answer.
31. What are eutectics and peritectics in the phase rule? Explain with the help of the phase rule.
32. Apply the phase rule to two components forming a solid solution. Illustrate with an example.
33. Discuss the phase diagram of $NaAlSi_3O_8$–$CaAl_2Si_2O_8$.
34. Write a note on the phase diagram of the silver–gold system.
35. Discuss the phase diagram of a system forming partially miscible solid solutions with special reference to the iron–carbon system.
36. Discuss the possible phase equilibrium of the iron–carbon system. Name the phases involved in each case.
37. Discuss the number of eutectics and peritectics in the iron–carbon phase diagram.
38. Describe the phase diagram of the lead and tin system.

15

Chemical Thermodynamics

Chapter Outline

Introduction. Concept of energy, heat and work in thermodynamics. Reversible isothermal work of an ideal gas. Terminology of thermodynamics—a system, surroundings. Types of systems—homogenous and heterogeneous. Thermodynamic properties of a system—thermodynamic state of a system, thermodynamic equilibrium of a state. Types of processes in thermodynamics—thermodynamic reversible and irreversible processes. First law of thermodynamics—energy change, heat and work relationship—internal energy. Thermodynamics problems. Enthalpy or heat content. Applications of first law of thermodynamics to different processes, heat capacity of system. Second law of thermodynamics and reversible processes, importance of the second law of thermodynamics, statement of the second law, entropy of a system, entropy changes, entropy change in irreversible process, entropy change in reversible process, entropy changes in an isothermal, isobaric reversible process, entropy changes in adiabatic reversible process, entropy change in isothermal expansion of a gas, entropy change with temperature change, entropy change of phase changes, entropy and thermodynamic equilibrium. Concept of free energy changes in thermodynamic processes, Helmholtz free energy, *A*, Gibb's free energy, *G*, Worked examples. The criterion of equilibrium between coexisting phases like ice and water, free energy change of processes with temperature and pressure, Gibb's Helmholtz equation, applications of Gibb's Helmholtz equation, Maxwell's relations in thermodynamics, Clapeyron–Clausius equation, applications, integrated form of Clapeyron–Clausius equation. Free energy change of a reaction at equilibrium: van't Hoff's isotherm, criteria for reversible and irreversible processes, calculation of Gibb's free energy change of reaction, worked examples. Thermochemistry—stoichiometry and energy balances, mole concept in thermochemistry, problems—thermochemical stoichiometric problems, thermochemical stoichiometry problems. Enthalpy of reaction—enthalpy change is an extensive property, enthalpy change of reaction at constant pressure and constant volume. Relation between the enthalpy change of reaction at constant pressure and constant volume, standard state enthalpies, worked examples. Enthalpies of formation, using enthalpies of formation, to calculate enthalpies of reaction under standard conditions. Enthalpies of combustion. Enthalpies of neutralization. Enthalpy of solution. Enthalpy of phase changes, bond enthalpy, Hess's law, effect of temperature on the heat or enthalpy change of reaction (Kirchhoff's equation)

15.1 Introduction

Innumerable physical and chemical processes are quite common in nature. In fact, these physical and chemical processes are invariably accompanied by energy changes. *Thermodynamics is the study of energy and its transformations; of particular interest is the relationship between heat and work and also the feasibility or spontaneity of chemical processes.*

A study of the interrelation of various forms of energy of systems (may be physical or chemical) under a set of conditions constitutes the subject of thermodynamics.

The three fundamental generalizations, generally known as the I, II and III law of thermodynamics, are of paramount importance to deduce and elucidate the following aspects of physical chemistry:

(i) van't Hoffs' law of dilute solutions,
(ii) Raoult's law of lowering of vapour pressure,
(iii) distribution law,
(iv) phase rule,
(v) the law of chemical equilibrium, and
(vi) the laws of thermochemistry.

The criteria of the feasibility or spontaneity of chemical reactions under a given set of conditions are explicitly explained by the application of the laws of thermodynamics. Also, thermodynamics determine the extent to which processes or chemical reactions can proceed before attainment of equilibrium. However, thermodynamic approach has few limitations. The laws of thermodynamics are applicable to substances (materials) of macroscopic aggregation or bulk; and are inapplicable for individual atoms or molecules of systems. The laws of thermodynamics predict the feasibility or spontaneity of reactions of systems but fail to account for the rate at which reactions or processes proceed at a given set of conditions.

15.2 Concept of Energy, Heat and Work in Thermodynamics

Energy is a state function of the system. Energy is defined as any property which can be produced from or converted into work. Energy is manifested as kinetic, potential, thermal, electrical, mechanical, chemical, etc. Kinetic energy is possessed by a body of a system by virtue of its motion. Potential energy is energy possessed by a body by virtue of its position in a field of force. If the attraction of atoms (or molecules) towards a more stable configuration, it is chemical potential energy; and so others.

Since we deal with chemical systems or processes, the energy change that occurs in chemical reactions is largely due to change in bond energy, that is, the change of potential energy that results from the breaking of bonds in reactants and formation of new bonds in products. Generally, these energy changes appear in the form of 'heat' which is liberated or absorbed in the systems. Energy that is transferred from one system to another by virtue of temperature difference is called '*heat'*.

'*Heat' is regarded as a form of energy, since most forms of work can be readily converted into heat, and heat can at-least partially be converted into work without some change in the system or its surroundings*.

The energy change accompanying the change of thermodynamic state of a system may be due to the performance of '*work*' on or by the system, and also due to the transfer of energy to or from the system in the form of '*heat*'. This statement has its own meaning and significance, especially, when gaseous systems are involved.

The accepted convention in calculation of 'energy transfer' in thermodynamics is that work done by a system, viz. the expansion work of a gaseous system (dw) is negative and the heat received (q) by a system is positive.

Work done in three dimensions involves a change in volume (especially for gaseous system). A positive amount of work done on a system will decrease its volume, i.e. $+dw$ is the infinitesimal work done on a system as its volume changes by $-dv$ and therefore

$$dw = -Pdv.$$

It is a matter of experience that when pressure–volume work is performed, the work done on the system by the surroundings is equal in magnitude but opposite in sign to the work done by

In the SI system, work, energy and heat are all expressed in joules [1 calorie = 4.184×10^7 erg].

15.3 Reversible Isothermal Work of an Ideal Gas

The work done reversibly and isothermally on a system by the surroundings can be stated in the integrated form as

$$w = -\int P(V_2 - V_1) = -\int P \cdot \Delta v$$

where V_1 is the initial volume of the system, V_2 is the final volume system and P is the pressure of the system or surroundings as the process occurs. The volume changes of the system are assumed to be equal in magnitude but opposite sign to the changes in the volume of the surroundings. The maximum work done by the system is equal in magnitude but opposite in sign to the work done on the system. That is, the work done by the system is $-w$.

$$-w = \int P \cdot \Delta v$$

There are two common ways in which isothermal reversible work can be done.

(i) *The pressure may vary as the change in volume occurs or*

(ii) *The pressure may remain constant as the volume changes.*

The latter situation arises in phase changes, as in equilibrium between vapour and condensed phases at a fixed temperature. The constant pressure is the vapour pressure at fixed temperature so that

$$w = -P\int dv = P \cdot \Delta v$$

where P is constant.

It is obvious that w is the function of state for P, and V_1 and V_2 themselves are the function of state.

Example: Condensation of water vapour at 100°C.

If water vapour at 100°C contains an infinitesimal drop of liquid water, the conditions for equilibrium between liquid water and vapour are satisfied. When the pressure exerted on the vapour by the surroundings through a piston is infinitesimally greater than 1 atmosphere, the piston will move inwards upon the water vapour and cause it to condense to liquid water at 100°C. The condensation could be stopped or reversed by an infinitesimal decrease in the pressure of the surroundings or an infinitesimal increase in the pressure of the water vapour. This increase in pressure could perhaps be caused by an infinitesimal increase of in the temperature of the water. *The process is reversible because its direction can be altered by an infinitesimal changes in the temperature or pressure*.

(a) Expansion work of an ideal gas at constant pressure

For expansion work by the system at a constant pressure of *I* atm, P is constant, the work done by the gas in going from a volume of V_1 to a volume of V_2 is given by

$$-w = P[V_2 - V_1] = P \cdot \Delta v$$

where $[V_2 - V_1] = \Delta v$, the increase in volume of the system and P is the pressure.

(b) Isothermal expansion work of an ideal gas

For one mole molecule of an ideal gas $PV = RT$.

Therefore, $P = RT/V$.

The work done by the gas reversibly and isothermally on surroundings is given by

$$-w = \int P \cdot \Delta v = \int RT \cdot \Delta v/V$$

$$-w = RT\int dv/V = RT \ln V_2/V_1$$

When the work is done very slowly by small changes and isothermally, it is the maximum work done. Alternatively, for an ideal gas $P_1V_1 = P_2V_2$ at constant temperature and so,

$$V_2/V_1 = P_1/P_2$$

Therefore, the maximum work done by an ideal gas at constant temperature

$$-w = RT \ln P_1/P_2$$

for an isothermal, reversible expansion of an ideal gas.

(c) Thermal expansion work of an ideal gas

The expansion work of ideal gas is given by

$$dw = P \cdot \Delta v \text{ at constant pressure.}$$

Since V and T are variables and related by $PV = RT$ for a mole of an ideal gas, it follows that

$$dw = R \cdot dT$$

or $$w = \int R \cdot dT = R\,[T_2 - T_1]$$

15.4 Terminology of Thermodynamics

During the investigation of the interrelationship of energies of systems, the following terms quite often appear and the understanding of their proper significance in thermodynamics is highly desirable. They are illustrated in the next section.

15.4.1 A System

A portion of the universe which is chosen for thermodynamic investigation is called a system.

Illustrations:

Consider the case of a piece of only ice.

It consists of a definite amount of a specific substance or substances for investigation, Fig. 15.1. A system has well-defined boundary that separates the other portion of the universe (surroundings).

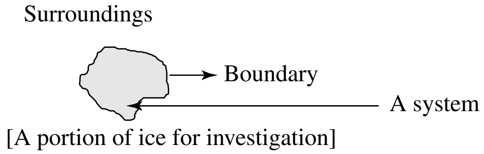


FIGURE 15.1 *A system indicating boundary and surroundings.*

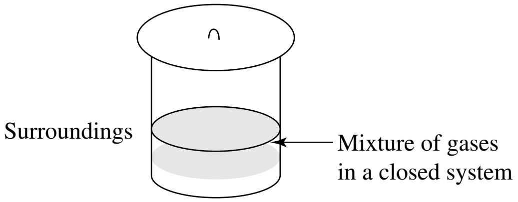


FIGURE 15.2 *A mixture of gases in a closed system.*

The existence of only ice is a state function which depends on both pressure (P) and temperature (T), i.e. $f(P, T)$

A homogenous mixture of two gases is shown in Fig. 15.2.

A homogenous mixture of gases at a set of condition of pressure and temperature [$f(P, T)$].

15.4.2 Surroundings

The rest of the universe remaining outside the boundaries of a system which can exchange both matter and energy is known as surroundings.

Consider the case of a portion of ice kept for investigation, Fig. 15.3.

The ambient surroundings of a portion of a system under investigation may be air, water bath or any particularly specified environment, as desired.

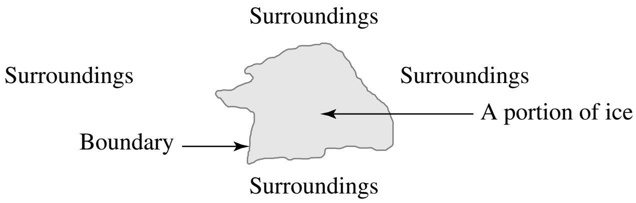


FIGURE 15.3 *Surroundings of a system.*

15.4.3 Types of Systems

The thermodynamic properties of various systems are carried out under different set of conditions and environments. The most important and commonly adopted systems for thermodynamic investigations are the following:

(i) An open system

A system under study can exchange matter as well as energy with its surroundings is called an open system.

Example: Water in a container.

Under thermodynamic investigation, when water in a container is heated, it takes up heat and gradually more and more water vapour escapes into the open surroundings. The open system can also receive an additional amount of water during studies. It is evident, therefore, that in an open system both matter and energy can be exchanged between the system and the surroundings, as shown in Fig. 15.4.

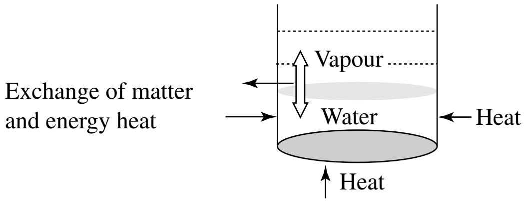


FIGURE 15.4 *An exchange of matter and energy in an open system.*

(ii) A closed system

A closed system is one in which no transfer of matter to and from the surroundings is possible but energy can be exchanged across the boundaries with the surroundings.

A definite amount of water (or any system) is taken in a closed or sealed container and heated, as shown in Fig. 15.5. The system absorbs heat and more and more water vapour is formed which does not escape into the surroundings. The total amount of water remains the same inside the closed system. It can be observed that only heat exchange is possible between the system and the surroundings in a closed system and the amount matter remains the same.

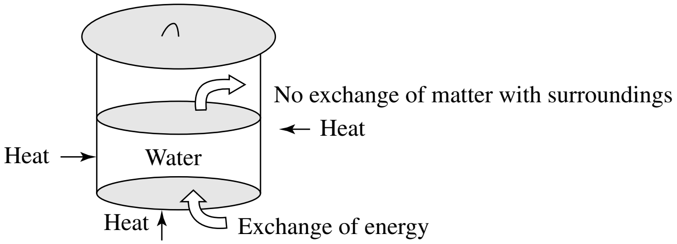


Figure 15.5 *A closed system in which there is exchange of energy and matter with surroundings.*

(iii) An isolated system

A system which can exchange neither energy nor matter with its surroundings is called an isolated system.

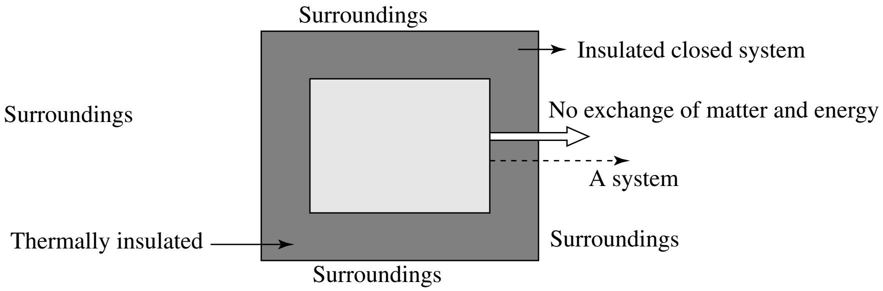


Figure 15.6 *No exchange of energy and matter in an isolated system.*

A definite amount of a system (say, water) is sealed or enclosed inside a container which is thermally insulated (Fig. 15.6). The system does not exchange heat and/or matter with the surroundings and retains the same state and thermodynamic properties.

Examples: Any substance or matter kept inside vacuum flask or refrigerator retains its characteristics.

15.4.4 Homogenous and Heterogeneous Systems

A system is said to be homogenous if it contains only one phase.

Examples: A pure solid, liquid or a mixture of gases or a homogenous solution.

A system is said to be heterogeneous when it contains two or more phases.

Examples: Immiscible liquids, different solids, solids and solution.

15.5 Thermodynamic Properties of a System

The characteristic physical properties such as pressure, temperature, volume, mass, density internal energy, enthalpy, etc. are known as thermodynamic properties, which define the state of systems.

Since the state of a system changes with change in any of the macroscopic properties, these properties are called ‘state variables’. It follows that when a system changes from one state (called initial state) to another (called final state), there is invariably a change in one or more of the thermodynamic properties.

An intensive property of a system is any property whose magnitude depends on the amount of the substance present.

Examples: Total mass, volume energy, etc.

Extensive properties are those whose value is independent of the total amount and are characteristic of the substance or substances present.

Examples: Pressure, temperature density, etc.

The properties of a system in thermodynamic equilibrium depend only on the state. The change in thermodynamic state depends only on the initial state and the final state of the system.

15.5.1 The Thermodynamic State of a System

The thermodynamic state of a system can be defined completely by specifying the state variables or properties.

They are the pressure (P), temperature (T), volume (V) and composition (C) properties of the system. Therefore, state of any system is a function of any two of the three properties, P, T, and V; since composition C is constant for homogenous single phase.

Example: Only ice.

The thermodynamic state of only ice in a closed system can be defined completely by specifying the two state variables, pressure (P) and temperature (T), since the composition (C) of ice is constant.

That is., $H_2O(s)$ is a $f(P, T)$.

Similarly, thermodynamic state of system consisting of homogenous liquid or solid of a single phase system can be completely defined by any two of the three state variables such as pressure, temperature and composition.

15.5.2 Thermodynamic Equilibrium of a State

The state of a system in which the macroscopic properties do not undergo any change with time is said to be in thermodynamic equilibrium.

Example: Ice in a refrigerator, a substance in a vacuum flask, etc.

Since none of the variables of such isolated systems change and therefore, the state of the system is in thermodynamic equilibrium.

15.6 Types of Processes in Thermodynamics

Many kinds of process, path or change are possible. *The operation by which a system changes from one state to another is called a process.*

The following types of processes are known in thermodynamic investigations:

(i) Isothermal process

A process said to be isothermal if the temperature of the system remains unaltered during each stage of the process.

(ii) Isobaric process

A process said to be isobaric if the pressure of the system remains constant during each step of the process.

(iii) Adiabatic process

A process is said to be adiabatic if no heat enters or leaves the system during each step of the process.

(iv) Isochoric process

A process is said to be isochoric if there is no change in volume of the system during each step of the process.

15.7 Thermodynamic Reversible and Irreversible Processes

All natural processes that occur spontaneously are irreversible.

In a reversible thermodynamic process, the change of state of the system appears to be in a state of temperature and pressure equilibrium with its surroundings during each step of the process; however, by an infinitesimal change in pressure in the surroundings, the process can be stopped or reversed. It is clear then that the driving and opposing forces in such a reversible process are everywhere a maximum and hence, the work done by (or on) the system is always maximum.

Example: Maximum pressure–volume (PV) type of work done by an ideal gaseous system under different process conditions.

15.8 First Law of Thermodynamics

The statement of the law is the same as that of 'the law of conservation of energy'.

It can be stated as the *total energy of a system and surroundings is always constant.*

This means that any loss or gain of energy by a system must be exactly equivalent to the gain or loss, respectively, of energy by the surroundings.

15.8.1 Energy Change, Heat and Work Done Relationship (first law)

Any type of energy change between a system and its surroundings involves 'heat' and 'work'.

When a system absorbs heat, q, this energy is not lost; it is stored and can be recovered. Absorption of heat increases the internal energy or the heat content, E, of the system. When a system performs work, w, this is done at the expense of the total energy of the system. This is applicable to any change in the state of the system.

$$E_2 - E_1 = \Delta E = q - w$$

E_1 and E_2 represent the values of the internal energy (*Energy, E, is the function of the state*) of the system in the initial and final states, whether for a reversible process or an irreversible process. This is the formulation of the first law of thermodynamics.

Thus, *the difference between the heat absorbed by a system and the total work done by the system may be defined as equal to the increase in property of the system called its internal energy.*

Generally, any one of the useful thermodynamic properties either 'internal energy' or 'enthalpy' of the state of a system changes during process depending on the type of process used.

15.8.2 Internal Energy

Every substance is associated with a definite amount of energy due to K.E, potential, chemical nature, etc. as well as upon its temperature, pressure and volume. This total energy of the system is known as 'internal energy' of the system.

The internal energy, E, of a system is of a definite quantity; and this thermodynamic property is a function of the 'state variables' of the system.

The change in energy of a system associated with the passage from one thermodynamic state to another depends only on the initial and final states and is independent of the path followed. If E_A represents the energy in the thermodynamic state A and E_B represents that in the state B, then the change of energy ΔE in passing from A to B is given by

$$\Delta E = E_B - E_A$$

Internal energy is a function of state, i.e. E is $f(V, T)$

15.8.3 Thermodynamics Problems

A system receives 325 kJ of heat from its surroundings and does 200 kJ of work on the surroundings. What is the change in its internal energy?

$$\Delta U = q + w$$
$$\Delta U = 325 \text{ kJ} + (-200 \text{ kJ})$$
$$\Delta U = 125 \text{ kJ}$$

Note that since work leaves the system, it is given a negative value.

The value for the ΔU of a system is –120 J. If the system is known to have absorbed 420 J of heat, how much work was done?

$$\Delta U = q + w$$
$$w = \Delta U - q$$
$$w = -120 \text{ J} - 420 \text{ J}$$
$$w = -540 \text{ J}$$

Again, the negative value indicates work was done by the system.

15.8.4 Enthalpy or Heat Content

Many chemical and physical processes are conducted at constant pressure. They are known as isobaric processes. In such cases, expansion work, $w = P\Delta v$, may occur.

Therefore, $\Delta E = q_p - w$ …… (First law)

or $q_p = \Delta E + w$

$q_p = \Delta E + P \cdot \Delta v$

From this relation, *it is seen that the heat absorbed in a process at constant pressure measures the change in internal energy plus the work of expansion.*

This defines a new thermodynamic property, the enthalpy or heat content 'H', by the relation

$$H = E + PV$$

Since E is defined in terms of properties of the state, 'H' is also a function of state i.e. 'H' is

$f(P, T)$.

It follows that ΔH for a cyclic process is zero.

For any process at constant pressure, the change in enthalpy is given by

$$H_2 - H_1 = E_2 - E_1 + (P_2V_2 - P_1V_1)$$
$$\Delta H = \Delta E + P \cdot \Delta v$$

and the value of ΔH depends only on the initial and final states of the system, and not the path.

For a process at constant pressure,

$$\Delta H_p = \Delta E_p + P \cdot \Delta v$$

substituting in equation $q_p = \Delta E + P \cdot \Delta v$,
we have

$$\Delta H_p = q_p$$

It is seen that the heat absorbed, q_p, in a process at constant pressure is a direct measure of the change in heat content or enthalpy of the system.

15.9 Application of First Law of Thermodynamics to Different Processes

The chemical or physical process of systems can be investigated under different thermodynamic conditions and the properties are unique in each case. The following are the few common important ones discussed in the light of the first law of thermodynamics:

1. Energy change in a cyclic thermodynamic process

A process or series of processes, as a result of which the system returns exactly to its original thermodynamic state, is referred to as a cyclic process.

In any cyclic process, $\Delta E = 0$.

Hence, from the first law, $\Delta E = q - w$, i.e. $q = w$.

That is *the heat absorbed, q, is equal to w, the work done by the system.*

The energy change is a function of the initial and final states of the system and is independent of the path by which the change has been brought about.

2. Energy change at constant volume

Expansion work due to changes in volume is perhaps the most common formwork in chemical processes. It would be a rare chemical or physical change in which the volume of the final state is identical to the volume of the initial state.

If the volume is kept constant for the process, then work done, $w = 0$; since $P \cdot dv$ is zero.

$$\Delta E = q_v$$

where q_v denotes the heat absorbed in the constant volume (isochoric) process.

For any process, be it physical or chemical, that involves only pressure–volume work and that is performed at constant volume, ΔE or q_v is the unique increase in energy that accompanies the change.

3. Energy change in adiabatic processes

The process undergone by the thermally insulated system is an adiabatic process; that is *neither heat is absorbed nor given out during the change or process.*

For an adiabatic process, $q = 0$

Therefore, $\Delta E = 0 - w$ i.e. $\Delta E = -w$.

In an adiabatic process, the internal energy change is equal to the work done by the system.

This is the basis of the definition that energy is the ability to do work.

4. Energy change in constant pressure processes

For any process at constant pressure, the change in enthalpy is given by

$$H_2 - H_1 = E_2 - E_1 + (P_2V_2 - P_1V_1)$$
$$\Delta H = \Delta E + P \cdot \Delta v$$

and the value of ΔH depends only on the initial and final states of the system, and not the path.
For a process at constant pressure,

$$\Delta H_\text{p} = \Delta E_\text{p} + P \cdot \Delta v$$

substituting in equation $q_\text{p} = \Delta E + P \cdot \Delta v$,
we have

$$\Delta H_\text{p} = q_\text{p}$$

It is seen that the heat absorbed, q_p, in a process at constant pressure is a direct measure of the change in heat content or enthalpy of the system.

Enthalpy is a convenient thermodynamic function for describing thermal effects that accompany isobaric changes, *for the increase in enthalpy, ΔH is equal to the heat absorbed by the system at constant pressure for changes that involve only expansion work.*

15.10 Heat Capacity of a System

When heat is transferred to an object, the temperature of the object increases. When heat is removed from an object, the temperature of the object decreases. The relationship between the heat (q) that is transferred and the change in temperature (ΔT) is

$$\text{Heat capacity} = \frac{\text{Heat energy in}}{\text{Change in temperature}} = \frac{\Delta q}{T} \text{ (units: J/kg°C)}$$

That is

$$q = C\,\Delta T = C\,(T_2 - T_1)$$

The proportionality constant in this equation is called the *heat capacity* (C).

The temperature change (ΔT) is the difference between the final temperature (T_2) and the initial temperature (T_1).

The heat capacity is the amount of heat required to raise the temperature of an object or substance one degree and has units of energy per degree.

The heat capacity is, therefore, an extensive variable since a large quantity of matter will have a proportionally large heat capacity.

(a) Heat capacity of a system at constant volume

The heat capacity of a system can be studied at constant volume and is given by

$$C_\text{v} = dq_\text{v}/dT.$$

Now, heat change at constant volume is equal to the change in internal energy, ΔE, of the system.

Therefore, $C_\text{v} = [dE/dT]_\text{v}$

The heat capacity at constant volume of a system is equal to the rate of increase of internal energy with the temperature or defined *as the increase in internal energy of the system per degree rise of temperature.*

(b) Heat capacity of a system at constant pressure

The heat capacity of a system at constant pressure is represented as $C_p = dq_p/dT$.

If, dq_p is the amount of heat absorbed, then dH is the enthalpy change at constant pressure of the system. The heat capacity of the system is

Therefore, $C_p = [dH/dT]_p$

The heat capacity, at constant pressure, of a system is given by the rate of change of enthalpy with temperature or *defined as the increase in enthalpy of the system per degree rise of temperature.*

(c) Heat capacities of gases

Heat capacity of a gas at constant pressure exceeds the heat capacity of that gas at constant volume because the gas performs work on its surroundings as it expands at constant pressure.

The difference between C_p and C_v for a perfect gas is easily evaluated thus,

$$C_p - C_v = [\delta H/\delta T]_p - [\delta E/\delta T]_v$$
$$C_p - C_v = [\delta (E + PV)/\delta T]_p - [\delta E/\delta T]_v$$
$$C_p - C_v = [dE/dT] + \delta (RT)/\delta T]_p - [dE/\delta T]$$

Therefore, $C_p - C_v = R$.

15.11 Second Law of Thermodynamics and Reversible Processes

The first law of thermodynamics *states the law of conservation of energy for heat and work; it systematized thermochemistry.* However, the emphasis has been on the initial and final states of a system. The reversibility of a process is of paramount importance in the second law and also stressed initial and final states of the system.

1. The first law establishes definite relationship between the heat absorbed and work performed by the system in a given process and does not emphasize on the direction of flow of heat in a system.
2. According to the first law, the energy of an isolated system remains constant during a specified change or a process. This fails to explain the spontaneity or feasibility of a process.
3. The first law states that the energy of one form can be converted into an equivalent amount of energy of another form. But it does not tell that heat energy cannot be converted into an equivalent amount of work without producing some change elsewhere.

15.12 Importance or Need for Second Law of Thermodynamics

The importance of the second law of thermodynamics can be enumerated thus:

- The law helps us *to determine the direction in which energy can be transferred*.
- It helps us to *predict the spontaneity or feasibility of a process or reaction*.
- Heat is a typical form of energy which cannot be converted into work without producing some change elsewhere. It helps us to calculate the maximum fraction of heat that can be converted into work in a given process.
- The law emphasizes clearly the need for the reversibility of thermodynamic processes, i.e. if, in the expansion of a gas, the opposing pressure is infinitesimally smaller than the pressure of the gas, the expansion takes place infinitesimally slowly, that is reversibly.

All natural processes are spontaneous and are irreversible and always tend towards equilibrium.

15.13 Statement of the Second Law

The total heat absorbed by a system cannot be completely converted into work.

or

It is impossible to convert a machine functioning in cycles which can convert heat completely into equivalent amount of work without producing changes elsewhere.

15.13.1 Entropy of a System

We have seen that spontaneous processes proceed in one direction only and that they are thermodynamically irreversible. One of the thermodynamic functions that serve as a reliable criteria for spontaneity is called the ‘entropy’ and is symbolized by the letter ‘s’. *Entropy is also a function of the state of a system*. As the entropy of a system is difficult to define directly, it is more convenient to define the ‘change of entropy’, *ds*.

If a process is carried out in a *thermodynamically reversible* manner, so that *dq* is the amount of heat absorbed by the system at constant temperature, then the entropy change, *ds*, of the system is given by the expression

$$ds = dq/dT$$

If S_2 is the entropy of the final state and S_1 is the entropy of the initial state of a system under investigation, the increase in entropy, Δs, is given by the equation

$$\Delta s = S_2 - S_1 = \int dq/dT$$

It is evident that $\int dq/dT$ is a definite quantity, independent of the taken for the change, and depends only upon the initial and final states of the system. This quantity, therefore, like ΔE or ΔH, should represent the change in the same single valued function of the initial and final states

Entropy is expressed in calories per degree or Jules per degree Kelvin (J/K).

Entropy may be regarded to as a measure of disorder or randomness of arrangement of atoms or molecules in a system.

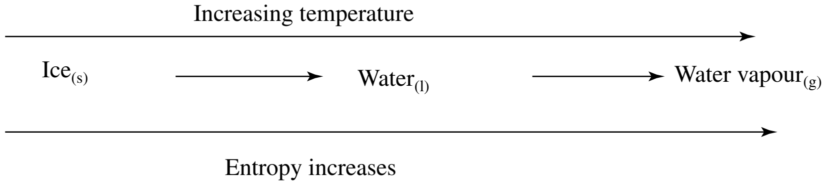


Figure 15.7 *Entropy increases with the increase of temperature of ice–water–vapour system.*

15.13.2 Entropy Changes

It is known that energy of a state of system is thermodynamically dependent on the variable properties such as volume, temperature, pressure and the new property, entropy.

Since a closed system by definition cannot gain or lose matter, the first law, $\Delta E = q - w$, involves only heat and work

That is $$dE = dq + dw \qquad \text{(i)}$$

By entropy definition, we have

$$dq \subseteq T \cdot ds \qquad \text{(ii)}$$

Addition of (i) and (ii) yields the fundamental relation of thermodynamics for the closed system.

$$dE \subseteq T \cdot ds + dw$$

15.13.3 Entropy Change of an Irreversible Process

The entropy of an irreversible process is given by

$$dq < T \cdot ds$$

That is entropy of the system increases with change of state during the irreversible process.

The second law also gives us a criterion for determining the direction of change which a system will follow in a spontaneous process. *The change will take place in that direction in which the entropy of the system will increase*.

15.13.4 Entropy Change in Reversible Process

The entropy change in a reversible thermodynamic process is given by

$$ds = dq/T \text{ or } dq = T \cdot ds$$

Of greater importance to the equation $dE = T \cdot ds + dw$, however, is the equality.

$$dE = T \cdot ds + dw$$

which applies to reversible changes and equilibrium.

When all the work is expansion work, $dw = -P \cdot dv$ and the equation assumes a very important form

$$dE = T \cdot ds - P \cdot dv$$

The independent variables in the above equation appear to be 's', the entropy and 'v' is the volume of the system. These are natural thermodynamic variables for the energy of a closed system when only expansion work is involved.

That is $E = f\,[S, V]$.

15.13.5 Entropy Changes in an Isothermal, Isobaric Reversible Process

For processes carried out at constant (isobaric) pressure, the enthalpy of state is given by

$$H = E + PV$$

dH is a complete differential

Therefore, $dH = d\,[E + PV]$

$$dH = dE + P \cdot dv + V \cdot dp \tag{1}$$

for any reversible process at constant pressure that involves only expansion work,

$$dp = 0 \text{ and } dE = T \cdot ds - P \cdot dv,$$

is true.

Hence, substitution of this into (1) yield (2)

$$dH = T \cdot ds - P \cdot dv + P \cdot dv + V \cdot dp \tag{2}$$

Therefore, $$dH = T \cdot ds + V \cdot dp \tag{3}$$

At constant pressure, $dp = 0$ and hence,

$$dH = T \cdot ds$$

Integration between states 1 and 2 is easy if T is constant, thus,

$$\int dH = \int T \cdot ds = T \cdot [S_2 - S_1] = T \cdot \Delta s$$

That is, $\Delta H = T \cdot \Delta s$ or

$$\Delta s = \frac{\Delta H}{T} \quad (4)$$

Equation (4) is applicable to isothermal isobaric reversible changes like phase transitions: fusion, vaporization and sublimation, etc.

It can be shown from Eq. (3) that

$$H = f(\text{S, P}).$$

15.13.6 Entropy Changes in Adiabatic Reversible Process

Any reversible process of a system is carried out under adiabatic condition,

then, $dq = 0.$

Therefore, $dq = T \cdot ds = 0$

Since, T is greater than zero ($T \neq 0$),

$$ds = 0 \quad (1)$$

$$\Delta s = \int ds = [S_2 - S_1] = 0 \quad (2)$$

The entropy of any system is constant in a reversible adiabatic process, for $S_2 = S_1$ by Eq. (2) and the initial and final states are restricted only in being joined by a reversible adiabatic path.

15.13.7 Entropy Change in Isothermal Expansion of a Gas

In isothermal expansion of a gas is carried out reversibly, there will be no change of internal energy, i.e. $\Delta E = 0$, and hence from the first law, $q_{\text{rev}} = -W$

In such cases, work done in the expansion of n moles of gas from volume V_1 to V_2 at constant temperature is given by

$$-W = nRT \ln V_2/V_1$$

$$q_{\text{rev}} = nRT \ln V_2/V_1$$

Hence, $\Delta S = q_{\text{rev}}/T = 1/T \cdot nRT \ln V_2/V_1$

$$\Delta S = nR \ln V_2/V_1$$

15.13.8 Entropy Change for Temperature Change

For a reversible process at constant pressure, entropy change for change in temperature is given as

$$\Delta S = \int dq/T = \int C_{\text{P}} \cdot dT/T$$

$$\Delta S = C_{\text{P}} \ln T_2/T_1$$

15.13.9 Entropy Change in Phase Changes

(i) From a solid phase to a liquid phase:

Consider a case when a solid changes into liquid state at its fusion point at constant pressure. This requires absorption of heat (enthalpy of fusion).

Consider melting of one mole of a substance reversibly at the fusion point T_f at constant pressure. Let ΔH_f be the molar enthalpy change of fusion. The entropy change of the process, ΔS_f, will be then given by,

$$\Delta S_f = \Delta H_f/T$$

(ii) From a liquid phase to a gaseous phase:

Suppose one mole of a substance changes from the liquid state to the vapour state reversibly at its boiling point T_b under constant pressure. If ΔH_b is the molar enthalpy change of vaporization, then the entropy change accompanying the process will be given by

$$\Delta S_b = \Delta H_b/T$$

Since ΔH_f and ΔH_b are both +ve, the processes of fusion and vaporization are both accompanied by increase of entropy.

15.13.10 Entropy and Thermodynamic Equilibrium

The evaluation of entropy changes affords a powerful tool for attacking the fundamental problem of physicochemical equilibrium.

The movement of a physicochemical system towards equilibrium is controlled by two factors, one involving 'entropy' and the other the 'energy' of the system.

We may, therefore, state that the following two criteria for the thermochemical equilibrium,

- In a system of constant energy, the position of equilibrium is that of maximum entropy.
- In a system of constant entropy, the position of equilibrium is that of minimum energy.

15.14 Concept of Free Energy Changes in Thermodynamic Processes

The chemist's interest in thermodynamics is primarily due to its application to the study of physicochemical reactions, which are almost never studied under constant energy or constant temperature.

For this reason, two new thermodynamic functions, called the 'work function', A, and 'Free energy, G, have been described and defined in terms of energy, entropy and enthalpy.

They are defined by these equations

$$A = E - T \cdot S \qquad \ldots\text{[Helmholtz free energy]}$$
$$G = H - T \cdot S \qquad \ldots\text{[Gibb's free energy].}$$

Since E, H and S depend upon the state of the system, it is evident that the functions A and G also depend upon the state only. Hence, changes in them, viz. ΔA and ΔG are independent of the path taken.

15.14.1 Helmholtz Free Energy, *A*

This property of a system is a function of the state only. The value of 'A' in a particular state of a system *is a measure of the work obtainable from the system in that state.*

If the isothermal reversible change of a system is carried out at a *constant volume*, there is no work ($w = -P \cdot dv$) done against pressure. That is, the whole decrease in Helmholtz energy ($-\Delta A$), of the system is available for work.

The relationship, $w = -\Delta A$, measures the maximum work a system can do in an isothermal reversible change at constant volume.

According to the first law of thermodynamics,

$$\Delta E = q - w$$

For a reversible isothermal change, we have

$$\Delta s = q/T \text{ or } q = T \cdot \Delta s$$

From substituting, $w = -\Delta A$, and $q = T \cdot \Delta s$ in the first law, we have

$$\Delta E = T \cdot \Delta s + \Delta A$$
$$\Delta A = \Delta E - T \cdot \Delta s$$

Therefore, $A = E - TS$ Where, 'A' is a function of state.

15.14.2 Gibb's Free Energy

If an isothermal change is carried out at constant pressure, P, the change in the system is accompanied by a change in volume, Δv.

The quantity, $P \cdot \Delta v$, is the work done by the system on expansion against the constant pressure, P. The work done other than that due to change in volume is called the *network*, thus

Net work = $[w - P \cdot \Delta v] = -\Delta G$

That is $-\Delta G$ = *is measure of the maximum network that can be obtained from a system at constant pressure and temperature*.

or $$w = -\Delta G + P \cdot \Delta v$$

This is for a reversible change at constant pressure.
From the first law of thermodynamics, $\Delta E = q - w$,

That is, $$\Delta E = T \cdot \Delta s - [-\Delta G + P \cdot \Delta v]$$
$$\Delta E = T \cdot \Delta s + \Delta G - P \cdot \Delta v \text{ or}$$
$$\Delta G = [\Delta E + P \cdot \Delta v] - T \cdot \Delta s$$
or $$\Delta G = \Delta H - T \cdot \Delta s$$
or $$G = H - TS.$$

where 'G' is a function of state and is called the Gibb's free energy.

The Gibb's free energy is a thermodynamic quantity which can be used to determine if a reaction is spontaneous or not. The definition of the Gibb's free energy is

$$G = H - TS$$

where G is the free energy, H is the enthalpy and S is the entropy. If we consider the change ΔG for a reaction where the temperature does not change, we have

$$\Delta G = \Delta H - T \cdot \Delta S$$

ΔH and ΔS are computed from the usual procedures. Temperatures for this equation must be in Kelvin. The above equation is often known as the 'Gibbs–Helmholtz equation'. Since the values of ΔH and ΔS are usually obtained from tables of standard entropy and enthalpy, the Gibb's free energy change usually computed is the standard Gibb's free energy change, $\Delta G°$ at 25°C for the different procedure is similar.

The sign of ΔG determines if a reaction is spontaneous or not.

- $\Delta G < 0$: the reaction is spontaneous
- $\Delta G > 0$: the reaction is not spontaneous
- $\Delta G = 0$: the reaction is at equilibrium.

When computing ΔG be careful about the units. Entropy is usually given in J/mol/K, whereas enthalpy is usually given in kJ/mol.

ΔH	ΔS	Low T	High T
+	+	G positive; not favoured	G negative; favoured
+	–	G positive; not favoured	G positive; not favoured
–	+	G negative; favoured	G negative; favoured
–	–	G negative; favoured	G positive; not favoured

Table 15.1

Temperature effect on the spontaneity of a process

The entropy of a system may decrease, but at the expense of large enthalpy.
These four relations apply to the closed systems that have *only expansion work done* on them. They are

(i) $dE \subseteq T \cdot dS - P \cdot dV$ (1)

(ii) $dH \subseteq T \cdot dS + V \cdot dP$ (2)

(iii) $dA \subseteq -S \cdot dT - P \cdot dV$ (3)

(iv) $dG \subseteq -SdT + V \cdot dP$ (4)

If the process undergone by the system *is reversible, the relations are equalities; if irreversible, that are inequalities.*

These four relations are the differential equations that express the four common thermodynamic functions as functions of the independent variables S, V, P and T. The thermodynamic functions depend on two of these four variables in the manner as given

1. $E = E\,(S, V)$
2. $H = H\,(S, P)$
3. $A = A\,(T, V)$
4. $G = G(T, P)$.

15.14 3 Examples

1. Is the following reaction spontaneous at 25°C?

$$2H_2(g) + O_2(g) \rightarrow 2H_2O(g)$$

Solution

We need to compute both ΔH and ΔS for the reaction, get the value of ΔG, and see if it is less than zero. First, look up the thermodynamic data.

Compound	$\Delta H_f°$ (kJ/mol)	$\Delta S°$ (J/mol/K)
$H_2(g)$	0.0	130.6
$O_2(g)$	0.0	205.0
$H_2O(g)$	−241.8	188.7

Table 15.2

Enthalpy of formation and entropy of molecules

Next, calculate both ΔH and ΔS

$$\Delta H = [(2\Delta H_{H_2O}{}^{\circ}) - (2\Delta H_{H_2}{}^{\circ}) + 1\Delta H_{O_2}{}^{\circ}$$

$$\Delta H = (2 \times -241.8) - (2 \times 0 + 1 \times 0)$$

$$\Delta H = -483.6 \text{ kJ/mol}$$

$$\Delta S = [(2\Delta S_{H_2O}{}^{\circ}) - (2\Delta S_{H_2}{}^{\circ}) + 1 \times \Delta S_{O_2}{}^{\circ}$$

$$\Delta S = (2 \times 188.7) - (2 \times 130.6 + 1 \times 205.0)$$

Combine the two of them into ΔG

$$\Delta G = \Delta H - T \cdot \Delta S$$

$$\Delta G = -483.6 \text{ kJ/mol} - 298\text{K} \times (-0.0886 \text{ kJ/mol/K})$$

Therefore, $\Delta G = -457$ kJ/mol

The reaction is spontaneous, since $\Delta G < 0$.

2. At what temperature does the following reaction become spontaneous?

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

Solution

Calculate the ΔG from the thermodynamic data from Table 15.3

Compound	ΔH_f° (kJ/mol)	ΔS° (J/mol/K)
$SO_3(g)$	−395.7	+256.7
$SO_2(g)$	−296.8	+248.1
$O_2(g)$	0.0	+205.0

Table 15.3

Enthalpy of formation and entropy of molecules

Next, calculate ΔH and ΔS

$$\Delta H = 2(\Delta H_{SO_2}{}^{\circ}) + 1 \times \Delta H_{O_2}{}^{\circ} - (2 \times \Delta H_{SO_3}{}^{\circ})$$

$$\Delta H = (2 \times -296.8 + 1 \times 0) - (2 \times 395.7)$$

$$\Delta H = +197.8 \text{ kJ/mol}$$

$$\Delta S = (2 \times \Delta S_{SO_2}{}^{\circ}) + 1 \times \Delta S_{O_2}{}^{\circ}) - (2 \times \Delta S_{SO_3}{}^{\circ})$$

$$\Delta S = (2 \times 248.1 + 1 \times 205) - (2 \times 256.7)$$

$$\Delta S = +187.8 \text{ J/mol/K} = 0.1878 \text{ kJ/mol/K}$$

Since both ΔH and ΔS are positive, this reaction will be spontaneous at high temperature. If we set ΔG equal to zero we can compute this temperature

$$\Delta G = \Delta H - T \cdot \Delta S$$

$$0 = 197.8 \text{ kJ/mol} - T \times 0.1878 \text{ kJ/mol/K}$$

$$\therefore \quad T = 1053 \text{ K.}$$

15.14.4 The Criterion of Equilibrium Between Coexisting Phases Like Ice and Water is $(\Delta G)_{P,T} = 0$

If for a process carried out at constant pressure and temperature and G_1, H_1 and S_1 represent the thermodynamic functions of the initial state and G_2, H_2 and S_2 represent that in the final state, the change of free energy, ΔG, of system is given by

$$G_2 - G_1 = [H_2 - H_1] - T(S_2 - S_2]$$

or $$\Delta G = \Delta H - T \cdot \Delta s$$

Note: $-\Delta G$ is the measure of decrease in free energy in any processes.

A special case of the application of $\Delta G = \Delta H - T \cdot \Delta s$ concerns an isothermal phase transition like fusion or evaporation. If ΔH is the enthalpy change of transition,

$\Delta H = T \cdot \Delta s$ yields for Δs for such a change

$$\Delta s = \Delta H / T$$

substitution in $\Delta G = \Delta H - T \cdot \Delta s$ yields

$$\Delta G = \Delta H - T \cdot [\Delta H/T] = 0$$

that is, for any isobaric reversible change in a closed system that, if it does only expansion work, there is no change in the Gibb's free energy, $\Delta G = 0$.

15.14.5 Free Energy Change of Processes with Temperature and Pressure

The free energy change of processes with variation of temperature, T, and Pressure, P, may now be considered. The Gibb's free energy for system is given by the relationship

$$G = H - TS$$
$$= [E + PV] - TS$$

Upon differentiation,

$$dG = dE + P \cdot dV + V \cdot dP - T \cdot dS - S \cdot dT \quad (1)$$

But $dE = TdS - P \cdot dV$ for a reversible process, since $dq = T \cdot dS$

Substituting in Eq. (1),

Therefore, $dG = V \cdot dP - S \cdot dT$ (2)

(a) Free energy change of a reversible process at constant pressure

$$dG = -S \cdot dT \quad \text{since } dP = 0 \text{ at constant pressure} \quad (3)$$

or $$[dG/dT]_P = -S$$

We get the variation of free energy with temperature of a reversible process at constant pressure.

(b) Free energy change of a reversible process at constant temperature

We know that for a reversible process,

$$dG = V \cdot dP \quad \text{since } dT = 0 \text{ at constant temperature} \quad (4)$$

or $$[dG/dP]_T = V$$

We get the variation of free energy of a reversible process at constant temperature.

Let the free energy change of a system be G_1 in the initial state and G_2 in the final state when an appreciable change of pressure has taken place, at constant temperature, then the integrating $dG = VdP$, the free energy change, ΔG, is given by

$$\Delta G = G_2 - G_1 = \int V dP$$

or

$$\Delta G = RT \ln P_2/P_1 = RT \ln V_2/V_1$$

where V_1 and V_2 are the initial and final volumes, respectively.

15.14.6 Gibb's Helmholtz Equation

A system undergoes a reversible process from state 1 to state 2 at constant pressure (isobaric process), then the change of free energy of the process is obtained, thus

Consider a system in state 1, then

$$dG_1 = -S_1 dT$$

and in state 2, it is

$$dG_2 = -S_2 dT.$$
$$d(G_2 - G_1) = -(S_2 - S_1)$$

this gives

$$d(\Delta G) = -\Delta S \cdot dT$$

Therefore, $[d(\Delta G)/dT]_P = -\Delta S$

Now the relationship,

$$\Delta G = \Delta H - T \cdot \Delta S$$

substituting ΔS, we get,

$$\Delta G = \Delta H - T \cdot [(d(\Delta G)/dT)_P]$$

This equation is known as Gibb's–Helmholtz equation. It is applicable for all processes occurring at constant pressure.

This equation also can be represented as

$$d(\Delta G)/dT)_P = \frac{\Delta G - \Delta H}{T}$$

This equation gives the variation of the change in Gibbs free energy with respect to variations in temperature at constant pressure for any process that can in principle be performed reversibly in a closed system.

15.14.7 Applications of Gibbs–Helmholtz Equation

(i) Thermodynamics of reversible electrodes and cells

The electrical energy of a reversible cell is given by the decrease of free energy ($-\Delta G$) of the electrode reaction occurring in the cell. This free energy change is equal to the electrical network done by the cell, i.e.

$$-\Delta G = -W_{net} = nEF$$

where

'nF' = quantity of electricity produced in coulombs and
'n' = the number of electrons liberated at one electrode,
E = Emf of the reversible cell in Volts.

(ii) ***Relation between electrical energy and enthalpy of cell reaction***

According to the well-known Gibbs–Helmholtz equation, decrease in free energy, $-\Delta G$, of the system at constant pressure is given by the expression

$$-\Delta G = -\Delta H - T\,[\delta(\Delta G)/\delta T]_P \qquad (1)$$

where $-\Delta H$ is the decrease in enthalpy of the cell reaction at constant pressure.

Substituting the value of $-\Delta G$ in Eq.(1), we get

$$nEF = -\Delta H - T\,[\delta(-nEF)/\delta T]_P$$
$$nEF = -\Delta H + nF \cdot T\,[\delta(E)/\delta T]_P$$

Since n and F are constants.

Evidently, whether the electrical energy, viz. nEF, is equal to or greater or less than the enthalpy of the cell reaction (ΔH), depends upon the sign of $[\delta(E)/\delta T]_P$, i.e. *upon the sign of the temperature coefficient of the emf of the cell*.

If it is zero, the electrical energy will be equal to the enthalpy of the cell reaction. If it is +ve, the emf of the cell increases with rise in temperature, the electrical energy will be greater than the enthalpy of the cell reaction. If $[\delta(E)/\delta T]_P$ is –ve, the electrical energy will be smaller than the enthalpy of the cell reaction.

Therefore, emf of the cell is very close to the enthalpy of the cell reaction

$$E_{\text{cell}} = -\Delta H/nF + T\,[\delta(E)/\delta T]_P$$

15.15 Maxwell's Relations in Thermodynamics

For processes in thermodynamics, Maxwell relations can be deduced from known thermodynamic functions of E, H, A and G of state.

1. For any reversible thermodynamic process, the energy change, dE, is given by

$$dE = T \cdot dS - P \cdot dV \qquad (1)$$

For a isochoric process, V is constant, i.e. $dV = 0$

$$[dE/dS]_v = T$$

Mixed second differentiation with respect to V at constant entropy, we have

$$\delta^2 E/\delta S\delta V = [\delta T/\delta V]_S \qquad (2)$$

(i) When entropy of a process is constant, $dS = 0$

$$[dE/dV]_s = -P$$

Second differentiation with respect to entropy, s, at constant volume, we have

$$\delta^2 E/\delta S\delta V = -\,[\delta P/\delta S]_V \qquad (3)$$

Since E is a state function, dE is an exact differential. That is

$$\delta^2 E/\delta S\delta V = \delta^2 E/\delta S\delta V$$

Therefore,

$$[\delta T/\delta V]_S = -\,[\delta P/\delta S]_V \qquad \text{(I)}$$

2. For any reversible thermodynamic change of state, enthalpy change, dH, is given by

$$dH = dE + P \cdot dV + V \cdot dP$$
$$= T \cdot dS - P \cdot dV + P \cdot dV + V \cdot dP$$

Since $dE = T \cdot dS - P \cdot dV$ for a reversible process.

$$dH = T \cdot dS + V \cdot dP \tag{4}$$

Now from Eq. (4) we have

(ii) When P is constant, then

$$[dH/dS]_\text{P} = T$$

Differentiating with respect to P at constant S

$$\delta^2 H/\delta P\delta S = [\delta T/\delta P]_S \tag{5}$$

When S is constant then

$$[dH/dP]_\text{P} = V$$

Differentiating with respect to S at constant P

$$\delta^2 H/\delta S\delta P = [\delta V/\delta S]_\text{P} \tag{6}$$

H is a state function. Hence, dH is a perfect differential

$$\delta^2 H/\delta S\delta P = \delta^2 H/\delta P\delta S$$

Therefore,

$$[\delta T/\delta P]_S = [\delta V/\delta S]_\text{P} \tag{II}$$

(b) From the definition of Helmholtz free energy relation

$$A = E - TS$$
$dA = dE - T \cdot dS - SdT$ is its exact differential.

That is $dA = T \cdot dS - P \cdot dV - T \cdot dS - SdT$

That is $dA = -P \cdot dV - SdT$ (7)

Equation (7) gives

(i) When V is constant, we have

$$dA = -SdT$$

That is $[dA/dT]_\text{V} = -S$

Differentiating with respect to V at constant T we get

$$\delta^2 A/\delta V\delta T = -[\delta S/\delta V]_\text{V} \tag{8}$$

(ii) When T is constant, then,

$$[dA/dV]_\text{T} = P$$

Differentiating with respect to T at constant V, we have

$$\delta^2 A/\delta V\delta T = -[\delta P/\delta T]_\text{V} \tag{9}$$

Since dA is an exact differential, therefore

$$\delta^2 A/\delta V\delta T = \delta^2 A/\delta V\delta T$$

$$-[\delta S/\delta V]_\text{V} = -[\delta P/\delta T]_\text{V}$$

or $[\delta S/\delta V]_\text{V} = [\delta P/\delta T]_\text{V}$ (III)

(a) Gibb's free energy gives

$$G = H - TS$$

dG is its exact differential, then

$$dG = dH - T \cdot dS - S\,dT$$

But, $H = E + PV$

$$dH = dE + P \cdot dV + V \cdot dP = T \cdot dS + V \cdot dP \quad [\text{since } T \cdot dS = dE + P \cdot dV]$$

Therefore,

$$dG = -SdT + V \cdot dP$$

(i) At constant T, $dT = 0$

then, $[dG/dP]_T = V$ (10)

Differentiating Eq. (10) with respect to T at constant P

$$[\delta^2 G/\delta T \delta P]_T = [\delta V/\delta T]_P)$$

(ii) At constant P, $dP = 0$, then,

$$[dG/dT]_P = -S \tag{11}$$

Differentiating Eq. (11) with respect to P at constant T

$$[\delta^2 G/\delta T \delta P]_T = -[\delta s/\delta P]_T$$

Since G is a function of state and hence

$$[\delta^2 G/\delta T \delta P]_T = [\delta^2 G/\delta T \delta P]_T$$

and therefore,

$$[\delta V/\delta T]_P = -[\delta s/\delta P]_T \tag{IV}$$

Equations (I)–(IV) are known as *Maxwell's thermodynamic equations*.

15.16 Clapeyron–Clausius Equation

This equation concerns the evaporation of a liquid or solid or the fusion of a solid.

(a) Solid–liquid melting point of the solid.
(b) Liquid–vapour boiling point of a liquid,
(c) Solid–vapour sublimation temperature.

Consider any two phases (e.g. liquid and vapour) of one and the same substance in equilibrium with each other, at a given temperature and pressure. It is possible to transfer any definite amount of substance from one phase to the other phase, in a thermodynamically reversible manner, i.e. infinitesimally slowly, the system remaining in a state of equilibrium all along. For example, by supplying heat infinitesimally slowly to the system, it is possible to change any desired amount of the substance from the liquid phase to the vapour phase at the same temperature and pressure. Similarly, by withdrawing heat infinitesimally from the system, it is possible to change any desired amount of the substance from the vapour state to the liquid state without any change of temperature and pressure. Since the liquid in the state of equilibrium, *the free energy change of either process will be zero*. We may conclude, therefore, that equal amounts of a given substance must have exactly the same free energy in the two phases at equilibrium with each other.

That is $G_1 = G_2$ or $\Delta G = G_2 - G_1 = 0$

According to thermodynamics,

$$dG = V \cdot dP - S \cdot dT \qquad (1)$$

This gives the free energy change when a system undergoes reversibly a change of temperature, *dT*, and pressure, *dP*.

For two phases in equilibrium, we have

$$dG_1 = V_1 dP - S_1 dT \qquad (2)$$

$$dG_2 = V_2 dP - S_2 dT \qquad (3)$$

Since $G_1 = G_2, dG_1 = dG_2$

$$V_1 dP - S_1 dT = V_2 dP - S_2 dT$$

Therefore, $\dfrac{dP}{dT} = \dfrac{S_2 - S_1}{V_2 - V_1}$

Hence,

$\dfrac{dP}{dT} = \dfrac{\Delta S}{\Delta V}$ where, $\Delta S = S_2 - S_1$, molar entropy change and

$\Delta V = V_2 - V_1$, molar volume change

If '*q*' is the heat exchanged reversibly per mole of the substance during the phase change at temperature, *T*, then the change of entropy ΔS, in this process is given by

$$\Delta S = q/T,$$

Hence, $dP/dT = q/T \cdot \Delta V$.

Thus,

$$dP/dT = q/T\,[V_2 - V_1]$$

This is *Clapeyron–Clausius* equation.

This equation, evidently, gives change in pressure *dP* which must accompany the change in temperature *dT* or vice versa, in the case of a system containing two phases of a pure substance in equilibrium with each other.

15.16.1 Applications

(a) Water–water vapour system

$$H_2O(l) \Leftrightarrow H_2O(v)$$

$$dP/dT = \Delta H_v/T\,[V_g - V_l]$$

where ΔH_v = molar heat of vaporisation

V_g = molar volume of vapour of water

V_l = molar volume of water in the liquid state.

(b) Water–ice system

$$H_2O(s) \Leftrightarrow H_2O(l)$$

$$dP/dT = \Delta H_f/T\,[V_l - V_s]$$

where, ΔH_f = molar heat of fusion of ice

V_l = molar volume of water

V_s = molar volume of ice in the liquid state.

15.16.2 Integrated form of Clapeyron–Clausius Equation

The integrated form of the Clapeyron–Clausius equation can be obtained for the liquid–vapour equation.

The molar volume of the substance in the gaseous state is greater than that in the liquid state. If the volume of liquid is neglected compared with that of the vapour

$$[V_\mathrm{g} > V_\mathrm{l}].$$

$$dP/dT = \Delta H_V/T\,V_\mathrm{g}$$

which now be integrated.

If the vapour behaves as an ideal gas, $V_\mathrm{g} = RT/P$, so that

$$dP/dT = \Delta H_\mathrm{V} \cdot P/R\,T^2$$

$$\frac{dP/P}{dT} = \frac{\Delta H_y}{RT^2} \quad \text{or} \quad \int d \ln P = \frac{\int \Delta H_y}{T^2}$$

If ΔH_V may be assumed not to vary with temperature

$$\ln P = \text{constant } \frac{-\Delta H_y}{R}\,[1/T]$$

Or between a pair of temperatures,

$$\ln(P_2/P_1) = 2.303 \log_{10}(P_2/P_1) = \frac{\Delta H_y}{R}\left|\frac{1}{T_1} - \frac{1}{T_2}\right|$$

$$= \frac{\Delta H_y}{R}\left|\frac{T_2 - T_1}{T_1 T_2}\right|$$

15.17 Free Energy Change of a Reaction at Equilibrium: van’t Hoff Isotherm

The free energy change of a reaction involving gaseous species in a state of thermodynamic equilibrium under a set of temperature and pressure can be understood from the van’t Hoff’s reaction isotherm.

Consider a gaseous reaction of the type in equilibrium is given by

$$a\mathrm{A} + b\mathrm{B} \Leftrightarrow c\mathrm{C} + d\mathrm{D}$$

when the rate of forward reaction equals the rate of backward reaction or in other words, when the free energy change of the reaction is zero then, the system is said to be in state of thermodynamic equilibrium; and the *equilibrium constant*, K_P, of the for the above reaction is given by

$$K_\mathrm{P} = [p^\mathrm{a} \times p^\mathrm{b}]/[p^\mathrm{c} \times p^\mathrm{d}]$$

where p^a, p^b, p^c and p^d are the partial pressures of gaseous reactants and products, respectively.

We know from thermodynamics that

$$dG = VdP - SdT$$

At constant temperature $dT = 0$

Therefore, $dG = V \cdot dP$ (1)

Or $dG = RT \cdot dP/P$ since $PV = RT$ for a mole

Upon integration, we get

$$\int dG = RT\int dP/P$$

or $$G = G° + RT \ln P \qquad (2)$$

where $G°$ is the standard free energy change of the reaction.
The free energy change of the reaction is given by

$$\Delta G = \Sigma G_{\text{products}} - \Sigma G_{\text{reactants}}$$

For a reaction at equilibrium, we have
From Eq.(2), the change in free energy,

$$\Delta G = \Delta G°+ RT \ln [(p^c \times p^d)/(p^a \times p^b)] \qquad (3)$$

At the thermodynamic equilibrium state of the reaction, $\Delta G = 0$, and therefore,

$$-\Delta G° = RT \ln [(p^c \times p^d)/(p^a \times p^b)] \qquad (4)$$

This expression relates the free energy change and the equilibrium constant of the reaction, thus, we have

$$-\Delta G° = RT \ln K_P \qquad \text{since } K_P = [p^a \times p^b]/[p^c \times p^d]$$

Substituting $\Delta G°$ value in Eq. (4), we have

$$\Delta G = -RT \ln K_P + RT \ln [(p^c \times p^d)/(p^a \times p^b)]$$

or $$-\Delta G = RT \ln K_P - RT \ln [(p^c \times p^d)/(p^a \times p^b)] \qquad (3)$$

This equation is known as van't Hoff's reaction isotherm.

That is $-\Delta G = RT \ln K_P - RT \ln K_C$.

or

$$K_C = e^{-\Delta G/RT}$$

where 'K_C' denotes the reaction quotient of the products and reactants in terms of their partial pressures at a particular pressure and temperature.

15.18 Criteria for Reversible and Irreversible Processes

1. In terms of entropy concept

The net entropy change of a process determines whether the process would proceed irreversibly(spontaneously) or not. If there is a *net increase in the entropy of the system and the surroundings taken together*, the process would proceed irreversibly, i.e. it would be *thermodynamically feasible*. If there is no net change in the entropy of the system and the surrounding put together, the process will be reversible, that is the *system will remain in a state of equilibrium.*

Now, a small change of state is brought about *irreversibly*, the heat absorbed by the system will be less (since $q_{\text{irr}} < q_{\text{rev}}$). But, the entropy change dS will have the same value. Hence, for an irreversible process,

$$TdS > dq_{\text{irr}}$$

We may write

$TdS = dU + PdV$ for reversible process
$TdS > dU + PdV$ for irreversible process

Combining the two we have

$$TdS \supseteq dU + PdV$$

The ‘equal to’ sign refers to a *reversible* process while the ‘greater’ than sign refers to an *irreversible* process.

If internal energy, U, and the volume, V, remain constant, then for an isothermal (constant T) process,

$$dS \supseteq 0$$

The ‘equal to’ sign refers to a *reversible* process while the ‘greater’ than sign refers to an *irreversible* process.

2. In terms of free energy change concept

The variation of free energy change with variation of temperature and pressure is given thus,

$$dG = dU + PdV + VdP - T \cdot dS - S \cdot dT$$

But, $\quad T \cdot dS \supseteq dU + PdV$

Substituting in the above equation we have

$$dG \subseteq VdP - SdT$$

Therefore, at constant temperature and pressure,

$$dG \subseteq 0$$

The ‘equal to’ sign refers to a *reversible* process while the ‘greater’ than sign refers to an *irreversible* process.

The criterion in terms of free energy change (viz. $(\delta G)_{T,P} \subseteq 0$*) is the most useful criterion to decide between reversibility and irreversibility of a process.*

If a process is thermodynamically irreversible, i.e. spontaneous or feasible, then

$$(\delta G)_{\mathrm{T,P}} < 0$$

If a process is thermodynamically reversible, i.e. there is a state of equilibrium, then

$$(\delta G)_{\mathrm{T,P}} = 0$$

If a process is thermodynamically not feasible, then

$$(\delta G)_{\mathrm{T,P}} > 0$$

In other words, if free energy change of a process has a negative value($-\Delta G$), *the process can take place spontaneously, i.e. it is feasible.*

If, on the other hand, the free energy change of a process is zero ($\Delta G = 0$), then, the process is reversible, i.e. *a state of equilibrium exists*. Finally, if the free energy change of a process is positive ($+\Delta G$), i.e. there is likely to be an increase in the free energy change, then *the process will not proceed.*

15.19 Calculation of Gibb’s Free Energy Change of Reaction

As mentioned, Gibb’s free energy, ΔG, determines whether a reaction is spontaneous or nonspontaneous. If $\Delta G < 0$, the reaction is spontaneous; if $\Delta G > 0$, the reaction is nonspontaneous; if $\Delta G = 0$, the reaction is at equilibrium. There are numerous ways to calculate Gibb’s free energy. One of the methods will be analogous to previous methods discussed for enthalpy and entropy. The equation is shown below:

$$\Delta G^\circ \text{ (reaction)} = \Sigma n\, \Delta G^\circ \text{ (products)} - \Sigma n\, \Delta G^\circ \text{ (reactants)}$$

where n represents the moles of each product or reactant is given by the coefficient in the balanced chemical equation. As with enthalpy, you will find that the standard Gibb’s free energy of formation of elements in their standard states is zero kJ/mol.

15.19.1 Examples

(i) Let us determine the change in the Gibb's free energy for the reaction,

$$16\ H_2S(g) + 8\ SO_2(g) \rightarrow 16\ H_2O(l) + 3\ S_8(s)$$

Given that the enthalpy change of formation of H_2O, $S_8(s)$, and $SO_2(g)$ are −237.13, 0, −33.56 and −300.19 kJ/mole, respectively.

(a) $\Delta G°\ (\text{reaction}) = \Sigma n\ \Delta G°\ (\text{products}) - \Sigma n\ \Delta G°\ (\text{reactants})$ (1)

Considering the stoichiometric equation, we have

$$\Delta G^o_{rxn} = [(16\ \text{mol})(\Delta G°\ \text{of}\ H_2O(l)) + (3\ \text{mol})(\Delta G°\ \text{of}\ S_8(s))] - [(16\ \text{mol})(\Delta G°\ \text{of}\ H_2S(g)) - (8\ \text{mol})(\Delta G°\ \text{of}\ SO_2(g))].$$

From the given data, substituting for the Gibb's free energies:

$$\Delta G°_{rxn} = [(16\ \text{mol})(-237.13\ \text{kJ/mol}) + (3\ \text{mol})(0\ \text{kJ/mol})] - [(16\ \text{mol})(-33.56\ \text{kJ/mol}) + (8\ \text{mol})(-300.19\ \text{kJ/mol})].$$

$$\Delta G°_{rxn} = -3794.1\ \text{kJ} + 0\ \text{kJ} + 537.0\ \text{kJ} + 2401.5\ \text{kJ}$$

$$\Delta G°_{rxn} = -\ 855.6\ \text{kJ}$$

As predicted, *the reaction is spontaneous, since* $\Delta G°_{rxn} < 0$.

(b) There is an alternative means for calculating $\Delta G°_{rxn}$, in which the Gibb's free energy is related to the enthalpy, entropy and temperature. If these values are known, then Gibb's free energy may be calculated by

$$\Delta G° = \Delta H° - T\,\Delta S° \qquad (2)$$

The enthalpy change, ΔH, of the above reaction is –1876 kJ, the entropy change, Δs, is –3.375 kJ/K and the temperature has a value of 298 K (standard temperature). Substituting these quantities into the above equation gives

$$\Delta G° = -1876\ \text{kJ} - (298\ \text{K})(-3.375\ \text{kJ/K}) = -870.2\ \text{kJ}$$

This answer agrees reasonably well with the result from the calculation from the Gibb's free energies of formation.

2. Is the following reaction spontaneous at 25°C?

$$2H_2(g) + O_2(g) \rightarrow 2H_2O(g)$$

Solution:

First calculate ΔH and ΔS for the reaction and finally find the value of ΔG of the reaction.

Compound	$\Delta H_f°$ (kJ/mol)	$\Delta S°$ (J/mol/K)
$H_2(g)$	0.0	130.6
$O_2(g)$	0.0	205.0
$H_2O(g)$	−241.8	188.7

Table 15.4

Thermodynamic data of molecules

Calculation of both ΔH, ΔS and ΔG of the reaction:

(i) $\Delta H = [\Delta H \text{ Product}] - [\Delta H \text{ reactant}]$

$\Delta H = [2(\Delta H_{H_2O}°)] - 2(\Delta H_{H_2}°) + 1\Delta H_{O_2}°$

$\Delta H = (2 \times -241.8) - (2 \times 0 + 1 \times 0)$

$\Delta H = -483.6$ kJ/mol

(ii) $\Delta S = [2(\Delta S_{H_2O}°) - [2(\Delta S_{H_2}°] + 1\ \Delta S_{O_2}°$

$\Delta S = (2 \times 188.7) - (2 \times 130.6 + 1 \times 205.0)$

Therefore $\Delta S = -88.6$ J/mol/K $= -\ 0.0886$ kJ/mol/K

Combine the two of them into ΔG

(iii) $\Delta G = \Delta H - T \cdot \Delta S$

$\Delta G = -483.6 \text{ kJ/mol} - 298\text{K} \times (-0.0886 \text{ kJ/mol/K})$

$\Delta G = -457$ kJ/mol

The reaction is spontaneous, since $\Delta G < 0$.

15.20 Thermochemistry

All physical or chemical processes are accompanied by energy changes; either energy is released or absorbed. This is the subject matter of 'thermochemistry'. The energy change during the process depends on many factors, viz. pressure, volume, quantity, physical state, etc.

15.20.1 Stoichiometry and Energy Balances

Stoichiometry deals with calculations about the masses (sometimes volumes) of reactants and products involved in a chemical reaction.

Let us consider the equation for the Haber reaction, the combination of nitrogen gas and hydrogen gas to make ammonia.

$$N_2 + 3H_2 \rightarrow 2NH_3$$

The formula for nitrogen is N_2 and the formula for hydrogen is H_2. The formula for ammonia is NH_3. The balanced equation requires one *mole* of nitrogen and three *moles* of hydrogen make two moles of ammonia. The suitable way to measure amounts by weight; and so 28 g of nitrogen and 6 g of hydrogen make 34 g of ammonia.

A mole ratio is just the ratio of one material in a chemical equation to another material in the same equation. The mole ratio uses the coefficients of the materials as they appear in the balanced chemical equation, i.e. 1 mole nitrogen combines with three moles of hydrogen to give two moles of ammonia.

$$1 \text{ mole } N_2 = \frac{\text{wt. in g}}{\text{Mol wt}} = \frac{28\text{g}}{28};\ 3 \text{ mole } H^2 = \frac{\text{wt in g}}{\text{Mol wt}} = \frac{3 \times 2\text{g}}{2};\ 2 \text{ mole } NH_3 = \frac{2 \times 17\text{g}}{17}$$

15.20.2 Mole Concept in Thermochemistry

The mole is a unit that is defined as 6.023×10^{23} particles. It is also equal to the formula mass of a substance expressed in grams. For example:

1 mole of $H_2O = 6.023 \times 10^{23}$ molecules of H_2O = 18 g of H_2O

This statement establishes the relationship between mole units and molecule units and the relationship between mole units and gram units.

The following are examples as to how the mole concept is applied:

- Mole to gram conversion
- Mole to molecule conversion
- Mole to atom conversion.

15.20.3 Practice Problems

1. Mole to gram conversion

Determine the number of grams in four moles of H_2O

Formula mass $H_2O = (2 \times 1.0) + (1 \times 16) = 18$

1 mole H_2O = formula mass H_2O = 18 g of H_2O

4 moles $H_2O \times 18$ g/1 mole = 72 g of H_2O

2. Gram to mole conversion

Determine the number of moles in 88 g of CO_2

Formula mass $CO_2 = (1 \times 12) + (2 \times 16) = 44$

1 mole of CO_2 = formula mass CO_2 = 44 g of CO_2

88 g of $CO_2 \times 1$ mole CO_2/44 g of CO_2 = 2 moles CO_2

3. Mole to molecule conversion

Determine the number of molecules of H_2O in three moles H_2O

1 mole $H_2O = 6.023 \times 10^{23}$ molecules H_2O

3 moles $H_2O \times 6.023 \times 10^{23}$ molecules H_2O/1 mole $H_2O = 18.069 \times 10^{23}$ molecules $H_2O = 1.8069 \times 10^{24}$ molecules H_2O.

4. Mole to atoms conversion

Example:

How many atoms are present in 17.0 mole of water?

In order to convert moles of a substance to atoms one must first convert moles to molecules

Step 1:

moles → molecules

The relationship between moles and molecules is

1 mole of any substance = 6.023×10^{23} molecules of that substance.

17.0 moles $H_2O \times 6.023 \times 10^{23}$/1 mole $H_2O = 102.4 \times 10^{23} = 1.02 \times 10^{25}$ molecules H_2O

Step 2:

molecules of substance → total atoms

That would depend upon the formula of the substance which in this case is H_2O

The subscripts that appear right after the symbol representing an element can be interpreted as atoms of that element in one molecule of the substance.

According to the formula for H_2O:

1 molecule H_2O = 2 atoms H + 1 atom O = 3 atoms total

So, once you have determined the number of molecules of H_2O from step 1 you then can use the above relationship between molecules and total atoms to determine the total atoms in the number of molecules determined in step 1

$$1.02 \times 10^{23} \text{ molecules } H_2O \times 3 \text{ atoms/1 molecule } H_2O = 3.06 \times 10^{23} \text{ atoms total}$$

(i) Mole relations in balanced equations:

One of the most common chemistry problems is figuring out how much product you will form if you do a reaction. The starting point with all of these types of problems is the *balanced chemical equation.*

The balanced chemical equation tells you how many molecules of a reactant are needed to produce some number of molecules of product.

For example, in the combustion of methanol

$$2CH_3OH(l) + 3O_2(g) \rightarrow 4H_2O(g) + 2CO_2(g)$$

Two molecules of methanol react with three molecules of oxygen to form four molecules of water and two of carbon dioxide. To figure out the amount of product from the amount of reactant,

- start with the number of moles of the reactant,
- write out the conversion factor to convert from reactant to product, and
- simply multiply.

Example: Four water molecules are formed for every two molecules of methanol from the above stoichiometric equation.

To produce one mole of water, stoichiometry of equation demands 2/4 moles of methanol.

Therefore, 0.5 mole methanol produces 1 mole of water.

15.20.4 Thermochemical Stoichiometry Problems

1. Two moles of hydrogen combines with one mole of oxygen to form two moles of water and the enthalpy change of reaction is $\Delta H = -571.7$ kJ.

   $$2H_2\,(g) + O_2\,(g) \rightarrow 2H_2O \qquad \Delta H = -571.7 \text{ kJ}$$

   How much energy is produced if 100 g of H_2 are converted to water?

   Convert 100 g of H_2 to moles H_2:

   $$100 \text{ g of } H_2 \times 1 \text{ mole } H_2/2.0 \text{ g of } H_2 = 50 \text{ moles } H_2$$

   - Identify the mole kJ relationship for the given H_2 using the balanced thermochemical equation
   - Convert moles H_2 to kJ

   Therefore, 50 moles $H_2 \times -571.7$ KJ/2 moles $H_2 = -14292.5$ kJ.
2. Given the following equation:

   $$2H_2(g) + O_2(g) \rightarrow 2H_2O(l)\ \Delta H = -571.7 \text{ kJ}$$

   How many grams of H_2 would be required to produce 5000 kJ?

   - Convert kJ to moles of requested H_2 using the balanced thermochemical equation:

     $$5000 \text{ kJ} \times 2 \text{ moles } H_2/-571.7 \text{ kJ} = 17.49 \text{ moles } H_2$$

   - Convert moles of requested H_2 to grams H_2

   Therefore, 17.49 moles $H_2 \times 2.0$ g of H_2/1 mole $H_2 = 34.98$ g of H_2
3. How many moles of nitrogen do we form If 0.35 moles of N_2H_4 react with $N_2O_4(l)$, how many moles of nitrogen is liberated?

   $$2N_2H_4(l) + N_2O_4(l) \rightarrow 3N_2(g) + 4H_2O(g)$$

Solution

It is evident from the above equation that for every two molecules (moles) of N_2H_4, three molecules (moles) of nitrogen are liberated and therefore, for the reaction of 0.35 moles N_2H_4, the number of moles of nitrogen liberated is

$$0.35 \text{ moles } N_2H_4 \times 3 \text{ moles } N_2/2 \text{ moles } N_2H_4 = 0.53 \text{ moles } N_2 \text{ formed.}$$

15.20.5 Thermochemical Stoichiometry

Thermochemical stoichiometry is stoichiometry problems with the energy factor involved in chemical reactions.

Thermochemical equations are balanced equations with the energy exchange indicated either within the equation or as an enthalpy, ΔH, term on the outside of the equation. For example:

$$2Na + 2H_2O(l) \rightarrow 2NaOH(aq) + H_2(g); \qquad \Delta H = -367.5 \text{ kJ.}$$

Using the coefficients and the ΔH, one can form mole kJ relationships for each component found in the balanced equation. The above thermochemical equation could be of use in any one of the following mole kJ relationships:

- 2 moles Na = −367.5 kJ
- 2 moles H_2O = −367.5 kJ
- 2 moles NaOH = −367.5 kJ
- 1 mole H_2 = −367.5 kJ

15.20.5.1 Thermochemical Stoichiometry Problems

(i) How much energy will be produced if 18 moles Na are reacted with enough water to consume all the Na?

Determine the mole kJ relationship for the given component using the balanced equation:

$$2 \text{ moles Na} = -367.5 \text{ kJ}$$

Convert given moles of Na to requested kJ of energy

$$18 \text{ moles Na} \times -367.5 \text{ kJ}/2 \text{ moles Na} = 9\ (-367.5 \text{ kJ}) = -3307.5 \text{ kJ.}$$

(ii) Another example illustrating a mass kJ problem:

How much energy will be produced when 69 g of Na react with excess water.

Convert grams Na to moles of Na:

$$69 \text{ g of Na} \times 1 \text{ mole Na}/23 \text{ g of Na} = 3 \text{ moles Na}$$

Convert moles Na to kJ using the balanced equation

$$3 \text{ moles Na} \times -367.5 \text{ kJ}/2 \text{ moles Na} = -551.25 \text{ kJ}$$

15.20.6 Enthalpies of Reaction

The overall change in the enthalpy between products and reactants in a reaction is called the *reaction enthalpy*. The reaction enthalpy can be either positive or negative.

1. As an example, consider the reaction:

$$CO(g) + \tfrac{1}{2}\, O_2 \rightarrow CO_2(g) \qquad \Delta H = -\ 283.0 \text{ kJ}$$

The change enthalpy when 1 mole of CO reacts completely with 0.5 mole of O_2 at 25°C can be measured by calorimetry and is observed to be

$$\Delta H = [H_{(\text{Products})} - H_{(\text{reactants})}] = q_P = -\,283.0 \text{ kJ}$$

The *negative* sign ($-\Delta H$) means that heat is given off by the system, rather than absorbed by it.

The enthalpy change, ΔH, of a reaction at constant pressure is defined as the difference in enthalpies (in kJ) of the products and the reactants, when stoichiometric g/moles of the reactants react completely in a chemical equation.

When heat is given off, the reaction is said to be *exothermic* and ΔH will be negative. If heat is absorbed, the reaction is said to be *endothermic* ($+\Delta H$). The reverse of the above reaction would be endothermic:

$$CO_2(g) \rightarrow CO(g) + \tfrac{1}{2}\, O_2(g) \qquad \Delta H = +283.0 \text{ kJ}$$

Because enthalpy is an extensive quantity, if we multiply the reaction by a constant, the enthalpy of the reaction is multiplied by the same constant.

Examples:

1. If we consider the decomposition of two moles of CO_2 to give two moles CO and one mole of O_2 according to,

   $2CO_2(g) \rightarrow 2CO(g) + O_2(g)$; the enthalpy of the reaction would be

   Therefore, $\Delta H = 2 \times 283.0 \text{ kJ} = 566.0 \text{ kJ}$.
2. Colourless nitric oxide, NO, combines with oxygen to form nitrogen dioxide, NO_2, a brown gas (photochemical smog): $2NO(g) + O_2(g) \rightarrow 2NO_2(g)$; $\Delta H = -114$ kJ
   What is the enthalpy per gram of nitric oxide?

Solution

This equation reads 'two moles of nitric oxide plus one mole of oxygen produces two moles of nitrogen dioxide and 114 kJ of heat'. To figure out the enthalpy per gram of nitric oxide you must determine how much heat is produced from 1 g of nitric oxide.

Since this is for 1 g of NO, the answer then is −1.9 kJ/gram.

Examples:

3. Calcium oxide (quick lime) reacts with water to produce calcium hydroxide (slaked lime).

   $$CaO(s) + H_2O(l) \rightarrow Ca(OH)_2(s); \qquad \Delta H = -65.2 \text{ kJ}$$

   How much heat is released from the reaction of 28.4 g of CaO with excess water?

Solution

28.4 g of CaO will liberate 33.0 kJ of heat.

15.20.6.1 Enthalpy is an Extensive Property

This fact means that *the magnitude of ΔH is directly proportional to the amount of reactant consumed in the process*. Consider the combustion of methane to form carbon dioxide and liquid water. It is found experimentally that 890 kJ of heat is produced when 1 mole of CH_4 is burned in a constant-pressure system:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O; \qquad \Delta H = -\,890 \text{ kJ}$$

Because the combustion of one mole of CH_4 with two moles of O_2 releases 890 kJ of heat, the combustion of twice this quantity (two moles of CH_4 with four moles of O_2) releases twice as much heat, 1780 kJ.

15.20.6.2 Enthalpy Change of Reaction at Constant Pressure and Constant Volume

Chemical reactions may be taking place at constant volume or at constant pressure. For solids and liquids, the effect of pressure is negligible. But, for reactions of gases, the enthalpy change of reaction is not the same.

(i) Enthalpy change of reaction at constant volume

According to the first law,

$$q = \Delta H = \Delta E + P \cdot \Delta V$$

when the volume is constant, $dV = 0$ and hence, $\Delta H = \Delta E$.

That is $q = \Delta E$

Thus, *the process taking place at constant volume, the enthalpy change is equal to the change in the internal energy of the system.*

(ii) Enthalpy change of reaction at constant pressure

If H_{Pr}, is the enthalpy of the products and H_{R} that of the reactants, then,

$$H_{\mathrm{Pr}} - H_{\mathrm{R}} = \Delta H$$

According to the first law,

$$q = \Delta H = \Delta E + P \cdot \Delta V$$

The enthalpy of reaction at constant pressure and at given temperature is given by the difference in the enthalpies of the products and the reactants in a stoichiometric chemical equation.

In other words, *for processes taking place at constant pressure, the heat absorbed is equal to the change in the heat content or enthalpy.*

15.20.6.3 Relation Between the Enthalpy Change of Reaction at Constant Pressure and Constant Volume

The enthalpy change of reaction at constant pressure has been identified as ΔH

And that at constant volume with ΔE and these quantities are related to one another as

$$\Delta H = \Delta E + P \cdot dV$$

where dV is the increase in volume at constant pressure.

Suppose n_1 and n_2 are the number of moles of gaseous reactants and products, respectively. Then, increase in the number of moles $= n_2 - n_1 = \Delta n$.

Assuming the gas law to hold good, if one mole occupies a volume V, then, Δn mole will occupy a volume of $\Delta n \cdot V$ at constant pressure.

Therefore, work done, W, by the system at constant pressure,

$$W = P \cdot \Delta V = \Delta n \cdot RT.$$

Since, $V = RT/P$.

$$dV = \Delta n \cdot RT/P$$

Substituting dV, we get

$$\Delta H = \Delta E + P \cdot \Delta n \cdot RT/P$$

$$\Delta H = \Delta E + \Delta n \cdot RT.$$

Examples:

1. The dissociation of gaseous ammonia at constant pressure at a particular temperature, the enthalpy change of the reaction is

   $$2NH_3 \rightarrow N_2 + 3\ H_2$$

   2 moles 1mole 3 moles

   Therefore, $\Delta n = 4 - 2 = 2$

   $\Delta H = \Delta E + 2 \cdot RT.$

   Hence, $\Delta H > \Delta E.$

2. The formation gaseous water from hydrogen and oxygen gases at constant pressure.

   $$2H_2(g) + O_2(g) \rightarrow 2H_2O(g)$$

   2 moles 1 mole 2 moles

   Therefore, $\Delta n = 2 - 3 = -1$

   $\Delta H = \Delta E - RT.$

   Hence, $\Delta H < \Delta E.$

15.20.6.4 Standard-State Enthalpies

Absolute enthalpies of substances, like absolute energies, cannot be measured—only changes in enthalpy can be measured. Thus, it is necessary to have a reference state, just as it is necessary to establish a reference point for computing potential energies.

The reference state for reaction enthalpies is called the *standard state*. This is defined as follows:

1. For liquids and solids, the standard state is the thermodynamically stable state at a pressure of 1 atmosphere and a specified temperature. The stable state could be liquid or solid.
2. For gases, the standard state is the gaseous state at a pressure of 1 atmosphere and a specified temperature, based on the assumption of ideal gas behaviour.
3. For dissolved species, the standard state is a 1 M solution at a pressure of 1 atmosphere and specified temperature, based on the assumption of ideal solution behaviour.

Given this definition of the standard state, a scale needs to be defined by arbitrarily setting the enthalpies of selected reference substances to zero. The choice of reference is the following: *the enthalpies of chemical elements in their standard states at 298.15 K are defined to be 0*. For elements that can exist in several forms, we choose that form and 298.15 K. For example, oxygen can exist as O(g) or O(g). The former is more stable, so it is defined to have a 'zero' enthalpy. Carbon can exist as graphite, diamond or fullerene. Graphite is the most stable of these at 1 atmosphere and 298.15 K, so it is assigned an enthalpy of '0'. For any reaction, the standard state enthalpy is denoted as $\Delta H°$.

In order to have useful enthalpy data to tabulate, one more concept is introduced.

The standard enthalpy of the formation of a compound is the enthalpy change for a reaction that produces 1 mole of the compound from its elements at 1 atmosphere pressure and 25°C.

For example, the standard enthalpy of formation of water is the standard enthalpy of the reaction:

$$H_2(g) + ½O_2(g) \rightarrow H_2O(l);\ \Delta H° = -\ 283.83\ kJ$$

For species that dissolve into ions in solution, ions of both positive and negative charge will be produced. Clearly, we cannot measure the enthalpy of formation of ions of only one type of charge, rather, only the sum of the enthalpies of formation of ions of both positive and negative charges can be measured. Thus, one more arbitrary assignment must be made, namely, that the standard enthalpy of formation of H^+(aq) is zero.

Example: Consider the following reaction, $C(s) + CO_2(g) \rightarrow 2CO(g)$
What is the standard state enthalpy of the reaction?

Consider this as a two-step process, wherein CO is broken down into its elements, then these elements recombine to form CO(g). The two steps of the reaction are

$$CO_2\,(g) \rightarrow C(s) + O_2\,(g)$$
$$2C(s) + O_2\,(g) \rightarrow 2CO\,(g)$$

It is easy to verify that the sum of these two reactions reproduces the desired reaction above.
The following standard enthalpies of formation are given:

$$C(s) + O_2\,(g) \rightarrow CO_2\,(g)\ \Delta H^\circ = -\,393.51\text{ kJ}$$
$$C(s) + \tfrac{1}{2}\,O_2\,(g) \rightarrow CO(g)\ \Delta H^\circ = -\,110.52\text{ kJ}$$

Thus, for the reaction considered, the standard enthalpy of formation is the sum of the standard enthalpies of formation for the two individual steps:

$$\Delta H^\circ = -\Delta H^\circ\,(CO_2) + 2\,\Delta H^\circ\,(CO) = 172.47\text{ kJ.}$$

As one final consideration of standard state enthalpies, *the bond enthalpy is defined as the enthalpy required to break 1 mole of a particular chemical bond in the gas phase.* Since energy is required to break a chemical bond, bond enthalpies are always positive.

An example of a bond enthalpy is the breaking of a C–H bond in methane:

$$CH_4(g) \rightarrow CH_3(g) + H(g);\ \Delta H^\circ = 438.0\text{ kJ}$$

Although the products in this reaction are not stable, i.e. they react quickly with other molecules, it is possible to produce such reactions in the lab, and study properties of the products over the period of their brief existence. This is how bond enthalpies are tabulated. Often, bond enthalpies are tabulated as averages of a particular bond breaking, e.g. a C–H bond, over many compounds. Given the relative constancy of bond enthalpies of a particular kind of bond over different molecules, such an averaging is a good measure of the bond enthalpy one might expect in an arbitrary specific situation.

15.20.6.5 Examples

1. Calculate the change in enthalpy for the above overall reaction, given:

$$N_2(g) + O_2(g) \rightarrow 2NO(g);\quad \Delta H = +181\text{ kJ}$$
$$2NO(g) + O_2(g) \rightarrow 2NO_2(g);\ \Delta H = -131\text{ kJ}$$

This problem is very simple. Adding up the two reactions keeping all the reactants on the left and all the products on the right gives the overall equation as stated below. The change in enthalpy of reaction is thus obtained:

$$\begin{array}{ll}
N_2(g) + O_2(g) \rightarrow 2NO(g); & \Delta H = +181\text{ kJ} \\
2NO(g) + O_2(g) \rightarrow 2NO_2(g); & \Delta H = -131\text{ kJ} \\
\hline
N_2(g) + O_2(g) + 2NO(g) + O_2(g) \rightarrow 2NO_2(g); & \Delta H = +\,68\text{ kJ}
\end{array}$$

15.20.7 Enthalpies of Formation

Enthalpy of formation (ΔH_f°) is the change in enthalpy when one mole of a compound is produced from the free elements. The units will be kilojoules per mole.

A negative sign represents an exothermic reaction. *The compound has less enthalpy than the elements from which it was formed.* A positive sign represents an endothermic reaction. *The compound has more enthalpy than the elements from which it was formed.*

Conditions which influence enthalpy changes include:

- temperature
- pressure
- physical state of reactants and products (*s, g, l, aq*)

The standard state of a substance is the form most stable at 298°K (25°C), or and 1 atmosphere of pressure.

When a reaction occurs with all reactants and products in their standard states, the enthalpy change is *the standard enthalpy of reaction* ($\Delta H°$).

Thus, *the standard enthalpy of formation* ($\Delta H_f°$) *of a compound is the change in enthalpy that accompanies the formation of 1 mole of that substance from its elements, with all substances in their standard states.*

The standard enthalpy of formation for ethanol (C_2H_5OH) is the enthalpy change for the following reaction

$$2C(s) + 3H_2(g) + ½\,O_2(g) \rightarrow C_2H_3OH(l);\ \Delta H_f° = -\,277.7\text{ kJ}$$

Notes:

- Elemental source of oxygen is O_2 and not O, because O_2 is the stable form of oxygen at 25°C and 1 atmosphere, likewise with H_2.
- Elemental source of carbon is specified as *graphite*, because graphite is the lowest energy form of carbon at room temp and 1 atmosphere.
- Why is the O_2 stoichiometry left at '1/2'?

The stoichiometry of formation reactions *always indicates the formation of 1 mole of a product.* Thus, $\Delta H_f°$ values are reported as kJ/mole of the substance produced.

If C (*graphite*) is the lowest energy form of carbon under standard conditions, then what is the $\Delta H_f°$ for C (*graphite*)?

By definition, *the standard enthalpy of formation of the most stable form of any element is zero* because there is no formation reaction needed when the element is *already* in its standard state.

Therefore, $\Delta H_f°$ for C (*graphite*), $H_2(g)$ and $O_2(g) = 0$.

15.20.7.1 Calculation of Enthalpies of Reaction under Standard Conditions ΔH°rxn) using Enthalpies of Formation ($\Delta H_f°$)

The standard enthalpy change for any reaction ($\Delta H°_{rxn}$) *can be determined by using standard enthalpies of formation* ($\Delta H_f°$) *and Hess's law.*

Examples

1. Consider the following combustion reaction of propane:

$$C_3H_8(g) + 5O_2(g) \rightarrow 3CO_2(g) + 4H_2O(l)$$

(a) *The reactants:*

The standard heat of formation ($\Delta H_f°$) of propane gas from its elemental constituents in the standard state is –103.85 kJ/mole

$$3C(graphite) + 4H_2(g) \rightarrow C_3H_8(g);\ \Delta H_f° = -103.85\text{ kJ}$$

The standard heat of formation ($\Delta H_f°$) for $O_2(g)$ is zero

$$O_2(g) \rightarrow O_2(g)\Delta H_f° = 0\text{ kJ}$$

and so.

$$5O_2(g) \rightarrow 5O_2(g);\ \Delta H_f° = 0\text{ kJ}$$

Overall, therefore, the standard heat of formation ($\Delta H_f°$) for the reactants is

$3C(graphite) + 4H_2(g) + 5O_2(g) \rightarrow C_3H_8(g) + 5O_2(g)$
$\Delta H_f° = -103.85$ kJ

(b) *The products:*
The standard heat of formation ($\Delta H_f°$) of $CO_2(g)$ from its elemental constituents in the standard state is –393.5 kJ/mole

$C\ (graphite) + O_2(g) \rightarrow CO_2(g)$; $\Delta H_f° = -393.5$ kJ

So, for three moles of CO_2 molecules the standard heat of formation would be

$3C\ (graphite) + 3O_2(g) \rightarrow 3CO_2(g)$; $\Delta H_f° = -1180.5$ kJ

The standard heat of formation ($\Delta H_f°$) of $H_2O(l)$ from its elemental constituents in the standard state is –285.8 kJ/mole

$H_2(g) + ½\ O_2(g) \rightarrow H_2O(l)$; $\Delta H_f° = -\ 285.8$ kJ

and so the $\Delta H_f°$ for 4 waters would be

$4H_2(g) + 2O_2(g) \rightarrow 4H_2O(l)$; $\Delta H_f° = -1143.2$ kJ

combining the $\Delta H_f°$ for both products yields

$3C\ (graphite) + 4H_2(g) + 5O_2(g) \rightarrow 3CO_2(g) + 4H_2O(l)$

with $\Delta H_f° = (-1180.5) + (-1143.2) = -2323.7$ kJ.

Summarizing:

Overall, the standard heat of the formation ($\Delta H_f°$) for *the reactants* is

$3C\ (graphite) + 4H_2(g) + 5O_2(g) \rightarrow C_3H_8(g) + 5O_2(g)$
$\Delta H_f° = -103.9$ kJ

Overall, the standard heat of formation ($\Delta H_f°$) for *the products* is

$3C\ (graphite) + 4H_2(g) + 5O_2(g) \rightarrow 3CO_2(g) + 4H_2O(l)$

Therefore, $\Delta H_f° = -2323.7$ kJ

2. Calculate the enthalpy change ($\Delta H°_{rxn}$) for the combustion of 1 mole of ethanol

$C_2H_5OH(l) + 3O_2(g) \rightarrow 2CO_2(g) + 3H_2O(l)$.

(a) *Heat of formation for reactants*

$2C(graphite) + 3H_2(g) + (1/2)O_2(g) \rightarrow C_2H_5OH(l)$; $\Delta H_f° = -277.7$ kJ

plus

$3O_2(g) \rightarrow 3O_2(g)$; $\Delta H_f° = 0$ kJ

gives

$2C\ (graphite) + 3H_2(g) + (7/2)O_2(g) \rightarrow C_2H_5OH(l) + 3O_2(g)$
$\Delta H_f° = -277.7$ kJ

(b) *Heat of formation for products*

$C\ (graphite) + O_2(g) \rightarrow CO_2(g)$; $\Delta H_f° = -393.5$ kJ

Therefore,

$2C\ (graphite) + 2O_2(g) \rightarrow 2CO_2(g)$; $\Delta H_f° = -787$ kJ

$H_2(g) + (1/2)O_2(g) \rightarrow H_2O(l)$; $\Delta H_f^\circ = –285.8$ kJ

Therefore,

$3H_2(g) + (3/2)O_2(g) \rightarrow 3H_2O(l)$; $\Delta H_f^\circ = −857.4$ kJ

combining gives

$2C\ (graphite) + 3H_2(g) + (7/2)O_2(g) \rightarrow 2CO_2(g) + 3H_2O(l)$
$\Delta H_f^\circ = (–857.4) + (–787) = –1644.4$ kJ
ΔH° of reaction

Therefore, $\Delta H^\circ_{rxn} = \Delta H_f^\circ$ (products) – ΔH_f° (reactants)
$= (–1644.4) – (–277.7) = –1366.7$ kJ

3. Calculate the ΔH_r° for the reaction: $P_4O_{10} + 6H_2O_{(l)} \rightarrow 4H_3PO_4$
The solution:
Use the equation $\Delta H_r^\circ = \Delta H^\circ_{f\,(products)} – \Delta H^\circ_{f\,(reactants)}$
From the table:
   - ΔH_f° for reactant P_4O_{10}; $\Delta H_f^\circ = –2980$ kJ/mole
   - ΔH_f° for reactant $H_2O(l)$; $\Delta H_f^\circ = –286$ kJ/mole
   - ΔH_f° for product H_3PO_4; $\Delta H_f^\circ = –1280$ kJ/mole

   Multiply each value by the coefficient from the balanced equation which represents the number of moles:

   ΔH_f° for reactant P_4O_{10}; $\Delta H_f^\circ = (–2980 \text{ kJ/mole}) \times (1 \text{ mole}) = –2980$ kJ
   ΔH_f° for reactant $H_2O(l)$; $\Delta H_f^\circ = (–286 \text{ kJ/mole}) \times (6 \text{ moles}) = –1716$ kJ
   ΔH_f° for product H_3PO_4; $\Delta H_f^\circ = (–1280 \text{ kJ/mole}) \times (4 \text{ moles}) = –5120$ kJ

   Plug the numbers into the equation and calculate:

   $\Delta H_r^\circ = \Delta H^\circ_{f\,(products)} – \Delta H^\circ_{f\,(reactants)}$
   $\Delta H_r^\circ = (–5120 \text{ kJ}) – [(–1716 \text{ kJ}) + (–2980 \text{ kJ})]$
   $\Delta H_r^\circ = (–5120 \text{ kJ}) – (–4696 \text{ kJ})$
   $\Delta H_r^\circ = –424$ kJ.

4. Calculate ΔH for the reaction $CH_4(g) + Cl_2(g) \rightarrow CH_3Cl(g) + HCl(g)$ from the following

Substance	ΔH_f (kJ/mol)
$CH_4(g)$	–74.8
$CH_3Cl(g)$	–82.0
HCl(g)	–92.3

Table 15.5

Enthalpies of compounds

Solution:

Remember products minus reactants and that the enthalpy of formation of $Cl_2(g)$ is zero (it is in its standard state):

$$\Delta H = [(1)(–82.0 \text{ kJ/mol}) + (1)(–92.3 \text{ kJ/mol})] – [(1)(0) + (1)(–74.8 \text{ kJ/mol})]$$

Therefore, $\Delta H = −99.5$ kJ/mol.

5. Enthalpy of formation calculation:
   Calculate the enthalpy change when one mole of B_5H_9 reacts with excess oxygen according to the following reaction

$$2B_5H_9(g) + 12\ O_2\ (g) \rightarrow 5B_2O_3(g) + 9H_2O\ (g)$$

Compound	ΔH_f(kJ/mol-K)
$B_5H_9(g)$	73.2
$B_2O_3(g)$	–1272.77
$O_2(g)$	0
$H_2O(g)$	–241.82

Table 15.6

Enthalpies of formation

In the reaction above two moles of B_5H_9 react with 12 moles of O_2 to yield five moles of B_2O_3 and nine moles of H_2O. The ΔH_f of the reaction is obtained by subtracting the sum of the enthalpies of the reactant from the sum of the enthalpies of the products.

The heat of formation of O_2 is zero because this is the form of the oxygen in its most thermodynamically stable state.

$$\Delta H_f = \Sigma nH^\circ\ (\text{products}) - \Sigma mH^\circ\ (\text{reactants})$$

$$= \frac{[\{2 \text{ mole } B_2O_3 \times -122.77 \text{ kJ}\}}{\text{mole}} + \frac{\{9 \text{ mole } H_2O \times -241.82 \text{ kJ}\}]}{\text{mole}}$$

$$- [\{2 \text{ mole } B_5H_9 \times 73.2 \text{ kJ}\} + \{12 \text{ mole } O_2 \times 0 \text{ kJ}\}]$$

$$= [(-6363.85 - 2176.38) - 146.4]$$

Therefore, $\Delta H_f = -8686.6$ kJ

15.20.8 Enthalpy of Combustion

Enthalpy of combustion, ΔH, of a substance is the enthalpy change accompanying the complete combustion of 1 mole of a substance with oxygen at a definite temperature and constant pressure.

The enthalpy change of combustion of substances also depends on the physical state, pressure, temperature, etc.

For example

$$\text{S(rhombic, solid)} + O_2(g); \quad \Delta H = -296.9 \text{ kJ/mol}$$
$$\text{S(monoclinic, solid)} + O_2(g); \ \Delta H = -297.2 \text{kJ/mol}$$

The most important of these are organic compounds, hydrocarbons (methane), and carbohydrates (glucose). The products are always, carbon dioxide and water.

For example,

$$CH_4(g) + 2O_2(g) \rightarrow CO_2(g) + 2H_2O(g) \qquad \Delta H = -882 \text{ kJ/mol}$$

Sucrose gives more energy per mole.

$$C_{12}H_{22}O_{11}(s) + 12O_2(g) \rightarrow 12CO_2(g) + 11H_2O(g) \quad \Delta H = -5644 \text{ kJ/mol}$$

The water is either produced as liquid or gas depending on the end conditions. This will give a slightly different result.

Often the enthalpy of formation and combustion can appear to be the same, especially if the formed compound is an oxide.

$$\text{C(graphite)} + O_2\,(g) \rightarrow CO_2(g) \qquad \Delta H = -393 \text{ kJ/mol}$$

As well as being the enthalpy of combustion of carbon in its standard state (graphite), this is also the enthalpy of formation of carbon dioxide.

Using diamond, however, would result in an incorrect value for the formation of CO_2, as diamond is less thermodynamically stable than graphite.

An enthalpy of formation must also be 1 mole of the substance, so the following would not be an enthalpy of formation.

$$2Li(s) + \tfrac{1}{2}O_2(g) \rightarrow Li_2O(s) \qquad \Delta H = -596 \text{ kJ/mol}$$

when a mole of propane is burnt completely with oxygen, the enthalpy change observed is

$$C_3H_8(g) + 5O_2(g) \rightarrow 3CO_2(g) + 4H_2O(l); \quad \Delta H = -2220 \text{ kJ};$$

Therefore, the heat of combustion of $C_3H_8(g)$ is –2220 kJ

Enthalpy of combustion—this is the heat released when 1 mole of a substance is completely burnt in oxygen.

$CH_4(g) + 2O_2(g) \rightarrow CO_2(g) + 2H_2O(l)$	$\Delta H = -882$kJ/mol
$CH_3OH(l) + 1\tfrac{1}{2}O_2(g) \rightarrow CO_2(g) + 2H_2O(l)$	$\Delta H = -715$kJ/mol
$C_2H_5OH(l) + 3O_2(g) \rightarrow 2CO_2(g) + 3H_2O(l)$	$\Delta H = -1371$kJ/mol
$C_3H_7OH(l) + 4\tfrac{1}{2}O_2(g) \rightarrow 3CO_2(g) + 4H_2O(l)$	$\Delta H = -2010$kJ/mol
$C_4H_9OH(l) + 6O_2(g) \rightarrow 4CO_2(g) + 5H_2O(l)$	$\Delta H = -2670$kJ/mol

Table 15.7

Enthalpies of combustion

The increase in enthalpy of combustion is regular as you proceed up the homologous series because each increase of a carbon and two hydrogen atoms produces the same amount of extra heat energy.

15.20.9 Enthalpy of Neutralization

Strong acids react with strong bases forming salt and water and the reaction is known as ‘neutralization reaction’.

The enthalpy of neutralization is defined as the enthalpy change when one gram equivalent of acid is neutralized completely with one gram equivalent of the base.

Enthalpy of neutralization is the heat released when 1 mole of water is formed during the neutralization of strong acid by an alkali,

For example,

1. $HCl + NaOH \rightarrow NaCl + H_2O; \qquad \Delta H = -57.0$ kJ/mol
2. $HNO_3 + NaOH \rightarrow NaNO_3 + H_2O; \qquad \Delta H = -57.0$ kJ/mol

The neutralization of any strong acid by any alkali always produces the same final equation and the same value for the change in enthalpy.

If, however, the acid or alkali is weak, the heat of neutralization will be different because the reaction will now involve ionization of the weak acid or base as well.

$$\begin{array}{ll}
CH_3COOH \rightarrow CH_3COO^- + H^+; & \Delta H = +\,2.08\text{ kJ/mol} \\
[CH_3COO^- + H^+] + [Na^+ + OH^-] \rightarrow H_2O & \Delta H = -55.20\text{ kJ/mol} \\
\hline
CH_3COOH + NaOH \rightarrow CH_3COONa + H_2O; & \Delta H = -57.28\text{ kJ/mol}
\end{array}$$

15.20.10 Enthalpy of Solution

Enthalpy change per mole of a solute when it is dissolved in a pure solvent to give a solution of a specified concentration is known as the enthalpy of solution.

1. Enthalpy of solution—this is the enthalpy change when 1 mole of substance is completely dissolved in water, e.g.

$$Na^+OH^-(s) + (aq) \rightarrow Na^+(aq) + OH^-(aq) \qquad \Delta H = -40\text{kJ/mol}$$
$$NH_4{}^+NO_3{}^-(s) + (aq) \rightarrow NH_4{}^+(aq) + NO_3{}^-(aq) \quad \Delta H = 26\text{ kJ/mol}$$

15.20.11 Enthalpy of Phase Changes

Although phase changes are not chemical reactions, a quantity of heat is still required, for example, to transform ice at 0°C to water at 0°C. This quantity of heat is called the *enthalpy of fusion*

$$H_2O(s)\rightarrow H_2O(l); \qquad \Delta H = 6.007\text{ kJ/mol}$$

Similarly, a quantity of heat is needed to transform water at 100°C to steam at 100°C. This quantity of heat is known as the *enthalpy of vaporization*

$$H_2O(l) \rightarrow H_2O(g); \quad \Delta H = 40.70\text{ kJ/mol}$$

15.20.12 Bond Enthalpy

Every bond between atoms has its own unique strength or bond enthalpy. The energy required to break a bond. Bond energy is always a *positive* number because the breaking of a bond requires an input of energy (endothermic). When a bond is formed, the amount of energy equal to the bond energy is released.

$$\Delta H^\circ = \Sigma(\text{bond energies of bonds broken}) - \Sigma(\text{bond energies of bonds formed}).$$

The bonds broken are the reactant bonds. The bonds formed are the product bonds.

Bond enthalpy is the energy required to break 1 mole of that bond into separate gaseous atoms. e.g.

$$H_2(g) \rightarrow 2H(g) \quad \Delta H = +436\text{ kJ/mol}$$
$$C(s) \rightarrow C(g) \qquad \Delta H = +715\text{ kJ/mol}$$

Bond enthalpy for a diatomic gaseous element is called its enthalpy of dissociation and bond enthalpy for a solid element is called its enthalpy of sublimation.

1. Total bond enthalpy of compounds

This is the energy required to break all of the bonds in a gaseous compound into separate gaseous atoms, e.g.

$$CH_4(g) \rightarrow C(g) + 4H(g)$$

From such a value, the mean bond enthalpy for one C–H bond with methane could be calculated =

$$(\text{total bond enthalpy})/4 = +414\text{ kJ/mol}$$

Different molecules, though, even containing the same kind of bond, will give different values, e.g. C–H in CH_4 is not the same as in $CHCl_3$.

2. Calculating mean bond enthalpies

(i) Calculate the mean bond enthalpy of C–H in methane, given that the enthalpy of formation of methane is –75 kJ/mol, enthalpy of sublimation of C is +715 kJ/mol and enthalpy of dissociation of H_2 is +436 kJ/mol.

Consider the equation,

$$CH_4(g) \rightarrow C(g) + 4H(g)$$

Data equations:

1. $C(s) + 2H_2(g) \rightarrow CH_4(g)$ $\Delta H = -75$ kJ/mol
2. $C(s) \rightarrow C(g);$ $\Delta H = +715$ kJ/mol
3. $H_2(g) \rightarrow 2H(g)$ $\Delta H = +436$ kJ/mol

Reverse 1 $CH_4(g) \rightarrow C(s) + 2H_2(g)$ $\Delta H = +75$ kJ/mol

2 $C(s) \rightarrow C(g);$ $\Delta H = +715$kJ/mol

2 × 3 $2H_2(g) \rightarrow 4H(g)$ $\Delta H = 2(+436$ kJ)

Add: $CH_4(g) \rightarrow C(g) + 4H(g)$ $\Delta H = +1662$ kJ/mol

Therefore, one C–H = total/4 = 1662/4 = 415.5 kJ/mol.

(ii) Calculate the bond enthalpy of C–C in ethane, given that the enthalpy of formation of ethane is –85 kJ/mol, the enthalpy of sublimation of carbon is +715 kJ/mol, the enthalpy of dissociation of hydrogen is +436 kJ/mol, and given that the mean bond enthalpy of C–H is +414 kJ/mol.

Target equation:

$$C_2H_6(g) \rightarrow 2C(g) + 6H(g)$$

Data equations:

$Na^+OH^-(s) + (aq) \rightarrow Na^+(aq) + OH^-(aq)$ $\Delta H = -40$ kJ/mol

1. $2C(s) + 3H_2(g) \rightarrow C_2H_6(g);$ $\Delta H = -85$ kJ/mol
2. $C(s) \rightarrow C(g);$ $\Delta H = +715$ kJ/mol
3. $H_2(g) \rightarrow 2H(g)$ $\Delta H = +436$ kJ/mol

Reverse **1** $C_2H_6(g) \rightarrow 2C(s) + 3H_2(g);$ $\Delta H = +85$ kJ/mol

2 × **2** $2C(s) \rightarrow 2C(g);$ $\Delta H = 2(+715$ kJ)

3 × **3** $3H_2(g) \rightarrow 6H(g);$ $\Delta H = 3(+436$ kJ)

Add $C_2H_6(g) \rightarrow 2C(g) + 6H(g);$ $\Delta H = +2823$ kJ/mol

Therefore, one C–C = total – 6(C–H) = 2823−6(414) = 339 kJ/mol.

The mean bond enthalpy of C–C in ethane is 339 kJ/mol.

Bond enthalpy can also be expressed as the energy released when 1 mole of a bond is formed from separate gaseous atoms,

(iii) e.g.

Bond breaking: $H_2(g) \rightarrow 2H(g);$ $\Delta H = +436$ kJ/mol

Bond forming: $2H(g) \rightarrow H_2(g);$ $\Delta H = -436$ kJ/mol

Similarly with total bond enthalpy for a gaseous compound, e.g.

Bond breaking: $CH_4(g) \rightarrow C(g) + 4H(g);$ $\Delta H = +1662$ kJ/mol

Bond forming: $C(g) + 4H(g) \rightarrow CH_4(g);$ $\Delta H = -1662$ kJ/mol

3. Find ΔH for the following reaction given the following bond energies

Bond	Bond Energy (kJ/mol)
H–H	436
O=O	499
O–H	463

Table 15.8

Different bond energies

We have to figure out which bonds are broken and which bonds are formed.

- 2 H–H bonds are broken
- 1 O = O bond is broken
- 2 O–H bonds are formed per water molecule, and there are two water molecules formed, therefore four O–H bonds are formed.

Now, we can substitute the values given into the equation:

$$\Delta H^\circ = [\underset{\text{H–H}}{2(436)} + \underset{\text{O–O}}{499}\ ] - \underset{\text{O–H}}{4(463)}$$

Therefore, $\Delta H^\circ = [1371 - 1852] = -481$ kJ

15.20.13 Hess's Law

The law states that the enthalpy change (i.e. heat evolved or absorbed) in a chemical process is the same whether the process takes place in one single step or through a series of steps.

A reaction is carried out in several steps. We have already seen that the *overall reaction* is the algebraic sum of the individual steps. Because energy/enthalpy is an additive quantity, *the overall enthalpy change for the overall reaction, obtained by adding the individual chemical equations, will be the sum of the enthalpy changes associated with each of the individual steps/chemical equations.*

The above statement is known as Hess's law.

Hess's law can be used to calculate the enthalpy of reactions that cannot be carried out straightforwardly in the laboratory.

Example

(i) Consider the reaction of solid carbon, in the form of graphite, with oxygen gas to produce carbon monoxide.

$$C(s) + \tfrac{1}{2} O_2\,(g) \rightarrow CO\,(g);$$

This reaction cannot be carried out directly experimentally. Rather, a two-step process is used, in which solid carbon is first converted to carbon *dioxide* via the reaction

$$C(s) + O_2\,(g) \rightarrow CO_2\,(g); \qquad \Delta H = -393.5 \text{ kJ/mol}$$

whose enthalpy can be easily measured at 25°C. Then the decomposition of CO to give oxygen gas and carbon monoxide can be added to this reaction according to:

$$\begin{array}{ll} C(s) + O_2(g) \rightarrow CO_2(g); & \Delta H_1 = -\,393.5 \text{ kJ/mol} \\ CO_2\,(g) \rightarrow CO(g) + \tfrac{1}{2}\, O_2(g); & \Delta H_2 = +\,283.0 \text{ kJ/mol} \\ \hline C(s) + \tfrac{1}{2}O_2\,(g) \rightarrow CO\,(g); & [\Delta H_1 + \Delta H_2] = \Delta H = -110.5 \text{ kJ/mol} \end{array}$$

(ii) Given the following reaction heats, find an equation describing the combustion of propane. Then find the total enthalpy change for the overall reaction.

1. $C(\textit{graphite}) + O_2(g) \rightarrow CO_2(g)$; $\Delta H = -394$ kJ/mol CO_2
2. $3C(\textit{graphite}) + 4H_2(g) \rightarrow C_3H_8(g)$; $\Delta H = -104$ kJ/mol C_3H_8
3. $2H_2(g) + O_2(g) \rightarrow H_2O(l)$; $\Delta H = -285.8$ kJ/mol H_2O
4. $2H_2O(l) + O_2(g) \rightarrow H_2O_2(l)$; $\Delta H = +394$ kJ/mol H_2O_2

   Combustion is the burning of a substance in the presence of oxygen. The results of combustion are carbon dioxide and water. So the general plan of the first part of the problem is to construct an equation with propane and oxygen on the left and carbon dioxide and water on the right. There is only one equation with propane in it, and that equation must be run in reverse in order to have propane as a reactant.

5. $C_3H_8(g) \rightarrow 3C(\textit{graphite}) + 4H_2(g)$; $\Delta H = +104$ kJ/mol CO_2

   Since neither graphite nor hydrogen is a final product, both intermediates must be completely consumed in subsequent reactions. Graphite is consumed in Eq. (1). The coefficients must be tripled so that the graphite completely cancels.

6. $3C(\textit{graphite}) + 3O_2(g) \rightarrow 3CO_2(g)$; $\Delta H = -1182$ kJ/mol CO_2

   Next, the hydrogen from the right side of Eq. (5) must be cancelled. Both Eqs. (3) and (4) use hydrogen. However, the product of Eq. (3) is water, a product of combustion, while the product of Eq. (4) is peroxide. Therefore, using Eq. (3) with the correct coefficients cancels the hydrogen.

7. $4H_2(g) + 2O_2(g) \rightarrow 4H_2O(l)$; $\Delta H = -1144$ kJ/mol H_2O

   Equations (5), (6), and (7) can then be combined. Substances that appear in equal amounts on the right of one equation and the left of a subsequent equation can be cancelled from both. The resulting reaction is then

8. $C_3H_8(g) = 5O_2(g)\ 3CO_2(g) + 4H_2O(l)$

   Since the enthalpies for each step were adjusted for any changes in direction or coefficients, the net enthalpy change is given by the sum of the enthalpy changes of Eqs. (5)–(7).

   This total is $\Delta H = -1650$ kJ/mole, a highly exothermic reaction.

(iii) One of the most useful applications of Hess's law is the calculation of the value of $\Delta H°$ for a reaction whose $H°$ is unknown or cannot be measured. Hess's law says that we can add thermochemical equations, including their values of $\Delta H°$, to obtain some desired thermochemical equation and its $\Delta H°$. Consider the following thermochemical equations:

$$\begin{array}{ll} C(s) + \tfrac{1}{2}O_2(g) \rightarrow CO(g); & \Delta H° = -110.5 \text{ kJ} \\ CO(g) + \tfrac{1}{2}O_2(g) \rightarrow CO_2(g); & \Delta H° = -283.0 \text{ kJ} \end{array}$$

Use them to find the $\Delta H°$in kilojoules for the reaction.

$$C(s) + O_2(g) \rightarrow CO_2(g)$$

Solution

This is a particularly simple problem, but it illustrates a few important points about all such problems. Let's add the two given thermochemical equations:

$$\begin{array}{lll}
C(s) + \frac{1}{2}O_2(g) & \rightarrow CO(g) & \Delta H^\circ = -110.5 \text{ kJ} \\
CO(g) + \frac{1}{2}O_2(g) & \rightarrow CO_2(g) & \Delta H^\circ = -283.0 \text{ kJ} \\
\hline
\multicolumn{3}{l}{C(s) + \frac{1}{2}O_2(g) + \frac{1}{2}O_2(g) + CO(g) \rightarrow CO(g) + CO_2(g)}
\end{array}$$

The resulting equations can be simplified. Cancel the CO(g) because it appears on both sides. You can do this as long as you have the same chemical in the same physical state. Add the two oxygen terms together. This gives us the target equation.

$$\begin{array}{lllll}
C(s) & + \frac{1}{2}O_2(g) & \rightarrow & CO(g) & \Delta H^\circ = -110.5 \text{ kJ} \\
CO(g) & + \frac{1}{2}O_2(g) & \rightarrow & CO_2(g) & \Delta H^\circ = -283.0 \text{ kJ} \\
\hline
C(s) & + O_2(g) & \rightarrow & CO_2(g); & \Delta H^\circ = -393.5 \text{ kJ} \\
\hline
\end{array}$$

(iv) Carbon monoxide is often used in metallurgy to remove oxygen from metal oxides and thereby give the free metal. The thermochemical equation for the reaction of CO with iron (III) oxide, Fe_2O_3, is

$$Fe_2O_3(s) + 3\,CO(g) \rightarrow 2\,Fe(s) + 3\,CO_2(g) \quad \Delta H^\circ = -26.74 \text{ kJ}$$

Use this equation and the equation for the combustion of CO

$$CO(g) + \frac{1}{2}O_2(g) \rightarrow CO_2(g) \qquad \Delta H^\circ = -283.0 \text{ kJ}$$

to calculate the value of ΔH° for the reaction

$$2Fe(s) + 1\frac{1}{2}O_2(g) \rightarrow Fe_2O_3$$

Solution

Combine the equations in such a way that we can add them to the final target equation. Then, we add the corresponding ΔH°'s to obtain the ΔH° of the target equation.

Step 1. The target equation must have 2 Fe on the left, but the first equation above has 2 Fe on the right. To move it left, reverse the entire equation and remember to reverse the sign of H°. When the equation is flipped over the Fe_2O_3 also falls into the correct position.

Step 2. There must be $1\frac{1}{2}O_2$ of the left, and we must be able to cancel *three* CO and *three* CO_2 when the equations are added. Multiply the second equation by 3 and we get the necessary coefficients. Multiply the H° values for the second equation by 3 as well. The adjusted equations are:

$$\begin{array}{ll}
2\,Fe(s) + 3\,CO_2(g) \rightarrow Fe_2O_3(s) + 3\,CO(g); & \Delta H^\circ = +26.74 \text{ kJ} \\
3\,CO(g) + 1\frac{1}{2}O_2(g) \rightarrow 3\,CO_2(g); & \Delta H^\circ = 3(-283.0 \text{ kJ}) = -849.0 \text{ kJ} \\
\hline
2\,Fe(s) + 1\frac{1}{2}O_2(g) \rightarrow Fe_2O_3; & \Delta H^\circ = -822.26 \text{ kJ}
\end{array}$$

(v) Calculate the enthalpy of formation of ethane given that the enthalpies of combustion of carbon, hydrogen and ethane are −394, −286 and −1542 kJ/mol, respectively.

Target Equation:

$$2C(s) + 3H_2(g) \rightarrow C_2H_6(g)$$

Data Equations:

$$\begin{array}{ll}
\text{1. } C(s) + O_2(g) \rightarrow CO_2(g); & \Delta H = -394 \text{ kJ/mol} \\
\text{2. } H_2(g) + \frac{1}{2}O_2(g) \rightarrow H_2O(l); & \Delta H = -286 \text{ kJ/mol} \\
\text{3. } C_2H_6(g) + (7/2)O_2(g) \rightarrow 2CO_2(g) + 3H_2O(l) & \Delta H = -1542 \text{ kJ/mol} \\
2 \times 1\ 2C(s) + 2O_2(g) \rightarrow 2CO_2(g) & \Delta H = 2(-394 \text{ kJ}) \\
3 \times 2\ 3H_2(g)\ (3/2)O_2(g) \rightarrow 3H_2O(l) & \Delta H = 3(-286 \text{ kJ}) \\
\text{Reverse 3 } 2CO_2(g) + 3H_2O(l) \rightarrow C_2H_6(g)\ (7/2)O_2(g) & \Delta H = 1542 \text{ kJ/mol} \\
\hline
\text{Add: } 2C(s) + 3H_2(g) \rightarrow C_2H_6(g) & \Delta H = -104 \text{ kJ/mol} \\
\hline
\end{array}$$

Therefore, $\Delta H_f = -104$kJ/mol

(vi) Calculate the enthalpy of formation of carbon monoxide, given the enthalpies of combustion of carbon and carbon monoxide are −394 and −283 kJ/mol, respectively.

Target equation:

$$C(s) + \tfrac{1}{2}O_2(g) \rightarrow CO(g)$$

Data equations:

1. $C(s) + O_2(g) \rightarrow CO_2(g)$; $\Delta H = -394$ kJ/mol
2. $CO(g) + \tfrac{1}{2}O_2(g) \rightarrow CO_2(g)$ $\Delta H = -283$ kJ/mol

$1\ C(s) + O_2(g) \rightarrow CO_2(g)$ $\Delta H = -394$ kJ/mol

Reverse

$2\ CO_2(g) \rightarrow CO(g) + \tfrac{1}{2}O_2(g)$ $\Delta H = 283$ kJ/mol

Add: $C(s) + \tfrac{1}{2}O_2(g) \rightarrow CO$ $\Delta H = -111$ kJ/mol

$\therefore$ $\Delta H_f = -111$ kJ/mol

15.20.14 Effect of Temperature on the Heat or Enthalpy (ΔH) of Reaction (Kirchhoff's equation);

The enthalpy change of any process, whether physical or chemical, varies with temperature. The exact influence of temperature can be worked out as follows:

Any process represented by

$$A\ (\text{reactant}) \rightarrow B\ (\text{product})$$

The internal energy change(ΔE) at constant volume is given by

$$\Delta E = E_B - E_A$$

The rate of change of internal energy with temperature is

$$[\partial \Delta E/\partial T]_V = [\partial E_B/\partial T]_V - [\partial E_A/\partial T]_V$$

$$= [C_V]_B - [C_V]_A$$

or $$[\partial \Delta E/\partial T]_V = \Delta C_V \qquad \ldots(1)$$

The rate of change of internal energy with respect to temperature is thus equal to the difference in heat capacities at constant volume of the products and reactants.

Similarly, for the reaction carried out at constant pressure, the enthalpy change of the process is given as

$$\Delta H = H_B - H_A$$

Therefore, the rate of change of enthalpy of the reaction at constant pressure will be

$$[\partial \Delta H/\partial T]_V = [\partial H_B/\partial T]_V - [\partial H_A/\partial T]_V$$

$$= [C_P]_B - [C_P]_A$$

or $$[\partial \Delta H/\partial T]_P = \Delta C_P \qquad \ldots(2)$$

Equations (1) and (2) are known as *Kirchhoff's equation* and are useful for the calculation of the effect of temperature on the ΔH and ΔE. This can be seen by integrating Eq. (2) we get,

$$[\partial \Delta H/\partial T]_P = \Delta C_P \quad \text{or} \quad \partial(\Delta H) = \Delta C_P \cdot dT$$

$$\partial \Delta H = \int \Delta C_P \cdot dT$$

Assuming ΔC_P to be independent of temperature, we get

$$\Delta H_2 - \Delta H_1 = \Delta C_P(T_2 - T_1) \quad \ldots(3)$$

Where ΔH_2 and ΔH_1 are the enthalpy change of reaction constant pressure and at temperatures T_2 and T_1. It is evident from Eq. (3) that *the extent of variation of enthalpy of reaction with temperature depends upon the magnitude and the sign of* ΔC_P.

Review Questions

1. Mention the scope and applications of thermodynamics.
2. Elucidate the concept of energy, heat and work in thermodynamics.
3. Explain the reversible isothermal work of an ideal gas.
4. Write notes on:
   (a) Isobaric expansion of an ideal gas (b) isothermal expansion of work.
5. Notes on a thermodynamics (i) A system (ii) Surrounding.
6. Mention the types of thermodynamic systems under consideration.
7. What is an open system? Illustrate.
8. What do you mean by a closed system?
9. A note on isolated system.
10. What do you mean by thermodynamic state of a system? Illustrate.
11. Explain, what are thermodynamic variables?
12. Mention the types of thermodynamic processes that can be executed to study a system.
13. State the first law of thermodynamics.
14. Give the energy, heat and work done relationship of a system.
15. What is internal energy of a system? Explain.
16. What is enthalpy of a system? Derive an expression for enthalpy change of an isobaric thermodynamic process.
17. Write notes on:
    (a) Energy change in a cyclic thermodynamic process.
    (b) Energy change at constant volume
    (c) Energy change in a adiabatic process
    (d) Energy change in constant pressure process.
18. Write a note on heat capacity of a system.
19. State the second law of thermodynamics.
20. Enumerate the need or importance for second law of thermodynamics.
21. Explain the concept of entropy of a system.
22. A note on entropy changes.
23. Notes on :
    (i) Entropy change of an irreversible process.
    (ii) Entropy change of a reversible process.
24. What is the entropy change in an isothermal, isobaric reversible process?
25. Give the thermodynamics of entropy change in adiabatic reversible process.
26. Explain the entropy change in an isothermal expansion of a gas.
27. Write a note on entropy change of phase change.
28. A brief note on the concept of free energy changes in thermodynamics.
29. Write notes on:
    (a) Helmholtz free energy
    (b) Gibb's free energy.
30. Discuss free energy and spontaneity of reactions.
31. Give the thermodynamic criterion for the existence of equilibrium between two phases, say, ice and water.
32. Explain the free energy change processes with temperature and pressure.
33. A system undergoes a reversible process from state-1 to state-2 at constant pressure. Discuss the free energy change of the process in the light of Gibbs–Helmholtz equation.
34. Write a note on the applications of Gibbs–Helmholtz equation.
35. Discuss Maxwell's relation in thermodynamics.
36. Derive Clapeyron–Clausius equation for a liquid or solid or the fusion of a solid.

37. Derive an integrated form of Clapeyron–Clausius equation for liquid-vapour system.
38. Derive an expression for the free energy change of a reaction at equilibrium (van't Hoff isotherm).
39. What is thermo-chemistry?
40. Discuss the applications of thermochemistry.
41. Define Enthalpy change of reaction. Illustrate with examples.
42. Derive an expression for enthalpy change of reaction at constant pressure and volume.
43. Derive a relationship between the enthalpy change of reaction at constant pressure and constant volume.
44. What do you mean by standard state enthalpies?
45. Write a note on enthalpy change of formation.
46. Define enthalpy change of combustion. Illustrate with examples.
47. Define enthalpy change of neutralization. Discuss with an example.
48. What is enthalpy change of solution?
49. Explain enthalpy change of phase changes.
50. What do you mean by bond enthalpy change? Illustrate.
51. Sate Hess's Law. Discuss the applications of Hess's law in chemistry.
52. What is the effect of temperature on the enthalpy of reaction?
53. Derive Kirchhoff's equation.
54. (a) Derive Gibbs–Helmholtz equation.
    (b) ΔG for a reaction at 300 K is −16 Kcal. ΔH for the reaction is –10 Kcal. What is the entropy of the reaction? What will be ΔG at 300K?

    [Annamalai University, Nov 2014]
55. What happens to the entropy of the following?
    (a) A gaseous nitrogen is converted to liquid nitrogen
    (b) Solid iodine is sublimed to its vapour
    (c) Calculate the entropy change involved in converting 1 mole of water at 373 K to its vapour at the same temperature (molar heat of vaporisation of water = 40.66 kJ/K/mole).

    [Annamalai University, May, 2015]
56. (a) Prove Maxwell relation

$$\left(\frac{\partial V}{\partial T}\right)_{\mathrm{P}} = \left(\frac{\partial S}{\partial P}\right)_{\mathrm{T}}$$

    (b) Derive Gibbs–Helmholtz equation.

    OR

    (a) Derive an expression for entropy change of an ideal gas at constant temperature.
    (b) Derive Clausius–Clapeyron equation.

    [Annamalai University, May, 2016]
57. (a) Define entropy of an ideal gas.
    (b) Derive Gibbs-Helmholtz equation and explain.
    (c) Compute the free energy change when 5 moles of an ideal gas expands reversibly and isothermally at 300 K from an initial volume of 50–1000 L.

    [Annamalai University, Dec, 2013]
58. (a) Explain the significance of Free energy.
    (b) Derive any one form of Gibbs–Helmholtz equation

    OR

    (a) Compare the reversible process with irreversible process.
    (b) Derive any two Maxwell relations.

    [Annamalai University, May, 2014]

Dynamics of Chemical Process

16

Chapter Outline

Introduction to kinetics—rate of a reaction, types of chemical reactions. Rate law for a reaction: single-step molecular reactions, multi-step molecular reactions. Factors affecting the rate of reactions—concentration of reactants, temperature. The collision model—the Arrhenius equation, presence of catalyst, surface area of the reactants. Order of a reaction: order of reaction for single step reaction, order of reaction for multi-step reaction, zero order reaction. Different types of order of reactions, molecularity of a reaction: pseudomolecular reaction. Rate equation and characteristics of first- and second-order reactions, integrated forms of rate laws: first-order reactions and second-order reactions. Complex reaction kinetics: reversible reactions, consecutive reaction kinetics, parallel reaction kinetics and chain reactions.

16.1 Introduction to Kinetics

Thermodynamics is concerned with the changes in entropy and energy that accompanies a change in a system. From the Gibb's free energy of a reaction, one can predict the direction in which a chemical change will take place and the amount of energy consumed/evolved. *Kinetics gives information about the rate at which a change takes place and the mechanism by which the reactants are converted to the products.*

Chemical kinetics deals with the rates of reactions. In order for a reaction to occur, a collision must occur; the collision must be of sufficient energy to break the necessary bonds and be of proper orientation. This is the basis of reaction rates. Rates of chemical reactions *depend on the nature of the reactants, temperature, presence of a catalyst and concentration.*

In chemical kinetics, *we investigate how and at what rate the reactants get transformed to products under given conditions and the factors affecting the rates of reactions*.

16.2 Rate of a Reaction

In a chemical process, reactant or reactants are consumed up as time elapses and the product or products are formed, e.g.

$$\underset{\text{(Reactants)}}{A} \rightarrow \underset{\text{(Products)}}{B}$$

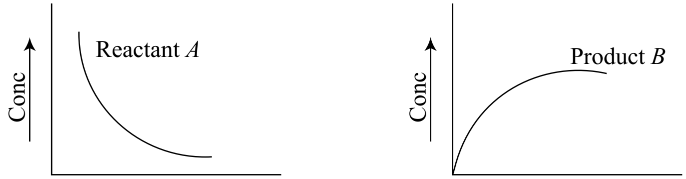
This is a typical example of a chemical reaction. There are innumerable chemical reactions and they offer wide scope for variety of investigations.

Every reaction has a 'rate' and is unique in respect of chemical kinetic study.

The time interval required for a chemical change to complete is called the '*reaction time*'.

In kinetics, the concentrations of reactants and products are expressed as *mol/dm*^{3}.

When a chemical reaction takes place, the concentrations of the reactants decrease with time while that of the products increase. This can be depicted in Fig. 16.1.



(a) Decrease or concentration of *A* (b) Increase or concentration of *B*

Figure 16.1 *Decrease and increase of concentration of reactant, A, and product, B, with time, respectively.*

As the reaction proceeds, the concentration of the reactants decreases and the rate of change of concentration (the slope of the curve) decreases. This rate, d(reactant)/dt, gives the change in concentration per unit time. Because *rate itself changes with time*, a differential rate expression is appropriate in kinetics.

The rate of a reaction is defined as a change in concentration of reactants or products per unit time.

Therefore,

$$\text{Rate of reaction} = dC/dt = \frac{\text{Amount of reactants consumed}}{\text{Time interval}} = \frac{\text{mol/dm}^3}{\text{Sec}}$$

The rate of reaction is conventionally represented as dC/dt, where –ve sign indicates a decrease of concentration of reactant, A (Fig. 16.1a) or dC moles of reactant A has consumed for time period of dt. Similarly, the rate of reaction for the formation of product is expressed as + dC/dt. Where +ve sign indicates an increase of concentration of products during the reaction (Fig. 16.1b). The unit of the rate of reaction dC/dt is *mol/dm*^{3}*/sec.*

The kinetics of a chemical reaction can be studied either with respect to the reactants or products.

Examples

1. Ammonia and hydrogen react at high temperature forming nitric oxide and water. If the rate of formation of nitric oxide is 2.4×10^{-3} mol/dm^3/sec. Calculate the rate of disappearance of ammonia and the rate of formation of water.

   $$4NH_3(g) + 5O_2(g) \rightarrow 4NO(g) + 6H_2O(g)$$

   Rate of reaction

   $$= -1/4\, d[NH_3]/dt = -1/5[O_2] = +\, 1/4 d[NO]/dt = +1/6 d[H_2O]/dt$$

Rate of disappearance of ammonia = Rate of appearance of nitric oxide.

$= 2.4 \times 10^{-3}$ mol/dm^3/sec

Rate of formation of water = d[H_2O]/d*t*

From rate equation = +1/6d[H_2O]/d*t* = +1/4d[NO]/d*t*

$$= \frac{2.4 \times 10^{-3}}{4}$$

$$\text{Therefore, } d[H_2O]/dt = 6 \times \frac{2.4 \times 10^{-3}}{4} \text{ mol/dm}^3\text{/sec}$$

$$= 3.6 \times 10^{-3} \text{ mol/dm}^3\text{/sec.}$$

2. For a reaction $2A + 3B \rightarrow C + 4D$, the rate of the reaction with respect to A, that is d(A)/d*t* = 0.0024 mol/dm^3/sec. Calculate the rates with respect to B, C and D.

   Rate of reaction, d*c*/d*t* = –1/2d(A)/d*t* = –1/3d(B)/d*t* = +d(C)/d*t* = +1/4d(D)/d*t*.

   Given: d(A)/d*t* = 0.0024 mol/dm^3/sec.

   Therefore,

   (i) $d(B)/dt = 3/2\, d(A)/dt = \frac{3 \times 0.0024}{2} = 0.0036$ mol/dm^3/sec.

   (ii) $d(C)/dt = 1/2\, d(A)/dt = \frac{0.0024}{2} = 0.0012$ mol/dm^3/sec.

   (iii) $d(D)/dt = 2d(A)/dt = 2 \times 0.0024 = 0.0048$ mol/dm^3/sec.

16.3 Types of Chemical Reactions

The types of chemical reactions are (i) ionic reactions, and (ii) molecular reactions.

(a) Ionic reaction

A reaction taking place between oppositely charged ions is called ionic reaction.

$$Na^+ Cl^- + Ag^+ NO_3^- \rightarrow AgCl + Na^+NO_3^-$$

The ionic reactions are spontaneous and rates are faster and hence, their rate of reaction cannot be measured by ordinary techniques and are inapplicable.

(b) Molecular reactions

A reaction taking place between the molecules is called molecular reaction.

$$CH_3COOCH_3 + H_2O \rightarrow CH_3COOH + CH_3OH$$
$$2N_2O_5 \rightarrow 2N_2O_4 + O_2$$

The rates of reactions of these molecular reactions can be easily measured.

A molecular chemical reaction, in fact, involves breaking of bonds in reacting molecules and formation of new bonds in product molecules. Since the number and nature of bonds are different in different substances, the rates of chemical reactions differ a lot from one another, i.e. *every reaction has a rate*.

In order to characterise the kinetic behaviour of a reaction, it is desirable to determine how the rate of reaction varies as the reaction progresses.

(c) Rate law or Rate equation

A rate law is a mathematical equation that describes the progress of the reaction. There are two forms of a rate law for chemical kinetics:

- The differential rate law, and
- The integrated rate law.

The differential rate law relates the rate of reaction to the concentrations of the various species in the system.

Differential rate laws can take on many different forms, especially for complicated chemical reactions. However, most chemical reactions obey one of three differential rate laws. Each rate law contains a constant, k, called the rate constant.

A rate law of a reaction is a mathematical expression relating the rate of a reaction to the concentration of either reactants or products.

One may not look at an overall reaction, and determine the rate law. The rate law may be theoretically determined from the rate-determining step (slow step) of the reaction mechanism. Many chemical reactions actually require a number of steps in order to break bonds and form new ones. These series of steps are called *the reaction mechanism.* The rate law of a reaction must be proven experimentally by looking at either the appearance of products or disappearance of reactants.

A mechanism for a reaction is a collection of elementary processes (also called elementary steps or elementary reactions) that explains how the overall reaction proceeds.

A mechanism is a proposal from which you can work out a rate law that agrees with the observed rate laws. The fact that a mechanism explains the experimental results is not a proof that the mechanism is correct. A mechanism is our rationalisation of a chemical reaction, and devising mechanism is an excellent academic exercise.

An elementary process is also called an elementary step or elementary reactions. It expresses how actually molecules or ions react with each other. The equation in an elementary step represents the reaction at the molecular level, not the overall reaction. Based on numbers of molecules involved in the elementary step, there are three kinds of elementary steps such as *unimolecular step (or process), bimolecular process and termolecular process.*

An elementary step is proposed to give the reaction rate expression. The rate of an elementary step is always written according to the proposed equation. This practice is very different from the derivation of rate laws for an overall reaction.

When a molecule or ion decomposes by itself, such an elementary step is called a unimolecular step (or process). A unimolecular step is always a first-order reaction. The following examples are given to illustrate this point:

$$O_3 \rightarrow O_2 + O, \text{ Rate} = k[O_3]$$

or in general

$$A \rightarrow B + C + D, \text{ Rate} = k[A]$$
$$A^* \rightarrow X + Y, \text{ Rate} = k[A^*]$$

where A* represents an excited molecule.

A bimolecular process involves two reacting molecules or ions. The rates for these steps are second-order, and some examples are given to illustrate how you should give the rate expression. The simulation illustrates a bimolecular process.

$$NO + O_3 \rightarrow NO_2 + O_2, \text{ Rate} = k[NO]\,[O_3]$$
$$Cl + CH_4 \rightarrow HCl + CH_3, \text{ Rate} = k[Cl]\,[CH_4]$$
$$Ar + O_3 \rightarrow Ar + O_3^*, \text{ Rate} = k[Ar]\,[O_3]$$
$$A + A \rightarrow B + C, \text{ Rate} = k[A]2$$
$$A + B \rightarrow X + Y, \text{ Rate} = k[A]\,[B]$$

It is a mathematical expression relating the rate of a reaction to the concentration of either reactants or products.

In a chemical process, *only the slow step of the molecular reaction is the rate-determining step* and hence, the rate law may be obtained.

The rate equation is defined as an expression for the rate or velocity of the rate-determining step of a molecular reaction.

16.3.1 Single-Step Molecular Reaction

In the case of single-step molecular process, *the stoichiometric or balanced chemical equation itself is the slow step which is the rate-determining step* of the reaction and is related to rate equation.

For example

$CH_3COOCH_3 + H_2O \rightarrow$ A trimolecular process involves the collision of three molecules.

For example

$O + O_2 + N_2 \rightarrow O_3 + N_2$, Rate = $k[O]\,[O_2]\,[N_2]$
$O + NO + N_2 \rightarrow NO_2 + N_2$, Rate = $k[O]\,[NO]\,[N_2]$

The N_2 molecules in the above trimolecular elementary steps are involved with energy transfer. They cannot be cancelled. They are written in the equation to give an expression for the rates. In general, trimolecular steps may be

$A + A + A \rightarrow$ products, Rate = $k[A]^3$
$A + A + B \rightarrow$ products, Rate = $k[A]^2\,[B]$
$A + B + C \rightarrow$ products, Rate = $k[A]\,[B]\,[C]$

Three molecules collide at an instant are rare, but occasionally these are some of the ways of reactions take place.

Elementary processes are written to show how a chemical reaction progresses leading to an overall reaction. Such a collection is called a *reaction mechanism*. In this mechanism, elementary steps proceed at various speeds. *The slowest step is the rate-determining step.* The order for that elementary process is the order of a reaction, but the concentrations of reactants in that step must be expressed in terms of the concentrations of the reactants.

In order to obtain a clear understanding of the concept of rate law in chemical kinetics, it is reasonable to classify the molecular reactions into two types:

(i) Single-step molecular reactions and
(ii) Multi-step molecular reactions.

A rate law of a reaction $CH_3COOH + CH_3OH$

The rate law for the reaction is

$$dC/dt\alpha\,[CH_3COOCH_3][H_2O]$$
$$= K[CH_3COOCH_3][H_2O]$$

16.3.2 Multi-Step Molecular Reaction

In multi-step molecular reaction, only one of the steps is slow while the others are fast. *It is this slow step which determines the rate of the reaction.*

For example

$$2N_2O_5 \rightarrow 2N_2O_4 + O_2\uparrow$$

The above reaction takes place as follows

$N_2O_5 \rightarrow 2N_2O_4 + ½O_2$ (1) slow
$O + O \rightarrow O_2$ (2) fast

Equation (1) is the slow and rate-determining step of the reaction, and hence the rate equation, while Eq. (2) is the fast reaction.

The rate equation of this multi-step molecular reaction is thus obtained by applying the law of mass action to the rate-determining step reaction.

That is,

$$dC/dt \propto [N_2O_5]^1 \text{ or } dC/dt = K[N_2O_5]^1$$

This is the rate equation for decomposition of N_2O_5 reaction.

The rate equation for the reaction,

$$N_2O_5 \rightarrow 2N_2O_4 + ½O_2$$
$$dC/dt = –d[N_2O_5]/dt = + ½d[N_2O_4]/dt = + 2d[O_2]/dt$$

where dC/dt is the rate or velocity of the rate-determining step of the reaction, –ve sign indicates the decrease concentration of N_2O_5 and 'K' is the *velocity constant* of the reaction at constant temperature and has the unit, *mol/dm^3/sec.*

If several steps are involved in an overall chemical reaction, the slowest step limits the rate of the reaction. *Thus, a slow step is called a rate-determining step*. The following examples illustrate the method of deriving rate laws from the proposed mechanism.

Other examples:

Examples 1 If the reaction, $2NO_2 + F_2 = 2NO_2F$, follows the mechanism,

(i) $NO_2 + F_2 = NO_2F + F$ (slow)
(ii) $NO_2 + F = NO_2F$ (fast)

What is the rate law?

Solution

Since step (I) is the rate-determining step, the rate law is

1. $$\frac{d[NO_2]}{2dt} = k[NO_2]\,[F_2]$$

   Since both NO_2 and F_2 are reactants, this is the rate law for the reaction.

   Discussion

   Addition of (i) and (ii) gives the overall reaction, but the step (ii) does not affect the rate law. Note that the rate law is not derived from the overall equation either.

Examples 2 For the reaction, $H_2 + Br_2 = 2HBr$, the following mechanism has been proposed

(i) $Br_2 \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} 2Br$ (both directions are fast)
(ii) $Br + H \xrightarrow[k_2]{} HBr + H$ (slow)
(iii) $H + Br_2 \xrightarrow[k_3]{} HBr + Br$ (fast)

Derive the rate law that is consistent with this mechanism.

Solution

For a problem of this type, the rate law according to the rate-determining (slow) elementary process is considered. In this case, step (ii) is the rate-determining step, and the rate law is

1. $\frac{d[HBr]}{2dt} = k_2[H_2]\,[Br]$

The factor 1/2 results from the 2HBr formed every time, one in step (ii) and one in step (iii). Since [Br] is not one of the reactants, its relationship with the concentration of the reactants must be sought. The rapid reaction in both direction of step (i) implies the following relationship,

$$k_1[Br_2] = k_{-1}[Br]^2$$

or

$$[Br] = ((k_1/k_{-1})\,[Br_2])^{(1/2)}$$

Substituting this in the rate expression results in

$$\text{rate} = k_2\,(k_1/k_{-1})^{(1/2)}\,[H_2]\,[Br_2]^{(1/2)}$$

The overall reaction order is 3/2, 1 with respect to $[H_2]$ and 1/2 with respect to $[Br_2]$.

Discussion:

The important point in this example is that the rapid equilibrium in step (i) allows you to express the concentration of an intermediate, [Br] in terms of concentrations of reactants $[Br_2]$ so that the rate law can be expressed by the concentrations of the reactants.

The ratio k_1/k_{-1} is often written as K, and it is called the *equilibrium constant* for the reversible elementary steps.

16.4 Factors Affecting the Rate of Reaction

Following are the main factors that influence the rate of a chemical reaction.

16.4.1 Concentration of Reactants

The concentration of reactants plays an important role in chemical kinetics. It is usually true that by increasing the amount of reactants, the rate of reaction is increased. As we know that an increase in concentration of reactants increases the number of molecules. According to collision theory, the greater the number of molecules the higher is the collision ratio, consequently faster is the rate of reaction.

For a general reaction,

$$aA + bB \rightarrow cC + dD$$

Rate of reaction = $-1/a\,d(A)/dt = -1/b\,d(B)/dt = +\,1/c\,d(C)/dt = +\,1/d\,d(D)/dt$

Since the rate of reaction *depends upon the concentration* of the reactants, hence

Rate of reaction α [concentration] = constant [concentration of A] [concentration of B].

This function is generally expressed in terms of *the products of the concentrations each raised to a certain power.*

Therefore, rate of reaction = $K[A]^n[B]^m$

where 'K' is called the *specific reaction rate* or *rate constant* at constant temperature and has the unit, *mol/dm^3/sec.*

Rate constant (k) *is the proportionality constant in the rate law equation that describes the relationship between the rate of a step in a chemical reaction and the product of the concentrations of the reactants consumed in that step.*

For a general reaction, $aA + bB \rightarrow cC + dD$.

- Rate of reaction w.r.t A = $-1/a\text{d(A)}/\text{d}t = K[A]^a[B]^b$.
- Rate of reaction w.r.t B = $-1/b\text{d(B)}/\text{d}t = K[A]^a[B]^b$.
- Rate of reaction w.r.t C = $+\ 1/c\text{d(A)}/\text{d}t = K[A]^a[B]^b$.
- Rate of reaction w.r.t D = $+\ 1/d\text{d(D)}/\text{d}t = K[A]^a/B]^b$.

The exponents '*a*' and '*b*' are generally integers or fractions.

The expressions, as given above, describing *the concentration dependence of the rate of reaction is called the differential rate laws or differential rate equations.*

Example:

Consider the case of hydrolysis of cane sugar which is a single-step molecular reaction at particular temperature.

$$C_{12}H_{22}O_{11} + H_2O \rightarrow C_6H_{12}O_6 + C_6H_{12}O_6$$

The rate equation for the process is

$$\text{Rate of reaction} = \text{d}C/\text{d}t \alpha [C_{12}H_{22}O_{11}]^1 [H_2O]^1$$
$$= K[C_{12}H_{22}O_{11}]^1 [H_2O]^1$$

where d*C*/d*t* is the rate of the reaction, K, the rate constant and concentration of cane sugar and water are expressed in mol/dm^3.

It is experimentally obtained that higher the concentration of reactants [$C_{12}H_{22}O_{11}$] and [H_2O], greater is the rate, d*C*/d*t*, of reaction or vice versa (Fig. 16.2).

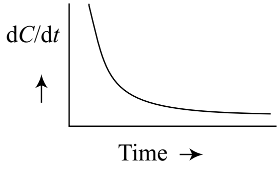


FIGURE 16.2 *Variation of rate of reaction with time.*

16.4.2 Temperature

The rates of most chemical reactions increase as the temperature rises. How do rate expressions reflect this experimental observation of faster reactions at higher temperatures? Because temperature does not affect concentration very much, *the rate constant must increase with increasing temperature*.

In chemical reactions, some bonds are broken, and some new bonds are formed. Bond formation is a favourable process, releasing energy, while bond breaking requires an investment of energy. The easiest energy to apply is usually thermal energy—upon heating a reaction occurs faster.

The rate of reaction increases with the increase of temperature. It can be explained by the fact that at higher temperature, a greater fraction of colliding molecules possess the necessary energy of activation.

As the temperature increases the velocity of molecules also increases which results in increase in the frequency of collision.

The increase in temperature raises the kinetic energy of each molecule. It has been found that, with increase of every 10 K in temperature the fraction of molecule possessing threshold or activation energy becomes double. As a result the number of effective collision is also double, hence the rate is doubled.

For example, let us consider the first-order reaction A + B → Products. Figure 16.3 shows the rate constant for this reaction as a function of temperature. The rate constant, and hence the rate of the reaction, increase rapidly with temperature, approximately doubling for each 10 K increase.

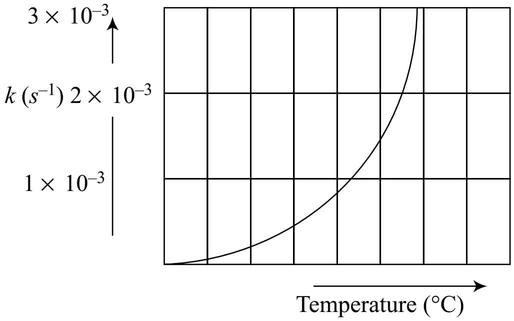


FIGURE 16.3 *Variation in the first-order rate constant for the re-arrangement of (A + B) as a function of temperature.*

16.4.3 The Collision Model

We have seen that reaction rates are affected both by the concentrations of reactants and by temperature. An appealing explanation that accounts for both of these effects is provided by the *collision model* of chemical kinetics. The central idea of the collision model is that *molecules must collide to react.* The greater the number of collisions occurring per second the greater the reaction rate. This idea allows us to understand the effect of concentration on rate. As the concentration of reactant molecules increases, the number of collisions increases leading to an increase in reaction rate. The collision model also allows us to understand the effect of temperature. We know from the kinetic–molecular theory of gases that increasing the temperature increases molecular velocities. As molecules move faster, they collide harder (with more energy) and more frequently, increasing reaction rates. However, the matter is not entirely this much simple. For most reactions, *only a small fraction of collisions actually lead to reaction.*

In 1888, the Swedish chemist Svante Arrhenius suggested that *molecules must possess a certain minimum amount of energy in order to react*. According to the collision model, this energy comes from the *kinetic energies of the colliding molecules*. Upon collision, the kinetic energy of the molecules can be used to stretch, bend and ultimately break bonds, leading to chemical reactions. If molecules are moving too slowly, with too little kinetic energy, they merely bounce off one another without changing (Fig. 16.4).

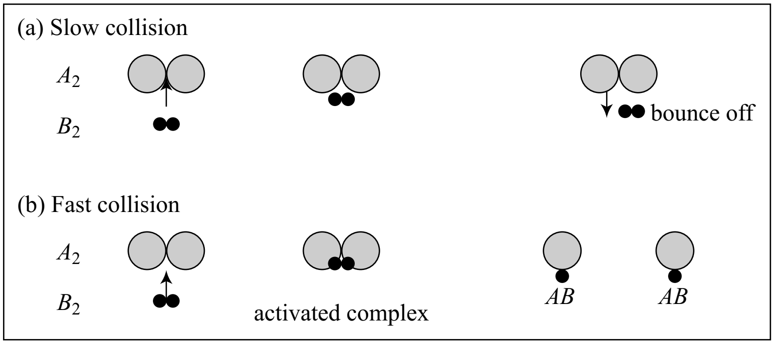


Figure 16.4 *Collisions between reactant molecules must be sufficiently energetic for reaction to occur, as illustrated by a hypothetical reaction $A_2 + B_2 \rightarrow 2AB$. (a) If the reactants are moving too slowly, they merely bounce off one another and no reaction occurs. (b) If the reactants are moving energetically enough and they are oriented correctly, reaction can occur.*

In order to react, colliding molecules must have a total kinetic energy equal to or greater than some minimum value. The minimum energy required to initiate a chemical reaction is called the *activation energy*, E_a. The value of E_a varies from reaction to reaction.

The boulder must be moved over the energy barrier before it can come to rest in the lower energy location, B.

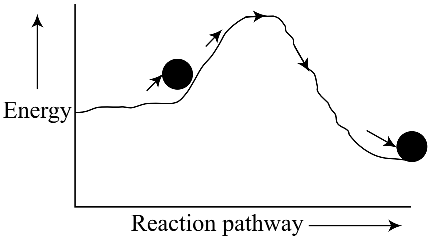


Figure 16.5 *Illustration of the potential–energy profile for a boulder.*

The situation during reactions is rather like that shown in Fig. 16.5. In the same way, molecules may require a certain minimum energy to break existing bonds during a chemical reaction. As an example, in the rearrangement of (A+ B), we might imagine passing through an intermediate state (AB*).

$$A + B \rightarrow [AB^*] \rightarrow \text{products}$$

The change in the energy of the molecule during the reaction is shown in Fig. 16.6. Thus, the barrier represents the energy necessary to force the molecule through a relatively unstable intermediate state to the final product. The energy barrier between the starting molecule and the highest energy along the reaction pathway is the activation energy, E_a. The particular arrangement of atoms at the top of the barrier is called the *activated complex or transition state.*

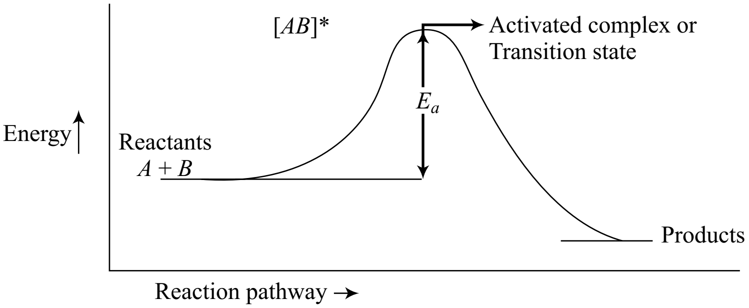


FIGURE 16.6 *Energy profile for the re-arrangement of (A + B). The molecule must surmount the activation–energy barrier before it can form the product, AB*.*

The conversion of A to B is exothermic. Figure 16.7, therefore, shows the product that having a lower energy than the reactant. The energy change for the reaction, E, has no effect on the rate of the reaction. The rate depends on the magnitude of E_a, generally, the lower the E_a the faster the reaction. Note that the reverse reaction is endothermic. The activation barrier for the reverse reaction is equal to the sum of E and E_a for the forward reaction. Energy is transferred between molecules through collisions. Thus, within a certain period of time, any particular A molecule might acquire enough energy to overcome the energy barrier and be converted into B. Recall that in a gas sample the molecules have a distribution of velocities and hence a distribution of kinetic energies. Figure 16.7 shows the distribution of kinetic energies for two different temperatures, comparing them with the minimum energy needed for reaction, E_a. Note that at the higher temperature a much greater fraction of the molecules has kinetic energy greater than E_a which leads to a much greater rate of reaction.

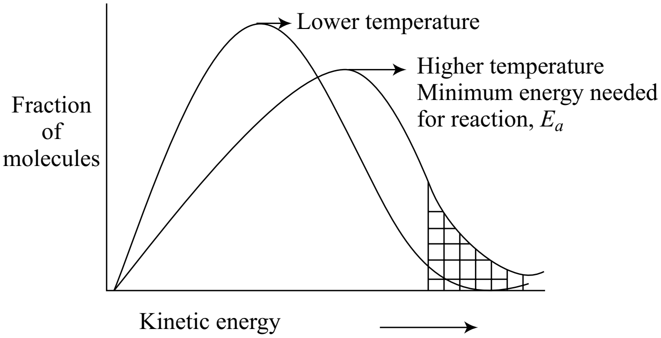


FIGURE 16.7 *Distribution of kinetic energies in a sample of gas molecules at two different temperatures. At the higher temperature, a larger number of molecules have higher energy. Thus, a larger fraction at any one instant will have more than the minimum energy required for reaction.*

In addition to the requirement that the reactant species collide with sufficient energy to begin to re-arrange bonds, an *orientation requirement exists*. The relative orientations of the molecules during their collisions determine whether the atoms are suitably positioned to form new bonds.

For example, consider the reaction of Cl atoms with NOCl:

$$\text{Cl} + \text{NOCl} \rightarrow \text{NO} + \text{Cl}$$

The reaction will take place if the collision brings Cl atoms together to form Cl_2, as shown in Fig. 16.8(a). In contrast, the collision shown in Fig. 16.8 (b) will be ineffective and will not yield products. Thus, collisions must occur not only with sufficient energy but also with suitable orientation in order to cause reaction.

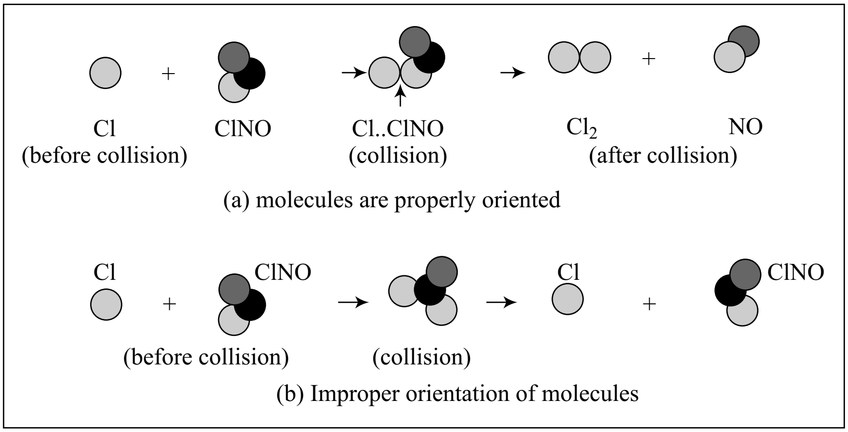


FIGURE 16.8 *Two possible ways by which Cl atoms and NOCl molecules can collide. (a) If molecules are oriented properly, a sufficiently energetic collision will lead to reaction. (b) If the orientation of the colliding molecules is wrong, no reaction occurs.*

16.4.4 The Arrhenius Equation

Arrhenius noted that for most reactions the increase in rate with increasing temperature is nonlinear, as in the example shown in Fig. 16.6. He found that most reaction–rate data obeyed the equation

$$k = A\,e^{-Ea/RT} \tag{1}$$

where k is the rate constant. This equation is called the *Arrhenius equation*. The term E_a is the activation energy, R is the gas constant (8.314 J/mol–K), and T is the absolute temperature. The term A is constant, or nearly so, as temperature is varied. Called the *frequency factor*, A is related to the frequency of collisions and the probability that the collisions are favourably oriented for reaction. Note that as the magnitude of E_a increases, k becomes smaller. Thus, reaction rates decrease as the energy barrier increases.

Taking the natural log of both sides of Eq. (1), we have

$$\ln k = \frac{-E_a}{RT} + \ln A \tag{2}$$

Equation (2) has the form of a straight line; it predicts that a graph of ln k versus $1/T$ will be a line with a slope equal to $-E_a/\text{R}$ and a y-intercept equal to ln A.

We can also use Eq. (2) to evaluate E_a in a nongraphical way if we know the rate constant of a reaction at two or more temperatures. For example, suppose that at two different temperatures, T_1 and T_2, a reaction has rate constants k_1 and k_2. For each condition, we have

$$\ln k_1 = \frac{-E_a}{RT_1} + \ln A \text{ and } \ln k_2 = \frac{-E_a}{RT_2} + \ln A$$

Subtracting $\ln k_2$ from $\ln k_1$ gives

$$\ln k_1 - \ln k_2 = \left|\frac{-E_a}{RT_1} + \ln \text{A}\right| - \left|\frac{-E_a}{RT_2} + \ln \text{A}\right|$$

Simplifying this equation and re-arranging it gives

$$\ln \frac{k_1}{k_2} = \frac{E_a}{R}\left|\frac{1}{T_2} - \frac{1}{T_1}\right| \tag{3}$$

Equation (3) provides a convenient way to calculate the rate constant, k_1, at some temperature, T_1, when we know the activation energy and the rate constant, k_2, at some other temperature, T_2.

Sample Exercise:

Table 16.1 shows the rate constants for the re-arrangement of methyl isonitrile at various temperatures (these are the data that are graphed in Fig. 16.3).

Temperature (°C)	k (s^{-1})
189.7	2.52×10^{-5}
198.9	5.25×10^{-5}
230.3	6.30×10^{-5}
251.2	3.16×10^{-5}

Table 16.1

Rate constant values at different temperatures

(a) From these data calculate the activation energy for the reaction. (b) What is the value of the rate constant at 430.0 K?

Solution

(a) We must first convert the temperatures from degrees Celsius to Kelvin. We then take the inverse of each temperature, $1/T$, and the natural log of each rate constant, $\ln k$. This is listed in Table 16.2.

T (K)	1/T (k^{-1})	ln k
4.69.9	2.160×10^{-3}	−10.589
472.1	2.118×10^{-3}	−9.855
503.5	1.860×10^{-3}	−7.370
524.4	1.907×10^{-3}	−5.757

Table 16.2

Reciprocal temperature and the corresponding ln k values

A graph of $\ln k$ versus $1/T$ results in a straight line, as shown in Fig. 16.9. The slope of the line is obtained by choosing two well-separated points, as shown, and using the co-ordinates of each:

$$\text{Slope} = \frac{\Delta y}{\Delta x} = \frac{-6.6-(-10.4)}{0.00195-0.00215} = -1.9 \times 10^4$$

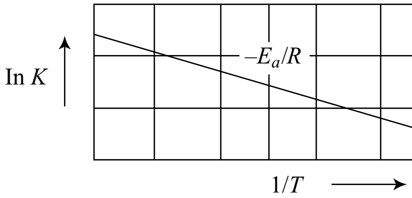


FIGURE 16.9 *The natural logarithm of the rate constant for the re-arrangement of methyl isonitrile is a function of 1/T. The linear relationship is predicted by the Arrhenius equation.*

Because logarithms have no units, the numerator in this equation is dimensionless. The denominator has the units of $1/T$, namely, K^{-1}. Thus, the overall units for the slope are K. The slope is equal to $-E_a/R$. We use the value for the molar gas constant R in units of J/mol · K (Table 16.2). We thus obtain

$$\text{Slope} = \frac{-E_a}{R}$$

$$E_a = -(\text{slope})\,(R) = -(-1.9 \times 10^4\ \text{K})\ [8.31\ 1/\text{mol/K}]\ \{1\text{kJ}/100\text{J}\}$$

$$= 1.6 \times 10^2\ \text{kJ/mol}$$

(b) To determine the rate constant, k_1, at 430.0 K, we can use Eq. (3) with E_a = 160 kJ/mol, k_2 = $2.52 \times 10^{-5}\ \text{s}^{-1}$, T_2 = 462.9 K and T_1 = 430.0 K.

$$\ln \cdot \frac{k_1}{2.52 \times 10^{-5}\ \text{S}^{-1}} = \left|\frac{160\ \text{kJ/mol}}{8.31\ \text{J/mol}}\right| \left|\frac{1}{462.9\ \text{K}} - \frac{1}{430.0\ \text{K}}\right|$$

$$= -3.18$$

Thus,

$$\frac{k_1}{2.52 \times 10^{-5}\text{s}^{-1}} = e^{-3.18} = 4.15 \times 10^{-2}$$

$$k_1 = (4.15 \times 10^{-2})(2.52 \times 10^{-5}\ \text{s}^{-1}) = 1.0 \times 10^{-6}\text{s}^{-1}$$

Note that the units of k_1 are the same as those of k_2.

Example A rate of reaction for a chemical process is investigated at two different temperatures. The rate of a reaction at 25°C is $1.55 \times 10^{-4}\ \text{s}^{-1}$. At 50°C the rate of reaction is $3.88 \times 10^{-4}\ \text{s}^{-1}$. Based on this data, what is the energy of activation for the chemical process expressed in J/mol?

Hint: Look at the units of the quantities given in the problem. The units for rate, s^{-1}, will cancel. Since the energy of activation is to be expressed in J/mol, the logical choice of the gas constant, R, will be 8.314 J/mol K. This choice of R will dictate the units for temperature. Since the R constant has temperature units expressed in Kelvin, K, the temperatures given in the problem must be converted in Kelvin, as well. To convert from Celsius temperature units to Kelvin units, add 273 to the Celsius temperature:

$$T_1 = 25°\text{C} + 273 = 298\ \text{K}$$

$$T_2 = 50°\text{C} + 273 = 323\ \text{K}$$

Substitute the rate constants and temperatures into the Arrhenius equation:

$$\ln k = \ln a + \frac{-E_a}{RT}$$

where k = rate constant at constant temperature
E = energy of activation
T = temperature in Kelvin,
a = collision factor.

Solve for E_a.

$$\ln \frac{K_2}{K_1} = \frac{E_a}{R}\left[\frac{1}{T_1} - \frac{1}{T_2}\right]$$

$$\log \frac{3.88 \times 10^{-4}}{1.55 \times 10^{-4}} = \frac{E_a}{(2.303)(8.314)}\left[\frac{1}{298} - \frac{1}{323}\right]$$

$$\log 2.50 = \frac{E_a \times 2.6 \times 10^{-4}\ \text{K}^{-1}}{(8.314\ \text{J/mol K})}$$

$$0.398 = (1.36 \times 10^{-5}\ \text{mol/J})\, E_a$$

$$E_a = 0.398/1.36 \times 10^{-5}\ \text{mol/J} = 2.93 \times 10^4\ \text{J/mol}$$

16.4.5 Presence of a Catalyst

A catalyst is a substance, which controls the rate of reaction without itself undergoing permanent chemical change. Catalysts are of two types:

(a) Positive catalyst

A positive catalyst increases the rate of reaction by lowering the energy of activation. Thus, in the presence of a positive catalyst, the greater fraction of the total molecule will possess lower energy of activation and collided successfully in a short period of time, thereby increasing the rate of reaction.

A positive catalyst functions by providing an alternate path to the reaction or by the formation of a transition (intermediate) compound having low energy of activation. The activation energy of this path is lower. As a result, rate of reaction is increased.

(b) Negative catalyst or inhibitor

A negative catalyst retards the rate of reaction. Negative catalysts do not lower the energy of activation rather they are combined with reactant molecule thus decreasing the number of colliding reactant molecules. This decreases the effective collisions, hence rate of reaction.

A negative catalyst does not lower the energy of activation rather it combines with the reactant molecules, thus decreasing the number of colliding reactant molecules. This decreases the effective collisions, hence rate of reaction.

16.5 Surface Area of the Reactants

In heterogeneous reactions, the rate of reaction depends upon the surface area of solid reactant. Greater the surface area, higher is the rate of reaction. For example, finely divided calcium carbonate (marble) reacts more quickly with hydrochloric acid than calcium carbonate chips. It is due to the fact that powdered calcium carbonate offers larger surface area to the reacting acid. In other words, *by increasing the surface area of reactant, rate of reaction increases due to greater contact between individual particles and also due to the fact that the surface molecules react more quickly.*

16.6 Order of a Reaction

The relationship between concentration of the reactants and reaction rate is expressed by an equation called a *rate equation or rate law*.

For a general chemical reaction: $\alpha A + \beta B + \gamma C \rightarrow xX + yY$, the rate equation has been found to have the form:

$$rate = k[A]^{\alpha}[B]^{\beta}[C]^{\gamma}$$

- *rate* is the reaction rate.
- k is the rate constant.
- α, β and γ are integers

This rate equation shows us that the reaction rate is proportional to the concentration of the reactants raised to some integer powers.

The power to which the concentration of a species (i.e. A, B or C) is raised to in a rate law is the order with respect to that species.

For example, 'A' has an order of α.

The *overall order* of the reaction is the sum of the individual orders of α, β and γ.

If the power β is one, then the reaction is first-order with respect to B. Likewise if $\alpha = 0$ and $\gamma = 1$, it means the reaction is zero-order with respect to A and first-order with respect to C. Thus, overall the order is second (0 + 1 + 1).

Order of a reaction is also defined as the number of reacting molecules whose concentration changes during the rate-determining step of the reaction.

16.6.1 For a Single-Step Reaction

The order of reaction can be defined as the number of reacting molecules whose concentration changes in the single-step stoichiometric reaction.

e.g. $nA \rightarrow mB$

The rate equation is $dc/dt = K \cdot [A]^n$ (single-step reaction).
Therefore, the order of reaction is '*n*'.

e.g. $CH_3COOCH_3 + H_2O \rightarrow CH_3COOH + CH_3OH$,

The rate law for the reaction is

$$\begin{aligned} dC/dt\ &\alpha\ [CH_3COOCH_3]^1[H_2O] \\ &= K\,[CH_3COOCH_3]^1[H_2O] \end{aligned}$$

The order of the reaction is one, since water is taken in excess.

16.6.2 For Multi-Step Reaction

In multi-step reaction, there are number of simple elementary reactions. It can be noted that many of such elementary steps of reactions are fast, while slowest step of one of the reactions is the *rate-determining step*.

Order of reaction in such a case can be defined as the number of reacting molecules whose concentration changes in the slowest step reaction or the rate-determining step of the reaction.

e.g.

$$2N_2O_5 \rightarrow 2N_2O_4 + O_2\uparrow$$

The above reaction takes place as follows

$N_2O_5 \rightarrow 2N_2O_4 + ½O_2$ (1) slow

$O + O \rightarrow O_2$ (2) fast

Equation (1) is the slow and rate-determining step of the reaction, and hence the rate equation, while Eq. (2) is the fast reaction.

The rate equation of this multi-step molecular reaction is thus obtained by applying the law of mass action to the rate-determining step reaction.

That is,

$$dc/dt = K \cdot [N_2O_5]^1$$

Therefore, the order of this reaction is 1.

Let the rate-determining step of a reaction be

$nA + mB \rightarrow$ products (slow step or rate-determining step)

Rate of reaction, $dc/dt = K \cdot [A]^n[B]^m$

where 'K' is the rate constant of the reaction at constant T.

'n' is the order of reaction with respect to reactant A and 'm' is the order of reaction w.r.t. reactant B. Therefore, overall order of the reaction is $(n + m)$.

If the concentration of one reacting molecule changes in the rate-determining step (or in the rate equation) is called the first-order reaction and so higher orders.

16.6.3 Zero-Order Reaction

If the change in concentrations in large excess of reactants does not affect the rate of reaction (constant rate) then such reactions are known to have *zero-order.*

$$dc/dt = K[HI]^0 \text{ or } dc/dt = K[NH_3]^0$$

that is decomposition of HI and NH_3 on the surfaces of gold and tungsten, respectively, are zero-order reactions.

16.7 Different Types of Order of Reactions

The empirical observations centred around establishment of reaction conditions under which the rate could be measured as a function of concentration of reactants. It was found that the rate was related to concentration in a predictable way, leading to the concept of the rate constant, and equations for simple reactions that fell into several classes:

1. First-order reactions

Reactions in which the rate varies with the concentration of single species, and the change in concentration is exponential, so that a plot of ln(concentration) versus time is linear. The stoichiometric coefficient is 1. Radioactive decay is an example usually cited, but many electron transfer processes, and most enzyme mechanisms contain intermediate reactions that are first-order.

2. Second-order reactions of Class I

In these reactions, the rate varies with concentration of a single species, but the stoichiometric coefficient is 2. The rate varies with the reciprocal of the concentration, so that a plot of 1/(concentration) versus time is linear.

3. Second-order reactions of Class II

In these reactions, the rate varies with the concentration of two substrates, each of which has a stoichiometric coefficient of 1. A plot of ln(concentration A/concentration B) versus time is linear.

4. Higher-order reactions

In these reactions, more than two species are involved, or one species reacts with stoichiometric coefficient greater than two. Many different classes of which the simplest involves a single species with stoichiometric coefficient greater than two, although many biochemical reactions might appear to be higher-order, in general, the enzyme mechanism functions through binding of multiple substrates to a single protein, and the empirical order is simpler, and often appears to be zero-order (see below).

5. Zero-order reactions

Many reactions in biochemistry appear to occur at a rate independent of substrate concentration. These are said to show zero-order behaviour. In general, this is because the reaction is enzyme–catalysed, and the rate is determined by the concentration of enzyme, as long as the substrate is in excess, so that the rate is saturated. Enzyme mechanisms will be discussed later in the course.

16.8 Molecularity of a Reaction

Stoichiometric equation for a reaction indicates the change in the reaction. It may have arisen from a combination of a number of individual reaction steps. Thus, it will not be reasonable to use a stoichiometric equation for deciding the molecularity of the reaction.

Therefore, *molecularity of a reaction is the number of reacting molecules taking part in a rate-determining step of a reaction.*

When one molecule is involved in a step reaction is called ‘*unimolecular reaction’*. On the other hand, when two molecules take part in a rate-determining step reaction such a reaction is called ‘*bimolecular reaction*’.

Examples:

(i) For a single-step reaction as given, the molecularity is obviously two.

$$CH_3COOC_2H_5 + NaOH \rightarrow CH_3COOH + C_2H_5OH,$$

(ii) Decomposition of N_2O_5 is a two-step process.

$$2N_2O_5 \rightarrow 2N_2O_4 + O_2\uparrow$$

The above reaction takes place as follows

$$N_2O_5 \rightarrow 2N_2O_4 + \tfrac{1}{2}O_2 \qquad \text{(1) slow}$$
$$O + O \rightarrow O_2 \qquad \text{(2) fast}$$

Equation (1) is the slow and rate-determining step of the reaction, and hence the rate equation, while Eq. (2), is the fast reaction.

Therefore, we need only this rate-determining step, i.e.

$$N_2O_5 \rightarrow 2N_2O_4 + \tfrac{1}{2}\, O_2$$

In this case, there is only one reacting molecule, N_2O_5.and hence the molecularity of the reaction is one.

16.9 Pseudomolecular Reaction

A reaction whose molecularity is two or more but whose order is one is called a pseudomolecular reaction

Consider the reaction hydrolysis of cane sugar.

$$C_{12}H_{22}O_{11} + H_2O \rightarrow C_6H_{12}O_6 + C_6H_{12}O_6$$

Experimental data indicates that it is a single-step process and is found to be *first-order*, since water is in large excess and its concentration does not change during the process.

$$\text{Rate of reaction} = dC/dt \alpha [C_{12}H_{22}O_{11}]^1$$
$$= K [C_{12}H_{22}O_{11}]^1$$

From the above rate-determining single-step reaction, it is found that the molecules involved in the reaction are one molecule of $C_{12}H_{22}O_{11}$ and one molecule of water and thus it is a *bimolecular reaction*. Such reactions are also referred to as pseudomolecular reactions.

16.10 Rate Equations and Characteristics of First- and Second-Order Reactions

Common types of rate laws:

1. First-order reactions

In a first-order reaction the rate is proportional to the concentration of one of the reactants. That is,

$$v = rate = k[B] \tag{1}$$

where B is a reactant. If we have a reaction which is known to be first-order in B, such as

$$B + \text{other reactants} \rightarrow \text{products},$$

we would write the rate law as

$$\frac{-d[B]}{dt} = k[B] \tag{2}$$

The constant, k, in this rate equation is the first-order rate constant.

2. Second-order reactions

In a second-order reaction the rate is proportional to the concentration squared. For example, possible second-order rate laws might be written as

$$Rate = k[B]^2 \tag{3}$$

or as

$$Rate = k[A][B]. \tag{4}$$

That is, the rate might be proportional to the square of the concentration of one of the reactants, or it might be proportional to the product of two different concentrations.

3. Third-order reactions

There are several different ways to write a rate law for a third-order reaction. One might have cases where

$$Rate = k[A]^3 \tag{5}$$

or

$$Rate = k[\mathrm{A}]^2[\mathrm{B}] \tag{6}$$

or

$$Rate = k[\mathrm{A}][\mathrm{B}][\mathrm{C}] \tag{7}$$

and so on.

We will see later that there are other, more 'interesting' rate laws in nature, but a large fraction of rate laws will fit in one of the above categories.

16.11 Integrated Forms of Rate Laws

In order to understand how the concentrations of the species in a chemical reaction change with time it is necessary to integrate the rate law (which is given as the time–derivative of one of the concentrations) to find out how the concentrations change over time.

16.11.1 First-Order Reactions

If a change in concentration is one molecule

If the number of reacting molecule whose concentration changes in the rate-determining step reaction is one, then it is called first-order reaction.

For the general case, consider a first-order reaction:

$$\mathrm{A} \rightarrow \mathrm{B}$$

Rate of reaction = $v = -d[\mathrm{A}]/dt = d[\mathrm{B}]/dt = k_1[a]^1$. Rate-determining step

where k_1 is the first-order rate constant for the forward reaction, ['a', mol/dm^3] is the initial concentration of the reactant 'A'. The rate of the reaction (v) is given by the rate of disappearance of reactant A.

We re-arrange this equation to bring terms with [a] to one side, and time 't' to the other:

$$\frac{-\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = k_1[a]$$

Then, using the standard integral,

$$\frac{\mathrm{d}[\mathrm{A}]}{[a]} = -k_1\mathrm{d}t$$

If the concentration A after a lapse of 't' seconds falls to $(a - x)$, then

$$\frac{\mathrm{d}[\mathrm{A}]}{[a-x]} = -k_1\mathrm{d}t$$

$$\frac{\int \mathrm{d}[\mathrm{A}]}{[a-x]} = \int -k_1\mathrm{d}t = -k_1\int \mathrm{d}t$$

On integration,

$$-\ln(a - x) = k \cdot t + C \tag{1}$$

where 'C' is the constant integration.

This equation should hold good for all conditions of experiment.

In order to find a value for the constant C, we note that the equation must apply for all concentrations, including that at $t = 0$.

When $t = 0$, $x = 0$ and therefore, Eq.(1) becomes

$$C = -\ln \text{'}a\text{'}$$

Replacing the constant of integration of Eq. (1), we have

$$kt = \ln \text{'}a\text{'} - \ln(a - x)$$

Therefore,

$$k = \frac{2.303}{t}\log_{10}\frac{a}{(a - x)}$$

The exponential form of this general equation is.

$$[a - x] = [a]e^{-kt}$$

Note:

(i) When the experimental value of 'a', 'x' and 't' are substituted in the above equation for first-order, we get a constant value of k.

(ii) In such reactions, the reaction rate is proportional to the first–power of concentration of reactant.

$$\frac{dc}{dt} = k[a–x]^1 \qquad \text{Rate law}$$

Therefore, plot of dc/dt versus $(a - x)^1$ is a straight line as shown in Fig. 16.10.

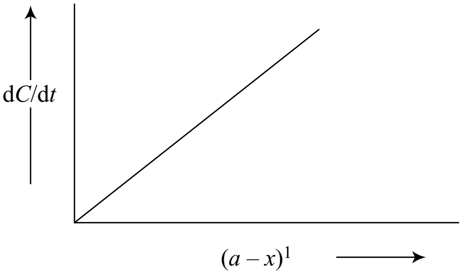


FIGURE 16.10 *Plot of dc/dt versus (a – x)¹.*

(iii) When $\log(a - x)$ of reactants at different intervals of time 't' against the corresponding time 't' is plotted, we get a straight line.
The slope of this straight line should be equal to $-2.303/k$, from which 'k' can be calculated (Fig. 16.11).

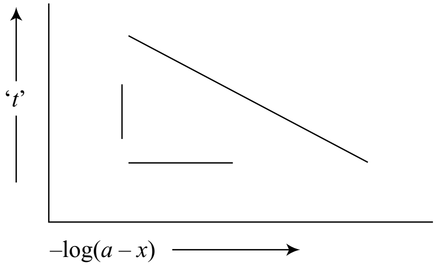


FIGURE 16.11 *Plot of –log(a – x) versus time.*

Characteristics of first-order reaction:

(a) Half-life

Half-life period or time taken for half the change of concentration of a first-order reaction is independent of its initial concentration.

The half-life, usually symbolised by $t_{1/2}$, is the time required for initial concentration 'a' mol/dm^3 to drop to $a/2$.

Using the integrated form of the first-order rate law we find that

$$a/2 = [a]\,e^{-kt} \text{ or,}$$
$$½ = e^{-kt}$$

Taking the logarithm of both sides gives,

$$\ln 1/2 = -kt_{1/2} \text{ or}$$
$$\ln 2 = kt_{1/2}$$

or $$k = \frac{0.693}{t_{1/2}} \text{ or } t_{1/2} = \ln 2/k$$

(b) The value of the velocity constant, 'k' of a first-order reaction is independent of the initial concentration unit.

Examples

1. The half–life for a first-order reaction is 100 seconds. Calculate the rate constant and determine what fraction will have reacted after 250 seconds.

$$t_{1/2} = 0.693/k$$
$$k = 0.693/t = 0.693/100$$
$$= 6.93 \times 10^{-3} s^{-1}$$

Asked then to calculate 'x', how much has reacted.

$$kt_{1/2} = \ln a/(a-x) = 6.93 \times 10^{-3} \times 250$$
$$= \ln a - \ln(a-x) = 6.93 \times 10^{-3} \times 250$$

Set a, the initial concentration to 1.00

$$= 0 - \ln(1-x) = 6.93 \times 10^{-3} \times 250$$
$$\ln(1-x) = 6.93 \times 10^{-3} \times 250 = 1768$$
$$x = 1 - 0.1768x = 0.8232 \therefore \text{(82.32 percent has reacted).}$$

2. Half–life of a first-order reaction is 2 minutes. How much will have reacted after 3 minutes? 70.3 percent)

$$t_{1/2} = 0.693/k$$
$$k = \frac{0.693}{t_{1/2}} = \frac{0.693}{180} = 5.775 \times 10^{-3}$$
$$kt = 2.303 \ln a/(a-x)$$
$$5.775 \times 10^{-3} \times 180 = 2.303 \ln \frac{a}{(a-x)}$$

Set a =100, then

$$5.775 \times 10^{-3} \times 180 = 2.303 [\ln 100 - \ln(100-x)] = 2.303 \ln(x)$$
$$\ln x = \frac{5.775 \times 10^{-3} \times 180}{2.303}$$

Taking antilogarithm,

Therefore, x = 70.3 percent

3. The half–life for carbon–14 is 5670 years. The value of 'a', the initial concentration is taken to be the same as for living material today. The value of $(a-x)$ can be given as a percentage (e.g. 67 percent of that for living material) or as a fraction ($^3/_5$th that of living material) or as a count 87 cpm compared 100 cpm for living material).

An atom was carbon–14 dated and found to have a carbon–14 content $^4/_9$th that of living material.

Given the t = 5760 years calculate the age of the item.

$$kt_{1/2} = \ln 2 = 0.693$$
$$k = 0.693/5760 = 1.20 \times 10^{-4} \text{ year}^{-1}$$
$$kt = \text{In}[a/(a-x)] = [1/(^4/_9)] = 0.810$$
$$t = 0.810/1.20 \times 1^{-4} = 6758 \text{ years.}$$

16.11.2 Second-Order Reactions

If the number of reacting molecules whose concentration changes in the rate-determining step of reaction is two, then it is called second-order reaction.

The rates of bimolecular reactions, which occur when two molecules come together, are frequently described by second-order equation. When the rate of reaction depends on the concentrations of A and B, with each term raised to the first power, the reaction is second-order.

Class I:

A general reaction for second-order reaction is:
Initial concentration, 'a' (mol/dm^3) 0 When time, 't' = 0

$$2A \rightarrow \text{products}$$

After a time lapse, $(a - x)$ xt seconds.

$$\text{Rate} = \frac{-d[A]}{dt} = k_2 [A]^2 \text{ ….. Rate-determining step}$$

When the concentrations of both A and B reactants are equal
Rate of reaction after a time lapse 't' seconds,

$$= -d[A]/dt = d[P]/dt = k_2[a - x]^2$$

After re-arrangement of terms, we get

$$\frac{-d[A]}{(a-x)^2} = k_2 dt \qquad \text{rate equation}$$

Integrating both sides of the equation, we get

$$\frac{1}{(a-x)} = k_2 t + C \tag{1}$$

where C = a constant of integration.
The expression is applicable for conditions.

When $t = 0$, $x = 0$ and therefore, $C = \frac{1}{a}$

Substituting in Eq. (1) we have

$$\frac{1}{(a-x)} = k_2 t + \frac{1}{a}$$

$$\text{Therefore } k_2 \cdot t = \frac{1}{(a-x)} - \frac{1}{a} = \frac{a-(a-x)}{a(a-x)} = \frac{x}{a(a-x)}$$

$$k_2 = \frac{1 \cdot x}{t(a-x)}$$

Half-life period, $t_{½}$ is given as when $t = t_{1/2}$, $x = a/2$

$$t_{1/2} = \frac{0.5a}{k_2 a(a-0.5a)} = \frac{1}{k_2 a}$$

Note:

The integrated rate equation shows that

- *a plot of 1/[a] against time will give a straight line for a second-order*, class I reaction, with an intercept at 1/[*a*].
- *a concentration term for [a] appears in the equation for $t_{½}$, so the half–time depends on initial concentration.*

Characteristics of second-order reaction:

- The velocity constant, k_2, of the second-order reaction is inversely proportional to the initial concentration, '*a*'.

  i.e. $t_{1/2} \alpha 1/a$ or ta = constant

- The velocity constant, k_2, of the second-order reaction varies inversely as the multiple unit of initial concentration '*a*'.

  i.e. $$k_2 = \left|\frac{m \cdot x}{m \cdot a(a-x)m}\right|$$

  The velocity constant, k_2, varies with the change of concentration unit, 'm'.

Class II

Let the concentration of A is 'a', mol/dm^3 and that of B is 'b', mol/dm^3, respectively.
The second-order reaction is only if the concentration of both A and B changes in the rate equation:
Initial concentration → '*a*' '*b*' 0

$$\text{A} + \text{B} \rightarrow \text{Product}$$

After '*t*' seconds → $(a-x)$ $(b-x)$

Rate of reaction = $v = -\text{d[A]}/\text{d}t = \text{d[P]}/\text{d}t$

$$= k_2[\text{A}]^1\,[\text{B}]^1$$

Therefore, it is a second-order ($n = 2$) reaction.

$$\frac{-d(a-x)}{\text{d}t} = \frac{-(b-x)}{\text{d}t} = \frac{\text{d}x}{\text{d}t} = k_2[a-x]^1[b-x]^1. \text{ Rate equation}$$

Integrating both sides,

$$\int \frac{dx}{(a-x)(b-x)} = \int k_2 \cdot \text{d}t$$

By partial fraction method, we have

$$\frac{-1}{(a-b)}\int\left[\frac{1-1}{[(a-x)(b-x)]}\right]dx = k_2 \cdot t = k_2 \cdot t \text{ if } \quad a>b$$

$$\frac{1}{(a-b)} \ln(a-x) - (\ln(b-x) = k_2 \cdot t$$

$$\frac{1}{(a-b)}\left[\frac{\ln(a-x)}{a} - \frac{\ln(b-x)}{b}\right] = k_2 \cdot t$$

$$\frac{1}{(a-b)}\frac{\ln b(a-x)}{a(b-x)} = k_2 \cdot t$$

Therefore,

$$k_2 = \frac{1}{t(a-b)} \frac{\ln b(a-x)}{a(b-x)}$$

or $$k_2 = \frac{2.303}{t \cdot (a-b)} \log_{10} \frac{b(a-x)}{a(b-x)}$$

16.12 Rate Equation for *n*th-Order Reaction

When the concentrations of the reactants are the same for *n*th-order reaction, the rate equation is given by $\mathrm{d}c/\mathrm{d}t = k_n \cdot (a-x)^n$. Rate equation

Where 'a' is the initial concentration of the reactant and 'x' is the amount of reactant reacted in time 't'.

$$\frac{\mathrm{d}c}{(a-x)^n} = k \cdot \mathrm{d}t$$

Integrating

$$\frac{1}{(n-1)} \cdot \frac{1}{(a-x)^{n-1}} = kt + C$$

when $t = 0$, $x = 0$, then

$$C = \frac{1}{(n-1)} \cdot \frac{1}{(a)^{n-1}}$$

$$K_n = \frac{1}{t(n-1)} \left| \frac{1}{(a-x)^{n-1}} - \frac{1}{(a)^{n-1}} \right|$$

The expression is not valid for $n = 1$.

16.13 Determination of the Order of Reaction

1. Integration method

The amount of reactants consumed after different intervals of time is found out by analysis and the data is then substituted for first-, second- or the *n*th-order reactions. The order of reaction is obtained by the constancy of velocity coefficient, k, of the reaction for any of the order of reaction.

2. Half-change method

It has shown that the time required to complete half of the reaction is independent of initial concentration for a first-order reaction, and is inversely proportional to the concentration for second-order reaction. The time taken for half the change is inversely proportional to square of the initial concentration for third-order reaction. Thus, if 't_1' and 't_2' be the time intervals for the completion of the same fraction of change with different initial concentrations 'a_1' and 'a_2' and if 'n' is the order of reaction, then

$$t_1 \alpha \frac{1}{a_1^{n-1}} \qquad t_2 \alpha \frac{1}{a_2^{n-1}}$$

$$\therefore \qquad \frac{t_2}{t_1} = \left| \frac{a_1^{n-1}}{a_2} \right|$$

By substituting the values in the above expression, the value of '*n*' can be calculated.

3. Ostwald's isolation method

If the number of reactants is more than two, Ostwald's isolation method can be adopted to determine the order of the reaction. The experiments are carried out by taking all reactants except one in excess, is said to be isolated. Evidently, the order of reaction will be determined by the number of molecules of the reactants isolated this way. The order of the whole reaction is then the sum of the individual isolation process.

$$\mathrm{d}c/\mathrm{d}t = k[\mathrm{A}]^n\,[\mathrm{B}]^m$$

Therefore, the order of reaction is $(n + m)$.

4. Graphical method

(i) In general, the rate of a reaction is given by the rate law as

$$\mathrm{d}c/\mathrm{d}t = k[a - x]^n$$

where '*n*' indicates the order of the reaction.

The rate of reaction, d*c*/d*t*, is obtained from the concentration-time graph (Fig. 16.12) for the same values of '*t*'.

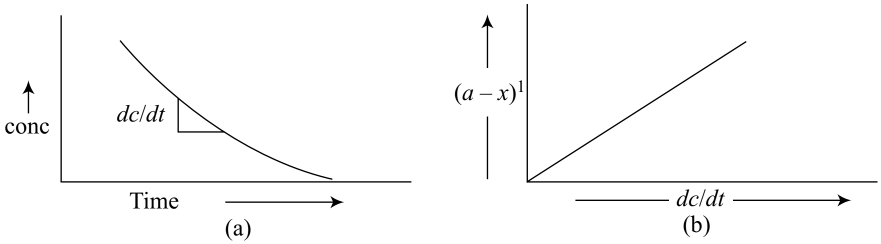


Figure 16.12 *Plots of (a) Variation of concentration versus time and (b) $(a - x)^1$ versus rate of reaction, dc/dt.*

The rate of reaction, d*c*/d*t* is then plotted as a function of different powers of $(a - x)$. The power required for obtaining a linear graph gives the order of reaction.

(ii) In case of first-order reaction, a straight line graph is obtained (Fig. 16.13) by plotting $\log(a - x)$ against time.

For second-order reactions, linear graphs are obtained by plotting $x/(a - x)$ against time for equal initial concentrations of reactants as shown in Fig. 16.14.

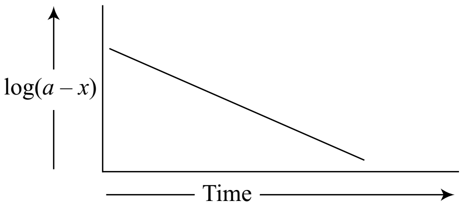


Figure 16.13 *Log(a – x) against time.*

5. Van't Hoff's differential method

The rate of reaction, dc/dt, is obtained from concentration–time plot. For concentration C_1, we have

$$dc_1/dt = kC_1^n. \qquad (1)$$

and for concentration C_2, we have

$$dc_2/dt = kC_2^n. \qquad (2)$$

where 'n' is the order of a reaction.

$$\frac{dc_1/dt}{dc_2/dt} = \left|\frac{C_1^n}{C_2}\right|$$

Taking logarithms, we have

$$\log dc_1/dt - \log dc_2/dt = n\,[\log C_1 - \log C_2]$$

Therefore, $n = \dfrac{\log dc_1/dt - \log dc_2/dt}{\log C_1 - \log C_2}$

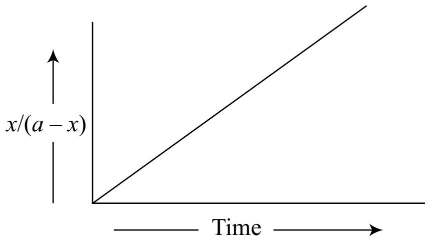


Figure 16.14 *A plot of x/(a – x) against time.*

16.14 Complex Reaction Kinetics

The study of chemical kinetics becomes complicated due to the occurrence of complex reactions which involve more than one step.

Important among such reactions are the following categories:

- Reversible reactions
- Consecutive reactions
- Parallel reactions
- chain reactions

16.14.1 Reversible Reactions

It can be remarked that when the products are formed in any reaction with too low concentration, to set up the opposing reaction at noticeable rate; the kinetics of such reaction are investigated in the initial stages. However, when the opposing reaction also takes place at a comparable rate, the problem becomes complicated and the rate constant obtained is not quite reliable.

(I) An opposing reaction is considered in which the forward as well as the reverse reactions are both first-order:

Initial concentration → 'a' mol/dm^3 = 0, when $t = 0$

$$A \xrightarrow[k_{-1}]{k_1} B$$

At 't' sec → $(a - x)$ x

Hence,

- The rate of the forward reaction, $\dfrac{dx}{dt} = k_1(a - x)^1$;
- The rate of the reverse reaction, $\dfrac{dx}{dt} = k_{-1}.x^1$.

Hence the net rate of formation of B is given by

$$dx/dt = k_1(a - x) - k_{-1}x \qquad \ldots(1)$$

If x_e is the concentration of B at equilibrium when the net rate is zero that is $dx/dt = 0$, then

$$k_1(a - x) - k_{-1}x = 0$$

$$k_{-1}x = k_1(a - x)$$

$$k_{-1} = \frac{k_1(a - x_e)}{x_e} \qquad \ldots(2)$$

Substituting for k_{-1} in Eq. (1), we have

$$dx/dt = k_1(a - x) - k_1\frac{(a - x_e)x}{x_e}$$

Therefore, $dx/dt = \dfrac{k_1(x_e - x)a}{x_e}$

Separating the variables, we have

$$\frac{dx}{(x_e - x)} = \frac{a}{x_e} \cdot k_1t$$

Integrating, we get,

$$-\ln(x_e - x) = \frac{a}{x_e}\ k_1t + I \qquad \ldots(3)$$

where 'I' is the constant of integration.

$$-\frac{x_e}{a}\ \ln(x_e - x) = k_1t + \mathrm{I}$$

At $t = 0$, $x = 0$, so that $I = -\dfrac{x_e}{a}\ln x_e$

Hence, substituting for I in Eq. (3), we have

$$\frac{-x_e}{a}\ \ln(x_e - x) = k_1t - \frac{x_e}{a}\ln x_e$$

$$k_1t = \frac{x_e}{a}\ln\frac{(x_e)}{(x_e - x)}$$

Therefore $k_1 = \dfrac{x_e}{at}\ln\dfrac{x_e}{(x_e - x)} \qquad \ldots(4)$

Equation (4) gives k_1 in terms of easily measurable quantities.

From Eq. (2), we have

$$(k_1 + k_{-1}) = \frac{x_e}{at}\ln\frac{x_e}{(x_e - x)} - \frac{x_e}{a}\ln x_e$$

$$(k_1 + k_{-1}) = k_1\frac{a}{x_e} = \frac{1}{t} \cdot \ln\frac{(x_e)}{(x_e - x)}$$

Thus, from the knowledge of k_1, a, x_e, the rate constant, k_{-1} for the reverse reaction can be calculated.

Examples:

1. Ester hydrolysis and
2. Conversion of ammonium thiocyanate into urea.

(II) In a reaction of the given type wherein, the first-order opposes the second-order reaction

$$A \underset{K_{-1}}{\overset{A}{\rightleftarrows}} B + C$$

The expression for this type of opposing reaction is

$$k = \frac{x_e}{t(2a - x_e)} \ln \frac{ax_e + x(a - x_e)}{a(x_e - x)}$$

Examples

(i) Decomposition of alkyl ammonium halides into tertiary amine and an alkyl halide
(ii) Decomposition of PCl_5 into PCl_3.
(iii) A reaction in which the second-order reaction is opposed by one of first order

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C$$

The rate equation for this type of opposing reaction is

mol/dm^3

$$K = \frac{x_e}{t(a^2 - x_e^2)} \ln \frac{x_e(a^2 - x_e x)}{a^2(x_e - x)}$$

Examples:

- Isomerisation of an alkyl ammonium cyanate to the corresponding substituted urea and
- Formation of ammonia from nitrogen and hydrogen.

16.14.2 Consecutive Reaction Kinetics

Consider the following reaction that takes place as stated below:

Thus, the sequence of reaction is

Initial concentration →	'x_0', mol/dm^3	0	0
	$A \xrightarrow{k_1}$	$B \xrightarrow{k_2}$	C
Concentration after 't' sec,	x_1	x_2	x_3

and this reaction can be regarded as a typical consecutive reaction scheme involving first-order steps. Quite often, it is observed that the products of one reaction itself act as a reactant. Thus, there may be a series of consecutive steps. The simplest consecutive reaction scheme may be the one involving one reversible first-order steps;

The rate of consumption of A, that is –dA/dt is given by

$$-d[A]/dt = k_1 a \text{ or } x_1 = x_0 e^{-k_1 t}. \quad (1)$$

that is the concentration of A falls off exponentially with time.

The net rate of consumption of B = rate of consumption – rate of formation.

Therefore, $-d[B]/dt = k_2 x_2 - k_1 x_1$

Substituting the value of 'a' from the above expression (1)

$$-d[B]/dt = k_2 x_2 - k_1 x_0 e^{-k_1 t}$$

this is the linear differential equation of the first-order, whose solution is

$$x_2 = x_0 \frac{k_1}{(k_2 - k_1)} . [e^{-k_1 t} - e^{-k_2 t}] \qquad \ldots(2)$$

There is no change in the number of moles during the entire reaction,

$$x_1 + x_2 + x_3 = x_0$$

or $$x_3 = x_0 - (x_1 + x_2)$$

Therefore,

$$x_3 = x_0 - \frac{x_0 k_1}{(k_2 - k_1)} [e^{-k_1 t} - e^{-k_2 \mathrm{t}}] - x_0 e^{-k_1 t}$$

$$= x_0 - x_0 \frac{k_1}{(k_2 - k_1)} e^{-k_1 \mathrm{t}} + x_0 \frac{k_1}{(k_2 - k_1)} e^{-k_2 t} - x_0 e^{-k_1 t}$$

$$x_3 = x_0 \left[1 - e^{-k_1 t} \left\{ \frac{k_1}{(k_2 - k_1)} + 1 \right\} + \frac{k_1 . e^{-k_2 t}}{(k_2 - k_1)} \right]$$

By simplifying the term

$$\left\{ \frac{k_1}{(k_2 - k_1)} + 1 \right\} = \frac{(k_1 + k_2 - k_1)}{(k_2 - k_1)} = \frac{k_2}{(k_2 - k_1)}$$ and substituting, we have

or $$x_3 = x_0 \left[1 - \frac{k_2}{(k_2 - k_1)} e^{-k_1 t} + \frac{k_1}{(k_2 - k_1)} e^{-k_2 t} \right] \qquad \ldots(3)$$

The variation of the concentration of the constituents with time is presented in Fig. 16.15.

$$\mathrm{A} \rightarrow \mathrm{B} \rightarrow \mathrm{C}$$

Concentration of A decreases exponentially [Eq. (1)] irrespective of the fate of its products. Concentration of B increases to a maximum and when the rate of decomposition, due to its accumulation, becomes greater than its rate of formation, its starts falling. The concentration of *C* increases steadily until it attains a limiting value.

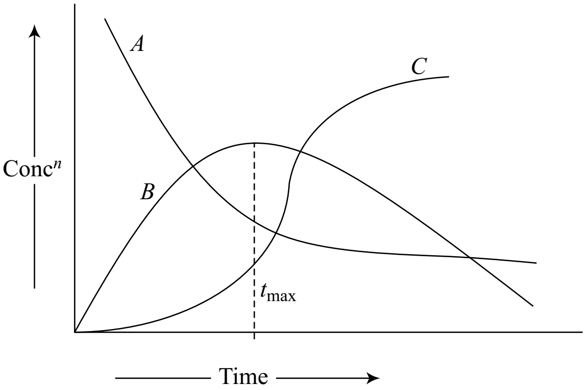


FIGURE 16.15 *Variation in the concentration of various species during the progress of the reaction.*

If $k_1 < k_2$, then in sufficiently large interval of time,

$$e^{-k_2t} < e^{-k_1t}$$

consequently, Eq. (2) becomes

$$x_2 = x_0 \frac{k_1}{(k_2 - k_1)} . e^{-k_1t} \qquad \ldots(4)$$

since, $x_1 = x_0\, e^{-k_1t}$, Eq.(4) gives

$$x_2/x_1 = \frac{k_1}{(k_2 - k_1)} \qquad \ldots(5)$$

that is the ratio of the amount of B and A formed in a definite interval of time after the start of the reaction becomes constant. In other words, the amounts of substances A and B diminish to the same extent. Such a state is called the *transient* equilibrium.

If $k_1 < k_2$, then

$$x_2/x_1 = k_1/k_2. \qquad \ldots(6)$$

For a first-order reaction, since

$$k \propto [t_{1/2}]^{-1}. \qquad \ldots(7)$$

Equation (6) may be written as,

$$\frac{x_2}{x_1} = \frac{[t_{1/2}]_B}{[t_{1/2}]_A} \qquad \ldots(8)$$

The equilibrium corresponding to Eq. (8) is called *secular* equilibrium.
Further, when $k_1 < k_2$, Eq. (3) becomes

$$x_3 = x_0\, [1 - e^{-k_1t}]. \qquad \ldots(9)$$

Thus, such a reaction shows a first-order kinetics, however in the very beginning of the reaction the concentration of *c* remains low almost constant over a period of time and cannot be detected analytically. This interval of time is called the *period of induction*.

Conditions for the maximum concentration of B:
The concentration of B at any instant is given by

$$x_2 = x_0 \frac{k_1}{(k_2 - k_1)} . [e^{-k_1t} - e^{-k_2t}]$$

Therefore, $$\mathrm{d}b/\mathrm{d}t = x_0 \frac{k_1}{(k_2 - k_1)} [k_2\, e^{-k_2t} - k_1\, e^{-k_1t}]$$

When the concentration of B is maximum,

$$\mathrm{d}b/\mathrm{d}t = 0$$

$$k_2 e^{-k_2t} = k_1\, e^{-k_1t}$$

$$k_1/k_2 = e - (k_2 - k_1)t_{\max}$$

or $$(k_2 - k_1)t_{\max} = \ln k_1/\mathrm{k}_2$$

$$t_{\max} = \frac{1}{k_1 - k_2} \ln k_1/k_2 \qquad \ldots(10)$$

and the maximum concentration of B ($x_{2\max}$) is given by

$$x_{2\max} = x_0[k_2/k_1]^{k_2/k_1 - k_2}. \qquad \ldots(11)$$

The maximum amount of the intermediate B depends not on the absolute values of the rate constants, k_1 and k_2, but only on the ratio k_2/k_1. The greater the value of k_2/k_1, greater is the ordinate (concentration) of the maximum of the curve B and closer is this maximum to the beginning of the reaction that is $t_{\max}$ shifts towards $t = 0$.

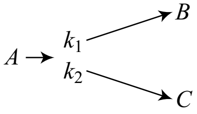
Examples

- Saponification of esters of dibasic acids with alkali that is amber ethyl ester with NaOH.
- The hydrolysis of a hydroxy ester in the presence of acids with the ultimate formation of a lactone.

16.14.3 Parallel Reaction Kinetics

In many cases, it can be observed that a given substance reacts or decomposes in more than one way. Then, the parallel reaction must also be taken into consideration while analysing the kinetic data.

In the case of an irreversible first-order reaction,



leading to the formation of two products through two different path ways of first-order.

The overall rate of reaction can be written as,

$$\begin{aligned} -\mathrm{d(A)}/\mathrm{d}t &= k_1(a) + k_2(a) \\ &= [k_1 + k_2](a) \end{aligned}$$

where (a) is the initial concentration of A.

$$\mathrm{d(B)}/\mathrm{d}t = k_1(a)$$
$$\mathrm{d(C)}/\mathrm{d}t = k_2(a)$$

This appears like first-order reactions.

$$\ln \frac{[a]}{[a_0]} = -[k_1 + k_2]\, t$$

Therefore, $[a] = [a_0]\; e^{-[k_1+k_2]t}$

Thus, $\mathrm{d[B]} = k_1[a] = k_1[a_0]\; e^{-[k_1+k_2]t}$

$$\mathrm{d[C]} = k_2[a] = k_2[a_0]\; e^{-[k_1+k_2]t}$$

If $[\mathrm{B}_0] = [\mathrm{C}_0] = 0$, we can integrate to get,

Therefore $[\mathrm{B}] = \dfrac{k_1[a_0]}{[k_1 + k_2]}[1 - e^{-[k_1+k_2]t}]$

Therefore $$[C] = \frac{k_2[a_0]}{[k_1 + k_2]}[1 - e^{-[k_1+k_2]t}]$$

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

they are formed at a constant ratio. If one reaction is much faster, then it dominates the overall reaction.

It means the ratio of the product concentrations at any instant is equal to the ratio of the corresponding rate constants.

The variation of the concentration of reactant A and products B and C with time for two first-order parallel reactions are shown in Fig. 16.16.

Thus, if the concentration ratio is determined at any stage and k is found from the kinetic equation, the individual rate constants can be determined.

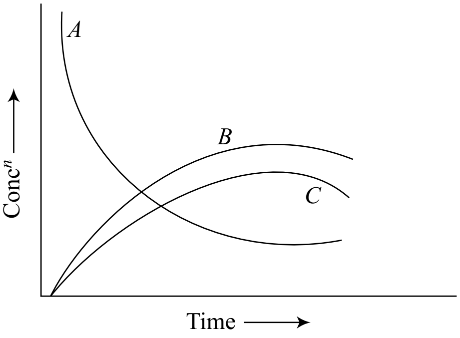


Figure 16.16 *Variation of the concentration of reactant A and products B and C with time for two first-order parallel reactions (for $k_1/k_2 = 2$).*

These types of reactions are very common in organic chemistry (substitution reaction). For example, the chlorination of methane may give mono, di and or tetra halogen derivative.

Other examples are (i) nitration of phenol and (ii) dehydration of ethanol.

16.14.4 Chain Reactions

The rate of reactions is measured with ease experimentally; however, in many cases, reaction rates are very high comparable to any that can be measured. The reaction between hydrogen and chlorine, in the presence of sunlight, proceeds at a rate which is many thousand times greater than that predicted from simple collision theory.

It appears that in exothermic reactions, the evolved heat is utilised in activating the fresh molecules or used in splitting the molecules into free atoms or free radicals. These free atoms or free radicals so formed participate in subsequent processes and thus give rise to a series of reactions. Such series of successive elementary processes is termed a *chain reaction*.

In general, such reactions consist of the following steps:

1. Chain initiation

The free radicals and or energised molecules of one of the products formed during the course of reaction *initiates* the chain reaction.

2. Chain propagation

These atoms, free radicals or energised molecules, now react with the molecules of the other reactants, resulting in the formation of the product, and the regeneration of the active species. Thus, re-starting the reaction leads to the formation of the products. Such steps are termed as *chain propagation* steps.

3. Chain inhibition

This step consists of the removal of the products with a simultaneous generation of more active species, with net decreasing effect on the overall rate of reaction. Such steps are termed as chain *inhibiting* steps.

4. Chain termination

Reaction steps which remove the active species needed for the initiation or propagation of the reaction are termed as chain termination steps.

Kinetics of Chain reactions:

Considering a chain reaction as given below, in which 'X' is a *reactant*; 'C' is a reactive *chain carrier*; 'Pr' is the product and 'α' is the *number of chain carriers* produced by one carrier in the propagation step.

(i) Initiation of reaction

$$X \xrightarrow{k_1} C$$

(ii) Propagation

$$C + A \xrightarrow{k_2} Pr + \alpha C$$

(iii) Termination

$$C \xrightarrow{k_3} \text{destruction}$$

It can be observed that the destruction of the chain carriers may be either due to collision with the walls of the vessel or with other molecules in the gaseous phase.

The rate of formation of the chain carrier is,

$$d[C]/dt = k_1[X] - k_2[C][X] + \alpha k_2[C][X] - k_3[C]$$

For a steady-state reaction rate,

$$\frac{d[C]}{dt} = 0$$

Thus,
$$[C] = \frac{k_1[X]}{k_2[X](1-\alpha) + k_3} \qquad \ldots(1)$$

The constant, k_3, can be written as the sum of two terms, k_w, for the wall reaction and k_g for the gas phase reaction.

Thus,

$$[C] = \frac{k_1[X]}{k_2[X](1-\alpha) + k_w + k_g} \qquad \ldots(2)$$

It would be of interest to analyse the limiting forms of the above expression.

- **Stationary chain reactions**:
  In a reaction where each propagating sequence leads to the formation of a molecule of product and regeneration of *only one chain carrier* (if, $\alpha = 1$). Such chain carriers are termed *nonbranched or stationary chain reactions*. In such cases, the radical concentration is proportional to the ratio of its rate of formation to the rate of destruction.
- **Nonstationary or branched chain reactions:**
  If more than one chain carriers are produced in the propagating sequence of a chain reaction (if $\alpha > 1$), the chain is termed a *nonstationary or a branched-chain*. In particular, a critical situation arises when α becomes so large that,

  $$k_2[X](1 - \alpha) = -(k_w + k_g)$$

  thereby making the denominator in Eq. (2) to approach zero and the carrier concentration infinite. Since, the reaction rate is proportional to the carrier concentration, the reaction proceeds so rapidly that an explosion results, that is the reaction is completed within fraction of a second. Such explosions are termed *isothermal explosions.*

The rate laws of chain reactions:

(i) The thermal decomposition of acetaldehyde

The thermal decomposition of acetaldehyde can occur as given

$$CH_3CHO\ (g) \rightarrow CH_4\ (g) + CO\ (g)$$

The rate of reaction is given as

$$\text{Rate} = k_i[CH_3CHO]^{1/2}. \qquad (1)$$

(i) Initiation reaction:

$$CH_3CHO \rightarrow \bullet CH_3 + \bullet CHO$$

(ii) Propagation

$$CH_3CHO + \bullet CH_3 \rightarrow CH_4 + CH_3CO\bullet \qquad k_p$$

(iii) Propagation

$$CH_3CO\bullet \rightarrow CH_3\bullet + CO \qquad k_p'$$

(iv) Termination $CH_3\bullet + \bullet CH_3 \rightarrow CH_3 - CH_3 \qquad k_t$

The net rates of change of the intermediates are

- $d[CH_3]/dt = k_i[CH_3CHO] - k_p\,[CH_3][\,CH_3CHO] + k_p'[CH_3CO] - 2k_t[CH_3]^2$
- $d[CH_3CO]/dt = k_p[CH_3][CH_3CHO] - k_p[CH_3CO]$

Applying the steady-state approximation to the above two equations

$$= k_i[CH_3CHO] - k_p\,[\bullet CH_3][CH_3CHO] + k_p[CH_3CO\bullet] - 2k_t[CH_3]^2$$

$$0 = k_p\,[\bullet CH_3][CH_3CHO] - k_p'[CH_3CO\bullet]$$

The sum of the above two equations is

$$k_i[CH_3CHO] - 2k_t[CH_3\bullet]^2 = 0$$

Thus the steady-state concentration of $[CH_3]$ is

$$[CH_3\bullet] = \left|\frac{k_i^{1/2}}{2k_t}[CH_3CHO]^{1/2}\right|$$

The rate of formation of CH_4 is

$$d[CH_4]/dt = k_p[CH_3\bullet][CH_3CHO] = \left|\frac{k_i^{1/2}}{2k_t}[CH_3CHO]^{3/2}\right|$$

Review Questions

1. Define rate of a reaction. Show how rate varies with time.
2. What are the types of chemical reactions? Give reactions why kinetics can be studied only in the case of molecular reactions.
3. What do you mean by rate law? Illustrate with any example.
4. How are molecular reactions classified in kinetics to get rate equation? Mention their importance.
5. Illustrate with example a single-step molecular reaction and the rate law.
6. Illustrate with example a multi-step reaction and the rate law.

7. Show how rate-determining step of a reaction is useful in deciding the rate law of a chemical reaction.
8. Enumerate the factors affecting the rate of a reaction.
9. Show that rate of a reaction is proportional to the concentration of the reactants. Illustrate with at least two examples to give the differential rate equations.
10. Write a note on the effect of temperature on rate of reaction.
11. A note on energy of activation of a reaction.
12. Write a note on Arrhenius equation.
13. How the energy of activation of a reaction can be determined experimentally?
14. What is the role of catalyst for a reaction rate?
15. Define order of a reaction. How is it related to rate law of a reaction?
16. Discuss the order of a reaction of a single-step reaction with an example.
17. If a reaction has number of elementary reactions, then how would you determine the order of reaction? Justify.
18. What is a zero-order reaction? Give its rate law.
19. What are the types of order of reactions? Give their rate law equations.
20. What do you mean by molecularity of a reaction? Give example.
21. Distinguish order and molecularity of a reaction.
22. What is a pseudomolecular reaction? Explain with an example.
23. Derive integrated form of rate law for a first-order reaction.
24. Enumerate the characteristics of a first-order reaction.
25. What is a second-order reaction? Derive an expression for the rate constant of a second-order reaction when the initial concentrations of the reactants are the same.
26. Derive an equation for the rate constant of second-order when the initial concentrations of the reactants are different.
27. Enumerate the characteristics of the second-order reaction.
28. What are complex reactions?
29. Discuss the kinetics of reversible reactions.
30. What is a consecutive reaction? Discuss the kinetics of consecutive reaction.
31. What do you mean by parallel reactions? Explain the kinetics of a parallel reaction.
32. What do you mean by chain reaction? Discuss.
33. For a first-order reaction, rate constant is found to be 7×10^{-7} at 7°C and 9.7×10^{-4} at 57°C. Calculate the energy of activation of the reaction. [UPTU, 2012]
34. Show that in the case of a first-order reaction, the time required for 99.9 percent of the reaction to take place is about 10 times that required for half the reaction. [UPTU, 2013]

17

Refractories

Chapter Outline

Introduction to refractories, classification, properties of refractories. Melting points of few refractory materials. Manufacture of refractories. Acid refractories—silica, silica bricks, fire clay/kaolin. Basic refractories—alumina, magnesia/magnesite. Types of magnesia—magnesia clinkers, dolamite, lime/limestone—lime, limestone, zirconia. Neutral refractories—silicon carbide. Other refractory materials—alumina, zirconia, aluminium silicate (sillimanite), bauxite, calcined and sintered alumina, calcium aluminate, cordierite, forsterite, mullite, spinel, quartz/fused silica.

17.1 Introduction

Refractories are generally ceramic materials that are *capable of withstanding high temperatures, resistant to thermal shock, chemically inert and have low thermal conductivities and coefficients of expansion.*

The materials employed for the construction of furnaces, kilns, crucibles, etc. are resistant to heat when they are subjected to the corrosive environment of gases and slags. *The refractoriness of a material is a measure of its ability to withstand exposure to elevated temperatures without undergoing appreciable deformation.* The bulk of refractory materials consist of single or mixed high melting point oxides of elements such as silicon, aluminium, magnesium, calcium and zirconium. Nonoxide refractories also exist and include materials such as carbides, nitrides, borides and graphite. The actual composition of a refractory material is dependent on operating factors such as temperature, atmosphere and what materials it will be in contact with.

The atmosphere in which a refractory is to operate can dictate what materials can be used in that particular application. For example, graphite refractories can operate at temperatures of up to several thousand degrees celsius under reducing conditions or oxygen-free conditions, such as vacuum. However, they may begin to sublime at approximately 1000°C under oxidizing conditions.

The type of materials that a refractory comes into contact with can also dictate which materials will be suitable. For instance, in steel making, basic refractories are used because the refractories often come into contact with basic slags containing magnesium and calcium oxides. If the refractory lining was made from acidic refractories, it would be eroded quickly by the chemical interaction of the basic slag and the acidic lining (e.g. silica) forming low melting point compounds.

Refractories are found in a number of different forms. These can be divided into three main forms:

- Refractory shapes: This grouping consists of preformed refractory shapes, such as bricks that are delivered from the manufacturer, ready to instal.
- Monolithics: Also known as refractory castables, these materials are essentially an analogue of concrete, consisting of refractory aggregates, refractory fines and refractory cements (generally calcium aluminate cements). They are obtained dry, mixed with water and formed and fired *in situ*.
- Ceramic fibre: Generally speaking, a very lightweight refractory material, found in a number of different forms to cope with different applications and temperatures, made from fibres of aluminosilicate materials, alumina and zirconia depending on the actual application.

Refractories are used to line furnaces (Fig. 17.1) and keep heat within the furnace hearth. By using refractories as insulation, the furnace efficiency is increased, the frame is protected and provides a safe protection to pass nearby the furnace.

Furnaces may use a number of layers of different refractories in their construction as shown in Fig. 17.2. For instance, a dense hot face material may be used on the inner surfaces which are exposed to the highest temperatures. On the outside, a low density highly insulating refractory layer may be employed to keep in as much of the heat as possible. Between the hot face refractory and the low density insulation, a number of different layers of intermediate materials may also be employed.

Each successive layer (moving away from the hot face) would have an increasingly lower density and more than likely a lower refractoriness compared to the previous layer. Applications of furnaces include anything from primary metal smelting, through to heat treatment, glass production and processing and ceramic component manufacture, many forms of chemical processing and testing.

FIGURE 17.1 *A typical brick kiln/furnace.*

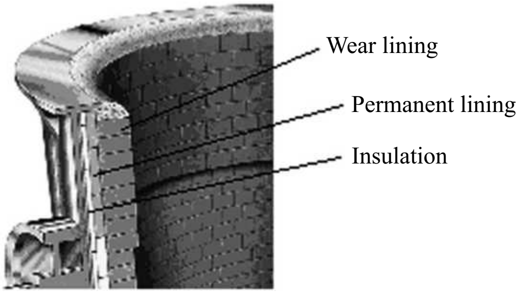


FIGURE 17.2 *No of layers of brick lining for furnace with refractories.*

17.2 Classification

The oxides of magnesium (magnesia) and calcium (lime) are the most important refractory materials, though fireclay is widely used as well. *Zirconia* is used when the material must withstand extremely high temperatures. Silicon carbide is another refractory material. It is very strong at high temperatures, but will burn in the presence of oxygen if the protective silica coating comes off. Refractories must be chosen according to the conditions they will face. For example, carbon cannot be used when it will be in contact with oxygen, as it will burn.

In brief, refractories are classified, broadly, as

1. Acid: Silica bricks, fireclay bricks.
2. Basic: Magnesia bricks, bauxite bricks
3. Neutral: Chromite, silicon carbide, graphite, mullite.

In addition to these, many single oxide bricks are used as refractory bricks.

- Alumina
- Zirconia
- Magnesia

Acidic refractories cannot be used in a basic environment, and basic refractories cannot be used in an acidic environment because they will be corroded. Zircon, fireclay and silica are acidic; dolomite and magnesite are basic and alumina, chromite, silicon carbide, carbon and mullite are neutral. Refractory metals are also frequently used.

17.3 Properties of Refractories

The enumerated properties of the refractories are as follows:

1. Any refractory material should withstand very high temperature without any deformation or softening. The fusion temperature of these materials is observed to vary between 1700 and 2500°C.
2. These materials withstand adequate load during working in kiln or furnace. It is therefore essential that refractory materials must also possess high mechanical load without breaking.
3. Refractories are able to withstand abrading action of the furnace charge and also the pressure load.
4. They should withstand spalling during any process, that is, cracking and flaking or peeling off of bricks due to uniform expansion at high temperature or vice versa.
5. They exhibit a dimensional stability without any change in the volume due to high temperature operation or over a prolonged heating time.
6. Chemical inertness is one of the essential factors to be considered for selecting a suitable refractory material in a kiln or furnace at high temperature and should not give a fusible product with fuel ashes, slags, etc. If one of the products in the furnace is acidic, then it is desirable to have basic brick lining in the furnace.
7. The refractory materials have either low or high thermal conductivities. The kinds of materials used in muffle furnaces are dependent on their thermal conducting behaviour and the selection of material is based on the specific purpose and use.
8. Refractories also come in a range of different *densities* and *porosities*. Generally, low porosity refractories have higher thermal conductivities compared to high porosity materials. High porosity materials are usually highly insulating due to high volume of air they envelope, since air is a very poor thermal conductor. However, they do not cope as well with higher temperatures and direct flame impingement and tend to shrink under these conditions. Thus, the low porosity materials are used in the hotter zones, while more porous materials are used as thermal backup materials.
9. The materials of low electrical conductivity are generally used for lining in furnaces. All refractories have low electrical resistances except in the case of graphite.
10. Heat capacity of refractory materials plays a major role for the selection for lining of furnace.

17.4 Melting Points of Few Pure Refractory Oxides

The melting points of few refractory oxide materials are given below indicating that they are thermally stable below these temperatures and can be best utilized as refractory lining for kilns and furnaces.

1. SiO_2	(Silicon Oxide)	1702°C
2. Al_2O_3	(Aluminium Oxide)	2050°C
3. Cr_2O_3	(Chrome Oxide)	2275°C
4. CaO	(Calcium Oxide)	2600°C
5. ZrO_2	(Zirconium Oxide)	2700°C
6. MgO	(Magnesium Oxide)	2800°C

17.5 Manufacture of Refractories

Generally, the following processes are adopted for the manufacture of refractory materials.

(1) Crushing and grinding

A good quality refractory material can be obtained by crushing the big lumps of material into smaller sizes; and the powders are ground well to anappropriate size and sieved to a particular uniform size.

(2) Screening

The materials so obtained are subjected to different physical, chemical and magnetic methods and the material became free from unwanted impurities.

(3) Mixing

It is subjected to a process of mixing which results in equal distribution of plastic materials throughout the mass so that moulding may be easy.

(4) Moulding

The moulding of the refractory material can be conducted by any of the two methods as follows:

- Hand moulding and
- Mechanical moulding.

It is observed that hand moulding gives refractories with low density and strength; while mechanical moulding produces refractories with high density and strength.

De-airing is desired to increase the density and strength. A standard procedure is used to de-air the materials to achieve high strength and density.

(5) Drying

The removal of water/moisture content of refractory material is executed through a process known as drying. Generally, drying is carried out very slowly and particularly under specific conditions of humidity and temperature.

(6) Firing and sintering

The very purpose of firing the refractory material at elevated temperature is to achieve the desirable properties such as vitrification, sintering and to obtain a stable mineral form. The usual temperature range is 1400–2000°C for any type of bricks.

17.6 Acid Refractories

17.6.1 Silica

(a) Silica is one of the most abundant oxide materials in the earth's crust. The main sources of silica are sandstone, silica sand or quartzite. The raw material for the manufacture of glass and ceramics is silica. It exists in an amorphous form (vitreous silica) or in a variety of crystalline forms. Often it will occur as a noncrystalline oxidation product on the surface of silicon or silicon compounds.

(b) **Crystalline forms of silica**

The three forms of crystalline silica are quartz, tridymite and cristobalite

Fused silica: 99.4–99.9 percent SiO_2 is known as a high purity fused silica which is produced by carbon arc, plasma arc, gas fired continual extrusion or carbon electrode fusion. Because of its good dielectric and insulating properties, it is used in the electronics industry; though it may also be used as a refractory material or in investment casting.

(c) **Properties**

Silica has good abrasion resistance, electrical insulation and high thermal stability. It is insoluble in all acids with the exception of hydrogen fluoride (HF).

Material	Quartz	Fused silica
Density (g/cm^3)	2.65	2.2
Thermal conductivity (W/mK)	1.3	1.4
Thermal expansion coeff. ($10^{-6}K^{-1}$)	12.3	0.4
Tensile strength (MPa)	55	110
Compressive strength (MPa)	2070	690–1380
Poisson's ratio	0.17	0.165
Fracture toughness (MPa)	—	0.79
Melting point (°C)	1830	1830
Modulus of elasticity (GPa)	70	73
Thermal shock resistance	Excellent	Excellent
Permittivity (ε')	3.8–5.4	3.8
Tan ($\delta \times 104$)	3	
Loss factor (ε'')	0.0015	
Dielectric field strength (kV/mm)	15.0–25.0	15.0–40.0
Resistivity (Ωm)	1012–1016	>1018

Table 17.1

Physical, mechanical, thermal and electrical properties of quartz and fused silica

Phase	Density (g/cm^3)	Thermal expansion ($10^{-6}K^{-1}$)
Quartz	2.65	12.3
Tridymite	2.3	21
Cristobalite	2.2	10.3

Table 17.2

Differences between the different crystal structures of silica

(d) Applications

Silica is widely used as a precursor material for the fabrication of other ceramic products and as a material on its own.

- **Refractory materials**
  Silica is used as a refractory material because of its low thermal expansion, a fairly high melting point and resistant to creep. It tends to be used in acid environments if used on its own or used as a starting material for the synthesis of other refractory products. Due to the fact that silica is insoluble in the majority of acids, it is used as a refractory material in acidic environments. Silica is classified as an acid refractory as it behaves like an acid at high temperatures reacting with bases.
- **Glass refractories**
  In the glass industry, it finds major application. Silica is an important material for use as a refractory in the production of glass. The two types of silica commonly used in glass refractories are as follows:
  - *Crystallised silica*, which is composed of pure crystalline quartz. This material is crushed, graded and pressed into bricks. Upon heating up to high temperatures, the quartz will transform to tridymite and cristobalite. In bricks, tridymite tends to be the most favourable of all crystalline forms of silica, as it has a smooth, predictable and low thermal expansion up to 600°C. Once beyond this temperature, its thermal expansion is practically zero. Bricks of this type are used in the melter crown of the glass furnace as they provide good resistance to creep and good mechanical strength at the operating temperature.
  - *Fused silica* is used because of its good resistance to thermal shock and there is no risk of contamination of the glass, since in glass furnaces where the glass is in direct contact with the refractory.

(e) Production of refractory materials from silica

- Synthetic mullite refractory bricks are obtained by reacting silica sand with Bayer alumina, which are used in furnace linings. Mullite imparts good refractoriness and has excellent creep resistance. Mullite can be used to line blast, copper roasting and sections of blast furnaces.
- Silicon carbide a high shock resistant refractory material is produced by the reaction of silica sand with coke.
- Silica is also used in the production of AZS by fusion casting, which is used in the glass industry.
- Fused silica is used in the refractory casting slurry, where it coats a wax replica of the component to be cast. The refractory stucco is fired and molten metal is poured into the mould. Finally, the fused silica shell is knocked out. Fused silica makes this knock out stage easier than other refractories such as zirconia and aluminosilicates.

17.6.2 Silica Brick

The silica brick can be manufactured by taking 90 percent crushed mineral with SiO_2 along with a flux material CaO; and heating at an elevated temperature of 1010° C. These bricks found use primarily in steel mills and coke by-products operations, primarily in strong phosphoric acid exposures where shale and fireclay brick have short life. The service of silica bricks can be extended to about 1093°C and are more resistant to thermal shock due to their greater porosity as high as 16 percent. The purity of the silica and its percentage of alkali, along with the manufacturing techniques, determine the uniformity or the wideness of ranges of the physical properties. The chemical composition of silica brick is given in Table 17.3.

Chemicals	Composition of silica brick (percent)
SiO_2	98.6–99.6
Al_2O_3	0.2–0.5
Fe_2O_3	0.02–0.3
MgO	0.02–0.1
CaO	0.02–0.03
Alkali metals (Na_2O, K_2O, Li_2O)	0.01–0.2

Table 17.3

Chemical composition of silica brick

A special kind of silica brick is manufactured exclusively for lining coke ovens and glass furnaces, which is obtained by thoroughly mixing ground silica (usually quartzite) with the addition of ½ to 3 percent milk-of-lime during sintering heating. This mixture is formed into various shapes and then fired in kilns.

Applications

- Silica bricks are used repeatedly for long periods in acid service without noticeable damage, showing greater resistance, especially to strong hot acids, and particularly phosphoric, than acid brick, and in halogen exposures (except fluorine), solvents and organic chemical exposures.
- Silica bricks are not recommended for service in strong alkali environments. Due to their porosity and the low thermal expansion, they have better shock resistance than shale or fireclay acid brick, but they have lower strength and abrasion resistance.

17.6.3 Fireclay/Kaolin

Generally, a good fireclay refractory material with less than 25 percent of Fe_2O_3 content in it has 24–26 percent plasticity and shrinkage, which reduces to 6–8 percent after a proper sintering process.

The refractory bricks made up of fireclay are the most common and extensively used in all places of heat generation; this is because of the abundant supply of fireclay and availability in the market at a cheap rate.

Fireclay is a group of refractory clays that can withstand temperatures above the pyrometric cone equivalent (PCE) value 19. If the PCE value of any clay is less than 19, then they are not used as a refractory material. Fireclay is essentially of kaolinite group and has a composition similar to that of china-clay. In nature, it is usually found to contain 24–32 percent Al_2O_3, 50–60 percent SiO_2 and LOI between 9 and 12 percent. Impurities like oxides of calcium, iron, titanium and magnesium and alkalies are invariably present, making it white, grey and black in colour. Fireclay is generally of sedimentary origin.

(i) Properties and tests

The most important properties of fireclay for its suitability as a refractory are its refractoriness and plasticity. A good fireclay should have a high fusion point and good plasticity. Depending upon their capacity to withstand high temperatures before melting, the fireclays are graded into the following:

1. *Low duty*: withstand temperatures between 1515–1615°C (PCE-19 to 28)
2. *Intermediate duty*: 1650°C (PCE 30)
3. *High duty*: 1700°C (PCE 32)
4. *Super duty*: 1775°C (PCE 35)

(ii) Industrial applications

Fireclay refractory is used in all places of heat generation extensively because of the abundant availability of fireclay and low cost material for boiler furnaces. It is extensively used in

- glass melting furnaces
- chimney linings
- pottery kilns blast furnaces
- reheating furnaces

Fireclay is an example of acid refractories. Acid refractories are those which are not attacked by acid slag. In blast furnaces, the lining is done almost entirely with fireclay bricks. Pouring refractories like sleeves, nozzles, stoppers and tuyers are made of fireclay.

(iii) Manufacturing process of fireclay

The process of manufacturing fireclay is as follows:

- The clay procured from mines is stacked in the factory yard and allowed to weather for about a year.
- For daily production of different types of refractories, this weathered clay is taken and mixed in different percentages with grog.
- The mixture is sent to the grinding mill from where it is transferred to the pug mill.
- In the pug, mill a suitable proportion of water is added so as to give it proper plasticity.
- The mould is supplied to different machines for making standard bricks or shapes. Intricate shapes are made by hand.
- The bricks, thus made, are then dried in hot floor driers and after drying they are loaded in kilns for firing.
- The firing ranges are, of course, different for different grades of refractories.
- After firing, the kilns are allowed to cool; then the bricks are unloaded. By burning fireclay is converted into a stone-like material, highly resistant to acid, water and most other solutions.
- While manufacturing high aluminous fire-bricks bauxite is added along with grog in suitable proportions

'*Kaolin*'-based refractories are using natural kaolin or a mixture of clay and other ceramics such as alumina, calcium aluminate or silicon carbide. Kaolin acts as a binder and provides plasticity. Kaolin is a hydrous aluminium silicate [$Al_2(Si_2O_5)(OH)_4$] based mineral clay. Kaolin is also referred to as clay, anhydrous aluminium silicate, aluminium silicate dihydrate, nacrite, dickite, kaolinite, calcined, kaolinite; china clay, porcelain clay, aluminium, silicate hydroxide or aluminium silicate (hydrated). The plate-like structure allows particles in a wet clay mass to slide across each other and maintain plasticity. Kaolin is a white soft plastic clay composed primarily of well-ordered kaolinite mineral [$Al_2(Si_2O_5)(OH)_4$] with minor amounts of quartz, feldspar and sheet silicate minerals (mica, illite, smectite and chlorite).

Geologically, there are two types of kaolin deposits, i.e. primary and secondary kaolin.

- Primary kaolin is formed through the alteration, or kaolinization, of *in situ* minerals of feldspar and other aluminium silicates to kaolinite.
- Secondary kaolin is laid down as sediments, usually in fresh water, far from the place of origin. Various types of secondary kaolin are ball clay, fireclay, or flint clay depending on kaolinite content and their properties.

17.7 Basic Refractories

Few basic refractories are listed below

(a) Alumina
(b) Lime
(c) Magnesia
(d) Dolomite and
(e) Zirconia.

17.7.1 Alumina

Aluminium oxide (Al_2O_3) or alumina is one of the most versatile refractory ceramic oxides and finds use in a wide range of applications. Alumina or aluminium oxide (Al_2O_3) is anoxide compound of aluminium and exists as α-alumina. In its pure form, alumina is a white, high hardness ceramic. Fully dense alumina can be translucent. It is found in nature as corundum in emery, topaz, amethyst and emerald and as the precious gemstones ruby and sapphire, but it is from the more abundant ores such as bauxite, cryolite and clays that the material is commercially extracted and purified.

Alumina has found a wide range of application due to its versatility and relatively low raw material cost. Depending on its purity and density, alumina is used in refractory tubes, industrial crucibles, analytical laboratory ware, dielectric substrates, wear components, refractory cements and abrasives. The main drawback of alumina is its poor thermal shock resistance which is because of its higher coefficients of thermal expansion and lower thermal conductivity.

Corundum exists as rhombohedral crystals with hexagonal structure. It is the close packing of the aluminium and oxygen atoms within this structure that leads to its good mechanical and thermal properties.

(i) Extraction

(a) **Alumina is extracted and purified by Bayer's process.**
The finely ground bauxite is mixed with NaOH solution. The bauxite slowly dissolves in the tank when steam is passed at high pressure. The alumina released reacts with the sodium hydroxide to form sodium aluminate. After the contents of the tank have passed through other vessels where the pressure and temperature are reduced and impurities are removed, the solution of sodium aluminate is placed in a special tank where the alumina is precipitated out. The precipitate is removed from the tank, washed and heated in a kiln to drive off any water present. The residue is a commercially pure alumina.

(b) Other extraction processes are used including pyrogenic treatment of bauxite with soda, and the extraction of aluminium hydroxide from metakaolin via either the chloride or sulphate.

The yield of alumina from these processes can approach 90 percent.

For advanced ceramics uses, the alumina manufactured by these processes requires further purification. This is often achieved by recrystallisation from ammonium alum.

(ii) Properties

The typical physical and mechanical properties for different purity grades of alumina are given in Tables 17.4 and 17.5.

Property	
Melting point (°C)	2015 ± 15
Refractive index	1.765
Molecular wt (g/mol)	101.96
Δ*Gf*° Free Energy of Formation (kJ/mol)	−1582.4

Table 17.4

Typical physical properties of alumina

Property	**Alumina grade**						
	86%	94%	97.5%	99.5%	99.9%	99% recry.*	Saph**
Density (g/cm^3)	3.5	3.7	3.78	3.89	3.9	3.9	3.985
Dielectric constant	8.5	9.2	9.5		9–10.1	9–10.1	7.5–10.5
Dielectric strength (kV/mm)	28		30–43		10–35	10–35	17
Volume resistivity (Ohm.cm)	>1014	>1014	>1014	>1014	>1014	>1014	>1016
Thermal conductivity (W/mK)	15	20	24	26	28–35	28–35	41.9
Thermal expansion coefficient (20–1000°C × $10^{-6}K^{-1}$)	7	7.6	8.1	8.3	8	8	5.8
Specific heat (JK/kg)	920	900		850			753
Compressive strength (MPa)	1800	2000	1750–2500		2200–2600	2200–2600	2100
Modulus of rupture (MPa)	250	330		262	320–400		260
Hardness(Vickers kgf/mm^2)			1500–1600		1500–1650	1500–1650	2500–3000

Table 17.5

Typical physical and mechanical properties of 86–99.9 percent alumina

(iii) Applications

Alumina refractories find a wide range of applications. Some of the major application areas can be grouped, which is shown in Table 17.6.

Percent Al_2O_3	Grain size	Porosity	Applications area
>99.6	Fine	closed	Electrical, engineering, biomedical
>99.8	Fine	zero	Lamp tubes, optical
>99.6* (recrystallised)	Medium	closed	High temperature uses
95–99.5	Fine	closed	General electrical, engineering
80–95	Fine	closed	Low duty electrical (spark plugs)
90–99.6	Fine/coarse	open	Filter media
80–90	Fine/coarse	open	Abrasive

Table 17.6

Example applications for a range of alumina

1. *High temperature and aggressive environments*: Its high free energy of formation makes alumina chemically stable and refractory, and hence it finds uses in containment of aggressive and high temperature environments.
2. *Wear and corrosion resistance*: The high hardness of alumina imparts wear and abrasion resistance and hence it is used in diverse applications such as wear resistant linings for pipes and vessels, pump and faucet seals, thread and wire guides, etc.
3. *Biomedical:* High purity aluminas are also used as orthopaedic implants particularly in hip replacement surgery.
4. *Metal cutting tools:* The high 'hot' hardness of alumina led to applications as tool tips for metal cutting and abrasives.
5. *Milling media*: Alumina is used as milling media in a wide range of particle size reduction processes.
6. *Microwave components*: The high dielectric constant coupled with low dielectric loss particularly at high frequencies leads to a number of microwave applications including windows for high power devices and waveguides.
7. *Electrical insulation*: The high volume resistance and dielectric strength make alumina an excellent electrical insulator which leads to applications in electronics as substrates and connectors, and in lower duty applications such as insulators for automotive spark plugs.

17.7.2 Magnesia/Magnesite

(a) Magnesia

Magnesia refractories are based on compounds of magnesium and oxygen. Magnesite or magnesia refractories are, generally, magnesium oxide, magnesium carbonate, dead burned magnesite, calcined magnesite, periclase or magnesia clinker.

Depending on the origin and processing, magnesia refractory is divided into the following:

- caustic
- dead-burnt
- fused
- precipitated
- sintered or calcined and
- synthetic magnesia forms.

The requisite of a magnesia (MgO) refractory is that this refractory material should have high melting temperature (2800°C) and should also have good heat resistance of 1700°C for reducing atmosphere and 2300°C in oxidizing atmosphere. Magnesite is the naturally occurring mineral or ore used to produce magnesium oxide-based refractories. Magnesite often contains iron, manganese or other activator elements. Magnesium oxide refractories with a carbon bond are frequently used in the steel industry. Magnesite refractories have good resistance to molten iron and steel.

Magnesia (or magnesium oxide) is an alkaline earth metal oxide. The majority of magnesium oxide produced today is obtained from the calcination of naturally occurring minerals, magnesite, $MgCO_3$, being the most common. Other important sources of magnesium oxide are seawater, underground deposits of brine and deep salt beds from which magnesium hydroxide, $[Mg(OH)_2]$ is processed. Magnesium is the eighth most abundant element constituting about two percent of the earth's crust and typically 0.12 percent of seawater. Both $MgCO_3$ and $Mg(OH)_2$ are converted to MgO by calcination. The thermal treatment of the calcination process affects the surface area and pore size and hence the reactivity of magnesium oxide formed. The source largely determines the level and nature of impurities present in the calcined material. Caustic calcined magnesia, which is used in a wide range of industrial applications, e.g. plastics, rubber, adhesives and acid neutralization is formed by calcining in the range 700–1000°C. By calcining in the range 1000–1500°C, the magnesium oxide is used where its lower chemical activity is required, e.g. fertiliser and animal feed. Dead-burned magnesia, which is produced in shaft and rotary kilns at temperatures over 1500°C, has reduced chemical reactivity; therefore, it is more suitable for refractory applications. Finally, fused magnesia which is produced in an electric arc furnace from caustic calcined magnesia at temperatures in excess of 2650°C is used for a variety of refractory and electrical applications.

Magnesite does not ordinarily form good crystals, but can make up a substantial portion of some rock types. It forms commonly from the alteration of magnesium-rich rocks during low grade metamorphism while they are in contact with carbonate-rich solutions. Magnesite has the same crystal structure of calcite, hence its inclusion into the calcite mineral group. Many of the properties of magnesite are either identical or similar to those of calcite. However, the magnesium ion does not allow the carbonate ion (CO_3^{-2}) to interact as easily with cold acids, as the calcium ion does in calcite. This provides the best means of distinguishing magnesite from calcite. However, dolomite $[MgCa(CO_3)_2]$ can be almost indistinguishable from magnesite.

(b) Physical characteristics

- Colour is white or grey, also tinted yellow or brown.
- Lustre is vitreous.
- Transparency crystals are translucent to transparent only in individual crystal.
- Crystal habits are usually massive forms such as lamellar, fibrous and coarse to fine grained rocks. Crystals are extremely rare, but when found are in the form of rhombohedrons or hexagonal prisms.
- Cleavage is perfect in three directions forming rhombohedrons.
- Hardness is 4–4.5.
- Specific gravity is approximately 3.0 (average). Streak is white.
- Associated mineral calcite, dolomite, aragonite, strontianite and serpentine.
- Other characteristics: effervesces easily only in hot dilute hydrochloric acid.

(c) Uses

1. Magnesite can be used as a slag former in steelmaking furnaces, in conjunction with lime, in order to protect the magnesium oxide lining.
2. It can also be used as a catalyst and filler in the production of synthetic rubber and in the preparation of magnesium chemicals and fertilizers.

3. Similar to the production of lime, magnesite can be burned in the presence of charcoal to produce MgO, otherwise known as periclase. Such periclase is an important product in refractory materials.

17.7.3 Types of Magnesia

(1) Fused magnesia

Fused magnesia (MgO) is normally manufactured by the electric arc melting of caustic calcined magnesia, dead-burned magnesia or raw magnesite in furnaces at temperatures in excess of 2750°C, producing a refractory product whose altered crystalline structure is such that its characteristics and performance are superior to competing materials. Magnesite ($MgCO_3$), the naturally occurring carbonate of magnesium (Mg), is one of the key natural sources for the production of magnesia (MgO) and subsequently fused magnesia. Magnesite occurs in two distinct physical forms: macrocrystalline and cryptocrystalline. Cryptocrystalline magnesite is generally of a higher purity than macrocrystalline ore, but tends to occur in smaller deposits than the macrocrystalline form.

Fused magnesia is superior to dead-burned magnesia in strength, abrasion resistance and chemical stability. Major applications are in refractory and electrical insulating markets. Producers of fused magnesia commonly fall into one of two categories as follows:

- Those producing refractory grades
- Those producing electrical grades
- Few producers serve both markets on a mainstream basis.

Refractory grade fused magnesia: The addition of fused magnesia grains can greatly enhance the performance and durability of basic refractories such as magcarbon bricks. This is a function of a higher bulk-specific gravity and large periclase crystal size, plus realignment of accessory silicates. Refractory grade fused magnesia has exacting specifications and is normally characterised by the following

- Generally high magnesia content (minimum 96 percent MgO and up to/exceeding 99 percent MgO)
- Low silica, lime: silica ratios of 2:1
- Densities of 3.50 g/cm^3 or more
- Large periclase crystal sizes (>1000 μm)
- Due to its excellent corrosion resistance, refractory grade fused magnesia is used in high wear areas in steel making, e.g. basic oxygen and electric arc furnaces, converters and ladles.

Ultra high purity (>99 percent MgO) grades have been used in high-tech applications such as optical equipment, nuclear reactors and rocket nozzles.

(2) High grade magnesia production

Historically, and due principally to the small size of most known cryptocrystalline deposits, production of high grade magnesia products was mainly by extraction from natural brines or seawater (synthetic MgO), a high cost and energy intensive process. High quality deposits, provide an alternative source of supply to the high cost seawater-sourced magnesia.

(3) Deadburned magnesia

It is used almost exclusively for refractory applications in the form of basic bricks and granular refractories. Dead-burned magnesia has the highest melting point of all common refractory oxides and is the most suitable heat containment material for high temperature processes in the steel industry. Basic magnesia bricks are used in furnaces, ladles and secondary refining vessels and in cement and glass making kilns.

(4) Fused magnesia production process

(i) Magnesite (magnesium carbonate $MgCO_3$) is converted into magnesia by the application of heat which drives off carbon dioxide (CO_2), thereby converting the carbonate to the oxide of magnesium (MgO).

(ii) Magnesite, from both natural sources (primarily magnesite) and synthetic sources (seawater, natural brines or deep sea salt beds), is converted into caustic calcined magnesia by calcining to between 700°C and 1000°C, driving off 96–98 percent of the contained carbon dioxide. Caustic calcined magnesia is both an end product and an intermediary step in the chain of magnesia products.

Further calcining of magnesite at higher temperatures between 1750–2200°C results in the largely inert product, dead-burned magnesia. Heating to this level drives off all but a small fraction of the remaining carbon dioxide to produce a hard crystalline nonreactive form of magnesium oxide known as periclase. Dead-burned magnesia exhibits exceptional dimensional stability and strength at high temperatures.

(iii) Fused magnesia is produced in a three phase electric arc furnace. Taking high grade magnesite or calcined magnesia as raw materials, 12 hours is required for the fusion process at temperatures in excess of 2750°C. The process promotes the growth of very large crystals of periclase (>1000 μm compared with 50–100 μm for dead-burned magnesia) with a density approaching the theoretical maximum of 3.58 g/cm^3.

In fused magnesia production, the main constraints on capacity are the size and number of electric arc furnaces, and the cost of energy. The manufacture of fused magnesia is very power intensive with electricity consumption varying between 3500 and 4500 kWh/ton.

(5) Properties

There are few dense engineering ceramics of the structural type made from pure magnesia. However, there is a wide range of refractory and electrical applications where magnesia is firmly established. The properties of major interest are as follows:

- Good refractoriness
- Good corrosion resistance, e.g. basic steelmaking slags, Sodium hydroxide, Fe, Co, Ni
- High thermal conductivity
- Low electrical conductivity
- Transparency to infrared

(6) Applications

1. Magnesia is widely used in the steel industry as a refractory brick often impregnated with carbon (tar, pitch, graphite) to give optimum properties for corrosion resistance in environments of basic slags, particularly in BOF furnaces or slag lines of treatment ladles.
2. Magnesia bricks often in combination with spinel or chrome are also used in ferroalloy, nonferrous, glass and cement industries.
3. Castables and sprayables based on magnesia are widely used for basic refractory linings for steel transfer applications. The lime to silica ratio present in the magnesia has a major influence on its properties.
4. Crucibles: Magnesia crucibles have found application in the super alloy industry, nuclear industry and chemicals industry where corrosion resistance is required. Manufacture is by firing isostatically pressed or slip cast shapes made from fused magnesia. Various purity grades are commercially available. Additives are employed to promote sintering or restrict crystal growth. These range from clay to yttria and alumina depending on application. Crucibles with stability of 2400°C in air, 1700°C in reducing atmospheres, 1600°C in vacuum and 1400°C in hydrogen have been reported in the literature.
5. Cements: Magnesia (or Sorel) cement is a refractory binder based on a magnesium oxychloride formulation. It is fast-hardening and has a number of refractory and general repair applications. Magnesia is also used as a room temperature curing agent for phosphate cements.
6. Heating elements: Magnesia powder is widely used as a filling for electrical heating elements for applications in contact with air or liquids such as electric cooker rings, storage heaters, washing machines

and diesel engine glow plugs. Fused magnesia has the ideal combination of electrical resistance and thermal conductivity. The MgO forms a layer between the element and the outer sheath. It is also used as mineral insulation in cables.

7. Thermocouple tubes: Extruded magnesia protective sheaths have been used to house thermocouples in aggressive environments.
8. Brake linings: Magnesia has been included in brake linings due to its thermomechanical properties. Its intermediate hardness gives sufficiently low wear on metal while conducting heat from the friction contact surfaces.

17.7.4 Magnesia Clinker

Magnesia clinker is basic refractory material and it has high corrosion resistance to basic slags. It has a high thermal conductivity and a large coefficient of thermal expansion. Calcining magnesite ($MgCO_3$) in a rotary kiln produces *magnesia clinker*. Iron oxide forms a solid solution with periclase (MgO) which acts as a sintering aid. Another magnesia raw material is fused magnesia, which is obtained by electrically fusing seawater magnesia clinker. Magnesia has a tendency to react with water to form magnesium hydroxide ($Mg(OH)_2$), therefore, magnesia-based castables must be stored carefully and the contact with steam avoided.

Sintered magnesia

The typical analyses for sintered magnesia refractory:

MgO: 85–99.8 percent
CaO: 0.1–3.5 percent
SiO_2: 0.1–3 percent
B_2O_3: 0.1–8 percent

17.7.5 Dolamite

The carbonate mineral is dolomite, $CaMg(CO_3)_2$. Often small amounts of iron, manganese or excess calcium replace some of the magnesium; cobalt, zinc, lead and barium are more rarely found. Dolomite is normally white or colourless with a specific gravity of 2.9 and a hardness of 3.5–4 on Mohs scale. It can be distinguished from calcite by its extremely slow reaction with cold dilute acid. Dolomite is a very common mineral, occurring in a variety of geologic settings. It is often found in ultrabasic igneous rocks, notably in carbonatites and serpentinites, in metamorphosed carbonate sediments, where it may recrystallize to form dolomite marbles, and in hydrothermal veins. The primary occurrence of dolomite is in sedimentary deposits, where it constitutes the major component of dolomite rock and is often present in limestones.

(a) Manufacturing processes

1. High purity refractory 'dolomite' (frequently called Doloma) and lower purity fettling grade 'dead-burned dolomite' (usually referred to as DBD) are both manufactured by calcining dolomitic limestone. The method is similar to the manufacture of ordinary lime, except that the burning time is longer and temperatures considerably higher (in the range of 1600–1850°C).
2. High purity doloma is fired in rotary or shaft kilns to the upper end of the temperature range without the addition of impurities. The lower purity DBD is fired in rotary kilns to the lower end of the temperature range, and iron oxides are added during calcinations to stabilize the resulting hard-burned quicklime against decomposition from moisture.

(b) Properties

The mineral dolomite crystallizes in the trigonal rhombohedral system. It forms white, grey to pink, commonly curved crystals, although it is usually massive. It has physical properties similar to those of the mineral calcite, but does not rapidly dissolve or effervesce in dilute hydrochloric acid. The Mohs hardness is 3.5–4 and the specific gravity is 2.85. Refractive index values are $n_\omega = 1.679 - 1.681$ and $n\varepsilon = 1.500$. Crystal twinning is common. A solid solution series exists between dolomite and iron-rich ankarite. Small amounts of iron in the structure give the crystals a yellow to brown tint. Manganese substitutes in the structure also up to about 3 percent MnO. A high manganese content gives the crystals a rosy pink colour. Lead and zinc also substitute in the structure for magnesium.

(c) Uses

- High purity doloma is used to manufacture refractory bricks employed in cement and lime rotary kiln linings, and in steel ladles and refining vessels. DBD is used for the manufacture of monolithic patching and repair materials for steel furnaces.
- Large quantities of light-burned dolomitic lime are also employed in the production of synthetic refractory grade magnesia (MgO).
- The quicklime is slaked in a magnesium chloride brine, precipitating magnesium hydroxide. The mag-hydroxide is calcined and fired into dense, high-purity magnesium oxide. Refractory magnesia is used in the production of linings for cement and lime kilns, in addition to steel ladles and furnaces.
- Dead-burned dolomite is produced when dolomite is calcined at very high temperatures. It is used both as a refractory product in granular form to repair linings and for making the bricks used in the refractory linings of casting ladles and cement kilns.
- In the steelrefining process, the use of dolime instead of pure quicklime will extend the life of refractory linings. In fact, adding dolime will create MgO in solution in the slag, which provides an excellent buffering capacity as MgO particles in suspension provide excellent coating protection. Any excess MgO precipitates, protecting refractories and tap holes, reducing the need for gunning and repairs.

17.7.6 Lime/Limestone

17.7.6.1 Lime

Lime/quicklime is a term commonly applied to a number of related materials. Pure lime is calcium oxide (CaO) formed by ‘burning’ a form of calcium carbonate such as limestone or marble ($CaCO_3$). Carbon dioxide gas (CO_2) is released and leaves lime behind. Dolomite, a calcium magnesium carbonate [$CaMg(CO_3)_2$] can also be calcined to form dolomitic lime, which has different reactivity due to the presence of MgO. Lime can be mixed with water to form hydrated lime ($Ca(OH)_2$), which is also known as slaked lime.

(a) Uses

Many applications for lime exploit the alkalinity that is created when the lime is formed. Lime is a strong base, so reacts with, and neutralizes, acids. In this capacity, it is used to treat wastewater, drinking water and other industrial acid streams. Lime is widely used to desulphurize stack emissions in coal- and oil-fired power plants. The lime reacts with sulphur in the coal or oil to form calcium sulphate (gypsum). The gypsum produced from desulphurization is now used as a raw material to create wallboard.

Lime is most extensive use as a flux in purifying steel in the both *electric arc furnaces* (EAF) and *basic oxygen furnaces* (BOF). It is particularly effective in removing phosphorus, sulphur and silica, and to a lesser extent, manganese. Lime also has important uses in the secondary refining of steel and in the manufacture of steel products.

(b) Applications

- Lime is one of the world's oldest components in cements and mortars for building purposes.
- Hydrated lime mortars were used in Roman times and are still performing as designed after all these years.
- Hydrated lime is widely used in asphalt, masonry, stuccos, cements and in soil stabilization.
- Finally, lime is widely used in the manufacture of precipitated calcium carbonate (PCC). Carbon dioxide gas is introduced into a lime slurry and $CaCO_3$ is precipitated out of solution. Precipitated calcium carbonates are produced in large quantities for paper, plastics, sealants, food and pharmaceuticals.

17.7.6.2 Limestone

Limestone is calcareous sedimentary rock composed of the mineral calcite ($CaCO_3$), which upon calcination yields lime (CaO) for commercial use. In its broadest interpretation, the term includes any calcareous material such as marble, chalk, travertine, limeshell, coral and marl each possessing different and distinct physical properties. The crystalline equivalents of limestone having the same chemical composition are calcite and aragonite. In nature, the limestone bed is found to occur in varying purity, generally a part of the calcium molecules being replaced by magnesium tending towards magnesium limestone or dolomitic limestone. Limestone with more than 10 percent of mineral dolomite is termed dolomite limestone and that with 5–10 percent, magnesium limestone.

Travertine is a banded, compact variety of limestone formed along streams, particularly where there are waterfalls and around hot or cold springs. Calcium carbonate is deposited where evaporation of the water leaves a solution that is supersaturated with chemical constituents of calcite. The rock containing more than 95 percent of calcium carbonate is termed high-calcium limestone. Recrystallized limestone takes good polish and is used as decorative and building stone. It is then called 'marble'.

A variety of other names is associated with limestone depending upon the modes of occurrence, texture, and the remains of foraminifera, molluscs and other shell-forming creatures, for example, oolitic and pisolitic limestones, reostone, crinoidal limestone, foraminiferal limestone, calcareous ooze, bryoza limes tone, argillaceous and hydraulic limestones and the like.

Oolitic limestone is the name given to granular limestone of which each grain consists of a series of concentric layers of calcium carbonate resembling the eggs of the fish. When it resembles the eggs of Roe fish, it is called roestone. The name pisolite-limestone is given when the grains are of the size of peas. Crinoidal limestone is a rock composed of encrinites with the remains of foraminifera-molluscs and their lime forming organism.

Such beds are commonly found in lower Carboniferous formations.

(a) Uses

- Limestone is quarried for roadbeds, building and landscape construction and cement manufacture.
- Limestone is a primary material used in architecture and many landmarks around the world.
- Limestone is readily available and relatively easy to cut into blocks or more elaborate carving.
- It is also long-lasting and stands up well to exposure.
- Though the limestone used for construction is good for humid climates, it is vulnerable to acids, making acid rain a problem when it occurs in places where limestone is used extensively.

17.7.7 Zirconia

Zirconia or zirconium oxide (ZrO_2) is an extremely refractory compound of zirconium and oxygen. Zirconia may have additions of calcia, magnesia or yttria to stabilize the structure into a cubic structure. Zirconia stabilized in the cubic crystal structure avoids cracking and mechanical weakening during heating and cooling.

Certain zirconia materials have the ability to transform structurally to tough material under applied stress and are frequently used in wear applications requiring improved fracture toughness and stiffness over alumina. Zirconia ceramics possess excellent chemical inertness and corrosion resistance at temperatures well above the melting point of alumina. Zirconia is more costly than alumina, so it is only used when alumina fail. Zirconia has low thermal conductivity and it has an electrical conductor above 800°C. Zirconia is used to fabricate oxygen sensors or fuel cell membranes because of its unique ability to allow oxygen ions to move freely through the crystal structure above 600°C. Zirconia products should not be used in contact with alumina above 1600°C. Depending on the purity and density, zirconia is used for refractory tubes, industrial crucibles, analytical labware, sensors, wear components, refractory cements, thermocouple protection tubes, furnace muffles, liners and high temperature heating element supports.

Stabilized zirconium oxide (ZrO_2) is produced by the addition of about 3 percent lime during the fusion or sintering process and is widely used for analysis crucibles and thermocouple tubes, in addition to crucibles and furnaces used for the production of high-temperature aerospace alloys.

17.8 Neutral Refractories

The following neutral refractories are, generally, used as furnace materials.

(i) Silicon carbide
(ii) Chromite
(iii) Carbon and others.

17.8.1 Silicon Carbide

(a) Silicon Carbide

Silicon carbide is a hard covalently bonded material predominantly produced by the carbothermal reduction of silica. Silicon carbide is made by heating silica sand and petroleum coke packed around electrodes in an electric resistance furnace to above 2200°C. Depending on the exact reaction conditions the resulting silicon carbide is either a fine powder or a bonded mass that requires crushing and milling to produce a usable feedstock. This material is very resistant to abrasion and to corrosion with a molten slag. It also has excellent resistance to thermal spalling. However as it is a carbide, it will oxidize readily silicon carbide has a fairly high conductivity.

Several hundred structures of silicon carbide (polytypes) have been identified which have different stacking arrangements for the silicon and carbon atoms. The simplest structure is a diamond structure which is designated β-SiC. Other structures are either hexagonal or rhombic and are referred to as α-SiC.

Silicon carbide is one of the very few totally manmade minerals used in refractory work. These are as follows:

- *Oxide-bonded:* (SiO_2, Al_2O_3, SiO_2 or silicate glass), silicon oxynitride (Si_2ON_2), silicon nitride (Si_3N_4). The first three of these four bonding systems result in a permeable product, and when failure occurs in such masonry systems due to chemical degradation, it is usually due to attack on the bond. Thus, permeable units (where the corrodent penetrates the mass) are far more rapidly damaged.
- '*Self-bonded*'– (silicon carbide to silicon carbide) impermeable ones, where the attack is limited to the surface.

  The self-bonded product can be manufactured by either of the two methods: reaction bonded or sintered. Both will produce an impermeable unit, and they have roughly comparable chemical resistances, but they do not have identical physical properties.

(b) Uses

Silicon carbide is a premium-priced unit which is employed in lining work for its uniformity, abrasion resistance and dimensional stability. It is resistant to most organics, inorganic acids, alkalis and salts in a variety of concentrations except to hydrofluoric acid and acid fluorides. The permeable units have the lowest resistance.

(c) Advantages as a refractory

The chief causes of failure of refractories in boiler furnaces are slag adhesion, erosion and failure of structure, dependent on the type of coals and feeds used. The use of bonded silicon carbide brick is preferable depending on the desired usage. Clinker trouble is eliminated by use of these bricks in furnaces using all kinds of present-day stoker equipment. The problems like, the failures due to chemical reaction between iron and silicon carbide, and torch action on a wall produced by a blast of flame under pressure together with medium amounts of iron in the ash can be minimized. Installations of air-cooled silicon-carbide blocks have also been taken up.

(d) Properties

The properties of silicon carbide are that it is a refractory material (high melting point), it has excellent thermal conductivity and low thermal expansion, consequently it displays good thermal shock resistance. In addition, the high hardness, corrosion resistance and stiffness lead to a wide range of applications where wear and corrosion resistance are primary performance requirements. Silicon carbide possesses interesting electrical properties due to its semiconductor characteristics, the resistance of different compositions varying by as much as seven orders of magnitude.

(e) Commercial grades of silicon carbide

Commercial silicon carbide products for engineering applications are commonly produced in three forms,

- sintered silicon carbide (SSC),
- nitride-bonded silicon carbide (NBSC) and
- reaction-bonded silicon carbide (RBSC).

However, there are also several other types such as clay bonded silicon carbide and sialon-bonded silicon carbide. The former is usually used for refractory applications, and many variations exist, depending on the manufacturer.

17.9 Other Refractory Materials

17.9.1 Alumina-Zirconia

Zirconia-toughened alumina (ZTA) and other zirconia-alumina ceramics are often used in wear applications as an intermediate solution between alumina and zirconia. ZTA offers increased fracture toughness over alumina at a lower cost compared to pure or high zirconia ceramics. Depending on its purity and density, alumina is used for refractory tubes, industrial crucibles, analytical labware,wear components, refractory cements and abrasives.

17.9.2 Aluminium Silicate (Sillimanite)

Sillimanite, lava, fibrolite, and aluminium silicate (Al_2SiO_5) are compounds of silicon, aluminium and oxygen. Sillimanite is also a naturally occurring mineral. Lava is machinable in the unfired state. After firing lava develops higher hardness and can be used up to 1150°C.

17.9.3 Bauxite

Bauxite is an ore which principally consists of either Boehmite (a monohydrate of alumina, $Al_2O_3.H_2O$) or Gibbsite (alumina trihydrate, $Al_2O_3.3H_2O$). Raw bauxite may contain other impurities such as titania, silica and ferrous oxide. Refractory grade bauxite has a high alumina and low iron content. Most of the material mined in Europe has a higher proportion of Boehmite whereas bauxites mined in Asia and South America have high proportions of Gibbsite. The Bayer is used to process bauxite to produce higher grades of alumina (which can be used in higher temperature applications).

Bauxite used in refractory applications is generally calcined in a rotary kiln producing a material mainly consisting of corundum (alpha-Al_2O_3), mullite ($3Al_2O_3.2SiO_2$) and a small amount of a glassy phase.

17.9.4 Calcined and Sintered Alumina

Calcined aluminas are produced from bauxite processed via the Bayer process. The resulting material has very low levels of impurities. Calcined alumina is made from heating bayerite ($Al(OH)_3$) in a rotary kiln. Calcined aluminas are stable to very high temperatures.

Sintered alumina is manufactured by sintering calcined alumina at 1800°C in a rotary kiln. Subsequently, it is crushed and classified according to grain size.

17.9.5 Calcium Aluminate

Calcium aluminate ($CaAlO_3$) refractories are usually derived from calcium aluminate, calcium or alumina-bearing minerals. Calcium aluminate is used in refractory cements and shapes as well as synthetic slag additions for metallurgical operations.

17.9.6 Cordierite

Cordierite ($2MgO \cdot 2Al_2O_3 \cdot 5SiO_2$) or cordierite porcelain is a magnesium-aluminium silicate produced by fusing a mixture of talc, clay and aluminium oxide. Cordierite and cordierite mineral precursors are also known as magnesium-aluminosilicate, dichroite and iolite. Cordierite has a low coefficient of thermal expansion, high mechanical strength and low dielectric loss. Cordierite is commonly fabricated into an insulator or insulating substrate due to its good dielectric properties.

Cordierite has excellent thermal shock resistance. It can withstand red heat to ice water quenching and then be returned to red heat. High-fire cordierite body will withstand a temperature increase from 70°C to 1800°C in 80 seconds, followed by an immediate room temperature air quench.

17.9.7 Forsterite

Forsterite is a stoichiometric magnesium orthosilicate (Mg_2SiO_4) utilized for applications requiring a high coefficient of thermal expansion. Forsterite has desirable electrical insulation properties and it is used as a layer on transformer steel sheets. The layer is formed by the reaction of magnesium oxide with the silicon additions of the steel during annealing. Forsterite is also used in bulk form to fabricate insulators.

17.9.8 Mullite

Mullite ($3Al_2O_3$-$2SiO_2$ or $Al_6Si_2O_{13}$) is a compound of aluminium, silicon and oxygen. Mullite can also be viewed as a phase in the alumina-silica binary system. Mullite is a synthetic, fused or calcined crystalline aluminium silicate produced in electric arc furnaces from alumina and silica. Mullite usually has an off-white or tan colour. Depending on the purity and density, mullite can have superior dielectric and thermal shock properties and resistance to slag and silicate refractory bonds. Mullite is used for refractory tubes, industrial crucibles, analytical labware, dielectric substrates, wear components and in refractory cements. Refractory grade mullite or alumina-mullite mixtures are often derived by calcining Kyanite minerals.

17.9.9 Spinel

Spinels which are used in refractories are generally synthesised from bivalent and trivalent oxides, mixed at equal mole ratios forming materials of the general formula AB_2O_4. A typical refractory spinel product is the double oxide of magnesia and alumina ($MgAl_2O_4$). Further examples of spinels are provided in Table 17.7.

Aluminate	Chromite	Ferrite
$MgAl_2O_4$	$MgCr_2O_4$	$MgFe_2O_4$
$FeAl_2O_4$	$FeCr_2O_4$	$FeFe_2O_4$
$MnAl_2O_4$	$MnCr_2O_4$	$MnFe_2O_4$
$ZnAl_2O_4$	$ZnCr_2O_4$	$ZnFe_2O_4$
$NiAl_2O_4$	$NiCr_2O_4$	$NiFe_2O_4$

Table 17.7

Examples of spinels

Both spinel ($MgAl_2O_4$) and dichromite ($MgCr_2O_4$) are used in refractory castable formulations. Their respective melting points are 2135°C and 2350°C. Spinel is more neutral at high temperature than alumina and its corrosion resistance against basic slags is high. Against Fe_2O_3 the Al^{3+} may be substituted by Fe^{3+} causing corrosion. Spinel's thermal conductivity and coefficient of thermal expansion is smaller than that of magnesia and it has good resistance to spalling.

17.9.10 Quartz/Fused Silica

It is a ceramic oxide a silicate system, such as fused silica, calcium silicate ($CaSiO_3$), aluminium silicate, magnesium silicate, aluminium silicate or quartz. Fused silica is a compound of silicon and oxygen. High purity amorphous fused silica is a high performance ceramic with very low expansion, remarkable thermal shock resistance, low thermal conductivity, excellent electrical insulation up to 1000°C and excellent resistance to corrosion from molten metal and glass.

Fused silica is a compound of silicon and oxygen. High purity amorphous fused silica is a high performance ceramic with very low expansion, remarkable thermal shock resistance, low thermal conductivity, excellent electrical insulation up to 1000°C and excellent resistance to corrosion from molten metal and glass. Quartz is typically a high purity form of silica. Quartz is found in a mined mineral formed as well as man-made fused quartz forms. Fused quartz is a high purity, crystalline form of silica used in specialized applications such as semiconductor wafer boats, furnace tubes, bell jars or quartz ware, silicon melt crucibles, high performance lamps such as mercury and quartz halogen, ultraviolet (UV) lamps, thermocouples protector, waveguide handles, analytical laboratory ware and other high temperature products. Single crystal quartz is also available for piezoelectric applications.

Review Questions

1. What are refractories?
2. What are the three main forms of refractories?
3. How are refractories classified? Give examples in each case refractory.
4. Enumerate the properties of refractories.
5. What are the processes adopted for the manufacture of refractory materials?
6. Enumerate the properties and applications silica as an acid refractory material.
7. Write a note on the applications of silica as a refractory material.
8. How other refractory materials can be produced from silica?
9. What is silica brick? Give its composition and applications.
10. What is fireclay? Give its properties and applications.
11. Give the manufacturing process of fireclay.
12. What is a basic refractory? Give examples.
13. Enumerate the properties of Alumina as a basic refractory material.
14. How alumina is extracted and purified by Bayer's process?
15. Give the wide range applications of Alumina refractories.
16. What is a magnesia refractory?
17. What are types of magnesia refractories?
18. Mention the properties and uses of magnesia refractory material.
19. Write a note on fused magnesia refractory material.
20. Give the production process of fused magnesia.
21. What are the applications of fused magnesia as refractory material?
22. Write a note on magnesia clinker.
23. What is lime? Give its application as a refractory material.
24. Write a note on lime stone.
25. A brief note on zirconia as a refractory material.
26. What are neutral refractories? Give few examples.
27. How silicon carbide refractory is manufactured?
28. Discuss the applications and advantages of SiC refractory material.
29. Write notes on the following: (i) Alumina-Zirconia, (ii) Aluminium silicate, and (iii) Bauxite.
30. Give a brief account of the following: (i) Calcium aluminate, (ii) Cordierite, and (iii) Mullite.
31. Write a note on spinel refractories.
32. What is quartz and how is it obtained as refractory material?

18

Abrasives

Chapter Outline

Introduction. What is an abrasive? Types of abrasives, abrasive requirements. How abrasives work, abrasive sizes. Classification of abrasives, natural abrasives, synthetic abrasives, types of synthetic abrasives, chemical composition of abrasives, sample preparation. Natural mineral abrasives, corundum, quartz, diamond, synthetic abrasives, silicon carbide, boron nitride, boron carbide, coated or bonded abrasives, emery abrasive powder. Types of sand papers. Sand paper classifications. Other abrasives products, different abrasive applications and processes. Abrasive cutting tools, raw materials, manufacturing process.

18.1 Introduction

Often, few naturally occurring minerals find extensive application as abrasive materials which are used to shape or finish a workpiece through rubbing. A fine finished surface of a material can be obtained by polishing it to gain a smooth, reflective surface; and it can also involve roughening as in satin, matte or beaded finishes.

Many abrasives are known and used very extensively in a wide variety of industrial, domestic and technological applications. The various applications of abrasive material give scope for the selection and manufacture of a wide variety of abrasives with large variation in the physical, chemical compositions and as well as the shape. The following common processes are adopted for fine finish which includes grinding, polishing, buffing, honing, cutting, drilling, shaping and sanding. Diamond files are a form of coated abrasive (as they are metal rods coated with diamond powder).

Abrasives are hard, having sharp edges that cause unevenness onto the surface of material called abrasion. Abrasions may arise following strong contact of abrasive with surfaces such as concrete, stone, wood, carpet and roads, though these surfaces are not intended for use as abrasives.

Abrasive materials are exclusively used to smooth or roughen another material through extensive rubbing. The main types of abrasives are as follows:

- Natural mineral abrasives
- Bonded abrasives
- Coated abrasives and
- Grains or 'loose' abrasives.

Quartz, alumina, flint, etc. are the common types of mineral abrasive. Bonded abrasives are abrasive grains that are bound together by a bonding agent. Some examples of bonded abrasives are conventional grinding wheels, segments, mounted points and honing stones. Coated abrasives are abrasives that are bonded to some type of paper or cloth. Some examples of coated abrasives are sandpaper belts, sheets and discs. Loose abrasives are loose grains that are used for tumbling and blasting. There are several different grains used in abrasives. Among the man-made grains are aluminium oxide, silicon carbide, zirconia aluminium and ceramic aluminium oxide. Aluminium oxide is perhaps the toughest abrasive. It works well on high tensile materials such as carbon steel, alloy steel and all ferrous materials. Silicon carbide is superior in speed of cutting, especially at high pressure. It works best for nonferrous materials such as aluminium, brass, plastic or fibrous wood. Zirconia-aluminium is a long-lasting abrasive and offers faster grinding on materials such as cast iron or stainless steel. Ceramic aluminium oxide gives cooler cutting and longer life. It is appropriate for use with ferrous materials, carbon steel and exotic alloys. There are also several natural abrasives: flint, emery, garnet, crocus and chrome oxide.

18.2 What is an Abrasive?

The abrasive material is a substance used for abrading, smoothing or polishing.

Abrasives are very hard, brittle substances that have been crushed to create particles with cutting edges. Hardness and sharp edges give abrasives the ability to grind and polish.

The hardness helps to retain sharpness of edges of abrasives, but with sufficient use the edges of any abrasive become rounded over. One of the desired qualities of an abrasive is that it must be hard and strong.

Abrasives can be naturally occurring minerals, man-made materials or by-products of another process. *The specific characteristics of the abrasives include hardness, shape, size, colour, weight, chemical composition, availability, cost and environmental concerns.*

1. Hardness

When an abrasive material is rubbed on a substrate material, the process leaves an etch or pattern on it, then this speaks about the hardness of an abrasive. The hardness of an abrasive material is expressed in the Mohs' hardness scale. The scale ranges from 1 to 10 (Table 18.1) with 1 being the softest (talc) and 10 being the hardest (diamond). Most abrasives that effectively achieve an anchor pattern on a surface have a Mohs' hardness of at least 6.0.

Some applications require softer abrasives ranging in the 3.0–4.5 scale range. They will not etch steel or glass, but will generally remove foreign debris and provide a clean surface. They are used in areas where the substrate's surface needs not be removed, or where abrasive particle or residue could damage the substrate or surrounding area. Soft abrasives are often used around bearings and other easily damaged machinery.

Mineral	Hardness
Diamond	10
Corundum	9
Topaz	8
Quartz	7
Orthoclase	6
Apatite	5
Fluorite	4
Calcite	3
Gypsum	2

Table 18.1

Hardness of different minerals

Mohs' scale of hardness

Mohs' hardness is a measure of the relative hardness and resistance to scratching between minerals. Other hardness scales rely on the ability to create an indentation into the tested mineral. The scratch hardness is related to the breaking of the chemical bonds in the material, creation of microfractures on the surface or displacing atoms (in metals) of the mineral. Generally, minerals with covalent bonds are the hardest while minerals with ionic, metallic or van der Waals bonding are much softer.

When doing the tests of the minerals it is necessary to determine which mineral was scratched. The powder can be rubbed or blown off and surface scratches can usually be felt by running the fingernail over the surface. One can also get a relative feel for the hardness difference between two minerals. For instance, quartz will be able to scratch calcite with much greater ease than you can scratch calcite with fluorite. One must also use enough force to create the scratch.

2. Shape

The etch pattern on a substrate material or nature of abrasion on material depends on the shape of an abrasive which may be angular, sharp pointed edges, semi-round or spherical. It is noted that an angular abrasive has sharp edges, which ensure the quickest cleaning rate when removing tightly adhering material or contamination from the substrate. The sharp edges of a hard, angular abrasive produce steep peaks and valleys in the anchor profile, thus increasing the surface area and providing an excellent condition for the mechanical bonding of a coating. A blocky abrasive has mostly flat edges. This abrasive provides good cleaning rates in most applications, except those where hard to remove contaminants are present on the substrate. Semi-round and spherical abrasives produce a dimpled surface. Blocky abrasives are generally used for surface hardening or stress relieving of the substrate.

3. Size

It is a notable fact that the size of the abrasive has a considerable effect on abrasive action on the material to be abraded. The cleaning rate and etch pattern produced always depend on the abrasive grit size. Abrasives are generally available from 4 to 325 mesh. The smaller the mesh size the larger the abrasive particle. The abrasive materials obtained using 8–16 mesh are very coarse and are used for hard to clean surfaces or surfaces that require a deep anchor pattern. A material graded 20–40 mesh is considered a general-purpose abrasive. The polishing grade abrasive materials are obtained using 80–120 mesh which are used for fine polishing surfaces. Abrasives should be graded to a uniform size. Hard abrasives must be uniformly graded to provide a consistent anchor profile on the substrate.

4. Colour

Colour of an abrasive is not an important factor for selecting a particular abrasive action. However, sometimes, abrasives can leave residue, which can affect the surface appearance of a blasted surface. Darker abrasives tend to be less dusty than lighter-coloured abrasives because they do not reflect light as easily. This is especially important when blasting inside areas that require lighting. If blasting is performed in a tidy area where spent material cannot be completely removed, the colour of the abrasive may be important for aesthetic reasons as well.

5. Specific gravity

The cleaning rate and etching pattern profile of an abrasive depends on its specific gravity or density. Density is measured in pounds per cubic foot; specific gravity is measured by the density of the particle relative to the density of water. If the specific gravity is high, then the abrasive is heavier. When blasted at the same pressure, a heavy abrasive achieves a deeper anchor pattern than that from a lighter one. Abrasives with a high specific gravity tend to be less dusty. Those with a low-specific gravity impact the substrate with less force and are used for light cleaning, polishing and deburring.

6. Chemical composition

An appropriate compatible abrasive selection is required for any material to be abraded. When blasting a substrate, particles may be embedded or leave a residue, which could damage the protective coating. In cases where an iron abrasive is used to blast a stainless steel substrate, the embedded particles will form corrosion cells on the substrate.

7. Availability

Ensure always that the abrasive material desired is always available in sufficient quantity and to obtain optimum results, it is best not to change abrasives in the middle of a project.

8. Cost

The cost of the abrasive is a critical part of any job. Often, the cost of freight to transport the abrasive from manufacturer to job site can determine your choice. Also important is its friability or breakdown rate. The cost of an abrasive that has a high recycle rate can be reduced substantially by reusing it in a blast cabinet or room with a reclaim system. The blasting pressure, hardness, malleability and size of the abrasive determine its recycle rate.

9. Environmental concerns

The final consideration is environmental. Environmental concerns include the abrasive's respiratory effects on the blaster and other workers in the area. Employees must be provided with approved respiratory equipment and monitored throughout the job. When blasting in areas with poor ventilation, proper dust collection equipment must be provided to minimize worker exposure limits. The toxicity of junk being removed sometimes makes disposal of the abrasive/junk mix difficult. The effect of spent abrasive on the soil must also be considered. In confined areas, the toxicity of the abrasive may require the use of additional safety and engineering controls.

18.3 Types of Abrasives

The following are the commonly used abrasive materials:

- Flint: used for normal polishing job
- Emery: for polishing works
- Silicon carbide: generally used for nonferrous metals
- Aluminium oxide or alumina: the most widely used abrasive, generally used for ferrous alloys, high tensile materials and wood
- Diamond: most often used in ceramic grinding or final polishing due to high hardness and cost
- Cubic boron nitride (CBN)
- Garnet: usually used for machining of wood
- Zirconia/alumina alloys: suited to carbon and stainless steels and welds
- Glass: usually used for pressure blasting operations
- Colloidal silica: generally used for finishing operations as it is only available as a suspension of extremely fine particles

18.4 Abrasive Requirements

1. The grain must have a shape that presents several sharp cutting surfaces.
2. High hardness at room and elevated temperatures.

3. Controlled toughness or, rather, ease of fracture allows fracture to occur under imposed mechanical stresses.
4. Low adhesion to the workpiece.

18.5 How Abrasives Work

It can be noted that each hard abrasive particle acts like a single point cutting tool. The effect they produce is quite significant when a number of such particles available in a small area. Coarser grades/grits of abrasive are used where high volumes of material need to be removed, such as in coarse polishing, large scratch removal or operations requiring significant shape or dimensional change. Finer grades are generally used after coarser grades to produce a higher quality surface finish than are possible with coarse grades.

18.6 Abrasive Sizes

The size of the particles is important for abrasive action on a substrate material. Abrasive particles cut away the higher areas of a surface as they travel over it, but in doing so they leave behind scratches whose depth depends on the size of the particles. To remove these scratches, a smaller particle size must be used, and then a still smaller size to remove the scratches produced by the new particles, and so on, until the scratches are so small that they do not matter. Abrasives are, therefore, almost always sold by the size of the grit.

Table 18.2 presents the coarseness of various abrasive grit sizes.

The usual way of grading grains of abrasive is to pass them through a series of sieves with ever-smaller holes. The names of the abrasive grades are taken from this means of grading, although grades finer than about 240 are actually sorted by the speed with which they settle in a liquid or a current of air.

Grit Size	Relative Coarseness
8–24	Coarse
30–60	Medium
70–180	Fine
220–1200	Very fine

Table 18.2

The coarseness of various abrasive grit sizes

18.7 Classification of Abrasives

The materials used to make abrasives can be broadly classified as either natural or synthetic.

18.7.1 Natural Abrasives

Widely used *naturally occurring abrasives* include garnet, cerium oxide, flint, emery, corundum (aluminium oxide) and diamond. They occur in natural deposits and can be mined and processed for use with little alteration of size.

These materials may have varying characteristics and chemical compositions depending on the specific geological source. Manufactured versions of these materials are usually more consistent in chemical composition and other characteristics.

Examples of natural abrasives

1. Flint

Flint material is used widely because it is cheap. It is used when the sandpaper will quickly become clogged, for example, in removing paint.

2. Garnet

A better performing natural abrasive mineral is garnet. This abrasive is made from the garnet species almandine, which is typically dark red—thus the orange colour of the paper.

Many woodworkers prefer garnet paper for sanding raw wood because of its 'sharper' action than other papers. Garnet particles are not as strong as those of aluminium oxide or silicon carbide, and so they break more easily, exposing fresh edges more frequently.

3. Aluminium oxide

Aluminium oxide sand paper abrasive is used popularly for all general applications because it is long-lasting and can be used on wood or metal.

4. Silicon carbide

Sandpaper made with silicon carbide is black; wet-or-dry sandpaper is an example. Hard enough to sharpen tool steel, but mostly used for finish sanding.

5. Alumina-zirconia

A newly available combination of aluminium oxide with zirconia is very hard, very tough and expensive; used to reduce thickness rather than to smooth surfaces.

18.7.2 Synthetic Abrasives

Synthetic abrasives, on the other hand, are the product of considerable processing of raw materials or chemical precursors.

18.7.2.1 Types of Synthetic Abrasives

1. Bonded abrasives

The bonded abrasives are obtained by mixing abrasive grits of particular size with binders and got in useful shapes by pressing. They are fired to high temperatures to induce bonding via materials that forms a glassy matrix; hence the name verified bonded abrasives. Common shapes include wheels, cylinders, blocks and cones.

2. Coated abrasives

Coated abrasives are those which are made up of abrasive grits bonded to flexible substrates such as paper, cloth, fibre or film. Bonding is via resins, glues of combinations of the two. Examples include belts, sheets and flap disks.

A coated abrasive product is made up of three basic raw materials:

- an *abrasive mineral*,
- the *backing* onto which the abrasive is applied, and
- an *adhesive bond.*

(a) Abrasive mineral hybrids

Abrasive minerals can be used by themselves or combined with other minerals to form a hybrid. Merit coated products typically use silicon carbide, aluminium oxide, zirconia alumina or ceramic alumina abrasives. The abrasive minerals are available in different sizes, called *grit sizes*. The lower the grit size the larger (coarser) the product.

(b) Backing types

Backings are the base for the abrasive minerals, and combined with the adhesive bond, support and anchor the abrasive mineral. The backings used in the manufacture of coated abrasives are paper, cloth and fibre. Backings come in different weights, and heavier weight (thicker) backings are typically used for more aggressive operations.

(c) Bond types

An *adhesive bond* system is required to secure the abrasive mineral to the backing. All coated abrasive products are made using a two-stage bonding process. The first layer of bond applied to the backing is called the *make coat*, which provides the adhesive base between the abrasive mineral and the backing. The second coat is called the *size coat*, which is applied over the abrasive mineral and make coat to anchor the abrasive material and provide the desired physical strength of the finished product. Glue, urea resin and phenolic resin are the three basic bonding agents most commonly used. There are many size coat/make coat combinations.

3. Nonwoven nylon abrasives

The nonwoven nylon abrasives have three-dimensional network action and are obtained by coating abrasive grits on an open nylon material. They are versatile items and have wide applications and are available in forms such as pads, wheels and brushes.

4. Metal-bonded abrasives

Certain precision grinding tools are available in which most notably diamond is held together in a metal matrix. They may take the form of cutting wheels or hole cutting tools for machining of tiles, glass and other ceramic materials. In these items, the metal matrix is designed to wear away to expose the abrasive media, although ‘dressing’ may be required to remove build up of abraded materials from the grinding surface of the tool.

5. Burs

Burs are generally small in size used for fine machining and precision operations. They are similar to metal bonded and usually have teeth on the cutting faces. Abrasives often employed include tungsten carbide and diamond. Typically used for enlarging of holes.

6. Lapping compound

A very fine grit material is used in fine polishing and finishing operations. It can be supplied in dry powder or paste forms, which is known as lapping or polishing powder.

18.8 Chemical Composition of Abrasives

Abrasives can be distinguished by various properties such as their *hardness, colour*, *chemical composition*, *crystal shape* and *friability*. Chemical composition is considered as the primary means of distinguishing one type of abrasive from another.

(a) **Alumina based (aluminium oxide, Al_2O_3)(Mohs’ hardness scale: 9.0):** White fused aluminium oxide: High chemical purity of greater than 99 percent Al_2O_3, generally used for applications where high purity is important (medical, dental or other industrial uses); generally softer or more friable than other abrasives and used in grinding applications where a more friable product is desired.

(b) **White calcined aluminium oxide:** High purity aluminium oxide greater than 99.5 percent Al_2O_3, manufactured by growing individual crystals from seed crystals at high temperature. Crystals are hexagonal platelet shaped with an expected ratio of 5:1. Used in a variety of lapping, buffing and polishing applications, incorporated into bars and pads and used in ceramics.

(c) **Aluminium oxide with chrome:** White aluminium oxide fused with Cr_2O_3 to enhance grinding; a pink abrasive used in grinding applications requiring slightly more toughness than the white fused aluminiumoxide.

(d) **Brown fused aluminium oxide**: Abrasive with a content of 2–4 percent TiO_2 to further enhance toughness; the 'workhorse' of the industry; used in a wide variety of uses including bonded, coated, refractory and industrial markets; probably the most widely used abrasive.

(e) **Low titania brown fused aluminium oxide:** Abrasive with a content of 1–2 percent TiO_2 to enhance toughness of the grain; generally used in bonded or coated applications requiring an abrasive slightly tougher than white aluminiumoxide.

(f) **Zirconia-alumina**: The toughest among all the alumina-based products, used in bonded, coated and sandblasting applications requiring an extra tough abrasive.

(g) **Hydrated alumina:** Aluminium oxide with water chemically bonded to the alumina. Crystals are small (typically 1 mm or less in size) and very soft; typically used for fine polishing applications and as a raw material for ceramics.

(h) **Ceramic aluminium oxide:** A high-purity fine-grained alumina obtained from sintering dispersed colloidal alumina. This provides a tough product primarily used for precision grinding of steels and hard alloys.

(i) **Silicon carbide (SiC)**: Silicon carbide has a hardness of 9.3 on the Mohs' scale, which is near the hardness of diamond. Silicon carbide abrasive particles are angular and jagged in shape and have very sharp edges and corners. Because of these characteristics, silicon carbide is very effective grinding abrasive and is preferred to other abrasives for metallographic grinding of almost all types of metals.

18.9 Processes of Sample Preparation

- Grinding
- Grinding mediums
- Hand grindings
- Belt, disks and surface grinders
- Lapping
- Automatic grinders

1. Grinding

The preparation of a specimen depends mainly on grinding operation. During grinding, the operator has the opportunity of minimising mechanical surface damage that must be removed by subsequent polishing operations. Even if sectioning is done in a careless manner, resulting in severe surface damage, the damage can be eliminated by prolonged grinding. However, prolonged polishing will do little towards eliminating severe surface damage introduced by grinding. Grinding process of specimen is subjected to various sequences of operations using progressively finer abrasive grit. Grit sizes from 40 mesh through 150 mesh are usually regarded as coarse (1) abrasives and grit sizes from 180 mesh through 600 mesh as fine abrasives. Initially, grinding is done with coarse grit size that will establish an initial flat surface and remove the effects of sectioning within a few minutes. An abrasive grit size 150 or 180 mesh is coarse enough to use on specimen surfaces sectioned by an abrasive cut-off wheels. Hack-sawed, band sawed or other rough surfaces usually require abrasive grit sizes in the range 80–150 mesh. The abrasive used for each succeeding grinding operation should be one or two grit size smaller than that used in the preceding operation. A satisfactory grinding sequence might involve grit sizes of 180, 240, 400 and 600 mesh.

If wet grinding is desired for any specimen to be polished, then all grinding processes are to be carried out at wet conditions preferably with water which has no adverse effects on any constituents of the microstructure. Wet grinding minimises loading of the abrasive with metal removed from the specimen being prepared. Water flushes away most of the surface removal products before they become embedded between adjacent abrasive particles. Thus the sharp edges of the abrasive particle remain exposed to the surface of the specimen throughout the operation. If the sharp edges are unexposed the result is smearing of the abraded surface rather than removal of surface metal. The operator must determine, by examining the specimen throughout the sequence of grinding steps, that the abrasive is actually cutting and not merely smearing or burnishing. Burnishing results primarily from using an abrasive beyond its effective limit. Use of worn-out abrasives and dulled cutting edges is detrimental to good preparation.

A noticeable advantage of wet grinding is the cooling effect of the specimen material. Considerable frictional heat can develop at the surface of a specimen during grinding and can cause alterations of the true microstructure; for example, tempering of martensite in steel—that cannot be removed during polishing. Wet grinding provides effective control of overheating. The abraded surface of a specimen may become embedded with loose abrasive particles during grinding. These particles may persist in the surface and appear to be nonmetallic inclusions in the polished specimen. The flushing action of the water removes many of loose particles that might otherwise become embedded. Some laboratories have found that dressing the abrasive material with a solid wax lubricant recommended for grinding and other machining operations can minimise the embedding of abrasive particles.

The purpose of grinding is to lessen the depth of deformed metal to the point where the last vestiges of damage can be removed by series of polishing steps. The scratch depth and the depth of cold worked metal underneath the scratches decrease with decreasing particle size of abrasive. However, the depth of cold worked metal is roughly inversely proportional to the hardness of the specimen and may be 10–50 times the depth of penetration of the abrasive particle. It is imperative that each grinding step completely removes the deformed metal produced by the previous step. The operator usually can assume that this is accomplished if he or she grinds more than twice as long as the time required to remove the scratches incurred by the previous step. To ensure the complete elimination of the previous grinding scratches found by visual inspection, the direction of grinding must be changed 45–90 degrees between the successive grit sizes. In addition, microscopic examination of the various ground surfaces during the grinding sequence may be worthwhile in evaluating the effect of grinding. Each ground surface should have scratches that are clean-cut and uniform in size, with no evidence of previous grinding scratches.

Success in grinding depends in part on the pressure applied to the specimen. A very light pressure removes insufficient metal. Somewhat heavier pressure produce polishing, while still heavier pressure brings about the desired grinding action. Very heavy pressure results in nonuniform scratch size, deep gouges and embedded abrasive particles. Generally, a medium to moderately heavy pressure applied firmly gives the best results.

Most grinding of metallographic specimen is performed by manually holding the specimen with its surface against a grinding material. To establish and maintain a flat surface over the entire area being ground, the operator must apply equal pressure on both sides of the specimen and avoid any rocking motion that will produce a convex surface. If grinding operation is interrupted; the operator must re-establish contact with grinding material carefully in order to resume grinding in the plane already established.

Specimen should be cleaned after each grinding step to avoid any carryover of abrasive particles to the next step. Water solutions containing detergents are excellent cleaners and ultrasonic cleaning is an effective technique. Cleanliness of the operator's hands is as important as cleanliness of the specimen. Contamination of the grinding equipment by flying abrasive particles must be avoided.

2. Grinding mediums

Silicon carbide (SiC), aluminium oxide (Al_2O_3), emery (Al_2O_3–Fe_3O_4), diamond particles, etc. are the commonly used abrasives in the preparation of specimen. These are generally bonded to paper or cloth backing material of various weights in the form of sheets, disks and belts of various sizes. Grinding wheels consisting of abrasives embedded in a bonding material have limited use. The abrasive may be used also in powder form by charging the grinding surfaces with loose abrasive particles.

3. Hand grinding

A hand grinding setup can be provided by a flat surfaced hard material on which an abrasive sheet rests. The specimen is held by hand against the abrasive sheet as the operator moves specimen in rhythmic style away from and towards him in a straight line. High pressure should be applied on the forward stroke than on the return stroke. The wet grinding can be done by sloping the plate glass surface towards the operator and providing a copious flow of water over the abrasive sheet.

4. Belt, disk and surface grinders

The most common types of motor-driven grinding equipment are the belt grinder and the disk grinder. By using either, the metallographic specimen is held by hand against a moving, fixed-abrasive grinding material supported by a platen. Belt grinders and disk grinders may be used in either a horizontal or vertical position. Abrasive belts are generally cloth-backed for strength, and the popular belt sizes are 4 by 36 inches and 4 by 54 inches. Although cloth-backed disks are available, paper-backed disks are preferred for disk grinding of metallographic specimen, because paper-backed disks rest flat against the platen whereas cloth-backed disks usually curl from the edge. Most metallographic grinding disks are 8 or 10 inch in diameter. Specimen can be belt or disk ground successfully through all the grinding sequences.

5. Lapping

A lapping process is similar to disk grinding. The lap (grinding surface) is a rotating disk whose working surface is charged with a small amount of hard abrasive material. Laps are made of wood, lead, nylon, paraffin, paper, leather, cast iron and laminated plastics. The abrasive charge may be pressed into lap material by means of a steel block, or the lap may be charged directly with a mixture of abrasive and distilled water during lapping. A groove in the form of a spiral is a direction counter to the lap rotation and is often cut in the surface of laps, particularly of lead and paraffin laps. The spiral groove aids retention of cooling water and abrasive.

6. Automatic grinding

Automatic grinding is done without hand assistance. All automatic grinding devices use lap surfaces on which paper-backed disks are placed, or abrasive powder is charged. The lap is either a rotating or a vibrating disk.

18.10 Natural Mineral Abrasives

18.10.1 Corundum: Aluminium Oxide (Al_2O_3)

(a) It is a natural occurring oxide of alumina (Al_2O_3) with 52.9 percent aluminium and 47.1 percent oxygen. Corundum is the second hardest natural mineral available on the Earth (Fig. 18.1). Its hardness in Mohs' scale is 9. Its abrasive property is quite appreciable. It is also a refractory material, the melting point being 2010°C and hence it is used in a sintered form in the manufacture of special refractory crucibles, rods and other materials.

Figure 18.1 *Corundum.*

Corundum is found in rocks containing a high percentage of alkalies, deficient in silica and excess of alumina. It is an original constituent of various igneous rocks. It may result from the metamorphism of high aluminous clay and is often found associated with and alusite, kyanite and sillimanite.

(b) **Industry specifications**

It is known that the industrial specifications for the industrial manufacture of superior grade abrasives demand that corundum should have the following qualities: bright, glassy lustre, splintery and devoid of cleavage plane. The minimum Al_2O_3 content in the mineral should be around 84–90 percent. The corundum is crushed, ground and screened to grain sizes varying from 100 to 300 mesh. The grains are utilised in the manufacture of grinding wheels for flour and rice mills and workshops. Grinding wheels are manufactured either by calcining a mixture of corundum, clay and feldspar or by using sodium silicate as a bonding agent. Finer grains are utilised for the preparation of grinding-pastes for the automobile industry. MgO and $MgCl_2$ are used as bonding agents. Corundum paper and cloth finished to belts, discs, rolls, sheets and other shapes are prepared by giving a suitable coating generally of sodium silicate.

(c) **Applications**

Corundum finds use in mortars, wire drawing dies, thread guides and gauge blocks because of its hardness. Quite often, gem varieties are used for pivot supporters in delicate scientific instruments as jewel bearing in watches. Corundum is also used in the manufacture of special sparking plugs, muffles, pyrometer tubes, rod and insulator for vacuum tubes of all kinds and crucibles.

18.10.2 Quartz: Silicon Dioxide, SiO_2

(i) The most abundant and common material available on the Earth is quartz (Fig. 18.2). It is found in almost every geological environment and also it is at least a component of almost every rock type. It has a hexagonal crystal structure and is made of trigonal crystallised silica. It is most varied in terms of varieties, colours and forms. There are two types of quartz materials: one is of macrocrystalline, which is individual crystal, and the other is microcrystalline or crypto-crystalline varieties, aggregates of crystals visible only under high magnification. The transparent variety tends to be macrocrystalline and the crypto-crystalline varieties are either translucent or mostly opaque.

Figure 18.2 *Quartz.*

(ii) **Properties**

- Colour: Clear (in pure form)
- Crystal habit: Six-sided prism ending in six-sided pyramid (typical)
- Crystal system: Trigonal
- Cleavage: None
- Fracture: Conchoidal
- Mohs' scale hardness: 7
- Lustre: Vitreous

(iii) **Uses of Quartz**

- Silica for glass
- Electrical components
- Optical lenses
- Abrasives
- Gemstones
- Ornamental stone
- Building stone

18.10.3 Diamond

FIGURE 18.3 *Diamond.*

(a) The hardest natural occurring material is diamond (Fig. 18.3). Diamond is a natural allotrope of carbon. The crystal bond structure of diamonds gives the stone its hardness and differentiates it from graphite which is the main allotrope of carbon. Increasing use is being made of diamond as a grinding medium as well as a polishing medium. Carefully sized diamond abrasive particles are available from 280 μm (about 60 mesh) to 0.25 μm in size. The coarser grades of diamond are used in the form of resin-bonded cloth-backed disks, metal bonded lapping surfaces and loose particles for charging of grinding surfaces. Diamond abrasives of all sizes are available as suspensions in oil-soluble and water-soluble paste vehicles known as diamond compounds. The extreme hardness (Mohs' 10) and sharp cutting edges of diamond particles impart at a high cutting rate to diamond abrasives. Diamond abrasives are particularly suitable for grinding the harder alloys and refractory materials.

As a gemstone, diamond's single flaw (perfect cleavage) is far outdistanced by the sum of its positive qualities. It has a broad colour range, high refraction, high dispersion or fire, very low reactivity to chemicals, rarity and of course, extreme hardness and durability. In terms of its physical properties, diamond is the ultimate mineral in several ways:

Hardness: The hardness of diamond is a perfect '10' on the Mohs' hardness scale.

Clarity: Diamond is transparent over a larger range of wavelengths than in any other solid or liquid substance.

Thermal conductivity: Diamond conducts heat better than anything: five times better than the second best element, silver.

Melting point: diamond has the highest melting point (3820°K)

Lattice density: The atoms of diamond are packed closer together than the atoms of any other substance

(b) **Physical characteristics**

- Colour is variable and tends towards pale yellow, brown, grey and also white, blue, black, reddish, greenish and colourless.
- Lustre is adamantine to waxy.
- Transparency crystals are transparent to translucent in rough crystals.
- Crystal habits include isometric forms such as cubes and octahedrons, twinning is also seen.
- Hardness is 10
- Specific gravity is 3.5
- Cleavage is perfect in four directions forming octahedrons.
- Streak is white.
- Other characteristics: refractive index is 2.4 (very high), dispersion is 0.044, fluorescent.
- Best field indicator is *extreme* hardness.

18.11 Synthetic Abrasives

18.11.1 Silicon Carbide

Silicon carbide is the only chemical compound of carbon and silicon. Silicon carbide was initially produced by a high temperature electro-chemical reaction of sand and carbon.

The hardness of SiC is 9.5 on the Mohs' scale which is near the hardness of diamond. Silicon carbide abrasive particles are angular and jagged in shape and have very sharp edges and corners. Because of these characteristics, silicon carbide is very effective grinding abrasive and is preferred to other abrasives for metallographic grinding of almost all types of metal. Silicon carbide is an excellent abrasive and has been produced and made into grinding wheels and other abrasive products for over one hundred years. Recently, the material has been developed into a high quality technical grade ceramic with very good mechanical properties. It is used in abrasives, refractories, ceramics and numerous high performance applications. The material can also be made an electrical conductor and has applications in resistance heating, flame igniters and electronic components. Structural and wear applications are constantly developing.

Key properties

- Low density
- High strength
- Low thermal expansion
- High thermal conductivity
- High hardness
- High elastic modulus
- Excellent thermal shock resistance
- Superior chemical inertness

18.11.2 Boron Nitride

(i) *Boron nitride* is a synthetic material. Boron and nitrogen are neighbours of carbon in the Periodic table. In combination, boron and nitrogen have the same number of outer shell electrons; the atomic radii of boron and nitrogen are similar to that of carbon. It is not surprising therefore that boron nitride and carbon exhibit similarity in their crystal structure.

In the same way, carbon exists as graphite and diamond, boron nitride can be synthesised in hexagonal and cubic forms.

By ammonoloysis or nitridation process of boric acid at elevated temperature, boron nitride can be obtained. Cubic boron nitride is formed at high pressure and high temperature treatment of hexagonal BN.

1. Hexagonal boron nitride (h-BN) is equivalent in structure of graphite. Like graphite, its plate-like microstructure and layered lattice structure give it good lubricating properties. h-BN is resistant to sintering and is usually formed by hot pressing.
2. Cubic boron nitride (C-BN) has the same structure as diamond and its properties mirror those of diamond. Indeed, C-BN is the second hardest material next to diamond. C-BN was first synthesised in 1957, but it is only in the last 15 years that commercial production of C-BN has developed.

(ii) **Properties**

(a) **Hexagonal boron nitride (h-BN)**

- The lubricating properties of h-BN are excellent
- h-BN is readily machinable in the hot pressed state using conventional metal cutting techniques. Therefore, complex shaped components can be fashioned from hot pressed billet.
- Surface oxidation can be prevented, h-BN is not wetted by most molten metals, glasses and salts and hence has a high resistance to chemical attack.
- High dielectric breakdown strength
- High volume resistance
- Good chemical inertness

(b) **Cubic boron nitride (C-BN)**

C-BN is the second hardest material known.

- High thermal conductivity
- Excellent wear resistance
- Good chemical inertness

Typical properties for hexagonal and cubic boron nitride are given in Table 18.3.

Property	h-BN	C-BN
Density (g.cm^{-3})	2.3	2.2
Melting point (°C)	3000 (dissociates)	
Hardness (Knoop 100g) (kg.mm^{-1})	400	
Modulus of rupture (MPa)	100 (ll to press dir) 50 (⊥ to press dir)	110
Young's modulus (MPa)	20–103	
Thermal expansion Coeff. (RT-1000°C × 10^{-6})(°C^{-1})	1 (ll to press dir) 4 (⊥ to press dir)	3.8
Thermal conductivity (W/mK)	20 (ll to press dir) 27 (⊥ to press dir)	
Dielectric constant		4.1

Table 18.3

Properties of hexagonal and cubic boron nitride

(iii) **Applications**

1. **Electrical insulators**
   The combination of high dielectric breakdown strength and volume resistivity leads to h-BN being used as an electrical insulator; however, its tendency to oxidize at high temperatures often restricts its use to vacuum and inert atmosphere operation.
2. **Crucibles and reaction vessels**
   Its chemical inertness leads to application as thermocouple protection sheaths, crucibles and linings for reaction vessels though as above oxidation must be avoided.
3. **Moulds and evaporating boats**
   h-BN is used in a bulk form or as a coating for refractory moulds used in glass forming and in super plastic forming of titanium. It is also used as a constituent in composite materials, e.g. TiB_2/BN composites for metal evaporation boats and Si_3N_4/BN for break rings in continuous casting of steel.
4. **Hot isostatic pressing**
   Its refractoriness combined with the fact that it is not wetted by molten glass leads to h-BN being used in the production of hot iso-statically pressed (HIP'ed) material, most notable ceramics. In this application preformed parts are coated in h-BN prior to glass encapsulation and HIP'ing. This protects the part being HIP'ed from actually coming into contact with the glass, which in turn makes it easier to remove after HIP'ing.

5. **Machine cutting tools and abrasives**
   Cutting tools and abrasive components particularly for use with low carbon ferrous metals have been developed using C-BN. In this application, the tools behave in a similar manner to polycrystalline diamond tools but can be used on iron and low carbon alloys without risk of reaction.
6. **Substrates for electronic devices**
   C-BN is used for substrates for mounting high density and high power electronic components where the high thermal conductivity achieved allows efficient heat dissipation.
7. **Wear resistant coatings**
   Due to its high-hardness and excellent wear-resistant properties, coatings of C-BN have been developed.

18.11.3 Boron Carbide

(a) Occasionally, boron carbide is also used as an abrasive material for grinding specimen, which has a hardness of nearly 10 on the Mohs' scale. Boron carbide is used primarily for grinding ceramic and other extremely hard materials. It is one of the hardest man-made materials available. Above 1300°C, it is even harder than diamond and cubic boron nitride. It has a four-point flexural strength of 50,000–70,000 psi and a compressive strength of 414,000 psi, depending on density. Boron carbide also has a low thermal conductivity (29–67 W/mK) and has electrical resistivity ranging from 0.1 to 10 Ω cm. With its strength to weight ratio and low thermal conductivity, it is ideal for a wide variety of applications.

(b) **Applications**
- Aerospace
- Armour
- Blasting nozzles
- Ceramic composites
- Honing stones
- Lapping
- Metal matrix composites
- Nuclear technology
- Refractory shapes, cements
- Tool and Die fabrication
- Ultrasonic machining

Physical and chemical compositions are given as follows:

1. Atomic properties

- Crystal structure: Rhombohedral
- Molecular weight: 55.3 gm/mol
- Oxidation temperature: 500°C
- Particle shape: Blocky, angular
- Colour: Black, sooty when powdered

2. Mechanical properties

- Mohs' hardness: 9.3+
- Bending strength: 30.50 kg/mm^{-2}
- Compressive strength: 200.300 kg/mm^{-2}

3. Thermal and electrical properties

- Melting point: 2450°C
- Thermal expansion: 5×10^{-6}/°C
- Thermal conductivity 25°: 07.0.16 cal/cm sec °C
- Electrical resistivity 25°: 01.10 ω cm

18.12 Coated or Bonded Abrasives

18.12.1 Emery Abrasive Powder

Composition: Al_2O_3 + an iron bearing mineral + trace impurities such as mullite, titania, silica and magnesia (Al_2O_3 = 58.0 percent, Fe_2O_3 = 24.0 percent SiO_2 = 4.0 percent TiO_2 = 3.0 percent, CaO = 2.0 percent).

(i) Emery abrasive powder

A natural occurring odourless mineral having its broadest is used as an abrasive in polishing applications. The hardness of emery is 8.0 on the Mohs' scale. *Emery* is an impure, fine-grained variety of natural corundum containing 25–45 percent of iron oxide. Emery abrasive particles have much smoother surfaces than either silicon carbide or aluminium oxide abrasive particles. For this reason, emery particles do not feel as coarse as silicon carbide or aluminium oxide particles of equivalent grit size and the cutting rate of emery is inferior to that of either of the two other abrasives. It has an intimate mixture of corundum and magnetite. The higher amount of impurities and the weaker internal structure than naturally occurring minerals have restricted its growth in new applications.

(ii) Grain Sizes available: 72, 80, 92, 100, 120, 150, 152, 180, 200, 220 and 240

Emery abrasive powder properties

- Mohs' hardness: 8
- Colour: black or dark grey
- Specific gravity: 3.5–3.8
- Particle shape: blocky with sharp edges
- Crystallography: polycrystalline

(iii) Typical applications

It is widely used for polishing applications, barrel finishing, pressure blasting, coated abrasives, tumbling and deburring media, lens grinding and lapidary work.

18.12.2 Types of Sandpapers

A wide variety of *sheet abrasives* are generally known as sandpaper. Sandpaper is an abrasive, but not all abrasives are sandpaper. For example, there is steel wool, glass, pumice, rottenstone—each product manufactured or innovated for a special smoothing job on wood, metal, plastic, glass, ceramic and even concrete. Flint, garnet, emery, aluminium oxide and silicon carbide are most common types of abrasives materials. The abrasive particles may be mounted on paper or cloth, in 'open-coat' or 'closed-coat' density.

1. Flint

Flint is compact, hard, brittle mineral, brown, black, or grey in colour, found as nodules in limestone or shale deposits. It consists of crypto-crystalline silica, SiO_2, principally in the crystalline form of quartz. Its hardness is 7 on the Mohs' scale. Because of this property, Mohs' scale flint splinters are used for abrasive purposes and, when ground into powder, added to clay during pottery manufacture. Flints have been used for making fire by striking the flint against steel, which produces a spark, and for discharging guns. Flints in cigarette lighters are made from cerium alloy.

This is the cheapest sandpaper. Flint is a grey mineral. Flint paper wears down quickly and is used mainly in removing paint finishes. Because paint will clog the grit on the paper and reduce its usefulness before the paper is worn smooth, pieces have to be discarded frequently. Use an open-coat when removing finishes.

The open coat does not clog or 'fill up' as quickly, and, therefore, you get more sanding action. Once the finish is off, one can use closed-coat flint paper for smoothing purposes, but it is recommended that for fine finishes another abrasive may be used.

2. Garnet

One of the advantages of the garnet is that the grit on garnet sandpaper is much harder than on flint paper and is more suitable for use in wood-working. The cost is slightly more than for flint paper, but the cost is not prohibitive.

3. Emery

This abrasive is recognised by its distinctive black colour. It is widely used as a metal abrasive, although more recently developed abrasives may be more effective. It is a low cost emery sandpaper.

4. Aluminium oxide

This probably is the most popular sandpaper abrasive for fine wood-working and furniture finishing. It is also the best choice for power sanding. The oxide is reddish in colour, and its grit is very sharp and much harder than flint, garnet and emery. While it is popular for use on wood, it also can be used on metal surfaces. Aluminium oxide is more expensive than other abrasive paper, but it lasts so long that it usually is cheaper to use in the long haul.

5. Silicon carbide

This is the hardest, sharpest sandpaper of all. The bluish-black material cuts extremely well and is commonly used for such tough jobs as finishing metal or glass, and for sanding hardwood floors before finishing/re-finishing.

18.12.3 Sandpaper Classifications

Apart from the type of material that gives sandpaper its abrasive quality, there are other characteristics that affect the performance of abrasives.

There is a paper for any job from the roughest stripping to the most delicate finishing. The following are the basic classifications of sandpaper.

(i) Grit

The particles that are glued to the backing material are identified by the number of the labelling systems. There is a retail system that carries word descriptions (coarse, medium, fine). The old system identified papers by numbers that run from 10/0 to 4 to reflect increasing coarseness. The currently used industrial system identifies grit with numbers from 600 down through 16. In the industrial system, the grit specified is coarser as the numbers become smaller.

(ii) Density

The density of the grit (that is, how close together the granules are on the backing) material is another factor that affects the way sandpaper functions.

There are two classifications as follows:

- Closed-coat and
- Open-coat

Closed-coat indicates that the grit material blankets 100 percent of the surface while open-coat indicates that the grit covers from 50 to 70 percent of the surface. Open-coat may not look very 'open' especially when the abrasive is a fine density. The backing looks 'full' of grit. As a rule, closed-coat sandpapers are designed for fine finishing. Closed-coat tends to clog quickly. Wood particles are caught and fill the spaces between the pieces of grit. Because open-coat sandpapers do not clog as easily they are a better choice when you do a first sanding to remove a lot of material from the wood. The usable life and cutting action of any sandpaper can be extended and improved by rapping the paper on a hard surface from time to time to dislodge wood particles or by cleaning the paper with an old toothbrush or stiff scrub brush.

(iii) Adhesives

The grit is held to the backing on sandpaper by any one of a variety of glues. Hide glue is used on sandpapers intended for light to medium work. Hide glue is not waterproof, so the paper cannot be used on a wet surface.

Thermosetting resin is used to secure grit on papers where the work is harder, such as floor sanding. Waterproof resin bonds grit to backings that are waterproof. Such glue allows the paper to be used with oil or water for extra fine and smooth finishing purposes.

(iv) Weight and backing

The backing of the sandpaper may range from paper or cloth to a combination of cloth and paper or even a plastic material. The paper comes in various weights, which lend certain advantages or disadvantages.

'A' weight paper is the lightest of all and the first to wear through. Use it only for lightly touching up wood prior to finishing it. Moving up the scale, 'C' weight is stiffer and stronger than 'A' weight and it is used for coarse machine sanding. 'D' weight is even stronger and is for heavy machine sanding. 'E' weight is to be used with floor sanding machines or belt sanders. In cloth backings, the weights are 'J' weight and 'X' weight. 'J' is the lighter of the two and is used on curves and other shapes. The 'X' weight cloth is designed for coarse grit abrasives used with a belt sander. Fibre weight backings are very heavy and are usually not needed. Various abrasives papers, their grades, grit sizes, description and uses are listed in Table 18.4.

Abrasive papers			
Grade (industrial system)	**Grit (old system)**	**Description (retail system)**	**Use**
16	4	very coarse	Very rough work, unplanned wood, initial machine sanding of floors
20	3½	very coarse	
24	3	very coarse	
30	2½	coarse	Initial sanding when necessary
36	2	coarse	
40	1½	coarse	
50	1	coarse	
60	½	medium	Intermediate sanding, especially of softwoods
80	1/0	medium	
100	2/0	medium	
120	3/0	fine	Preparatory sanding of hardwoods; final smoothing, especially of softwoods
150	4/0	fine	
180	5/0	fine	

(*Contd.*)

(*Contd.*)

Abrasive papers			
Grade (industrial system)	**Grit (old system)**	**Description (retail system)**	**Use**
220	6/0	very fine	Final sanding; sanding between finish coats
240	7/0	very fine	
280	8/0	very fine	
320	9/0	extra fine	Sanding between finish coats; smoothing the final coat of finish
360		extra fine	
400	10/0	extra fine	
500		superfine	Sanding metal, plastic, ceramics
600		superfine	

Table 18.4

Various abrasives papers, their grades, grit sizes, description and uses

18.12.4 Other Abrasive Products

There are two powdered stone abrasives commonly used in fine finishing. These abrasives are applied with oil. Depending on the situation, one can mix the powdered stone with lubricating oil and apply with a cloth, or you may dip the cloth in the oil and then the stone before applying it.

(a) Pumice

This is lava. The stone is porous and relatively soft. When pumice is crushed to a powder, the stone is a medium to fine abrasive.

(b) Rottenstone

This is decomposed limestone, a form of sedimentary rock. The material crushes to a very fine abrasive powder. Applied with oil, rottenstone is used to create a smooth finish to projects such as furniture. Either pumice or rottenstone may be used to create the glassy finish of a French polish.

(c) Steel wool abrasive

Steel wool is also used for smoothing wood. Like sandpaper, it comes in a variety of grades. Each has a number and name description.

Specially-packages steel wool for furniture finishing costs more than the regular stuff and you will know it by price. The reason for the cost is that the steel wool has been degreased.

18.13 Different Processes of Abrasive Application

1. Sanding

Fine-grained quartz material is used for the sanding process. Quartz sand is hard enough for wood working (Mohs' hardness 7) but not very tough or sharp. The virtue of sand sandpaper is its cheapness. Fine woodworkers do occasionally use flint sandpaper, though.

Flint is a rock made of microcrystalline quartz. It is no harder than quartz but it is tougher, so its sharp edges last longer. Garnet paper is still widely available. The garnet mineral is harder than quartz (Mohs' 7.5), however its real virtue is sharpness, giving it cutting power without scratching wood too deeply.

The best utilised sand paper is corundum. Extremely hard (Mohs' 9) and sharp, corundum is also usefully brittle, breaking into sharp fragments that keep on cutting. It's great for wood, metal, paint and plastic. All sanding products today use artificial corundum and aluminium oxide. If you find an old stash of emery cloth or paper, though that probably has the real mineral: emery is a natural mix of fine-grained corundum and magnetite.

2. Polishing

Three natural abrasives are commonly used for polishing and cleaning of metal, enamel finish, plastic and tile.

Pumice has very fine grain and is a stone, not a mineral. Its hardest mineral is quartz, so it has a gentler action than sanding abrasives.

Feldspar is still softer (Mohs 6), which is the most famously used in household cleaner.

Tripoli (also called rottenstone) is used for the most delicate polishing and cleaning work, such as jewellery and fine crafts. Tripoli is microscopic, microcrystalline quartz that is mined from beds of decomposed limestone.

3. Sandblasting

Applications of this industrial process range from scrubbing rust off of steel girders to inscribing gravestones, and there is a wide range of blasting abrasives in use today. Sand is one, of course, but airborne dust from crystalline silica which is a health hazard. Safer alternatives include olivine (Mohs' 6.5), garnet and staurolite (Mohs' 7.5). Many artificial sandblasting abrasives are in use, too, as well as exotic things like ground walnut shells and solid carbon dioxide.

4. Diamond grit

The hardest mineral of them all is diamond (Mohs' 10), and diamond abrasive is a large part of the world diamond market. Diamond paste is available in many grades for sharpening hand tools, and you can even buy nail files impregnated with diamond grit for the ultimate grooming aid. Diamond is best suited for cutting and grinding tools though, and the drilling industry uses lots of diamond for drill bits.

18.14 Abrasive Cutting Tools

Abrasive cutting tools are any number of tools which cut by rubbing a hard material quickly over the object to be cut. One of the most basic types of abrasive cutting tools is a *grinding wheel*. An abrasive material is hard and rough, and wears down whatever it is rubbed against. Abrasive cutting tools are embedded with an abrasive material.

Because abrasive cutting is efficient and relatively inexpensive, abrasive cutting tools come in a great variety. For any solid that needs to be cut, an abrasive cutting tool exists for the job.

The abrasives in abrasive cutting tools are usually very hard crystals. In order to work well, the abrasive cutting tools must use an abrasive that is harder than the material to be cut. Because diamond is the hardest substance on the Earth, it is used in abrasive cutting tools which cut or polish very hard substances.

Other abrasives used in abrasive cutting tools include aluminium oxide, called corundum in its crystal form, garnet and emery. Abrasive cutting tools use the friction of these abrasives to cut materials ranging from wood and metal to stone and gems. The softer abrasive cutting tools, such as an iron file, are used to cut softer materials like wood. Harder materials such as gems require abrasive cutting tools with very hard abrasives, like diamond.

Abrasive cutting tools can range from the very simple to the very complex. A very simple abrasive cutting tool is a metal file. The rough surface of the file can be used to cut or rasp wood or soft metals, but it requires a lot of physical work. Adding power to abrasive cutting tools gives us slightly more complicated cutting or grinding wheels. These abrasive cutting tools are often used to cut and polish stone, and work much faster than hand tools.

A more complex abrasive cutting tool is abrasive water jet cutters. These abrasive cutting tools are often computer programmed to cut complex shapes from metal sheets in manufacturing. They operate by shooting a high-powered stream of water at the metal. Inside the water are abrasive crystals, such as garnet, and the pressure of these abrasives can cut metal with incredible accuracy and speed. From the most simple roughened piece of metal to the most complex and accurate computer-driven machine, abrasive cutting tools find their place in almost every situation in which something hard needs to be cut.

18.14.1 Raw Materials

The grinding wheels are manufactured with two important components, *abrasive grains* and *bonding materials*. Often, additives are blended to create a wheel with the properties necessary to shape a particular material in the manner desired.

Abrasive grains constitute the central component of any grinding wheel, and the hardness and friability of the grinding materials will significantly affect the behaviour of a given wheel. *Friability refers to how easily a substance can be fractured or pulverised.* People who design grinding wheels consider the friability of their abrasives—which can differ with the nature of the materials being ground very carefully. For example, while diamond is the hardest known material, it is an undesirable steel abrasive because it undergoes a destructive chemical reaction during the cutting process; the same is true for silicon carbide. On the other hand, aluminium oxide cuts irons and steels better than diamond and silicon carbide, but it is less effective for cutting nonmetallic substances.

If selected correctly, an abrasive chosen to shape a particular substance will retain its friability when ground against that substance: because the grinding will cause the abrasive to continue fracturing along clean, sharp lines, it will maintain a sharp edge throughout the grinding process. This gives the grinding wheel a unique characteristic of being a tool that sharpens itself during use.

Although bonded abrasives began as tools made from natural minerals, modern products are made almost exclusively with *synthetic materials*. A bonding material holds the abrasive grits in place and allows open space between them. Manufacturers of grinding wheels assign hardness to the wheel, which should not be confused with the hardness of the abrasive grain. Bonds that allow abrasives grains to fracture easily are classified as *soil bonds.* Bonds that restrict the fracturing of the grains and allow a wheel to withstand large forces are classified as *hard bonds.* Generally, soil wheels cut easily, produce poor surface finishes, and have a short useful life. On the other hand, harder wheels last longer and produce finer surface finishes, but cut less well and produce more heat during grinding.

The bonding matrix in which the abrasive grains are fixed may include a variety of organic materials such as rubber, shellac or resin; inorganic materials such as clay are also used. Inorganic bonds with glass-like or vitreous structures are used on the tool-sharpening wheels for the home workshop grinder, while resin bonds are used in masonry or steel-cutting wheels. Generally, vitrified bonds are used with medium to fine grain sizes in wheels needed for precision work. Resin bonds are used generally with coarse grains and for heavy metal removal operations such as foundry work. In addition to their abrasive and bonding materials, grinding wheels often contain additional ingredients that produce pores within the wheel or assist chemically when a particular abrasive is used to grind a special material. One important aspect of a grinding wheel that can be created or altered through additives is *porosity*, which also contributes to the cutting characteristics of the grinding wheel. Porosity refers to the open spaces within the bond that allow room for small chips of metal and abrasive generated during the grinding process. Porosity also provides pathways that carry fluids used to control heat and improve the cutting characteristics of the abrasive grains. Without adequate porosity and spacing between abrasive grains, the wheel can become loaded with chips and cease to cut properly.

A variety of products are used as additives to create proper porosity and spacing. In the past, sawdust, crushed nut shells, and coke were used, but today materials that vapourise during the firing step of manufacturing (for example, naphthalene-wax) are preferred. Some grinding wheels receive additional materials that serve as aids to grinding. These include sulphur and chlorine compounds that inhibit microscopic welding of metal particles and generally improve metal-cutting properties.

18.14.2 The Manufacturing Process

Most grinding wheels are manufactured by the cold-press method, in which a mixture of components is pressed into shape at room temperature. The details of processes vary considerably depending upon the type of wheel and the practices of individual companies. For mass production of small wheels, many portions of the process are automated.

1. **Mixing the ingredients**

   Preparing the grinding wheel mixture begins with selecting precise quantities of abrasives, bond materials, and additives according to a specific formula. A binder, typically a water-based wetting agent in the case of vitrified wheels, is added to coat the abrasive grains; this coating improves the grains' adhesion to the binder. The binder also helps the grinding wheel retain its shape until the bond is solidified. Some manufacturers simply mix all materials in a single mixer. Others use separate steps to mix abrasive grains with binder.

   Wheel manufacturers often spend considerable effort to develop a satisfactory mixture. The blend must be free-flowing and distribute grain evenly throughout the structure of the grinding wheel to assure uniform cutting action and minimal vibration as the wheel rotates during use. This is particularly important for large wheels, which may be as big as several feet in diameter, or for wheels that have a shape other than the familiar flat disk.

2. **Moulding**

   For the most common type of wheel, an annular disc, a predetermined amount of grinding wheel mixture is poured into a mould consisting of four pieces: a circular pin the size of the finished wheel's arbour hole (its centre hole); a shell with a 1-inch (2.5-centimetre) wall, about twice as high as the desired grinding wheel is thick; and two flat,

   After moulding and final shaping, the wheel is fired in an oven or furnace. Firing melts the binder around the abrasives and converts it to a form that will resist the heat and solvents encountered during grinding. Finishing steps that follow firing may include reaming the arbour (centre) hole to the proper size, correcting the thickness of the wheel sides, balancing the wheel, and adding labels.

   Circular plates with diameter and arbour hole sizes equal to those of the wheel. A variety of methods are used to distribute the mixture evenly. Typically, a straight edge pivots about the centre arbour pin to spread the mixture throughout the mould.

3. Using pressures in the range of 100–5000 pounds per square inch (psi) for 10–30 seconds a hydraulic press then compact the mixture into the grinding wheel's final shape. Some manufacturers use gage blocks between the two face plates to limit their movement and establish uniform thickness. Others control wheel thickness by closely monitoring the consistency of the mix and the force of the press.

4. After the mould has been removed from the press and the wheel stripped from the mould, the wheel is placed on a flat, heatproof carrier. Final shaping of the wheel may take place at this time. All work at this stage has to be done very carefully because the wheel is held together by only the temporary binder. Lighter wheels can be lifted by hand at this stage; heavier ones may be lifted with a hoist or carefully slid on a carrier to be transported to the kiln.

5. **Firing**

   Generally, the purposes of the firing are to melt the binder around the abrasives and to convert it to a form that will resist the heat and solvents encountered during grinding. A wide range of furnaces and kilns are used to fire grinding wheels, and the temperatures vary widely depending upon the type of bond. Wheels with a resin bond are typically fired at a temperature of 300–400 degrees Fahrenheit (149–204 degrees Celsius), and wheels with vitrified bonds are fired to temperatures between 1700 and 2300 degrees Fahrenheit (927–1260 degrees Celsius).

6. **Finishing**

   After firing, wheels are moved to a finishing area, where arbour holes are reamed or cast to the specified size and the wheel circumference is made concentric with the centre. Steps may be necessary to correct thickness or parallelism of wheel sides, or to create special contours on the side or circumference of the wheel. Manufacturers also balance large wheels to reduce the vibration that will be generated when the wheel is spun on a grinding machine. Once wheels have received labels and other markings, they are ready for shipment to the consumer.

Review Questions

1. What is an abrasive?
2. What are the specific characteristics of abrasives?
3. Write a note on harness of an abrasive.
4. What are the types of abrasives? Give examples in each case.
5. Mention the requirements of an abrasive.
6. How do abrasives work?
7. Write a note on the sizes of abrasives.
8. How are abrasives classified?
9. Write a note on natural abrasives.
10. Write a note on synthetic abrasives.
11. A note on coated abrasives.
12. Write notes on (i) Abrasive hybrids (ii) Backing types of abrasive and (iii) Bond type of abrasive.
13. Give a brief account of the following:
    (a) Metal bond abrasives (b) burs and (c) Lapping of compound
14. Write a note on chemical composition and abrasives.
15. Write a note on sample preparations.
16. What is grinding process?
17. Write a note on grinding mediums and the role of abrasives.
18. What do you mean by hand grinding?
19. Write a brief note on Belts, disks and surface grinders.
20. What is lapping?
21. Write a note on corundum, mention its properties and applications as an abrasive.
22. What is quartz? Give its properties and uses.
23. Write a note on abrasive diamond. Mention its properties and uses.
24. Write a note on silicon carbide as an abrasive material.
25. Write about boron nitride as an abrasive material.
26. What are the applications of boron nitride abrasive?
27. Write a note on boron carbide. Give its applications.
28. What is an emery abrasive paper? Give its properties and uses.
29. Mention the types of sand papers. Give examples.
30. How are sand papers classified? Describe their characteristics.
31. Mention the different abrasive application processes.
32. What do you mean by sanding?
33. What is polishing?
34. What is sand blasting?
35. What is abrasive cutting tool? Describe how abrasive cutting tools are manufactured.
36. Describe the manufacturing process of grinding wheels.
37. What are adhesives? Explain the synthesis and properties of epoxy resin. [VTU, 2014]

19

Nanoscience and Technology

Chapter Outline

Introduction. Scope of nanotechnology. Nanomaterials—types of nanomaterials. Classification of nanomaterials on dimension—zero-dimensional nanomaterials, one-dimensional nanomaterials, 2D nanomaterials, three-dimensional nanomaterials. Nanomaterial categories—nanotools, nanodevice, nanostructured materials, nanocrystalline materials, carbon-based nanomaterials. Single nanostructured particles—carbon black/amorphous carbon/vitreous carbon, graphene, graphite, fullerene, carbon onion, nanocluster. Carbon nanotubes—single-wall CNT (SWCNT) and double-wall CNT (MWCNT). Carbon nanohorns (SWNHs).Carbon nanowires. Nanorods. Carbon nanofibre, films, coating, nanostructured surfaces, carbon-based nanofilms—DLC. Nanostructured bulk materials—nanoporous carbon, carbon nanofoam, carbon aerogel, bulk carbon nanocrystals, nanocrystalline diamond. Quantum dots, fabrication methods for quantum dots, applications of quantum dots. Organic nanoparticles. Dendrimers, methods of preparation of dendrimers, applications of dendrimers. Inorganic–organic hybrid nanoparticles—polyhedral silsesquioxanes. Nanointermediates, nanocomposite materials—the basic ingredients of nanocomposite, processing of nanocomposite—carbon nanotube–polymer composite, polymer matrix composite materials, applications of polymer matrix nanocomposite materials. Clay–polymer nanocomposites—structure of clay material, characteristics of clay material, fabrication methods, applications of clay–polymer nanocomposite. Metal matrix nanocomposite materials—cryomilling process of manufacture of nanocomposite, applications of MMC nanocomposite.

19.1 Introduction

Nanoscience is an emerging area of science which concerns itself with the study of materials whose size is 1–100 nanometres (nm) on the nanometre scale. One nanometre is billionth of a metre (10^{-9} m*)*. Comparative sizes of the bulk and nanomaterials are shown in Fig. 19.1. Bulk materials always exhibit macroscopic physical properties. The same material at the nanoscale can have properties (e.g. optical, mechanical, electrical, etc.) which are very different from the properties the material that has at the macro (bulk) scale. Nonintentionally made nanomaterials, which refers to as nanosized particles or materials that belong naturally to the environment (e.g. proteins, viruses, nanoparticles produced during volcanic eruptions, etc.) or that are produced by human activity without intention (such as nanoparticles produced from diesel combustion, chimney soot, etc.). Intentionally produced nanomaterials, which means nanomaterials produced deliberately through a defined fabrication process.

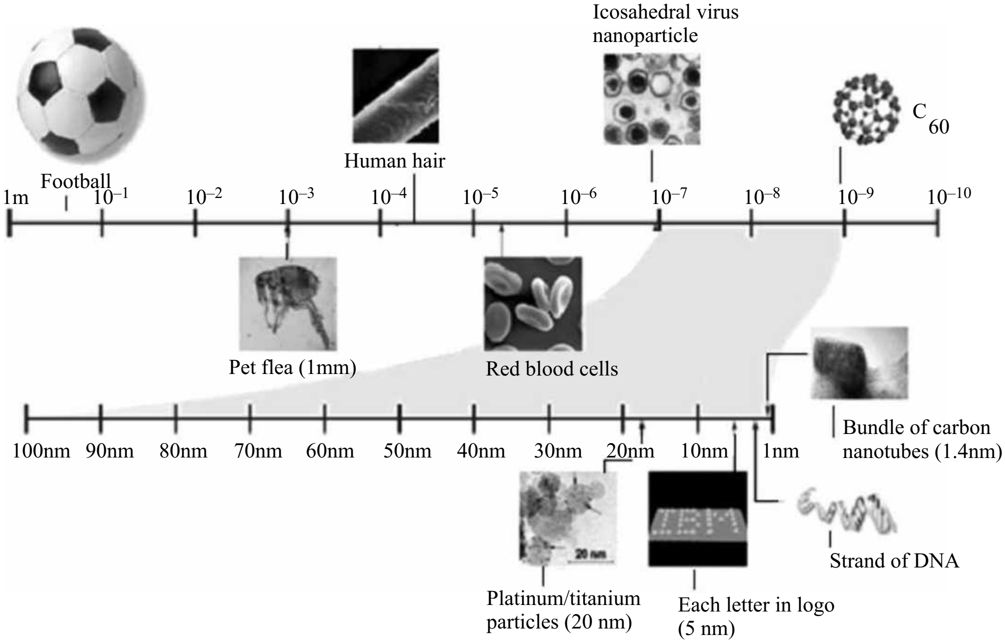


Figure 19.1 *Comparative sizes of bulk and nanosized materials.*

Nanotechnology is a rapidly emerging technology with vast potential to manipulate and control a substance at the nanometre level (1 nm = 10^{-9} m) and create new useful materials and devices with fascinating functions making the best use of the special physical properties of nanosized substances and improved products for numerous applications.

It is also concerned with materials and systems whose structures and components exhibit novel and significantly improved physical, chemical and biological properties, phenomena and processes, because of their small nanoscale size. Generally, nanoscience aims to understand the novel properties and phenomena of all nano-based entities.

Nanotechnology involves research and technology development at the 1–100 nm range. Nanotechnology creates and uses structures that have novel properties because of their small size. Nanotechnology builds on the ability to control or manipulate at the atomic scale.

Nanotechnology is being applied to almost every field imaginable, including electronics, magnetics, optics, information technology, materials development and biomedicine. Because of their small size, nanoscale devices can readily interact with biomolecules on both the surface and inside cells. However, there is concern that the unique properties of nanomaterials might pose substantial risks, which have been largely unexplored, to both human health and the environment. Moreover, understanding issues related to current technologies will serve to promote less harmful and/or safer alternative technologies for the future.

19.2 Scope of Nanotechnology

19.2.1 Nanotechnology in Electronics

- Nanomaterials with less than 100 nm in size are used in microprocessors in the electronic industry—smaller sizes allow faster processing times and also more processing power to be packed into a given area.
- Some of the areas under the nanotechnology development include:
  The qualities of screens are greatly improved by the use of nanomaterials in electronic device display. This is achieved by reducing power consumption due to decreasing the weight and thickness of the screens containing nanomaterials. This can be achieved using carbon nanotubes (CNTs). They can be used as field emitters with extremely high efficiency for field emission displays (FED).
- The density of memory chips can be increased by nanomaterials. Researchers are developing a type of memory chip with a projected density of one terabyte of memory per square inch or greater. Integrated nanosensors are used for collecting, processing and communicating massive amounts of data with minimal size, weight and power consumption.
- The transistors size in ICs can be reduced reasonably smaller dimensions.
- The use of nanomaterials in refrigerator allows the refrigeration without the need of refrigeration fluids. This can be done if nanoparticles with large magnetic moments and adequate coercivity can be obtained then the magnetocaloric effect may allow refrigeration on a practical nanoscale.

19.2.2 Nanotechnology in Health and Medicine

- The health conditions and nonavailability of proper medicines to human beings is a big threat. There is growing demand for new kinds of medicines for high number of serious and complex illnesses-like cancer, cardiovascular diseases, multiple sclerosis, Alzheimer's and Parkinson's disease and diabetes as well as different kinds of serious inflammatory or infectious diseases (e.g. HIV).
- Nanotechnology has also its applications in field of health and medicine called nanomedicine.
- The approaches to nanomedicine range from the medical use of nanomaterials, to nanoelectronic biosensors, and even possible future applications of molecular nanotechnology. The medical area of nanoscience application is one of the most potentially valuables, with many projected benefits to humanity.
- Nanomedicine has the potential to enable early detection and prevention, and to essentially improve diagnosis, treatment and follow-up of diseases.
- Biological tests measuring the presence or activity of selected substances become quicker, more sensitive and more flexible when certain nanoscale particles are put to work as tags and labels.
- Nanodevices can make gene sequencing more efficient. Gold nanoparticles tagged with short segments of DNA can be used for detection of genetic sequence in a sample.

- Nanotechnology can help to reproduce or to repair damaged tissue. This so called tissue engineering makes use of artificially stimulated cell. It might replace today's conventional treatments, e.g. transplantation of organs or artificial implants.
- Carbon nanotubes have recently become promising functional materials for the development of advanced biosensors with novel features. These sensors are being used for astrobiology to study origins of life. The technology is also being used to develop sensors for cancer diagnostics. CNTs, although inert, can be functionalized at the tip with a probe molecule. Their study uses AFM as an experimental platform.
    (i) Probe molecule to serve as signature of leukaemia cells identified
    (ii) Current flow due to hybridization will be through CNT electrode to an IC chip.
    (iii) Prototype biosensors catheter development.

19.2.3 Nanotechnology in Transportation

- The extended benefits may be reaped in transportation industry adopting nanomaterials and technology.
- Nanomaterials will allow cars and planes to become safer and cheaper.
- The light nanomaterials replace the weight of heavy structural materials in transport industry; and the new fuel sources will permit travel to become farther and more financially feasible, by reducing the cost. It may be possible to make carbon-based fibres which are 100 × stronger than steel and only one sixth of weight.
- Nanotechnology will enhance aerospace application and space flight as new materials will allow space shuttles to become lighter and tougher.
- Nanotechnology may be able to reduce health risks due to emissions from cars containing many different noxious gases. Nanotwin technology has recently released an air filter, which uses nanotechnology principles, to remove hazardous chemicals from the air in car cabins.
- Improved catalysts could reduce or eliminate the emission of pollutants from engines.
- Nanocoating of metallic surfaces to achieve super-hardening, low friction and enhanced corrosion protection; 'tailored' materials for infrastructure and vehicles; and 'smart' materials that monitor and assess their own status and repair any defects resulting from fatigue, fire, etc.
- Cerium oxide nanoparticles are used in diesel fuel to greatly increase fuel efficiency. Introducing the nanoparticles reduces fuel deposits on pistons and cylinders, increasing fuel efficiency by about 10 percent.
- Nanoparticles of inorganic clays and polymers will replace carbon black tyres and therefore, we will have environmental friendly, wear resistant tyres.

19.2.4 Nanotechnology, Energy and Environment

- Nanotechnology will provide sufficient energy for a growing world and to protecting the environment in which we live. Under all practical solutions nanotechnology will play a critical role in any successful outcome.
- The most advanced nanotechnology projects related to energy are as follows: storage, conversion, manufacturing improvements by reducing materials and process rates, energy saving by better thermal insulation and enhanced renewable energy sources.
- Using nanotechnology more clean and less expensive ways for energy production is desired. Nanotechnology has great contribution on renewal energies from solar technology, to nanocatalysis, fuel cells and hydrogen technology.

- Carbon nanotube fuel cells are being used to store hydrogen. These are the environmentally friendly form of energy. Researchers are trying to increase effectiveness of carbon nanotube at storing hydrogen. This has the potential to power cars.
- Research on photovoltaic is being done to make them cheap, light weight and more efficient.
- Nanotechnology can contribute to the further reduction of combustion engine pollutants by nanoporous filters, which can clean the exhaust mechanically, by catalytic converters based on nanoscale noble metal particles or by catalytic coatings on cylinder walls and catalytic nanoparticles as additive for fuels.
- Nanotechnology can help in developing new environmental safe and green technologies that can minimize the formation of undesirable by-products or effluents.
- Solid state lightening can reduce total electricity consumption by 10 percent and cut carbon emission by the equivalent of 28 million tons/year.
- A reduction of energy consumption can be reached by better insulation systems, by the use of more efficient lighting or combustion systems, and by use of lighter and stronger materials in the transportation sector.
- Currently used light bulbs only convert approximately 5 percent of the electrical energy into light. Nanotechnological approaches like LEDs (light-emitting diodes) or QCAs (quantum-caged atoms) could lead to a strong reduction of energy consumption for illumination.

19.2.5 Nanotechnology in Space Exploration

- Rocket scientists are actively researching new forms of space propulsion systems because today’s rocket engines rely on chemical propulsion.
- Electric propulsion (EP) is the heavily researched area which includes field emission electric propulsion (FEEP), colloid thrusters and other versions of field emission thrusters (FETs). EP systems significantly reduce the required propellant mass compared to conventional chemical rockets, allowing increasing the payload capacity or decrease the launch mass.
- A new EP concept proposes to utilize electrostatically charged and accelerated nanoparticles as propellant. Millions of micron-sized nanoparticle thrusters would fit on one square centimetre, allowing the fabrication of highly scalable thruster arrays. The greatest challenges in the space craft are performance reliability and safety cost. Nanotechnology can improve the situation. Some specific benefits nanotechnology can bring are:
- Nanotechnology can make the structure of space planes much lighter thus can greatly improve their viability.
- Nanotechnology can improve the performance of laser sails. Using nanotechnology sails with 20 nm thickness can be constructed making them light and more durable.
- Performance can also be increased using solar powered ion engines with nanotechnology.
- Using a combination of AI and nanorobotics, computer controlled manufacturing systems will reduce the time and cost of developing new technologies.

19.3 Nanomaterials

Nanomaterials can be metals, ceramics, polymeric materials or composite materials. Their defining characteristic is a very small feature size in the range of 1–100 nm. Nanomaterials are not simply another step in miniaturization, but a different arena entirely; the nanoworld lies midway between the scale of atomic and quantum phenomena, and the scale of bulk materials.

A typical nanomaterial is summarized in Table 19.1, which include

1. Zero-dimension nanostructures, such as metallic, semiconducting and ceramic nanoparticles
2. One-dimension nanostructures, such as nanowires, nanotubes and nanorods
3. Two-dimension nanostructures, such as thin films, magnetic materials
4. Three-dimension nanomaterials, such as metals, semiconductors, magnetic materials.

1. Nanocrystals and clusters (quantum dots)	Diameter 1–10 nm	Metals, semiconductors, Magnetic materials, ceramic oxides
2. Nanowires and nanotubes	Diameter 1–100 nm Diameter 1–100 nm	Metals, semiconductors, oxides, sulphides, carbon layer metal chalcogenides
3. Two-dimensional arrays surface and thin films	Several nm^2 to pm^2 Thickness 1–1000 nm	Metals, semiconductors, magnetic materials
4. Three-dimensional arrays (superlattices)	Several nm in all three dimensions	Metals, semiconductors, magnetic materials

Table 19.1

Typical nanomaterials

Besides these individual nanostructures, ensembles of nanostructures form high dimension arrays, assemblies and superlattices.

The structural features of nanomaterials are in between of those of atoms and the bulk materials. While most microstructured materials have similar properties to the corresponding bulk materials, the properties of materials with nanometre dimensions are significantly different from those of atoms and bulk materials. These properties are mainly due to the nanometre size of the materials which render them

- large fraction of surface atoms
- high surface energy
- spatial confinement
- reduced imperfections, which do not exist in the corresponding bulk materials.

The 'surface'-dependent material properties of nanomaterials are due to their small dimensions and have extremely large surface area to volume ratio, which makes a large fraction of atoms of the materials to be the surface or interfacial atoms. Especially when the sizes of nanomaterials are comparable to the Debye length, the entire material will be affected by the surface properties of nanomaterials. This in turn may enhance or modify the properties of the bulk materials. For example, very active catalysts are obtained from metallic nanoparticles. The sensitivity and sensor selectivity of chemical sensors can be improved by using nanoparticles and nanowires. The nanometre feature sizes of nanomaterials also have spatial confinement effect on the materials, which bring the quantum effects. Nanoparticles can be viewed as a zero dimension quantum dot while various nanowires and nanotubes can be viewed as quantum wires. The profound effects on the properties of nanomaterials are due to their quantum confinement. The energy band structure and charge carrier density in the materials can be modified quite differently from their bulk counterpart and in turn will modify the electronic and optical properties of the materials.

Lasers and light-emitting diodes (LEDs) from both of the quantum dots and quantum wires are very promising in the future optoelectronics. High-density information storage using quantum dot devices is also a fast developing area. Reduced imperfections are also an important factor in determination of the properties of the nanomaterials. Nanostructures and nanomaterials favour a self-purification process in that the impurities and intrinsic material defects will move near the surface upon thermal annealing. This increased material perfection affects the properties of nanomaterials. For example, the chemical stability for certain nanomaterials may be enhanced, the mechanical properties of nanomaterials will be better than the bulk materials. The superior mechanical properties of carbon nanotubes are well known. Due to their nanometre size, nanomaterials are already known to have many novel properties. Many novel applications of the nanomaterials rose from these novel properties have also been proposed.

19.4 Types of Nanomaterials

Nanomaterials are cornerstones of nanoscience and nanotechnology. Nanomaterials can exist in one dimension (e.g. surface films), two dimensions (e.g. strands or fibres), or three dimensions (e.g. nanostructured material (NsM) particles). They can exist in single, fused, aggregated or agglomerated forms with spherical, tubular and irregular shapes.

19.4.1 Classification of Nanomaterials on Dimension

Nanostructured materials are classified as zero-dimensional, one-dimensional, two-dimensional, three-dimensional nanostructures, Fig. 19.2(a)–(d).

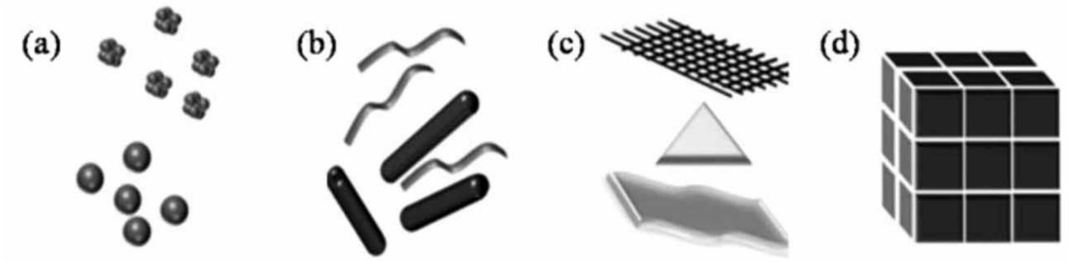


Figure 19.2 *Classification of nanomaterials: (a) zero-dimensional spheres and clusters, (b) one-dimensional nanofibres, wires and rods, (c) two-dimensional films, plates, and networks, (d) three-dimensional nanomaterials.*

(a) Zero-dimensional nanomaterials

Materials wherein all the dimensions are measured within the nanoscale (no dimensions, or 0-D, are larger than 100 nm).

The best example of zero-dimensional (Fig. 19.3) nanomaterials are 'nanoparticles'.

Nanoparticles can be

- amorphous or crystalline
- single crystalline or polycrystalline
- composed of single or multi-chemical elements
- exhibit various shapes and forms
- exist individually or incorporated in a matrix
- metallic, ceramic or polymeric.

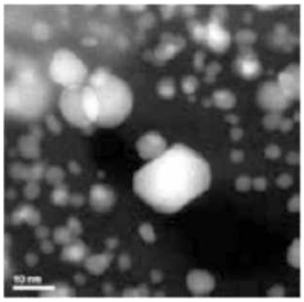


Figure 19.3 *Zero-dimensional nanomaterial.*

(b) One-dimensional nanomaterials

Nanomaterials exist in one-dimensional structure. This leads to needle-like-shaped nanomaterials.

One-dimensional materials include 'nanotubes, nanorods and nanowires' (Fig. 19.4).

One-dimensional nanomaterials can be

- amorphous or crystalline
- single crystalline or polycrystalline
- chemically pure or impure
- standalone materials or embedded in within another medium metallic, ceramic or polymeric.

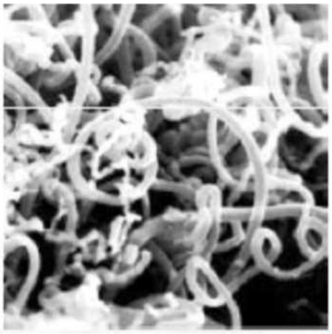


Figure 19.4 *1D nanomaterial.*

(c) Two-dimensional nanomaterials

Two-dimensional nanomaterials exhibit plate-like shapes (Fig. 19.5). Two-dimensional nanomaterials include 'nanofilms, nanolayers and nanocoatings.

Two-dimensional nanomaterials can be:

- amorphous or crystalline
- made up of various chemical compositions
- used as a single layer or as multilayer structures
- deposited on a substrate
- integrated in a surrounding matrix material metallic, ceramic or polymeric.

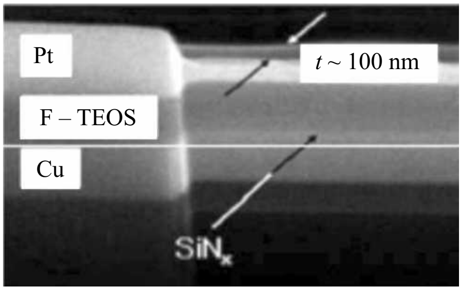


Figure 19.5 *Two-dimensional nanomaterials.*

(d) Three-dimensional nanomaterials

'Bulk nanomaterials' are materials that are not confined to the nanoscale in any dimension. These materials are thus characterized by having three arbitrarily dimensions above 100 nm. Materials possess a nanocrystalline structure or involve the presence of features at the nanoscale. In terms of nanocrystalline structure, bulk nanomaterials can be composed of a multiple arrangement of nanosize crystals, most typically in different orientations. In a nanoscale, three-dimensional nanomaterials can contain dispersions of nanoparticles, bundles of nanowires and nanotubes as well as multinanolayers.

19.5 Nanomaterial Categories

Nanomaterials are also classified into 10 main material categories in order to give a good overview of the different nanomaterials. The main material categories are defined as follows:

1. Nanotubes: carbon-based nanomaterials
2. Nanocomposites
3. Metals and alloys
4. Biological nanomaterials
5. Nanopolymers
6. Nanoglasses
7. Nanoceramics
8. Nanocatalysts
9. Nanofilms
10. Nanomachines

In addition, on the basis of phase composition, nanomaterials in different phases can be classified as follows:

- Single phase solids include crystalline, amorphous particles and layers, etc.
- Multiphase solids include matrix composites, coated particles, etc.
- Multiphase systems include colloids, aero gels, ferro fluids, etc.

19.6 Nanotechnology

Nanotechnology may be divided into the following three broad categories:

- Nanotools
- Nanodevices
- Nanostructured materials

19.6.1 Nanotools

Nanotools are devices that manipulate matter at the nano or atomic regime of materials.

Devices such as

- atomic force microscopes,
- scanning probe microscopes, and
- atomic layer-deposition devices and nanolithography tools can manipulate matter in the atomic or molecular regime.

Some other nanotools include fabrication techniques such as

- analysis and metrology instruments and
- software for nanotechnology research and development.

19.6.2 Nanodevices

Nanodevices are any complete system with nanostructured components that carries out as assigned function other than manipulating nanomatter.

The first nanodevices on the market were the following:

- Quantum dot fluorescent biodetectors
- MEMS devices are used as accelerometers in automotive airbags

Many other promising applications are in development such as

- nanoelectric memory devices,
- nanosensors, and
- drug delivery systems

Components to nanodevices will include nanomaterials, semiconducting organic molecules, polymers and high-purity chemicals and materials.

Nanodevices based on carbon nanotubes

The carbon nanotubes have extensive commercial applications. The applications of CNTs could include field emission-based flat panel displays, novel semiconducting devices, chemical sensors and ultra-sensitive electromechanical sensors.

The commercial use of CNTs for molecular electronics or computers is now being explored through experiments to fabricate and conceptualize new devices based on simulations.

Preferentially, carbon nanotubes are replacing silicon chips. Calculations show that nanotubes can have metallic or variable semiconducting properties with energy gaps ranging from a few meV to a few tenths of an eV. Experiments probing the density of states confirm these predictions.

Conductivity measurements on single nanotubes have shown rectification effects for some nanotubes and ohmic conductance for others. These properties suggest that nanotubes could lead to a new generation of electronic devices.

Simulations to investigate the interaction of water molecules with a nanotube tip revealed an atomistic understanding of the interaction, which is critical in designing commercial-quality flat panel displays around carbon nanotubes.

Their use as ultra-sensitive electromechanical sensors has also been explored.

19.6.3 Nanostructured Materials

Nanostructured materials are materials with a microstructure, the characteristic length scale of which is on the order of 1–100 nm. The microstructure refers to the chemical composition, the arrangement of the atoms (the atomic structure), and the size of a solid in one, two, or three dimensions.

The effects of controlling the properties of nanostructured materials are mainly due to size effects, changes of the dimensionality of the system, changes of the atomic structure and the alloying of components (e.g. elements) that are not miscible in the solid and/or the molten state.

In the paragraphs below, the various classes of nanoparticles that serve as the building blocks of nanostructured materials will be discussed. The nanostructured materials may be grouped under the following categories:

1. Nanocrystalline materials
2. Carbon nanotubes/fullerenes
3. Dendrimers (organic nanoparticles)
4. Polyhedral silsesquioxanes (inorganic-organic hybrid nanoparticles)
5. Nanocomposites
6. Nanointermediates

The synthesis, characterization and processing of nanostructured materials are part of an emerging and rapidly growing field. The main emphasis is stressed to generate new nanomaterials with controlled microstructural characteristics. The processing of these new nanomaterials as bulk materials with engineered properties and technological functions; and their application to new device technology.

19.6.3.1 Nanocrystalline Materials

The nanocrystalline materials are ceramics, metals and metal oxide nanoparticles. An assembly of nanosized crystallites (building blocks) gives rise to nanocrystalline materials. The building blocks may differ in their atomic structure, crystallographic orientation or chemical composition. In cases where the building blocks are crystallites, incoherent or coherent interfaces may be formed between them, depending on the atomic structure, the crystallographic orientation, and the chemical composition of adjacent crystallites.

In other words, materials assembled of nanometre-sized building blocks are microstructurally heterogeneous, consisting of the building blocks (e.g. crystallites) and the regions between adjacent building blocks (e.g. grain boundaries).

It is known that glasses, gels, etc. are microstructurally homogeneous while for other nanocrystallites, their inherent heterogeneous structure is very crucial for their properties.

There are grain boundaries that make up a major portion of the material at nanoscales, and strongly affect properties and processing. The properties of NsM deviate from those of single crystals (or coarse-grained polycrystals) and glasses with the same average chemical composition. This deviation results from the reduced size and dimensionality of the nanometre-sized crystallites as well as from the numerous interfaces between adjacent crystallites.

'Nanoparticles' are the simplest form of structures with sizes in the nm range. In principle, any collection of atoms bonded together with a structural radius of <100 nm can be considered a nanoparticle. Fullerenes, metal clusters (agglomerates of metal atoms), large molecules, such as proteins and even hydrogen-bonded assemblies of water molecules, which exist in water at ambient temperatures are the best known examples.

There are innumerable nanoparticles which are very common in nature—for example, proteins exist in almost all biological systems, metal-oxide nanoparticles are easily produced, etc. Nanocrystallites of bulk inorganic solids exhibit the size-dependent properties such as lower melting points, higher energy gaps, and nonthermodynamic structures. Increased ductility has been observed in nanopowders of metal alloys in comparison to macro-scale powders. In nanoparticles, quantum effects become significant leading to such phenomena as quantum 'dots lasers'.

'Nanocrystalline metals' are extensively used as heterogeneous catalysts in a variety of reactions. The surface-dependent properties of heterogeneous catalyst activity of nanocrystalline metal over macro-scale materials, nanometals and oxides are due to their vastly increased surface area of nanocrystalline structure and these nanocrystalline metals behave as ultra-high activity catalysts. They are also used as desirable starting materials for a variety of reactions, especially solid-state routes. Nanometals and oxides are also widely used in the formation of nanocomposites. Aside from their synthetic utility, they have many useful and unique magnetic, electric and optical properties.

Another important application concerned with nanocrystalline metal is to identify light metals with improved mechanical properties. The specific mechanical properties of nanostructured materials which are desired are as follows: high strength that for some special methods of production can be combined with ductility, high fatigue limits, elevated temperature strength, corrosion or wear resistance, etc.

Present research is focused on 'magnetic materials' and their applications. The material may become magnetically soft for very fine grain size, thus a decrease of energy losses when it is applied as a transformer core or in other applications with oscillating magnetic fields is considered. This leads to energy saving during energy transmission. Recently, research work was carried out on bulk metals and nanopowders and the observations are similar. The bulk metals nanoparticles are divided into magnetic materials and structural materials. The application of nanopowders is strongly focused on the antibacterial and catalytic activity.

The application based on most exploited properties of 'nanopowders' is dependent on the following three phenomena:

- High specific surface area
- Soft magnetic properties
- New deformation mechanisms

There are a few other specific phenomena of nanomaterials which are to be exploited:

Surface treatment of metals—desirable technique to produce a nanostructure on their surface, increase terminological properties, chemical properties, reduce friction, increase biocompatibility, etc.

Mechanical and multifunctional properties

Metals in the form of single crystals, amorphous materials or nanomaterials should be used in order to obtain structures of the size of some microns. In the case of conventional polycrystals, single grain boundaries would cross the structure element and impair its reliability.

Applications of nanocrystalline metals

Innumerable applications of nanoparticles are as follows:

- Hydrogen storage
- Catalysis
- Production of carbon nanotubes
- Sensors

- Getters
- Optical and electronic devices
- Biological imaging
- Antibacterial applications

In microsystems, there are a wide variety of applications of nanometals especially in microelectromechanical systems (MEMS), bioMEMS, nanoelectromechanical systems (NEMS), optical, electronic, electrochemical microsystems, for multifunctional devices and systems for chemical and biological analysis/detection, drug delivery/discovery, tissue engineering, chemical and materials synthesis, energy conversion and storage. Nanometal coatings by electrodeposition as wear and corrosion-resistant coatings on heat-exchangers decreased wear and friction. Besides coating, the surface treatment of metal parts to produce a nanostructured layer on the surface will play an important role.

In summary, nanometals have an enormous potential for applications in electronics, construction, power transformation, energy storage, telecommunications, information technology, medicine, catalysis and environment protection, with high possible impact in the areas of technology related to energy, health and materials.

19.6.3.2 Carbon-based Nanomaterials

Carbon-based nanomaterials are synthesized and investigated extensively. The possible applications are exploited. Carbon-based nanomaterials are defined as materials in which the 'nanocomponent' is pure carbon. Table 19.2 shows a list of carbon-based nanomaterials.

Single nanostructures			
Nanoparticles (1)	**Nanotubes (2)**	**Films, coating, nanostructured surfaces (3)**	**Nanostructured bulk materials (4)**
(a) Carbon black	(i) SWCNT/MWCNT	(a) Carbon films	(i) Nanostructured carbon
(b) Graphene		(b) DLC	(ii) Nanoporous carbon
(c) Graphite	(ii) Nanohorns	(c) Covalent carbides (SiC)	(iii) Carbon foams
(d) Fullerenes	(iii) Nanowires	(d) Metallic carbides (TiC)	(iv) Carbon arogels
(e) Carbon onion	(iv) Nanorods	(e) Nanocarbon nitrides	(v) Carbon nanocrystals
(f) Nanocluster	(v) Nanofibre		

Table 19.2

Most prominent carbon-based nanomaterials

19.6.3.2.1 Single Nanostructure Particles

(a) Carbon black

A pure elemental carbon in the form of colloidal particles is carbon black. An incomplete combustion or thermal decomposition of gaseous or liquid hydrocarbon under controlled conditions produces carbon black. The furnace black and thermal black processes are used to make carbon black, but the furnace black process is the most common commercial process. The average particle size of carbon black is few 'nm'. The primary use of carbon black is in rubber products (approximately 90 percent), mainly tyres and other automotive rubber products such as hoses, gaskets and coated tubes. About 9 percent of carbon black is used for pigment in toners and inks, and the remaining 1 percent is used as an ingredient in hundreds of diverse applications. Much smaller amounts of carbon black are used in the manufacture of dry-cell batteries.

Carbon black is currently the most widely used carbon nanomaterial and it has found application in car tyres, antistatic textiles and is used for colour effects.

Amorphous carbon

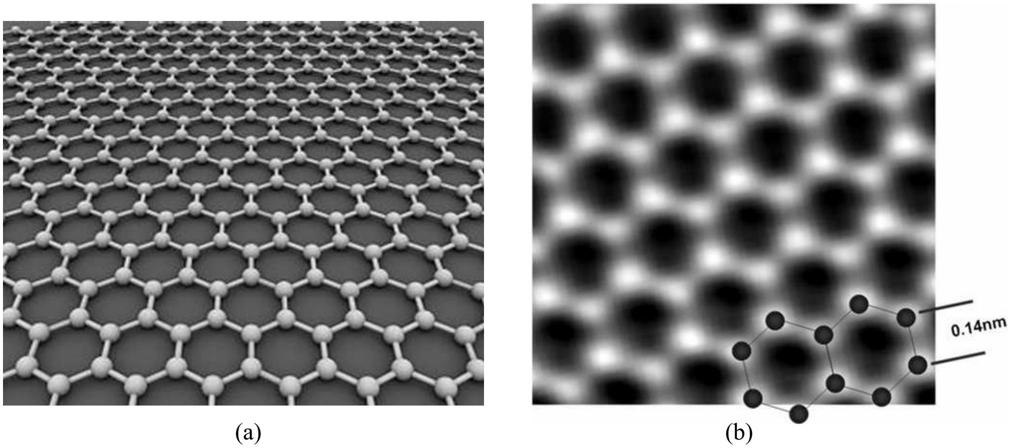
There is no crystalline structure for amorphous carbon. Amorphous carbon can be synthesized/obtained, but there exist some microscopic crystals of graphite-like or diamond-like carbon. Therefore, the properties of amorphous carbon depend on the ratio of sp^2- to sp^3-hybridized bonds present in the material. Graphite consists purely of sp^2-hybridized bonds, whereas diamond consists purely of sp^3-hybridized bonds. Materials that are high in sp^3-hybridized bonds are referred to as tetrahedral amorphous carbon (owing to the tetrahedral shape formed by sp^3-hybridized bonds) or as diamond-like carbon (owing to the similarity of many of its physical properties to those of diamond).

Glassy or vitreous carbon

Another form of carbon is glassy or vitreous carbon which is widely used as an electrode material in electrochemistry as well as in prosthetic devices and high-temperature crucibles. The most important properties are high temperature resistance, hardness (7 Mhos), low density, low electrical resistance, low friction, low thermal resistance, extreme resistance to chemical attack and impermeability to gases and liquids.

(b) Graphene

A layer of one-atom-thick planar sheet of carbon atoms, densely packed together into a honeycomb shaped crystal lattice is graphene (Fig. 19.6a). It is an allotrope of carbon whose structure is a single planar sheet of sp^2-bonded carbon atoms that are densely packed in a honeycomb crystal lattice.



(a) (b)

Figure 19.6 *(a) A sheet of thin graphene layer and (b) C–C bond distance in graphene.*

It looks similar to atomic-scale wire mesh made up of carbon atoms and their bonds. It is very light and 1-square-metre sheet weighs only 0.77 mg. The carbon-carbon bond length in graphene is about 0.142 nm, Fig. 19.6(b). Graphene is the basic structural unit of several carbon allotropes including graphite, carbon nanotubes and fullerenes. An isolated atomic plane of graphite gives graphene. Graphene is a material of interest due to its high electron mobility and its possible applications in electronics.

Graphene is a flat monolayer of carbon atoms tightly packed into a two-dimensional (2D) honeycomb lattice, and is a basic building block for graphitic materials of all other dimensionalities. It can be wrapped up into 0D fullerenes, rolled into 1D nanotubes or stacked into 3D graphite.

Exfoliated graphene

The micro-mechanical alleviation of graphite is the process to obtain graphene from graphite. In the process, an adhesive tape is used to repeatedly tear off graphite crystals into very thin sheets. The optically transparent thin sheets obtained by the above process which is attached to the tape is dissolved in acetone and, after a few further steps, the flakes including monolayers were sedimented on a silicon wafer. Individual atomic planes of graphene were then viewed in an optical microscope and separated.

Using a simplified technique, a dry deposition of graphene was done avoiding the stage when graphene floated in a liquid. Relatively large crystallites were obtained by the technique. It is often referred to as a scotch tape or drawing method. To obtain ultra-thin films of graphene from graphite, epitaxial growth production technique was adopted.

Graphene is obtained by heating silicon carbide (SiC) to high temperatures (>1100°C) under low pressures ($\sim 10^{-6}$ Torr). This process produces epitaxial graphene with dimensions dependent upon the size of the SiC substrate (wafer). The face of the SiC used for graphene formation, silicon- or carbon-terminated, highly influences the thickness, mobility and carrier density of the graphene.

A new example of isolated graphene is graphene on silicon carbide (SiC) after its passivation with hydrogen. Carbon nanotubes (buckytubes) are cylindrical atoms of carbon (graphene hexagons rolled up) with novel properties that are potentially useful in a wide variety of applications including nanoelectronics, optics, materials applications, etc. They exhibit extraordinary tensile strength, a unique range of electrical properties, and are efficient thermal conductors.

(c) Graphite

Graphite is the most stable form of carbon under standard conditions. Graphite is another allotrope of carbon. The other allotropes of carbon are (a) diamond, (b) graphite, (c) lonsdaleite, (d) C60 (Buckminsterfullerene), (e) C54, (f) C70, (g) amorphous carbon, and (h) single-walled carbon nanotube.

The basic structure in other allotropes of carbon such as graphite, charcoal, carbon nanotubes and fullerenes is graphene. A single layer of graphite is called graphene. The graphite displays extraordinary electrical, thermal and physical properties. It is an allotrope of carbon whose structure is a single planar sheet of sp^2-bonded carbon atoms that are densely packed in a honeycomb crystal lattice. Graphene is a semimetal or zero-gap semiconductor, allowing it to display high electron mobility at room temperature.

Graphite has a layered, planar structure (Fig. 19.7). In each layer, the carbon atoms are arranged in a hexagonal lattice with separation of 0.142 nm, and these sheets stack to form graphite with an interplanar spacing of 0.335 nm.

In graphite, the carbon atoms are only bonded in two dimensions. The carbon atoms form layered sheets of hexagons. Since there are no bonds between the layers, the layers can easily slip off one another.

(d) Fullerenes

Fullerenes are single structured nanomaterials and have wide variety of applications. Fullerenes are a class of carbon allotropes in which carbon takes the form of a hollow sphere, ellipsoid or tube. This class of materials includes carbon nanotubes, buckyballs and the nanobuds. Carbon soot contains many types of fullerenes which are tubes of carbon many thousands of times long as they are wide, with the same icosahedral structure. The fullerenes have diameters as small as 2 nm.

Fullerene is an elongated sphere of carbon atoms formed by interconnecting six-member rings and twelve isolated five-member rings forming hexagonal and pentagonal faces. The typical structure of fullerenes are a class of cage-like carbon compounds composed of fused, pentagonal and/or hexagonal sp^2 carbon rings, Fig. 19.8(a)–(d).

C60 was the first isolated and characterized fullerene, contains 20 hexagonal faces and 12 pentagonal faces and possesses perfect icosahedral symmetry.

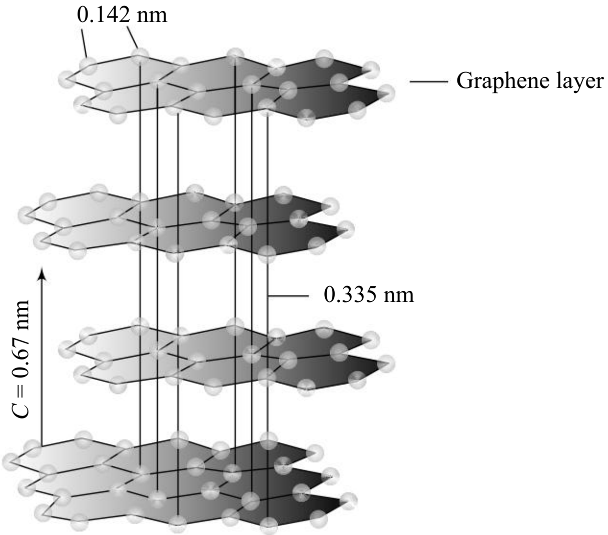


Figure 19.7 *Structure of graphite.*

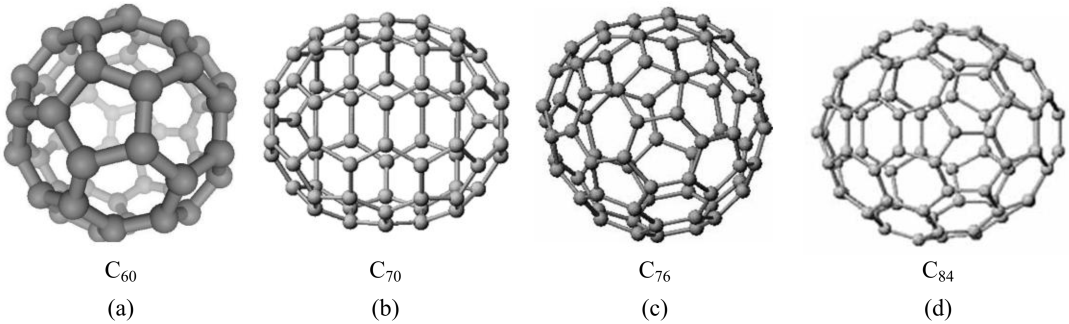


Figure 19.8 *Representation of fullerenes: (a) C_{60} (buckminsterfullerene), (b) C_{70}, (c) C_{76} and (d) C_{84} molecules.*

It is well characterized that fullerenes are cage-structured carbon molecules in a spherical shape with a wide range of sizes and molecular weights such as C_{60}, C_{70}, C_{76} and C_{84}. The C_{60} fullerene (buckminster fullerene) is the representative and approximately 0.7 nm in diameter. In the fullerene structure, all C sites are equivalent and the bond lengths are 0.14 nm for the double bond and 0.146 nm for the single bond. Figure 19.8(a) is a representation of a C_{60} fullerene molecule.

Fullerene chemistry is an exciting field with promising new applications. The properties of fullerenes may be tuned as desired to suit an application via the addition of element and molecular species into the fullerene lattice ($C_{59}N$), within the cage (N at centre of C_{60}), or coating the surface of fullerene with transition metals. These properties can be used for drug delivery, facial creams, moisturizers, lubrications, trace monitors and electronic circuits.

Magnetic nanoparticles (nanomagnetic materials) show great potential for high-density magnetic storage media. It is observed that C_{60} dispersed into ferromagnetic materials such as iron, cobalt or cobalt-iron alloy can form thin films with promising magnetic properties. A number of organometallic-fullerene compounds have recently been synthesized. Of particular note are a ferrocene-like C_{60} derivative and pair of fullerenes bridged by a rhodium cluster.

Superconducting character has been shown by some fullerene derivatives.

A superconducting field-effect device containing fullerene has been found with a *Tc* as high as 117 K. There is potential use of fullerenes as fillers. In the longer term, applications as drug deliverers either by the attachment of functional ligands to the carbon cage or by trapping molecules inside has been predicted. The trapping of metals inside the cage such as lanthanides could enable the application of fullerenes as tracer molecules.

The production of fullerenes

A production of microscopic amounts of fullerenes can be achieved through laser vaporization of carbon technique in an inert atmosphere.

In 1990, Kratschmer and Huffmann (Fig. 19.9) used a new type of apparatus using an arc to vaporize graphite.

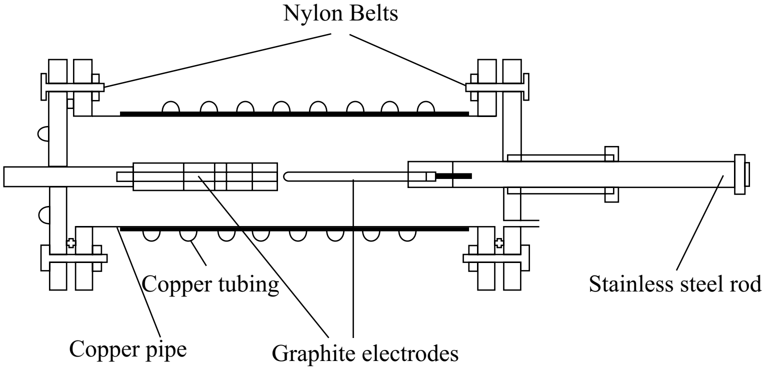


Figure 19.9 *Kratschmer–Huffman apparatus.*

Fullerene synthesis

Helium gas at 100 Torr is fed into the bell jar after evacuating all the air of the apparatus.

A welding kit is attached with power supply. The power supply is switched on for about 10–15 seconds. A plenty of black soot-like material is produced inside the bell jar. After a 5–10 minute cool down period, the bell jar is cooled down to atmospheric pressure. The bell jar is removed and the glass surface is scraped to collect all the material. Ten percent of the soot should be made up of C_{60}. The fullerenes in the soot are then extracted by a small amount of toluene. After extraction, the solvent (toluene) is removed using a rotary evaporator, leaving behind a solid mixture of mostly C_{60} with small amounts of larger fullerenes. Pure C_{60} is obtained by liquid chromatography. The mixture is dissolved in toluene and pumped through a column of activated charcoal mixed with silica gel. The magenta C_{60} comes off first, followed by the red C_{70}. The different colour solutions are collected separately and the toluene removed using the rotary evaporator.

(e) Carbon onion

One of the fullerene-related materials having concentric graphitic shells is ‘Carbon onion’. The particle size of carbon onion is about 10 nm in diameter. It can be seen that the centre shell of carbon onion is a C_{60} molecule. Carbon onions consist of carbon cages one inside the other (Fig. 19.10). There are numerous carbon atoms inside carbon onion, and

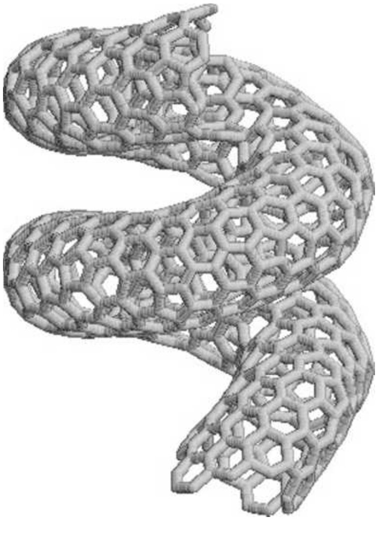


Figure 19.10 *A helically coiled carbon nanotube.*

many of such structures have been observed with dozens of concentric shells. The hypothetical structures of carbon cages of carbon onion are consisted of not only pentagonal and hexagonal rings (like C_{60}) but also heptagonal (seven membered) rings.

These carbon onion particles can be made by high-energy electron bombardment on carbon soot, heat treatment of soot or diamond clusters, carbon implantation into copper or silver substrate, plasma chemical vapour deposition, and arc discharge methods. To make carbon onion with carbon implantation, 120 keV carbon ions implant into crystalline copper substrate at 700–1000°C. During the process, 10–5 Pa of vacuum is maintained. Carbon onions are formed on the surface of substrate. Schematic diagram of chamber and image of carbon onion are shown in Fig. 19.11. These carbon onions are 5–10 nm in size and can be used for solid lubrication.

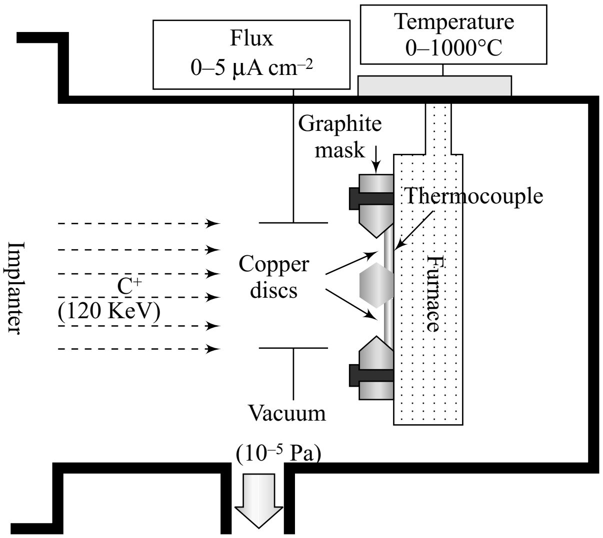


FIGURE 19.11 *Schematic diagram for implantation of carbon.*

(f) Nanocluster

One can visualize that an intermediate state of matter between molecules and solids are 'nanoclusters' (Fig. 19.12). Nanocluster sizes range from subnanometre to about 10 nm in diameter and are of technological interest in numerous areas of applied science (e.g. materials science, catalysis, optoelectronics).

The nanoclusters have to be differentiated between the following:

(i) nanoclusters consisting up to a couple of hundred atoms and
(ii) larger aggregates containing 103 or more atoms which are more often called nanoparticles

The properties of nanoparticles gradually approach those of bulk materials or extended surfaces, i.e. are scalable with size. Nanoclusters, however, have properties and structures which are very sensitive to their composition and size (i.e. 'every atom counts') which can lead to new and interesting properties not realised in the corresponding bulk material.

The research is confined to determine the properties of nanoclusters of technologically important inorganic bulk materials (e.g. ZnO, SiO_2, CdS), and nanostructures and materials formed by their assembly. Research interest is confined to non-bulk-like properties of nanoclusters and how they may provide the inspiration for novel low density (nanostructured) materials for uniquely tailored applications.

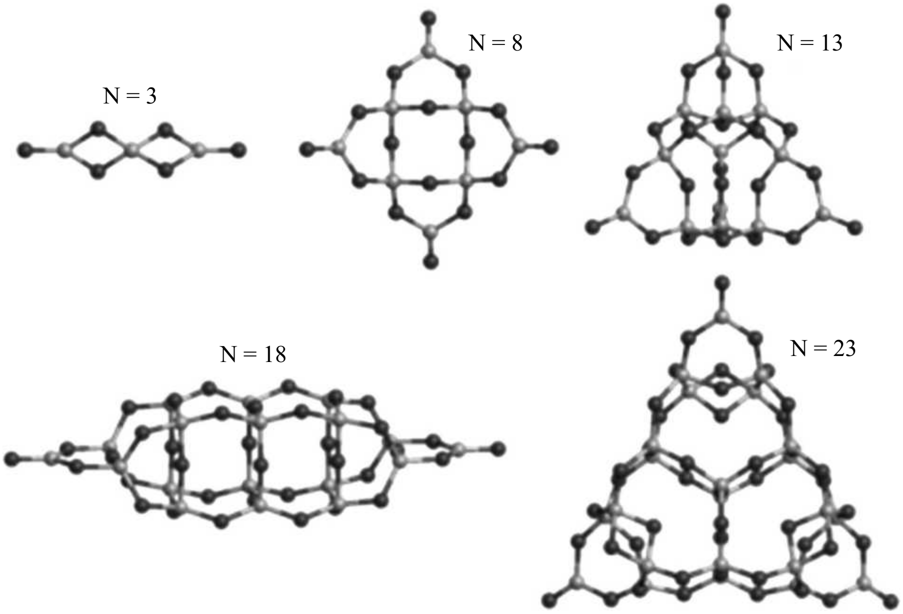


Figure 19.12 *Low-energy* $(SiO_2)N$ *nanocluster structure changes with increasing size.*

19.6.3.2.2 Nanostructure Materials/Particles

Single-wall carbon nanotube (SWCNT) and multi-wall carbon nanotube (MWCNT).

There are two main types of nanotubes obtained from graphene:

- single-walled nanotubes (SWNTs), and
- multi-walled nanotubes (MWNTs).

The diameter of most single-walled nanotubes (SWNTs) is close to 1 nm, with a tube length that can be many millions of times longer.

A one-atom-thick layer of graphene is wrapped into a seamless cylinder to give the structure of a SWNT, Fig. 19.13(a)–(c). The way the graphene sheet is wrapped is represented by a pair of indices.

A carbon nanotube is a tubular form of carbon with a diameter as small as 1 nm and length of a few nanometres to a micron. CNT is configurationally equivalent to a two-dimensional graphene sheet rolled into a tube (Fig. 19.14).

Depending on the direction of rolled up of graphene hexagons, carbon nanotubes can be classified as either zigzag, armchair or chiral. The different types of SWNTs are defined by their diameter and chirality, Figs. 19.15 and 19.16. When nanotubes are synthesized, they tend to get a mixture of several types. Synthesis of one type of CNT is always desirable for application.

In other words, carbon nanotubes are hollow cylinders of carbon atoms. Their appearance is that of rolled tubes of graphene such that their walls are hexagonal carbon rings and are often formed in large bundles. The ends of CNTs are domed structures of six-membered rings capped by a five-membered ring (Fig. 19.16).

The dimensions of CNTs are variable generally down to 0.4 nm in diameter. CNTs also exist as nanotubes within nanotubes, leading to a distinction between single-walled carbon nanotubes and multi-walled carbon nanotubes.

As their names imply, SWNTs consist of a single, cylindrical graphene layer, whereas MWNTs consist of multiple graphene layers telescoped about one another. Multi-walled nanotubes consist of multiple rolled

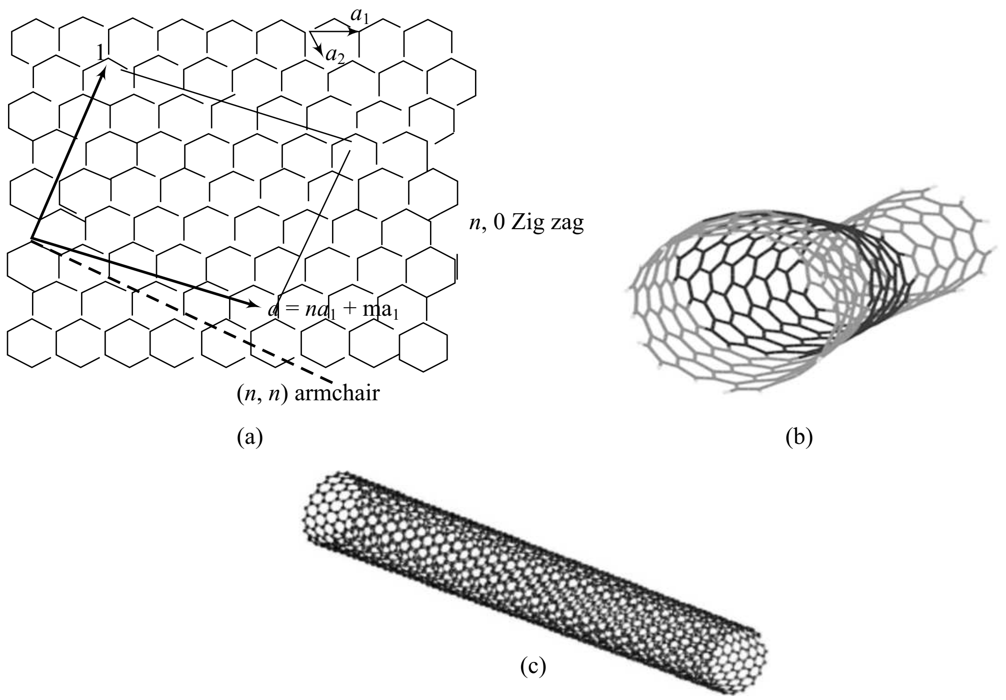


Figure 19.13 *(a) A sheet of graphene in two dimensions and (b) and (c) rolled-up graphene sheet.*

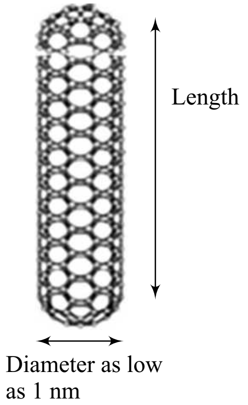


Figure 19.14 *A typical structure of a carbon nanotube.*

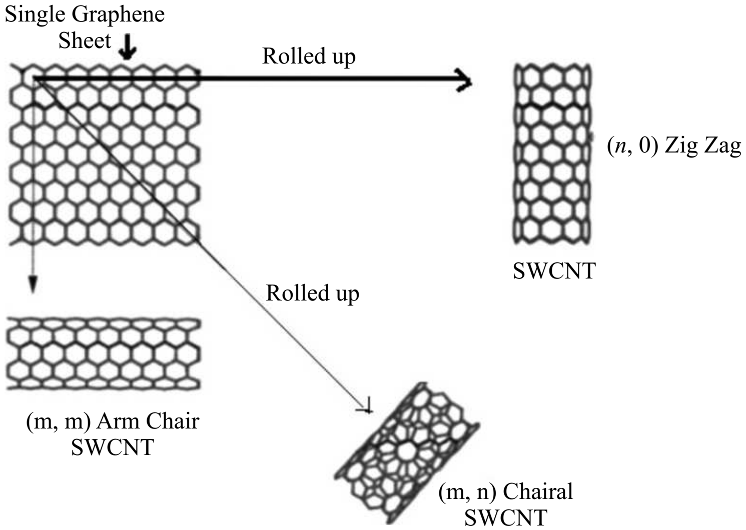


Figure 19.15 *A two-dimensional graphene sheet rolled into a nanotube.*

layers (concentric tubes) of graphene (Fig. 19.17). The interlayer distance in multi-walled nanotubes is close to the distance between graphene layers in graphite, approximately 3.3 Å.

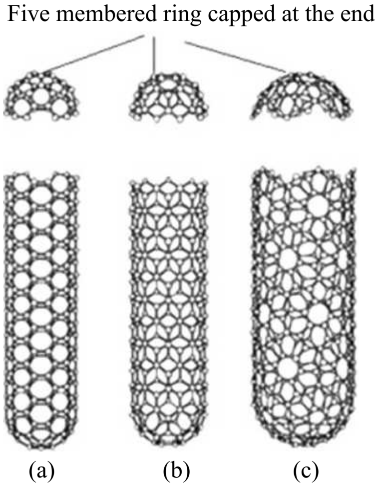


FIGURE 19.16 *Structure of single-walled carbon nanotubes with (a) armchair, (b) zig-zag and (c) chiral chirality.*

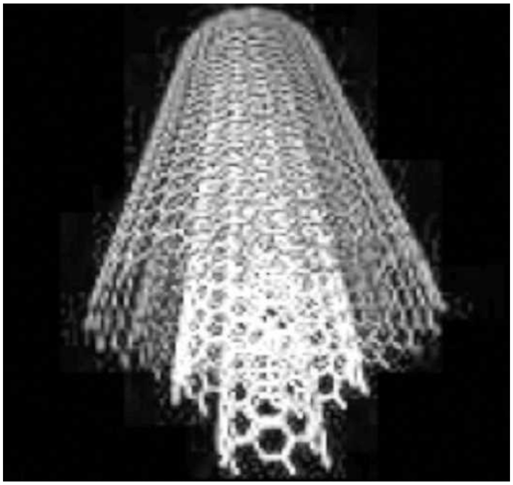


FIGURE 19.17 *Multiwall carbon nanotubes where the several nanotubes have different diameters.*

Graphene sheets are rolled up into a tube to give a CNT structure and capped by half a fullerene. The carbon atoms in a SWNT are sp^2-hybridized like those in a fullerene. The electronic properties of a particular SWNT structure are based on its chirality or twist in the structure of the tube. The chirality in turn affects the conductance of the nanotube, its density, its lattice structure, and other properties. CNT's can also act as either conductors or semiconductors depending on their chirality (twist), possess an intrinsic superconductivity, are ideal thermal conductors and can also behave as field emitters. Carbon nanotubes show extraordinary mechanical, electrical and thermal properties because of their C–C covalent bonding and seamless hexagonal network architecture. Figure 19.18 shows the image of single- and multi-walled carbon nanotubes.

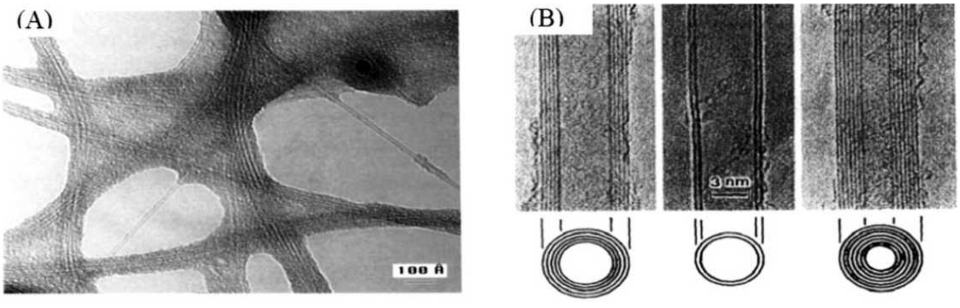


FIGURE 19.18 *Single- and multi-walled carbon nanotubes. (A): Single-walled carbon nanotube, (B): Multi-walled carbon nanotube.*

A nanotube and buckyballs are members of the fullerene structural family. Buckyballs are spherical in shape, whereas a nanotube is cylindrical with at least one end typically capped with a hemisphere of the buckyball structure. Depending on their size, the diameter of a nanotube is of the order of a few nanometres or, while they can be up to several centimetres in length. Inorganic nanotubes have also been synthesized.

Buckypaper is extremely lightweight. In the form of a composite, buckypaper is potentially stronger than steel when sheets are stacked. Buckypaper is one of the most thermally conductive materials known (Fig. 19.19). It is flexible like paper very electrically conductive and fire retardant.

Buckypaper's potential applications include use as an aerospace material, fire protection, novel television screens, heat sinks, electromagnetic interference shielding, filtration, armour and even artificial muscles.

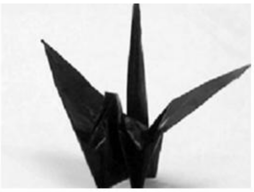


Figure 19.19 *A macroscopic aggregate of carbon nanotubes is called 'buckypaper'.*

(a) Synthesis of carbon nanotubes

Carbon-based nanomaterials are produced and finds industrial applications. There are many techniques to produce nanotubes through arc discharge, laser ablation, CVD, high pressure carbon monoxide (HiPco) and vapour liquid solid (VLS) catalysed growth. The following nanomaterials, carbon nanotubes, fullerenes, carbon nanofibres, carbon black and carbon onions can be synthesised; and by changing fabrication conditions the properties of carbon-based nanomaterials can be changed. The types of carbon materials deposited on the substrate are different in the CVD technique that depends on the changing chamber conditions. All these deposition processes take place in vacuum or at low pressure with gas circulation. However, VLS growth can take place at atmospheric pressure. Nanotubes can be synthesized in large quantities by these methods. However, to use carbon nanotubes, pretreatment such as purification and/or surface modification of the carbon nanotubes is necessary.

1. Arc discharge process

Carbon nanotubes and fullerenes are formed during carbon soot formation in arc discharge process. The high temperatures because of the discharge caused the carbon contained in the negative electrode to sublime and the CNTs are deposited on the opposing electrode. CNTs produced by this method were initially multi-walled tubes (MWNTs).

However, vaporized carbon in the presence of cobalt, it is possible to grow single-walled nanotubes. A mixture of components are produced in this method, and requires further purification to separate the CNTs from the soot and the residual catalytic metals. Producing CNTs in the high yield depends on the uniformity of the plasma arc, and the temperature of the deposit forming on the carbon electrode.

2. Dual-pulsed laser method

Dual-pulsed laser method produces pure SWNTs with a higher yield. With direct vaporization of a Co/Ni-doped graphite rod with a high-powered laser in a tube furnace operating at 1200°C, SWNTs can be grown easily. The material produced by this method appears as a mat of 'ropes', 10–20 nm in diameter and up to 100 μm or more in length. Each rope consists of a bundle of SWNT's, aligned along a common axis. By varying the process parameters such as catalyst composition and the growth temperature, the average nanotube diameter and size distribution can be varied.

The arc-discharge and laser vaporization techniques have drawbacks for the synthesis of small quantities of high quality SWNTs. The first is that they involve evaporating the carbon source, making scale-up on an industrial level difficult and energetically expensive. The second issue relates to the fact that vaporization methods grow SWNTs in highly tangled forms, mixed with unwanted forms of carbon and/or metal species. The SWNTs thus produced are difficult to purify, manipulate and assemble for building nanotube-device architectures for practical applications.

3. Chemical catalysis method

In order to overcome the drawbacks of synthesis of SWNTs, chemical catalytic method is adopted, in which a hydrocarbon feedstock (e.g. ethylene) is used in combination with a metal catalyst. The feedstock such as ethylene is heated under reducing conditions in the presence of a catalyst like iron, cobalt, or iron/molybdenum to produce a nanotube. This method can be used for both SWNTs and MWNTs; the formation of each is controlled by the identity of the catalyst and the reaction conditions. A convenient laboratory scale apparatus is available for the synthesis of highly uniform, consistent, research sample that uses preweighed catalyst/carbon source ampoules. This system, allows for 200 mg samples of MWNTs to be prepared for research and testing. The use of CO as a feedstock, in place of a hydrocarbon, led to the development of the high-pressure carbon monoxide (HiPco) procedure for SWNT synthesis. By this method, it is possible to produce gram quantities of SWNTs, unfortunately, efforts to scale beyond that have not met with complete success.

4. Vapour liquid solid growth method

Large-scale production of nanotubes can be produced from the vapour liquid solid growth. The carbon-containing gas is broken apart at the surface of the catalyst particle in the VLS growth method, and the carbon is transported to the edges where it forms the nanotube at sites of the metal catalyst. The length of the tube grown in surface supported catalyst VLS systems appears to be dependent on the orientation of the growing tube with the surface. By properly adjusting the surface concentration and aggregation of the catalyst particles it is possible to synthesize vertically aligned carbon nanotubes, i.e. as a carpet perpendicular to the substrate.

(b) Wet chemical synthesis of nanomaterials

In principle, the wet chemical synthesis of nanomaterials is classified into the following two broad groups:

1. The top-down method where single crystals are etched in an aqueous solution for producing nanomaterials, for example, the synthesis of porous silicon by electrochemical etching.

Lithography

At the moment, the most used top-down approach is photolithography. It has been used for a while to manufacture computer chips and produce structures smaller than 100 nm.

2. The bottom-up method is consisted of the sol–gel method, precipitation, etc. where materials containing the desired precursors are mixed in a controlled fashion to form a colloidal solution.

(a) Sol–gel process

A colloidal suspension (sol) is obtained from desired precursors mixed in a controlled fashion and gelation of the sol to form a network in a continuous liquid phase (gel). This is known as the sol–gel process. These are synthesized from precursors, a metal or metalloid element surrounded by various reactive ligands.

Initially, a dispersible oxide sol in contact with water or dilute acid is got. Removal of the liquid from the sol yields the gel, and the sol/gel transition controls the particle size and shape. Calcination of the gel produces the oxide.

Nanomaterials can be produced maintaining different conditions and procedures in the sol–gel method. A schematic representation of the sol–gel process of synthesis of nanomaterials is described in Fig. 19.20.

The precursors such as $Si(OEt)_4$ (tetraethyl orthosilicate or TEOS) is involved in the sol–gel process for the synthesis of nanomaterials. The reactions involved in the sol–gel chemistry based on the hydrolysis and condensation of metal alkoxides M(OR)z can be described as follows:

$$\text{MOR} + H_2O \rightarrow \text{MOH} + \text{ROH (hydrolysis)}$$

$$\text{MOH} + \text{ROM} \rightarrow \text{M-O-M} + \text{ROH (condensation)}$$

Nanomaterials are obtained from the sol–gel method and are widely used to prepare nanooxide materials.

The sol–gel process is unique and follows a series of distinct steps.

Alkoxide or solvated metal precursor solutions to be obtained.

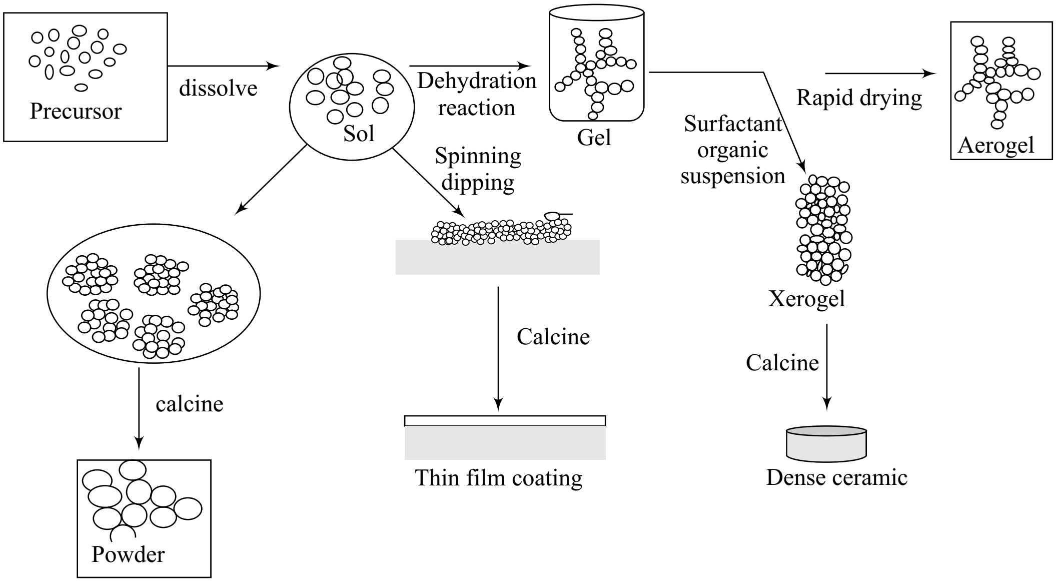


Figure 19.20 *Schematic representation of the sol–gel process of synthesis of nanomaterials.*

Gel formation from an oxide- or alcohol-bridged network by a polycondensation reaction that results in a dramatic increase in the viscosity of the solution.

Aging process of the gel is known as syneresis, during which the polycondensation reactions continue until the gel transforms into a solid mass, accompanied by contraction of the gel network and expulsion of solvent from gel pores. The ripening of gel where smaller particles are consumed by larger particles during the growth process; and there could be phase transformations process with syneresis. The aging process of gels can exceed 7 days and is critical to the prevention of cracks in gels.

Water and other volatile liquids are removed from the gel network in drying process of the gel. This process is complicated due to fundamental changes in the structure of the gel. The drying process takes place in the following four distinct steps:

- Constant rate period
- Critical point
- Falling rate period
- Second falling rate period

The thermal evaporation is done under isolated condition; the resulting monolith is termed a xerogel. Again, it can be noticed that if the solvent (such as water) is extracted under supercritical or near supercritical conditions, the product is an aerogel.

The elimination of surface-bound M–OH groups are achieved by dehydration process and stabilized against rehydration. This is normally achieved by calcining the monolith at temperatures up to 8000°C.

The densification of gel is done by heating them at high temperatures ($T > 8000°C$). The pores of the gel network are collapsed, and remaining organic species are volatilized. The interest in this synthetic method arises due to the possibility of synthesizing nonmetallic inorganic materials like glasses, glass ceramics or ceramic materials at very low temperatures compared to the high-temperature process required by melting glass or firing ceramics.

(b) Gas phase synthesis of nanomaterials

A very precise way of controlling size, shape and chemical composition of nanomaterials with controlled structures are achieved through the gas-phase synthetic (Fig. 19.21) methods.

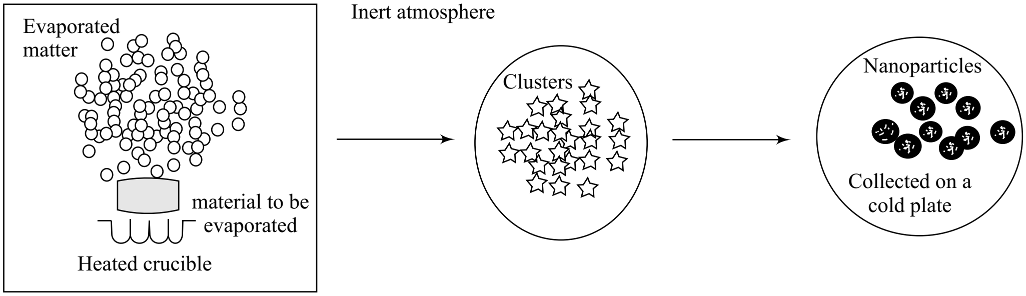


FIGURE 19.21 *Schematic representation of gas phase process of synthesis of single-phase nanomaterials from a heated crucible.*

Gas phase processes have inherent advantages, some of which are as follows:

- An excellent control of size, shape, crystallinity and chemical composition
- Highly pure materials can be obtained
- Multi-component systems are relatively easy to form
- Easy control of the reaction mechanisms

The synthetic routes adopted in gas phase are based on the production of small clusters that can aggregate to form nanoparticles (condensation). Condensation occurs only when the vapour is supersaturated and in these processes homogeneous nucleation in the gas phase is utilised to form particles. This can be achieved both by physical and chemical methods.

Furnace for gas phase synthesis

The procedure adopted to synthesize nanoparticles is to heat the desired material in a heat-resistant crucible. This technique is appropriate only for materials that have a high vapour pressure at the elevated temperatures (>2000°C).

Arc heating, electron beam heating or Joule heating supplies adequate source of energy to the precursor material. The atoms are evaporated into an atmosphere, which is either inert (e.g. He) or reactive (so as to form a compound). Materials with high vapour pressure have to be fed into the furnace in the form of a suitable precursor such as organometallics, which decompose in the furnace to produce a condensable material. Due to condensation process, these precursor atoms form small clusters through homogeneous nucleation.

In case a compound is being synthesized, these precursors react in the gas phase and form a compound with the material that is separately injected in the reaction chamber. The clusters would continue to grow if they remain in the supersaturated region. To control their size, they need to be rapidly removed from the supersaturated environment by a carrier gas. The cluster size and its distribution are controlled by only three parameters:

1. The rate of evaporation (energy input)
2. The rate of condensation (energy removal)
3. The rate of gas flow (cluster removal)

The above process is simple and it is possible to scale up this process from laboratory (mg/day) to industrial scales (tons/day).

(c) Flame assisted ultrasonic spray pyrolysis

The precursor materials are heated in a nebulizer using piezoelectric transducer (Fig. 19.22) and then the unwanted components of the reaction are got rid of by burning in flame to obtain the required material, e.g. ZrO_2 has been obtained by this method from a precursor of $Zr(CH_3CH_2CH_2O)_4$.

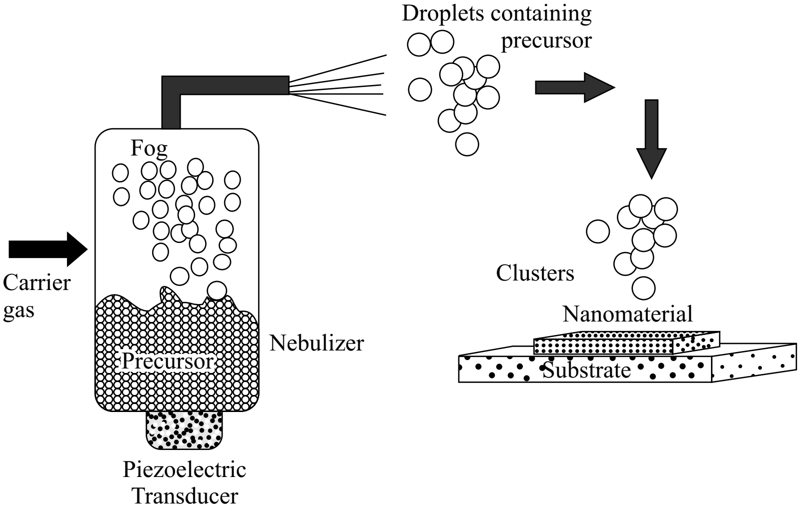


Figure 19.22 *Flame-assisted ultrasonic spray pyrolysis.*

The fused silica is manufactured from flame hydrolysis (variant) process. In the process, silicon tetrachloride is heated in an oxy-hydrogen flame to give a highly dispersed silica. The burning of a gas mixture, e.g. acetylene and oxygen or hydrogen and oxygen, supplies the energy to combustion flame synthesis to initiate the pyrolysis of precursor compounds; and is widely used for the industrial production of powders in large quantities, such as carbon black, fused silica and titanium dioxide. The resulting white amorphous powder consists of spherical particles with sizes in the range 7–40 nm. The basic idea of low pressure combustion flame synthesis is to maintain the desired pressure as used in gas phase synthesis and thus, to reduce or avoid the agglomeration.

A flat flame front in the synthesis of nanomaterial could lead to the formation of narrow size distributions of particles.

Under these conditions of thermal history, i.e. time and temperature, of each particle formed is identical and narrow distributions result. However, due to the oxidative atmosphere in the flame, this synthesis process is limited to the formation of oxides in the reactor zone.

(d) Gas Condensation Processing (GPC)

Nanomaterials are synthesized using gas condensation method (Fig. 19.23). It is also a unique technique in which a metallic or inorganic material [e.g. a suboxide] is vaporised using thermal evaporation sources such as crucibles, electron beam evaporation devices or sputtering sources in an atmosphere of 1–50 mbar helium.

Cluster of material is formed in the vicinity of the source by homogenous nucleation in the gas phase and grow by coalescence and incorporation of atoms from the gas phase.

The cluster (or particle) size of the material of interest depends critically on the time of residence the particles in the growth system, which can also be influenced by the gas pressure, the kind of inert gas and on the evaporation rate/vapour pressure of the evaporating material. With increasing gas pressure, vapour pressure

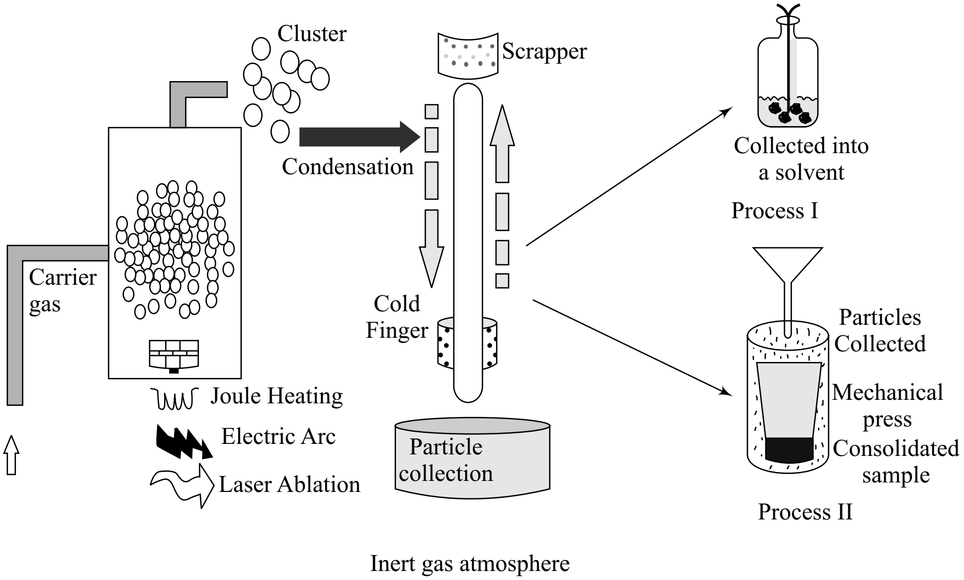


FIGURE 19.23 *Schematic representation of typical set-up for gas condensation synthesis of nanomaterials followed by consolidation in a mechanical press or collection in an appropriate solvent media.*

and mass of the inert gas used the average particle size of the nanoparticles increases. Originally, a rotating cylindrical device cooled with liquid nitrogen was employed for the particle collection: the nanoparticles in the size range from 2 to 50 nm are extracted from the gas flow by thermophoretic forces and deposited loosely on the surface of the collection device as a powder of low density and no agglomeration. Subsequently, the nanoparticles are removed from the surface of the cylinder by means of a scraper in the form of a metallic plate.

In the process, if evaporation of material can be done using metal crucibles (W, Ta or Mo), the synthesis of nanocrystalline pure metals is a simple process.

If the melting points of the metals are high or if the metals can react with the crucibles, such metals can be prepared by sputtering technique (i.e. for W and Zr), or laser ablation or electron beam evaporation. Synthesis of alloys or intermetallic compounds by thermal evaporation can only be done in the exceptional cases that the vapour pressures of the elements are similar. Cu/Bi or W/Ga composite material is synthesized by simultaneous evaporation from two separate crucibles onto a rotating collection device. In this process, it is noticed that excellent intermixing on the scale of the particle size can be obtained. However, control of the composition of the elements has been difficult and reproducibility is poor. A controlled post oxidation of nanoparticles of a pure metal (e.g. Ti) yields nanocrystalline oxide powders (e.g. TiO_2) or a sub oxide (e.g. ZrO to ZrO_2). Although the gas condensation method including the variations have been widely employed to prepare a variety of metallic and ceramic materials, quantities have so far been limited to a laboratory scale.

(e) Chemical Vapour Condensation (CVC)

The chemical vapour condensation process is schematically shown in Fig. 19.24. A 'hot wall reactor' is used in the chemical vapour condensation or the CVC process. Depending on the processing parameters of chemical vapour deposition (CVC), nucleation of thin film nanoparticles are observed. A major problem in the process is to obtain good film qualities.

The very intention of the CVC process is to adjust the parameter field during the synthesis in order to suppress film formation and enhance homogeneous nucleation of particles in the gas flow.

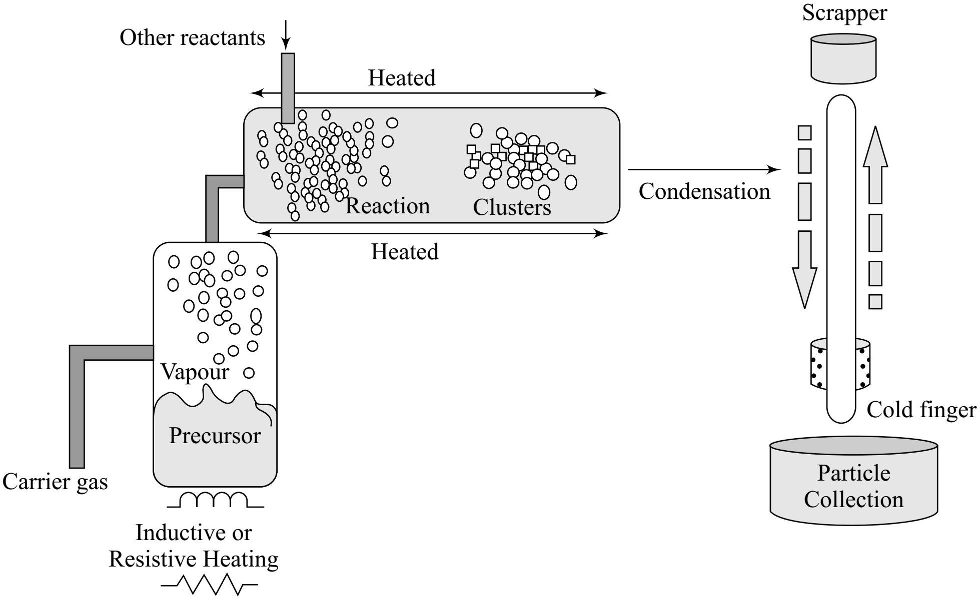


Figure 19.24 *A schematic of a typical CVC reactor.*

It is readily found that the 'residence time' of the precursor in the reactor determines if films or particles are formed. In a certain range of residence time both particle and film formation can be obtained. However, the residence time of the precursor molecules can be adjusted by changing the gas flow rate, the pressure difference between the precursor delivery system and the main chamber. Then, the temperature of the hot wall reactor results in the fertile production of nanosized particles of metals and ceramics instead of thin films. A metal organic precursor is introduced into the hot zone of the reactor using mass flow controller. A wider range of ceramics including nitrides and carbides can be synthesised with increased production quantity.

Similarly, complex oxides such as $BaTiO_3$ or composite structures can be synthesized, as well. In addition to the formation of single phase nanoparticles by CVC of a single precursor, the reactor allows the synthesis of

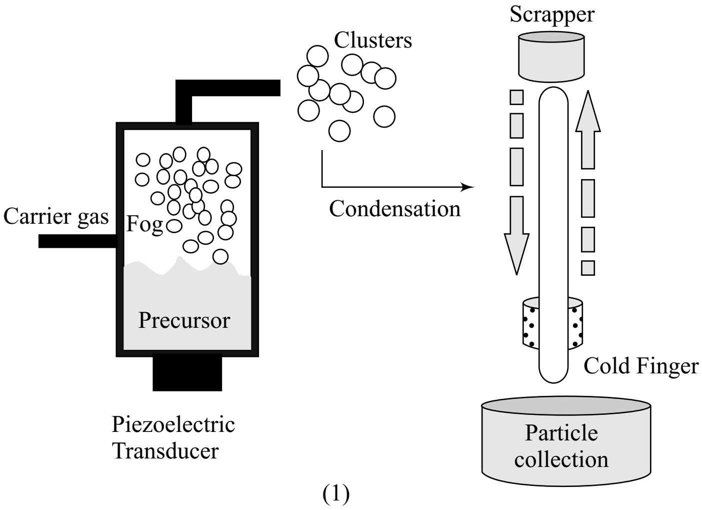
1. mixtures of nanoparticles of two phases or doped nanoparticles by supplying two precursors at the front end of the reactor, and
2. coated nanoparticles, i.e. n-ZrO_2 coated with n-Al_2O_3 or vice versa, by supplying second precursor at a second stage of the reactor. In this case nanoparticles which have been formed by homogeneous nucleation are coated by heterogeneous nucleation in a second stage of the reactor.

(f) Sputtered plasma processing

Sputtered plasma processing is again a variation of the gas-condensation method excepting the fact that the source material is a sputtering target and this target is sputtered using rare gases and the constituents are allowed to agglomerate to produce nanomaterial. Both dc (direct current) and rf (radio-frequency) sputtering has been used to synthesize nanoparticles. Again reactive sputtering or multitarget sputtering has been used to make alloys and/or oxides, carbides, nitrides of materials. This method is specifically suitable for the preparation of ultrapure and nonagglomerated nanoparticles of metal.

(g) Particle precipitation aided CVD

Materials in the form of colloidal clusters are used to prepare nanoparticles. The conditions of CVD are set in a such way that the clusters are formed by condensation in the gas phase and collect on to a substrate. The



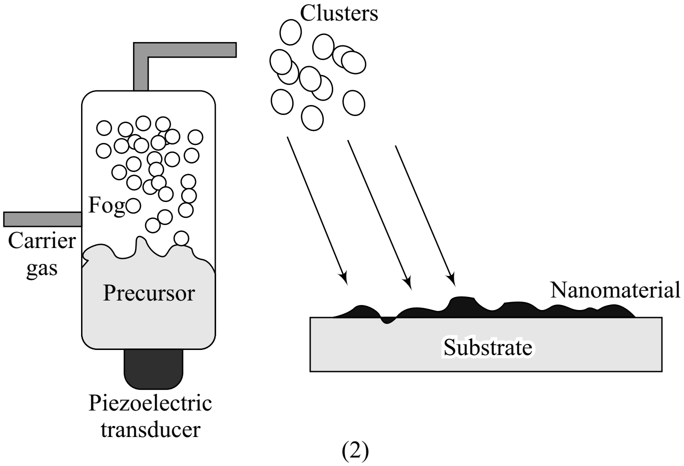


Figure 19.25 *Schematic representation of (1) particulate film and (2) nanoparticle formation.*

conditions maintained at the substrate are favourable for heterogeneous nucleation process. By this method (Fig. 19.25) both nanoparticles and particulate films can be prepared.

SnO_2 nanomaterials are prepared by this method known as pyrosol deposition process. In this process, clusters of tin hydroxide are formed into small aerosol droplets, following which they are reacted onto a heated glass substrate.

(h) Laser ablation

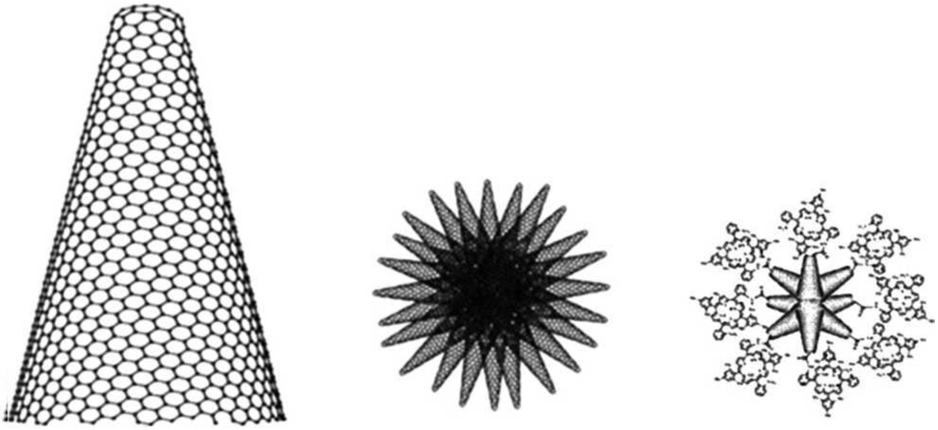
Nanoparticles and particulate films are prepared by a laser ablation technique. In this process, a laser beam is used as the primary excitation source of ablation for generating clusters directly from a solid sample in a wide variety of applications. An efficient tool is used for the production of ceramic particles and coatings by this process because of small dimensions of the particles and the possibility to form thick films. It is used as an ablation source for analytical applications such as the coupling to induced coupled plasma emission spectrometry and ICP. The formation of the nanoparticles has been explained following a liquefaction process which generates an aerosol, followed by the cooling/solidification of the droplets which results in the formation of fog. The conditions of the above process favour both the aerosol and the fog aggregation process and micrometre-sized particles are formed. The laser spark atomizer can be used to produce highly mesoporous thick films and the porosity can be modified by the carrier gas flow rate. ZrO_2 and SnO_2 nanoparticulate thick films were also synthesized successfully using this process with quite identical microstructure. Synthesis of other materials such as lithium manganate, silicon and carbon has also been carried out by this technique.

19.6.3.2.3 Other Nanostructured Materials/Particles

(i) Single-walled nanohorns (SWNHs)

The structure of single-walled carbon nanohorns has horn-shaped sheath aggregate of graphene sheets.

Carbon nanohorns (SWNHs) are consisted of single-walled, multi-walled carbon nanotubes (SWNTs and MWNTs), carbon onions and cones and belong to the family of carbon nanotubes [Fig. 19.26(a)]. The length of the SWNHs are about 40–50 nm and about 2–3 nm in diameter. They are derived from SWNTs and ended by a five-pentagon conical cap with a cone opening angle of ~20°. 'Dahlia-like' and 'bud-like' structured aggregates have an average diameter of about 80–100 nm are obtained when thousands of SWNHs are united together (Fig. 19.26b) The former consists of tubules and graphene sheets protruding from its surface-like petals of a dahlia, while the latter is composed of tubules developing inside the particle itself.



(a) Nanohorn (SWNH) (b) Aggregates of Nanohorns

FIGURE 19.26 *(a) Schematic representation a single carbon nanohorn and (b) carbon nanohorn aggregates.*

Their unique structures with high surface area and microporosity make SWNHs become a promising material for gas adsorption, biosensing, drug delivery, gas storage and catalyst support for fuel cell.

The following types of nanohorn materials are available:

- Carbon nanohorns (CNHs)
- Single-walled carbon nanohorns (SWCNHs)
- Double-walled nanohorns (DWNHs)
- Multi-walled nanohorns (MWNHs)
- Single-walled carbon nanohorn aggregates, boron nitride nanohorns (BNNHs)

Characteristics of nanohorns (SWNHs, DWNHs and MWNHs)

(a) Carbon nanohorns have the same graphene carbon atom structure like the CNTs.
(b) The main characteristic of the carbon nanohorns is that many nanohorns group together as an aggregate and form about 100 nm. The advantage is that when used as an electrode for a fuel cell, the surface area not only becomes extremely large, but it also becomes easy for the gas and liquid to permeate to the inside. In addition, compared with normal nanotubes, the nanohorns are easily prepared with high purity.
(c) It is expected to become a low-cost raw material.

Nanohorns properties

1. Electrical conductivity—probably the best conductor of electricity on a nanoscale level that can ever be possible.
2. Thermal conductivity—comparable to diamond along the tube axis.
3. Mechanical—probably the stiffest, strongest and toughest fibre that can ever exist.
4. Chemistry of carbon—can be reacted and manipulated with the richness and flexibility of other carbon molecules. Carbon is the basis of most materials we use every day.
5. Molecular perfection—essentially free of defects.
6. Self-assembly—strong van der Waals attraction leads to spontaneous roping of many nanotubes.

Synthesis

The techniques used to prepare single-walled nanohorns are high purity by CO_2 laser ablation and arc discharge processes without a metal catalyst. The size and purity of the SWNHs can be changed by varying the parameters such as temperature, pressure, voltage and current.

1. CO_2 laser ablation

The technique used to produce SWNHs is CO_2 laser ablation procedure at room temperature in absence of a metal catalyst. The CO_2 laser ablation generator is composed of a high-power CO_2 laser source and a plastic-resin reaction chamber attached with a vacuum pumping system, inlet and outlet gas valves and a ZnSe lens system to adjust beam intensity. The products of the reaction are removed by Ar gas which is circulated inside and flushed through the collection filter under the pressure of 760 Torr at room temperature. In the middle of the chamber, a graphite rod continuously rotates and advances along its axis so that a new surface could be exposed to the laser beam that is vertical to the rod and thus SWNHs are produced.

2. Arc discharge

A simple pulsed arc discharge between pure carbon rods in an atmospheric pressure of air and He (and Ar) for arcing period of 30 seconds produce single-wall nanohorns. The production of nanohorns is set with an arc current at 120 A and voltage between the electrodes is 15 V. The quality of SWNHs can be improved by preheated carbon rod at 10,000°C just before ignition. The arc soot deposited on the surface of the chamber is collected and characterized. By this method, the purity of obtained SWNHs is higher than 90 percent. The mean size of SWNH particles is about 50 nm, which is smaller than those prepared by the CO_2 laser method.

Applications

The manifold applications of nanohorns are as follows:

- Carbon nanohorn facilitates electron transfer process and is used as very promising.
- Material for chemical and biosensors.
- Functionalized carbon nanohorns show better dispersity and when bioconjugated, they can serve biomedical applications such as probing, imaging and drug delivering.
- Carbon nanohorns possess strong catalytic property, which can be applied to fuel cell fabrication.

- Due to their tremendous porosity, they are great materials for gas storage.
- They have high current capacity and stability, and they are applied in field emission.

Sensor materials

A single-wall nanohorn gas sensor could be fabricated with an electrokinetic method using dielectrophoresis (DEP). Upon exposure to ppm-levels of NO_2 or NH_3, respectively onto this sensor material, the sensor is active indicating an increased or decreased conductivity. The SWNH aggregate behaves as a *p*-type semiconductor. The comparison of SWNH and CNT gas sensor reveals that intrinsic NO_2 sensitivity of the SWNHs is lower than that of single-wall CNTs but comparable with the intrinsic sensitivity of multiple-wall CNTs.

A coated film of SWNHs is used as a gas sensor to detect ozone in water. This sensor is based on the phenomena that the electric resistance of the SWNH film decreased with the adsorption of ozone molecules due to charge transfer from the surface of SWNHs to O_3 molecules. The shift of the electric resistance of the SWNH film was correlated with the ozone concentration and temperature based on monolayer adsorption model with consideration of activation energies of the relevant adsorption, desorption and sensitivity of charge transfer.

An amperometric sensor containing SWNHs is prepared to detect hydrogen peroxide using carbon paste electrode. Instead of high surface area platinum electrode, SWNHs paste electrode is used for determination of hydrogen peroxide, demonstrating a metal-free and user-friendly electrochemical sensing method.

Hydrogen peroxide biosensor was fabricated using soybean peroxidase decorated SWNH-modified electrode based on the realization of direct electrochemistry of enzyme. In the absence of a mediator, this H_2O_2 biosensor exhibited a high sensitivity and a wide linear range.

- A glass carbon electrode modified with SWNHs displayed excellent electrochemical catalytic activities and could be used to simultaneously determine uric acid, dopamine and ascorbic acid in urine samples.
- A specially designed noncomposites modified with SWNHs have versatile biosensing applications.
- A sandwich containing nanohybrid of SWNHs–TiO_2–porphyrin is obtained via the dentate binding of TiO_2 nanoparticles to carboxylate groups. The nanohybrid showed an excellent electrocatalytic activity towards the reduction of chloramphenicol in neutral media. This is a highly sensitive and stable amperometric biosensor for chloramphenicol.
- A specific peptide-functionalized SWNH nanocomposite was used to fabricate an immunosensor towards microcystin LR. SWNH is a better nanomaterial indicating increased sensitivity of immunoassay.

2. Carbon nanohorn reinforced nanocomposites

To improve the mechanical properties such as modulus, tensile strength and failure strength of nanocomposites, they are reinforced with carbon nanotubes.

It is observed that adding CNT to polyacrylonitrile (PAN)/CNT fibre composites decreases the fibrillation tendency of these fibres. CNH-reinforced nanocomposites show larger surface area performance compared to CNT-reinforced nanocomposites. However, higher surface of CNHs compared to the CNTs results in the higher tendency of these structures for agglomeration which impedes the wide application of CNHs as a reinforcement for nanocomposites. It can be remarked that the tendency to agglomerate act as stress concentration sites which reduce the overall strength of nanocomposites. Furthermore, there are some reports on the adverse effect of presence of agglomerates on the physicochemical properties of nanocomposites.

3. Biomedical Engineering

Single-wall nanohorns possess distinctive dahlia-flower-like structure and desirable sizes (usually <100 nm). SWNHs are used as potential vehicles for intracellular delivery. They could be successfully isolated by a copolymer (Gum Arabic) through steric stabilization and in vitro study showed that the modified SWNHs are nontoxic and might be used as a promising vehicle for intracellular delivery.

In order to study the issue related to biomedical applications, the toxicity of SWNHs is critical, which has been investigated in vitro and in vivo extensively.

The SWNHs were found to be a nonirritant and a nondermal sensitizer through skin primary and conjunctival irritation tests and skin sensitization test.

Negative mutagenic and clastogenic potentials suggest that SWNHs are not carcinogenic.

The acute peroral toxicity of SWNHs was found to be quite low. Intratracheal instillation tests revealed that SWNHs rarely damaged rat lung tissue for a 90-day test period, although black pigmentation due to accumulated nanohorns was observed.

While further toxicological assessments, including chronic (repeated dose), reproductive, and developmental toxicity studies, are still needed, yet the present results strongly suggest that as-grown SWNHs have low acute toxicities.

Single-wall nanohorns find applications in antitumour drug delivery and therapy.

Oxidized form of SWNHs could entrap cisplatin, an anticancer agent, which was slowly released from the SWNHs in aqueous environments. The released cisplatin was effective in terminating the growth of human lung-cancer cells, while the SWNHs themselves had no such effect, showing that cisplatin-incorporated oxidized SWNHs are a potential drug delivery system.

Cisplatin was incorporated inside SWHNox by a new nanoprecipitation method involving dispersion of cisplatin and SWNHox in a solvent followed by the solvent evaporation. The incorporated cisplatin quantity increased from the previously reported value of 15– 46 percent, and the total released quantity of cisplatin also increased from 60 to 100 percent by changing the solvent from dimethylformamide to water. Concurrently, in vitro anticancer efficiency of cisplatin@SWNHox increased to 4–6 times greater than that of the intact cisplatin. In vivo, cisplatin@SWNHox intratumorally injected to transplanted tumours of mice suppressed the tumour growth more than the intact cisplatin. Cisplatin@SWNHox adhered to the cell surfaces in vitro and stayed within the tumour tissues in vivo. Therefore, the cisplatin released from SWNHox realized high concentrations locally at the cells in vitro and in the tissues in vivo and could efficiently attack the tumour cells.

An advantage of the interactions between VCM and SWNHox is beneficial to incorporate Vancomycin hydrochloride (VCM) into SWNHox structure for controlled release of VCM. Phospholipid–poly(ethylene glycol) was used to modify the hydrophobic surface of SWNHox to improve its dispersion in aqueous systems. In the release study using this complex, a stable release of VCM was achieved for an extended period.

The binding of polyethylene glycol to the hydrophobic surface of SWNHs is achieved to enhance their dispersibility in water for further application in drug delivery. Adsorption of polyethylene glycol-doxorubicin (PEG-DXR) conjugate onto SWNHox could form a water-soluble nanocomposite. SWNHs served as drug carriers to realize local cancer chemotherapy. When injected intratumorally, PEG-DXR SWNHs caused significant retardation of tumour growth associated with prolonged DXR retention in the tumour, showing that water-dispersed SWNHs were useful drug carriers for local chemotherapy.

It is essential to quantitatively determine biodistribution and ultrastructural localization in drug delivery. To achieve this, Gd_2O_3 nanoparticles were embedded within SWNH aggregates (Gd_2O_3@SWNHag) to facilitate detection and quantification. Gd_2O_3@SWNHag was intravenously injected into mice, and the quantities of Gd in the internal organs were measured by inductively coupled plasma atomic emission spectroscopy: 70–80 percent of the total injected material accumulated in liver. The high electron scattering ability of Gd allows detection with energy dispersive X-ray spectroscopy and facilitates the ultrastructural localization of individual Gd_2O_3@SWNHag with transmission electron microscopy. In the liver, the Gd_2O_3@SWNHag was localized in Kupffer cells but was not observed in hepatocytes. In the Kupffer cells, most of the Gd_2O_3@SWNHag was detected inside phagosomes, but some were in another cytoplasmic compartment that was most likely the phagolysosome.

4. Fuel cell

Pt nanoparticles can be impregnated on SWNH material to have great catalytic activity.

Pt nanoparticles (less than 5 nm) could be well dispersed on SWNHs. This nanohybrid with greater catalytic properties can be used for the power generation by polymer electrolyte fuel cell.

Electropolymerization of methylene blue (MB) onto SWNHs modified electrode can be used to build a fuel cell.

For the oxidation of glucose, glucose dehydrogenase was immobilized on the poly MB–SWNH-modified electrode. Employing Pt nanoparticles supported on functionalized TiO_2 colloidal spheres with nanoporous surface as cathode catalyst, the assembled glucose/O_2 biofuel cell operate at the physiological condition with good performance.

5. Gas storage

(a) Hydrogen storage

For the advancement of fuel cell power systems in transportation applications, hydrogen storage plays a key role. In order to protect environmental concern, there is growing demand for solid adsorbents which can store fuel gases such as hydrogen and methane at high density. The reason is that such vehicles have low emissions of CO_2 and environmental friendly. However, it is difficult to store these gases in a highly dense state, because supercritical gases do not condense to liquid at room temperature even under high pressure. Carbon materials such as graphite nanofibres (GNF), single-wall carbon nanotubes and modified carbon nanotubes are hopeful hydrogen storage candidates. The gas storage mechanism includes the following four different concepts:

- Physical adsorption
- Chemisorption
- Absorption
- Occlusion

Since, physical adsorption is completely reversible in nature and both adsorption and desorption rates are very large, it is the best suited mechanism to the application of fuel cells.

However, the ordinary storage capacity by physical adsorption is limited because of weak hydrogen–hydrogen and hydrogen–carbon interactions. Adsorption process through chemisorption is realized to have the high adsorption capacity, it is not reversible. On the other hand, absorption and occlusion are usually difficult in carbon materials because the carbon structure is rigid. Due to its high purity (>95 percent) of SWNH without any metal catalyst, SWNH has been thought to be an ideal candidate for hydrogen storage study, without any possible effect by metallic particles as catalyst on hydrogen storage capacity.

(b) Methane storage

Single-walled carbon nanohorns were synthesized via a gas-injected arc-in-water method. The duration of arc discharge and the electrode compositions are altered to increase the yield and methane adsorption properties of SWNHs. The nanohorn particles in the SWNH aggregates increased in size due to modified parameters of the experiment. Again, the thermal stability of SWNHs increased in an oxidative environment accordingly. SWNHs obtained using the above modified conditions adsorbed a larger amount of methane than did SWNHs obtained from the conventional synthetic conditions. Increased micropores with flexibility are created because of the effect of a mild oxidation treatment on SWNHs and the methane adsorption is more. By comparing the methane adsorption on the oxidized SWNHs to that on the pristine SWNHs, one can see that the oxidation of the SWNHs significantly increased the amount of methane adsorbed per apparent volume of SWNH bulk.

(6) Field emission

In the field emission process, the emission of electrons is induced by an electrostatic field. The development of large scale/low price production methods is one of the key problems among the tasks for optimization of the field emission. Pure carbon nanohorns are synthesized in large quantities. The carbon nanohorn thin films show good field emission characteristics due to the sharp horn-like structures, in particular, a low turn-on field and good long-term stability. One can observe a marked difference in the behaviour of nanotube films and the nanohorns when current densities higher than1 mA/cm^2 is passed through them. It is noticed that permanent damage will be caused to nanotube films whereas nanohorns can withstand densities that are higher by at least two orders of magnitude. This again may be due to the very peculiar structure and high resistivity of nanohorns. Since their long-term stability is comparable to that of nanotubes, nanohorns could represent an enticing alternative for field emission applications that do not require high current densities.

(ii) Nanowires

A nanowire has a nanostructure, with the diameter of the order of a nanometre (10^{-9} m).

The structure of nanowires has a thickness or diameter limited to tens of nanometres or less and an unconstrained length.

At these critical sizes, ‘quantum mechanical effects’ are important. These are known as quantum wires. Many different types of nanowires exist. They include the following:

- Metallic (e.g. Ni, Pt, Au)
- Semiconducting (e.g. Si, InP, GaN, etc.)
- Insulating (e.g. SiO_2, TiO_2)

Molecular nanowires are composed of repeating molecular units either organic (e.g. DNA) or inorganic (e.g. Mo_6S_{9-xIx}). In the future, miniature circuits can be obtained from the nanowires that are linked to tiny components. Utilizing nanotechnology processes, the following nanowires can be created out of chemical compounds.

Inorganic molecular nanowires (Mo_6S_{9-xIx}, $Li_2Mo_6Se_6$), which can have a diameter of 0.9 nm and be hundreds of micrometres long.

Other important examples are based on semiconductors such as InP, Si, GaN, etc., dielectrics (e.g. SiO_2, TiO_2), or metals (e.g. Ni, Pt).

Applications

There are innumerable applications of nanowires. They are

- electronic,
- optoelectronic and nanoelectromechanical devices,
- as additives in advanced composites, for metallic interconnects in nanoscale quantum devices, as field-emitters, and
- as leads for biomolecular nanosensors.

Syntheses of nanowires

(a) VLS growth

A nanowire is obtained by vapour–liquid–solid synthesis. Crystalline nanowires of some semiconductor materials can be obtained by the VLS method. In the process, the source material is either laser-ablated particles or a feed gas such as silane. VLS synthesis requires a catalyst. For nanowires, the best catalysts are liquid metal (such as gold) nanoclusters, which can either be self-assembled from a thin film by dewetting, or purchased in colloidal form and deposited on a substrate.

The source material enters these nanoclusters and begins to saturate them. On reaching supersaturation, the source particle solidifies and grows outward from the nanocluster. The final length of the nanowire can be adjusted to the desired length by closing the source. Switching sources while still in the growth phase can create compound nanowires with super-lattices of alternating materials.

A single-step vapour phase reaction at elevated temperature synthesises inorganic nanowires such as $Mo_6S_{9-x/x}$. From another point of view, such nanowires are cluster polymers.

(b) Solution-phase synthesis

The nanowires grow in solution-phase synthesis technique. This process can produce nanowires of many types of materials. The advantage of solution-phase synthesis is that it can produce very large quantities, compared to other methods. Ethylene glycol is both solvent and reducing agent in polyol synthesis. The process is particularly versatile at producing nanowires of lead, platinum and silver.

Semiconductor nanowires of Si and Ge are synthesized by supercritical fluid-liquid-solid growth method.

Si and Ge organometallic precursors are sent into a reactor filled with a supercritical organic solvent, such as toluene. On heating the degradation of the organic metallic precursor of Si and Ge occurs, releasing of Si or Ge. They dissolve into the introduced metal nanocrystals (as seeds). As more of the semiconductor solute is added from the supercritical phase (due to a concentration gradient), a solid crystallite precipitates, and a nanowire grows uniaxially from the nanocrystal seed.

(iii) Nanorods

Nanorods are materials having dimensions in nanoscale range (1–100 nm). These nanorods may be synthesized from metals or semiconducting materials. The standard aspect ratio (length divided by width) is 3–5 for nanorods. Nanorods are synthesised by a process called direct chemical synthesis. To control the shape of this nanomaterial, a combination of ligands acts as shape control agents and bond to different facets of the nanorod with different strengths. This allows different faces of the nanorod to grow at different rates, producing an elongated object.

Applications

A potential application of nanorods is in display technologies. This is because the reflectivity of the rods can be changed by changing their orientation with an applied electric field.

In microelectromechanical systems, nanorods, along with other noble metal nanoparticles, also function as theragnostic agents. Nanorods absorb in the near IR, and generate heat when excited with IR light. This property has led to the use of nanorods as cancer therapeutics. Nanorods can be conjugated with tumour targeting motifs and ingested. When a patient is exposed to IR light (which passes through body tissue), nanorods selectively taken up by tumour cells are locally heated, destroying only the cancerous tissue while leaving healthy cells intact.

Nanorods based on semiconducting materials are being used for application as energy harvesting and light emitting devices.

It is observed to have electric-field mediated tunable photoluminescence from ZnO nanorods, with potential for application as novel sources of near-ultraviolet radiation.

(i) ZnO nanorods

Zinc oxide nanorods are produced in the nanoscale; and found that its direct band gap energy is 3.37 eV, which is similar to that of GaN. It has an excitation binding energy of 60 meV. It is of interest to note that the optical band gap of ZnO nanorods can be tuned by changing the composition, morphology, size, etc. ZnO nanorods have wide variety of applications and are used to fabricate nanoscale electronic devices, including field effect transistor, ultraviolet photodetector, Schottky diode and ultra-bright light-emitting diode.

There are many methods available to fabricate the single crystalline, ZnO (Wurtite) nanorods. Among the methods, the most developed approach is to grow nanorods from vapour phase.

In a typical growth process, ZnO vapour is condensed onto a solid substrate. ZnO vapour can be generated by the following three methods:

- Thermal evaporation
- Chemical reduction
- Vapour–liquid–solid method

Commercial ZnO powder is mixed with SnO_2 in the 'thermal evaporation process' and heating the mixture is evaporated at elevated temperature.

In a typical chemical reduction method, the ZnO is reduced to zinc vapour and is sent to the growth zone followed by reoxidation of Zn vapour to ZnO.

The VLS process is the most commonly used process to synthesize single crystalline ZnO nanorods. In a typical process, catalytic droplets are deposited on the substrate and the gas mixtures, including Zn vapour and a mixture of CO/CO_2, react at the catalyst-substrate interface, followed by nucleation and growth. Typical metal catalysts involved are gold, copper, nickel and tin. ZnO nanowires are grown epitaxially on the substrate and assemble into monolayer arrays. ZnO nanorods are synthesized by using the metal-organic chemical vapour deposition (MOCVD) technique. There is no necessity of catalyst to execute the process and the growth temperature is maintained at 400–500°C, i.e. considerably milder conditions compared to the traditional vapour growth method.

(ii) Gold nanorods

A high-quality gold nanorod is synthesized commonly through a seed-mediated growth process. A citrate-capped gold nanosphere serves as seeds are added to the bulk $HAuCl_4$ growth solution. The reduction of $HAuCl_4$ solution is achieved with ascorbic acid in the presence of cetyl trimethyl ammonium bromide (CTAB) surfactant and silver ions which results in the formation of nanorod growth solution. In the absence of silver nitrate, longer nanorods can be obtained by use of a three-step addition procedure. Seeds are sequentially added to growth solution in order to control the rate of heterogeneous deposition and thereby the rate of crystal growth. The shortcoming of this method is the formation of gold nanospheres, which requires separations and cleanings. In one modification of this method, sodium citrate is replaced with a stronger CTAB stabilizer in the nucleation and growth procedures. Another improvement is to introduce silver ions to the growth solution, which results in the nanorods of aspect ratios less than five in greater than 90 percent yield. Silver of a lower reduction potential than gold, can be reduced on the surface of the rods to form a monolayer by underpotential deposition. Here, silver deposition competes with that of gold, thereby retarding the growth rate of specific crystal facets, allowing for one-directional growth and rod formation.

(iv) Carbon nanofibre

Carbon fibre (graphite fibre, or carbon graphite) is a material consisting of extremely thin fibres about 0.005–0.010 mm in diameter and composed mostly of carbon atoms.

Carbon nanofibres (CNF) are cylindrical nanostructures with graphene layers arranged as stacked cones, cups or plates. Carbon nanofibres with graphene layers wrapped into perfect cylinders are called carbon nanotubes.

The carbon atoms in carbon nanofibres are bonded together more or less aligned parallel to the long axis of the fibre. The alignment of these carbon atoms in crystal makes the fibre very strong for its size.

Several thousand carbon fibres are twisted together to form a 'yarn', which may be woven into a fabric. Carbon fibre produced by a process has many different weave patterns. Carbon fibre is very strong when stretched or bent, but weak when compressed or exposed to high shock (e.g. a carbon fibre bar is extremely difficult to bend, but will crack easily if hit with a hammer).

Carbon nanofibres have a very narrow inner cavity or practically no hollow channel in their centre, and the graphitic layers at the tips are defective or partially broken.

The diameter of the innermost carbon fibres is smaller (less than 1 nm) than MWCNTs and the radius of the curvature at the tip is less than 3 nm.

Catalytic chemical vapour deposition (CCVD) is a commercial technique for the fabrication of vapour grown carbon fibres (VGCF) and vapour grown carbon nanofibres (VGCNF). In the process, gas-phase molecules are decomposed at high temperatures and carbon is deposited in the presence of a transition metal catalyst on a substrate where subsequent growth of the fibre around the catalyst particles is produced. In general, this process involves separate stages such as gas decomposition, carbon deposition, fibre growth, fibre thickening, graphitization and purification. The process produces hollow fibres. The nanofibre diameter depends on the size of the catalyst. Carbon nanofibres can also be fabricated by thermal CVD using templates. Anodic aluminium oxide (Al_2O_3) with a thickness of 30 μm and pores of about 200 nm in diameter is used as the template. After deposition of carbon materials through the pores, carbon nanofibres remain after removal of the template. Carbon nanofibre can be used for filler in composite materials or electron emitters such as CNTs.

A transmission electron microscopy (TEM) image of nanofibre produced by hydrogen gas arc discharge is shown in Fig. 19.27. It shows an extremely narrow channel in the centre. The diameter of the innermost layer is only about 1 nm.

Fishbone carbon nanofibres can be manufactured in large quantity and at low costs. Their strength and graphite structure make them suitable for application as catalysts support materials.

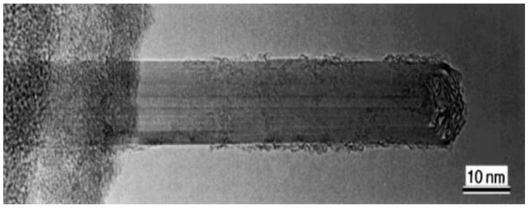


Figure 19.27 *Carbon nanofibre produced in a hydrogen gas arc discharge.*

A composite material known as carbon fibre reinforced plastic is synthesized by the combination of the carbon fibre weave patterns with a plastic resin, which are wound or moulded to get the composite material. This procedure provides carbon fibre reinforced plastic a high strength-to-weight ratio. The density of carbon fibre is also considerably lower than the density of steel, making it ideal for applications requiring low weight. The properties of carbon fibre such as high tensile strength, low weight, and low thermal expansion make it very popular in aerospace, civil engineering, military, and motorsports, along with other competition sports. However, these compounds are expensive because of the enhanced features when compared to similar materials such as fibreglass or plastic.

19.6.3.2.4 Films, coating, nanostructured surfaces

(a) Carbon-based nanofilms

Carbon-based nanofilms play a major role in coating industry and are very promising materials. Several properties can be tailored separately to suit desired specific applications so that the coatings can be optimized. They can be used for wear protection and to reduce friction. There are several elements which can be used to build up carbide nanofilms and nanocrystalline carbon films like 'diamond-like carbon' (DLC). Generally, the following elements such as C, Si, N, B, Ti are used in combination with nanocarbon. The combination of these elements with nanocarbon/others leads to the formation of several materials such as C_3N_4 and Si_3N_4 with different characteristics. The applications of these two groups of binary compound materials are of extra ordinary interest. These are covalent carbides like SiC and B_4C and metallic carbides like TiC or LaC_2.

Diamond-like carbon

There are seven different forms of diamond-like carbon exists. They exhibit some of the typical properties of diamond. Diamond-like carbon is amorphous carbon. The hardest of these is called tetrahedral amorphous carbon (ta-C), which ideally consists of purely sp^3- bonded carbon atoms.

The reason for DLC to be hard and tough is due to the inherent sp^3 bond configuration in the structure. In the sp^3 bond configuration, each carbon atom is connected by four strong (covalent) bonds to its neighbours. The sp^3 bonds are arranged symmetrically in three dimensions to form a tetrahedral lattice, similar to the structure of diamonds. In contrast, sp^2 bonds are covalent bonds arrayed at 120° around an atom in a plane, producing a two-dimensional lattice. Much weaker electronic forces hold each plane of the lattice together; this is the structure of graphite. However, some impurities are present in DLC containing some fraction of sp^2 bonds.

A highly useful film material tetrahedral amorphous carbon for controlling wear is used with a low friction coefficient having sp^3 content of more than 50 percent has a hardness of over 40 GPa. Some of the properties of DLCs are exploited by coating this material onto other materials. All seven contain significant amounts of sp^3-hybridized carbon atoms. Even diamond lattice can be found in two crystalline poly types. The usual one has its carbon atoms arranged in a cubic lattice, while the very rare one (lonsdaleite) has a hexagonal lattice. By mixing these poly types in various ways at the nanoscale level of structure, DLC coatings can be made that are amorphous, flexible and yet purely sp^3- bonded 'diamond'.

The tetrahedral amorphous carbon, or ta-C is the hardest, strongest and slickest in such a mixture. Such ta-C can be considered to be the 'pure' form of DLC, since it consists only of sp^3- bonded carbon atoms. Fillers such as hydrogen, graphitic sp^2 carbon, and metals are used in the other six forms to reduce production expenses or to impart other desirable properties. The various forms of DLC can be applied to almost any material that is compatible with a vacuum environment.

Production

To synthesize DLC material in processes like in plasmas, in filtered cathodic arc deposition, in sputter deposition and in ion beam deposition, high energy precursive carbons are used. They are rapidly cooled or quenched on relatively cold surfaces producing DLCs.

There may be some chances that the cubic and hexagonal lattices can be randomly intermixed layer by atomic layer, because there is no time available for one of the crystalline geometries to nucleate and grow at the expense of the other. Coatings on all materials are amorphous in nature. Without long range order, there are no brittle fracture planes in the coatings on the surface of material. Such coatings are flexible and conformal to the underlying shape being coated, while still being as hard as diamond. In fact this property has been exploited to study atom-by-atom wear at the nanoscale in DLC.

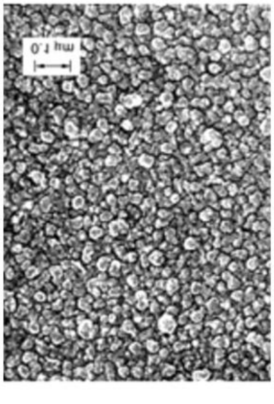


FIGURE 19.28 *SEM image of a ta-C 'diamond-like' coating.*

Structural elements in DLCs are not crystallites but are nodules of sp^3- bonded carbon atoms. The grains are so small that the surface appears mirror smooth to the eye. A classic 'medium' morphology for a ta-C film is shown in Fig. 19.28.

Properties

The diamond-like carbon coatings have abilities to provide some of the properties of diamond to surfaces of almost any material and these properties are of considerable interest for many applications. The primary desirable qualities are hardness, wear resistance and slickness. The properties of DLC film coating onto a surface depend upon which type of amorphous carbon (7 forms) is applied and the amounts and types of diluents added to reduce the cost of production. These [sp^3] bonds are also present in amorphous carbon where the atoms are in a random arrangement. In such a situation, there will be bonding only between a few individual atoms and not in a long-range order extending over a large number of atoms. The bond types in carbon materials have a considerable influence on the material properties of amorphous carbon films. If the sp^2 type is predominant the film will be softer, if the sp^3 type is predominant the film will be harder. Another contribution to the properties of carbon is the presence of fractional content of hydrogen in the amorphous carbon. A considerable percentage of hydrogen is found in the finished DLC material during the process of its production. It is not surprising to learn that fractions of hydrogen remaining in DLC films degrade them almost as much as do residues of sp^2- bonded carbon.

It is observed that 'pure' DLC material is ta-C and the materials obtained are degraded by diluents such as hydrogen, sp^2- bonded carbon and metals.

Tribological properties of DLC

- Due to its excellent tribological properties, DLC coatings are often used to prevent wear and tear.
- DLC is very resistant to abrasive and adhesive wear making it suitable for use in applications that experience extreme contact pressure, both in rolling and sliding contact.
- DLC coating is used on razor blades to prevent wear and metal cutting tools, including lathe inserts and milling cutters.
- DLC is used in bearings, cams, cam followers and shafts in the automobile industry. DLCs may also be used in chameleon coatings that are designed to prevent wear during launch, orbit, and reentry of land-launched space vehicles.
- DLC provides lubricity at ambient atmosphere and at vacuum, unlike graphite which requires moisture to be lubricious.
- Despite the favourable tribological properties of DLC, it must be used with caution on ferrous metals. If it is used at higher temperatures, the substrate or counter face may carburize, which could lead to loss of function due to a change in hardness. This phenomenon prevents the use of DLC coated machine tool on steel.

Applications

Typical applications of DLC are as follows:

- To utilize the ability of the material to reduce abrasive wear. Component of tools, such as endmills, drill bits, dies and moulds often use DLC coatings.
- DLC is also used as a coating on hard disk platters and hard-disk read heads to protect against head crashes.
- All of the multi-blade razors have the edges coated with hydrogen-free DLC to reduce friction, preventing abrasion of sensitive skin.
- DLC coats the cutting edges of tools for the high-speed, dry shaping of difficult exposed surfaces of wood and aluminium, for example on automobile dashboards.
- The implantable human heart pump can be considered the ultimate biomedical application where DLC coating is used on blood contacting surfaces of the key components of the device.

Unique medical applications of DLCs are found in aspercutanory inventions employing brachytherapy, which is due to benefits from the unique electrical properties of DLC. At low voltages and low temperatures electrodes coated with DLC can emit enough electrons to be arranged into disposable, micro-X-ray tubes as small as the radioactive seeds that are introduced into arteries or tumours in conventional brachytherapy. The same dose of prescribed radiation can be applied from the inside, out with the additional possibility to switch on and off the radiation in the prescribed pattern for the X-rays being used.

19.6.3.2.5 Nanostructured Bulk Materials

One of the most versatile elements is carbon, which has been widely studied in many of its common forms, such as graphite, diamond and its nanoscale materials. Few among the important nanoscale materials of carbon are carbon nanotubes, nanofibres, fullerenes and graphene. Carbon and nanostructured carbon is one of the most active and interdisciplinary research areas.

(a) Nanoporous carbon

Nanoporous carbon is used for separations of small molecules and shape selective catalysis; activated carbon nanospheres for electrodes; activated carbon to adsorb mercury from power plant flue gases; graphene for electronics; cellulosic carbon as a biofuel; nanoporous carbon derived from coal tar pitch for ultracapacitors; metal and carbon in electrocatalysts; multi-walled carbon nanotubes as frictionless bearings for NEMS and

MEMS, among many other research areas. Nanoporous carbon can be used as storage media for electrical energy because of its large surface area. The large surface area also makes it suitable for use as an anode in lithium ion batteries.

(b) Carbon nanofoam

Carbon nanofoam is the fifth known allotrope of carbon, the others being graphite, diamond, fullerene (e.g. C-60 molecules), and carbon nanotubes. The foam is, along with aerogel, one of the lightest known solid substances (with a density of ~2 mg/cm^3).

Carbon nanofoam is a low-density cluster assembly of carbon atoms strung together in a loose three-dimensional web; pure atomic and diatomic carbon; and linear acetylenic carbon, which is a one-dimensional carbon polymer with the structure –(C:::C)*n*–.

A method of synthesis of carbon nanofoam is to blast a high-power laser at disordered solid carbon which results in the formation of nanomaterial made of randomly interconnected clusters of carbon atoms with an average diameter of 6–9 nm. Carbon nanofoams are less denser than carbon and silicon aerogels.

The carbon nanofoam has other interesting properties:

- It is also a semiconductor, making it attractive for device applications.
- The most salient property of carbon nanofoam is its magnetism.
- The most unusual property, 'ferromagnetism', is displayed by carbon nanofoam. This ferromagnetic effect is temporary and vanishes a few hours after the nanofoam is made, though it can be preserved by cooling the nanofoam to extremely low temperatures of about –183°C. Consequently this 'ferromagnetic semiconductor' might have very useful applications for spintronics, an emerging field of devices based on a material's magnetic properties.

The complex microstructure of carbon nanofoam leads to this novel magnetic behaviour which is an intrinsic property of the material. Carbon atoms in the foam forms heptagon structures, seven-corner, seven-edge polygons that have an unpaired electron, one that does not form a chemical bond and has a magnetic moment which may lead to the magnetism. One possible application of the carbon nanofoam is in biomedicine, as tiny ferromagnetic clusters that could be injected in blood vessels may significantly increase the quality of magnetic resonance imaging pictures.

The carbon nanofoam is produced by firing a high-pulse, high-energy laser at graphite or disordered solid carbon suspended in some inert gas such as argon. Like aerogels, carbon nanofoam has extremely high surface area and acts as a good insulator, capable of being exposed to thousands of degrees Fahrenheit before deforming. It is practically transparent in appearance, consisting of mostly air and fairly brittle.

(c) Carbon aerogel

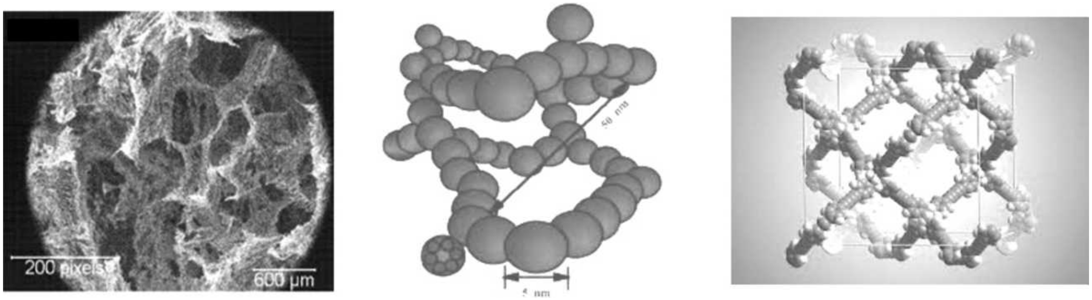
Carbon aerogels are composed of particles with sizes in the nanometre range, covalently bonded together.

A nanophase manmade material known as aerogels are synthesized with high porosity and surface area, varied chemical structures. They have fascinating useful properties.

An aerogel is formed when a wet sol–gel upon drying, releases entrapped liquid while maintaining a gel structure, thus resulting in a highly porous, air-filled gel. Different forms of carbon aerogels are presented in Fig. 19.29.

They have very high porosity (over 50 percent) with pore diameter under 100 nm and surface areas ranging between 400 and 1000 m^2/g. They are manufactured as nonwoven composite paper made of carbon fibres impregnated with resorcinol-formaldehyde aerogel and pyrolysed.

Carbon aerogels may be electrically conductive, depending on their density. Composite aerogel papers are useful for electrodes in capacitors or deionization electrodes. Carbon aerogels are used to create supercapacitors because of their high surface area, with values ranging up to thousands of farads based on a capacitance



(a) (b) (c)

Figure 19.29 *(a–c) Different forms of carbon aerogels.*

of 104 F/g and 77 F/cm³. Carbon aerogels are also extremely 'black' in the infrared spectrum, reflecting only 0.3 percent of radiation between 250 nm and 14.3 μm, making them efficient for solar energy collectors.

Few characteristic properties of silica aerogel are presented in Table 19.3.

Extremely low thermal conductivity	9–12 mW/MK
Extremely low thermal conductivity	9–12 mW/MK
High porosity	>90% air
Nanosized pores	20–40 nm
High surface area	~750 m^2/g
Very low tap density	30–100 kg/m^3
High oil absorption capacity (DBP)	540 g/100 g
Specific heat capacity	KJ/kg 7–1.15
Variety of particle sizes	5 μm to 4 mm
Surface chemistry	Completely hydrophobic
Opacity	Translucent, IR opacified and opaque

Table 19.3

Silica aerogel characteristics

Carbon aerogel monoliths and powders

There is need for different particle sizes of carbon aerogel powders for various powdered carbon applications, as desired. The standard material has a density of 0.6 g/cm^3 and a surface area of 400 m^2/g.

Unique high surface carbon powders

1. A high surface area carbon powders are inherent to the sol–gel process. However, such procedures have been used to produce carbon aerogel powders with approaching 2500 m^2/g. The characterization of the primary carbon particles that make up the aerogel network shows that they are composed of nanocrystalline, graphite-like ribbons that are inter-twined to form the particles.
2. Apart from high surface area carbon powders, carbon aerogels have oxygen-free surfaces that can increase their effectiveness in some applications.

Carbon WOI DOI papers

This carbon WOI DOI papers consist of a nonwoven carbon paper impregnated with a carbon aerogel. They typically have a density of 0.4 g/cm^3 and a surface area of 400 m^2/g. The pyrolysis from the RF aerogel sheet composite was conducted in N_2 at 1050°C. These composite sheets are suitable for direct use as capacitors or deionization electrodes, etc. and have some degree of flexibility.

Carbon aerogel typical applications

Carbon aerogels have found use as electrode materials in several electrochemical devices, most notably, aerocapacitors and electrosorption processes.

They find applications as: composite structures, reinforcing agents for organic rubbers, pigments for ink jet printers, electro-chemical storage devices, media for gas separations or storage, high temperature insulation, carriers for controlled release agents, electrodes for fuel cells and noble metal catalyst supports.

(d) Bulk carbon nanocrystals

Bulk carbon nanocrystals are crystals of nanometre dimensions, usually consisting of aggregates of a few hundred to tens of thousands of atoms combined into a cluster. A carbon nanocrystal is a material composed of atoms in polycrystalline arrangement.

Luminescent nanocrystals have generated considerable interest because of their potential applications in biolabelling and medical diagnosis. However, many luminescent nanocrystals contain heavy metals, causing concerns about their toxicity. A water-soluble carbon nanocrystal was fabricated from graphite which exhibits electrochemiluminescence. The favourable properties of these carbon nanocrystals, including low cytotoxicity, low environmental impact and strong optical luminescence, make this material ideal for building biosensors.

Nanocrystals were obtained from graphite rods using an electrochemical cell as depicted in Fig. 19.30.

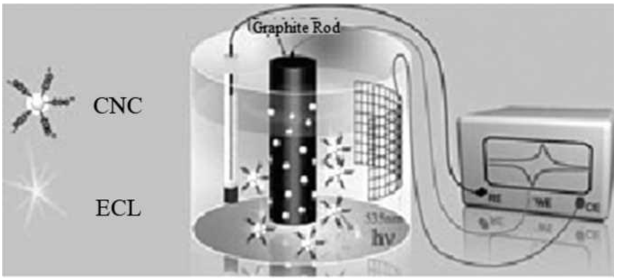


Figure 19.30 *Schematic illustration of electrochemical synthesis of nanocrystal.*

Within the porous graphite material, it is found that the nanocrystals are immobilized. On immersion in a buffer solution, the outer nanocrystals solution was oxidized. This process makes them soluble in water and allowing the nanocrystals to be released into solution. This process produced two sizes of nanoparticles: well-defined spherical nanoparticles of 20 nm in diameter, and smaller crystals of around 2 nm in size (Fig. 19.31). The surfaces of the smaller crystals were found to be covered in carboxyl groups, which facilitate labelling in biological applications.

Electrochemiluminescence is a property of a carbon nanocrystal in solution, which could be seen as optical emissions excited by the application of a voltage across an electrochemical cell at voltages more negative than –0.6 V and more positive +1.0 V, with a peak wavelength of emission of 535 nm. This

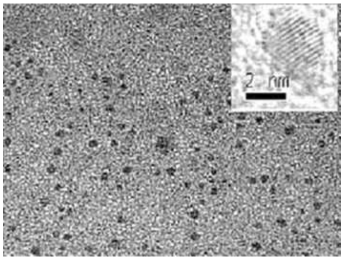


Figure 19.31 *Transmission electron microscopy image of carbon nanocrystals.*

emission peak (535 nm) differs from the photoluminescence peak of carbon nanocrystals (455 nm) because the electrochemiluminescence is due to excited states formed at the surfaces of the nanocrystals, whereas photoluminescence originates from band gap emissions inside the crystals. It throws light to the fact that the combination of positively and negatively charged nanocrystals in the electrochemical cell produces electronically excited nanocrystals that can then emit light by electrochemiluminescence. This mechanism is also observed in inorganic nanocrystals such as zinc selenide and cadmium telluride.

A new type of electrochemiluminescent nanocrystal was prepared with strong light emission, low cytotoxicity, excellent water solubility, good stability, easy labelling, low cost, and environmental friendliness. It should find exciting applications in the construction of highly sensitive and 'biocompatible electrochemiluminescent biosensors'.

(e) Nanocrystalline diamond

Nanocrystalline diamond (NCD) in an amorphous carbon matrix deposited in a nanocomposite film (carbide nanofilms) combines the properties of diamond with very low friction, high toughness and biocompatibility.

Nanocrystalline diamond offers great potential for many micromechanical systems, particularly where low fiction, thermal mismatch and mechanical integrity are of great importance. The reduction of grain size in NCD films to a few nanometres is, however, accompanied by a significant change in morphological, chemical, mechanical and wear properties. The evaluation of these properties is therefore necessary before designing NCD-coated micromechanical components for which reliability, wear and long lifetime are critical issues. Applications of nanocrystalline diamond are in the medical and health area of replacement joints. A further advantage of these materials is that the adjustment of the properties of the films can be done without the need to adapt the production process.

The well-established CVD, PECVD, PVD techniques have been adopted for the preparation of nanocrystalline diamond material. There is need for the investigation of the composition and structure of the films produced as well as the properties of the film. Applications are also expected in scratch-resistant surfaces with a low friction coefficient where e.g. the thin layers have a good balance of grip hardness and friction.

Conducting polymers may be able to be optimise electrical energy storage applications and may have higher electrical capacitance than pure CNT films or pure electrically conducting polymers.

19.7 Quantum Dots

Quantum dots (QDs) are extremely small semiconductor structures, usually ranging from 2 to 10 nm in diameter.

Within the regime of these critical nanoscale dimensions, quantum dots exhibit distinctly different behaviour from their bulk form in optical, electronics and magnetic properties. Generally, quantum dots fall into the category of nanocrystals and include 'quantum rods and quantum wire'. Quantum dot nanocrystalline semiconductors are the most widely used materials. A quantum dot is a nanoparticle made of any semiconductor material such as silicon, cadmium selenide, cadmium sulphide or indium arsenide. At these small nanosizes, materials behave differently giving quantum dots unprecedented tunability and wide variety of applications. Quantum dots have many unique properties and show interesting phenomena, such as size-dependent emission wavelength, narrow emission peak and broad excitation range.

Quantum dots are tiny man-made crystals. Quantum dots are actually very powerful devices and its size that gives them a unique ability to convert light into nearly any colour in the visible spectrum with very high efficiency. Each quantum dot is actually a tiny semiconductor—which means it can convert incoming energy. The electronic characteristics of quantum dots are determined by their size and shape. This means we can control the colour of light given off by a quantum dot just by changing its size. Bigger size quantum dots emit longer wavelengths like red, while smaller quantum dots emit shorter wavelengths like green. Quantum dots are semiconductor nanoparticles easily tunable by changing the size and composition of the nanocrystal.

The energy band gap of quantum dots depends on the particle size of the material. This aspect is the most predominant feature of quantum dots. By changing the particle size, the band gap will change, and so will the colour of light emitted. This occurs because the specific wavelength of light emitted is a function of the band gap energy.

When the bulk material is reduced to the size of nanoscale, the energy band is broken into 'discrete levels' so that the energy band gap is widened as shown in Fig. 19.32. This means that both absorption and emission wavelength of light become shorter. The smaller the size of quantum dots, larger is the band gap energy. Thus, by simply altering the size of materials, the energy band gap can be changed, e.g. accordingly, the colour of light emitted can be changed. This is because the emission colour from quantum dot depends on band gap energy. Varying the size and composition of quantum dots determines the colour of emitted light. Cadmium sulphide (CdS), zinc selenide (ZnSe), cadmium selenide (CdSe), indium phosphor (InP) and indium arsenic (InAs) are examples of quantum dots. Figure 19.32 shows schematic relationships between band gap and photon energy and colour of quantum dots according to size of dots.

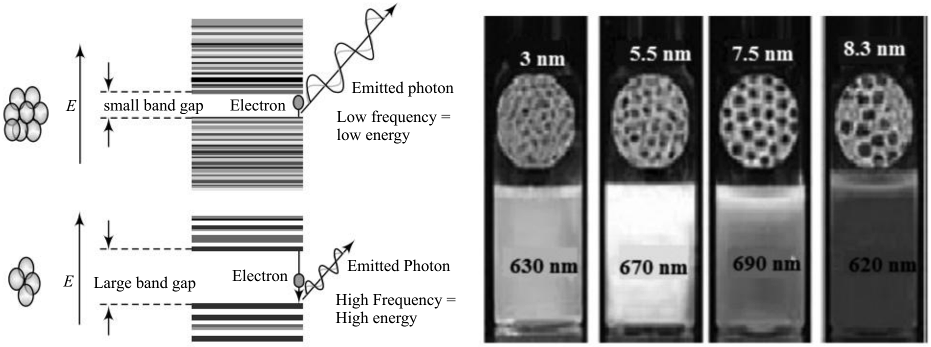


Figure 19.32 *Schematic diagrams of band gap and photon energy, and changing colour of quantum dots according to size of dots.*

Generally, quantum dots have a core/shell structure. The diameter of the semiconductor core is smaller than the material's exciton Bohr radius (>2–10 nm), which is a material-specific property. For example, the exciton Bohr radius of CdSe quantum dot is 5.3 nm. The shell is composed of several atomic layers of an inorganic wide band gap semiconductor. The core/shell structure results in brighter emission when the shell is of a different semiconductor material with a wider band gap than the core semiconductor material.

As such, the emissions of wavelengths of light from semiconductor cores are relatively weak and quite unstable. By encapsulating these cores with a high band gap material that is structurally related, the emission intensity and stability are enhanced. For example, the CdSe/CdS core/shell structure quantum dots enhance the property of plain CdSe quantum dots.

The biological applications are limited because these core/shell quantum dots are not soluble in water. To overcome this, an outer coating, such as a polymer layer, is added to make the quantum dot water-soluble and allow for conjugation to the biomaterial. Figure 19.33 shows the schematic diagram of a polymer-coated core/shell quantum dot for conjugation to biomaterials.

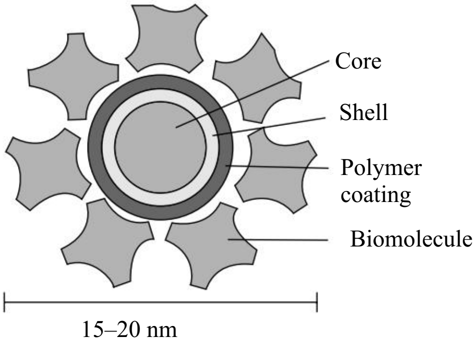


Figure 19.33 *The schematic diagram of polymer coated core/shell quantum dot.*

19.7.1 Fabrication Methods for Quantum Dots

There are several methods for fabrication of quantum dots:

- molecular beam epitaxy (MBE) method,
- electron beam lithography and molecular beam epitaxy (MBE), metal organic chemical vapour deposition (MOCVD) method, and
- colloidal synthesis.

Each method has some advantages and different applications.

(i) Colloidal process

A low-cost mass production process creates quantum dots in a liquid. Due to size-specific properties and flexible processing chemistry, colloidal semiconductor quantum dots have great interest for applications. The size control of quantum dots are achieved by the time duration of the process in which quantum dots remain in the solution.

Cadmium oxide, selenium powder, tributylphosphine (TBP), sulphur powder, trioctylphosphine oxide (TOPO), 1-octadecene (ODE), oleic acid (OA), octyldecylamine (ODA) and stearic acid are the materials taken for the synthesis of the CdSe/CdS core/shell quantum dots.

For synthesis of CdSe core nanocrystals, mixtures of CdO, stearic acid, and ODE are first heated up to 200°C to produce a colourless solution. After cool-down to room temperature, ODA and TOPO are added and then heated up to 280°C with argon (Ar) flow. A solution of dissolved selenium (Se) in TBP and ODE is then added and CdSe nanocrystals (3.5 nm) are precipitated. The second step is fabrication of CdS shell on the core surface. Solutions of cadmium oxide (CdO) dissolved in oleic acid and sulphur dissolved in ODE are sequentially added to finish the core/shell structure. Washing with methanol and hexane, and centrifuging and are two methods used to purify and extract quantum dots. Figure 19.34 shows the process flow for synthesis of CdSe/CdS structure.

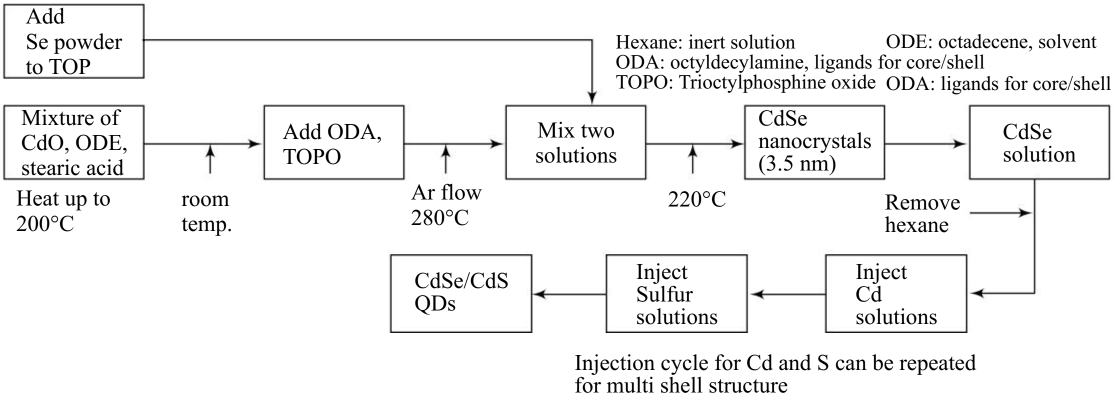


FIGURE 19.34 *Chart flow for synthesis of CdSe/CdS structure.*

(ii) Molecular beam epitaxy (MBE)

Molecular beam epitaxy is a process that deposits layer by layer of crystals on a substrate. Pure elements are evaporated under a vacuum condition, then condensed and allowed to settle on the surface of the substrate. In the process, evaporated atoms of the precursor do not interact with each other or gases of vacuum chamber until they reach the substrate. For indium arsenic/gallium arsenic (InAs/GaAs) systems, In and As atoms are deposited on GaAs substrate to form an epitaxial layer (Fig. 19.35). A large band gap material, such as InGaAs, is then deposited on top of the InAs epitaxial layer. Through post deposition heat treatment, quantum dots are formed by self-assembly.

A combination of electron beam lithography and molecular beam epitaxy are used to fabricate quantum dots. The electron beam is used to etch a pattern onto the substrate, and then MBE is used to deposit quantum dot materials on top. Usually the pattern is an array of holes, and these holes act as preferential nucleation sites for dots. The cost is high, but the position of dots can be controlled.

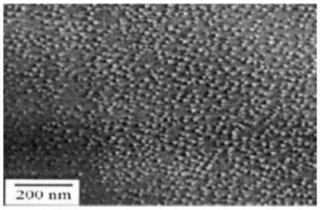


Figure 19.35 *Quantum dot arrays of InGaAs/AlGaAs.*

19.7.2 Applications of Quantum Dots

Semiconductor quantum dots (CQ) can be tuned and adjusted for many applications, including optics, medicine, and quantum computation. Quantum dots are used in biolabelling, in anticounterfeiting applications to create special inks, in dyes and paints, in light displays, and in chemical sensing.

(a) Life science

Innumerable applications of quantum dots are found in life science. Cadmium telluride (CdTe) and CdSe nanocrystals are of great industrial interest for developing photoluminescence-based biomedical labelling reagents. Noncadmium quantum dots can be used for in vivo imaging (Fig. 19.36). Quantum dots preferentially stain the vascular and lymphatic system, tumours, etc. in large multicellular animals such as mice. Western blotting, cell staining, flow cytometry, and cellular uptake are other examples of quantum dot applications. Figure 19.36 shows an application of noncadmium quantum dots for life science: in vivo imaging with preferential staining in specific parts.

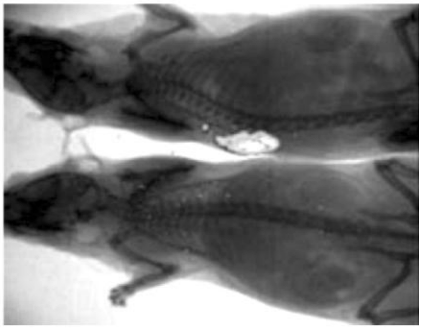


Figure 19.36 *Quantum dots application for in vivo imaging.*

(b) Display

The applications of quantum dots are both in organic and inorganic light emitter devices such as electroluminescence (EL) devices.

The quantum dot material is dispersed within or between phosphor layers and emits a specific colour of light. Quantum dots can be applied over large areas using liquid phase deposition techniques such as roll-to-roll printing and spin coating. The emission of colours of nanophosphors based on quantum dots, which can be tuned by changing the size of the quantum dot. Also, quantum dots are utilized to enhance light-emitting diodes and can emit almost any colour. Figure 19.37 shows an application of quantum dots for a colour display device.

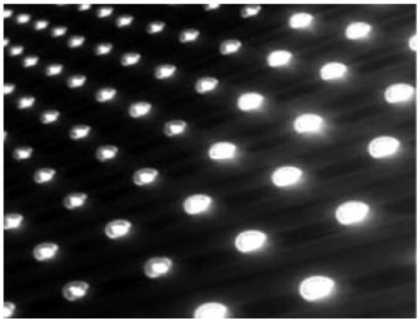


Figure 19.37 *Quantum dots application for display device.*

(c) Photovoltaic devices

The properties of quantum dots can be tuned to desired sizes and band gaps. Interestingly, it is possible to get multiple energy band gaps for quantum dots containing different size materials. Different layers of different band gaps can then be used to reduce loss of incident light. Multiple band gaps can absorb photons from several different wavelengths according to their band gap energies. This allows increased photon-to-electron conversion efficiency in photovoltaic devices. Also, quantum dots are cost-effective and can be incorporated into semiconductor polymers or inks and deposited onto low-cost substrate by roll-to-roll printing techniques.

(d) Photonics and telecommunications

The properties of quantum dots provide an opportunity to exploit and develop optical switches, modulators, and other nonlinear optic devices. The reason for this behaviour is that quantum confinement enhances nonlinear absorption and nonlinear refractive index.

The optical switches and logic gates that work faster than 15 terabits per second can be achieved when quantum dots are being used.

The quantum dot laser device is another potential application. The tunable band gap of quantum dots makes the laser wavelength changeable. Also, lasers with quantum dots have potential for very low-threshold current density, which is the current required to stimulate the laser or create output power from the device. The most studied material for the quantum dot laser are InGaAs/AlGaAs, GaAs/InAs. For laser application, single- or multiquantum dot layers are used. Figure 19.38 shows the structure of a quantum dot laser with multiple active layers.

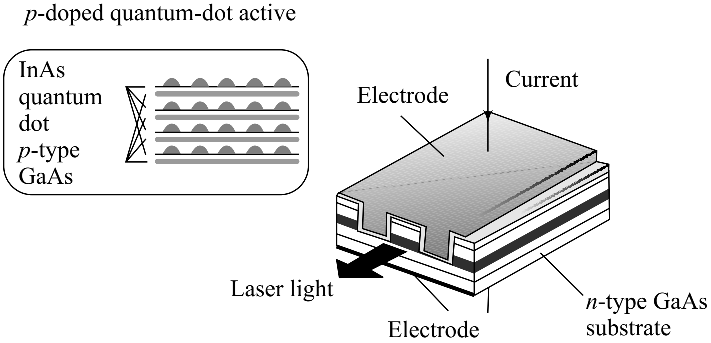


Figure 19.38 *Structure of a typical quantum dot laser.*

(e) Security inks

For security and anticounterfeiting applications, inks and paints incorporated with quantum dots can be applied to many types of surfaces, including papers, plastics and metals. The wide range of colour combinations of multiple quantum dots and other pigment creates a unique visible image that identifies any subject or document when exposed to UV light (Fig. 19.39).

Adjusting the quantum dots can produce info-ink that consists of CdSe quantum dots, toluene and polystyrene.

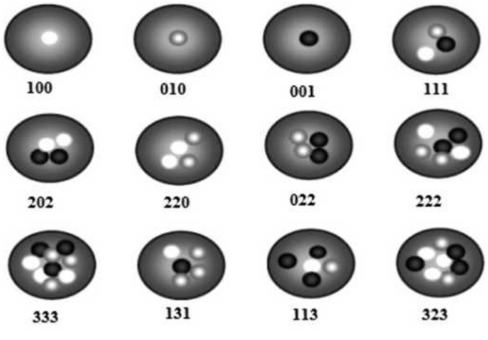


Figure 19.39 *Illustrates schematically the samples of info-inks consisting of three different quantum dots with different emission wavelengths.*

19.8 Organic Nanoparticles

Dendrimers

Dendrimers are macro nanomaterials consisting of particles ranging from 1 to 100 nm. A variety of dendrimers have been created with 3D structures having novel predetermined properties for device technology.

A dendrimer is a macromolecule, which is characterizes by its high branched 3D structure that provides a high degree of surface functionality and versatility. Its structure is always built around a central multifunctional core molecule with branches and end groups.

A nanoscale fabrication process is used for synthesises of three-dimensional macromolecule. A dendrimer is built up from a monomer, with new branches added in steps until a tree-like structure is created (dendrimer, meaning tree), Fig. 19.40. A dendrimer is technically a polymer.

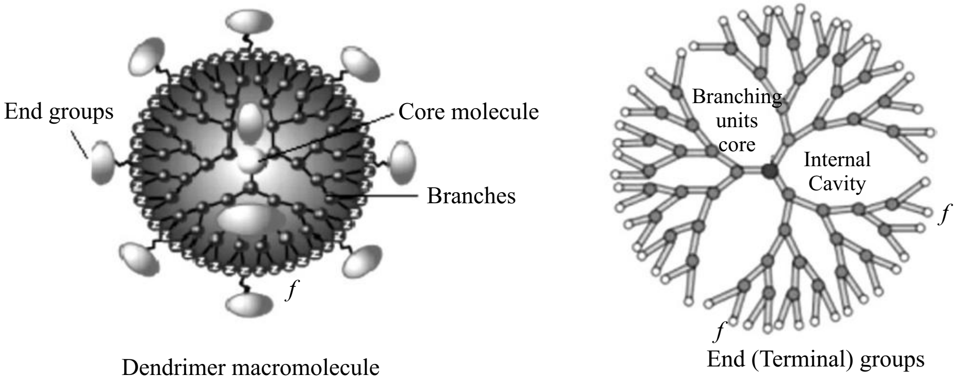


Figure 19.40 *A typical dendrimer assembly.*

For the formation of dendrimers, the successive addition of layers of branching groups is carried to construct them. Each new layer added is called a 'generation'. The final generation incorporates the surface molecules that give the dendrimer the desired 'function' for pharmaceutical, life science, chemical, electronic and materials applications.

The nanometre-sized, polymeric systems are hyperbranched materials having compact hydrodynamic volumes in solution and high, surface, functional group content. They may be water-soluble but, because of their compact dimensions, they do not have the usual rheological thickening properties that many polymers have in solution. Dendrimers, the most regular members of the class, are synthesized by step-wise convergent or divergent methods to give distinct stages or generations.

Dendrimers are globular macromolecules that consist of two or more tree-like dendrons emanating from either a single central atom or atomic group called the core.

The main building blocks of dendrimer molecular architecture are the branch cells which can be considered as three-dimensional dendritic repeat units that always contain at least one branch junction (Fig. 19.41).

Dendrimers are defined by their three components: a central core, an interior dendritic structure (the branches), and an exterior surface (the end groups) (Fig. 19.42).

In an ideal dendrimer structure (Fig. 19.42), these branch cells are organized in a series of regular, radially concentric layers (called generations) around the core. Each of these layers contains a mathematically precise number of branch cells which increases in a geometrically progressive manner from the core to the dendrimer exterior (often also referred to as the dendrimer surface). The outermost branches end with the end-groups, which can be either chemically inert or reactive. Their number depends on the particular dendrimer composi-

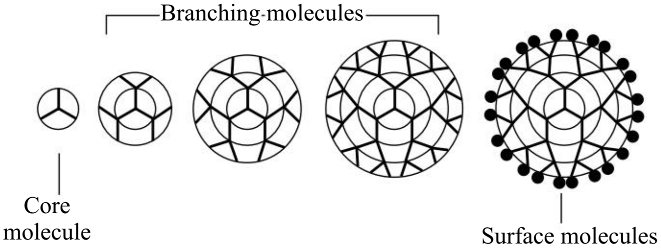


FIGURE 19.41 *Formation of dendrimer with core, branches and end functional groups.*

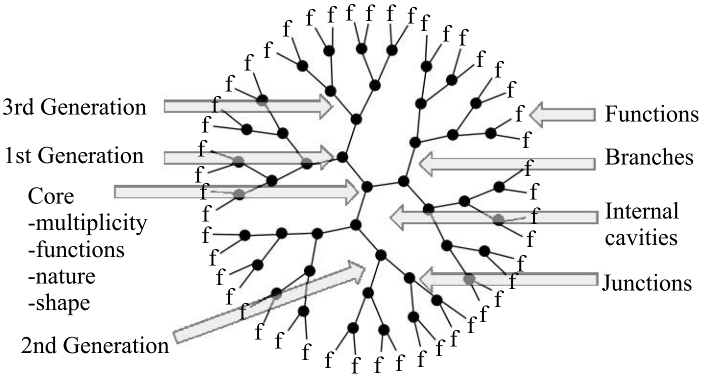


FIGURE 19.42 *A Dendrimer consisting of a central core, branches and functions.*

tion (i.e. the branching functionality at each generational level) and generation, and may range from only a few (i.e. three or four functional groups of common cores) to several hundreds or even thousands at very high generations. Because of this, high generation dendrimers may be viewed as globular, reactive, nanoscopic entities with some of the highest density of functionality known to chemistry.

Dendrimers provide a high degree of surface functionality and versatility. Since dendrimers are synthesized from branched monomer units in a stepwise manner, it is possible to conduct a precise control on molecule size, shape, dimension, density, polarity, flexibility and solubility by choosing different building/branching units and surface functional groups. Moreover, they can use small organic molecules and polymers as structural components, and thus acquire special physical and chemical properties.

Dendrimers of lower generations (0, 1 and 2) have highly asymmetric shape and possess more open structures as compared to higher generation dendrimers. As the chains growing from the core molecule become longer and more branched (in four and higher generations) dendrimers adopt a 'globular structure'. Dendrimers become densely packed as they extend out to the periphery, which forms a closed membrane-like structure. When a critical branched state is reached dendrimers cannot grow because of a lack of space. This is called the 'starburst effect'.

In divergent method, Dendrimers are synthesized by the repetitive addition of branch cells in the radial molecular direction, one layer or generation at a time, either from the core to the outer surface or from the end-groups to the core (convergent method). With rigorous process control and purification procedures after each iterative growth step, dendrimers can be obtained in very high purities and monodispersities of shapes and sizes, even at truly 'polymeric' high molecular weights, which can reach into several hundreds of thousands, or even a million. Their sizes increase with generations in a manner that leads to the crowding of the end-groups, which in turn, forces dendrimers to adopt spheroidal or globular shapes at sufficiently high generations.

Dendritic polymers are recognized as the fourth major class of macromolecular architecture (together with linear, cross-linked and branched architectures) (Fig. 19.43), consisting of four subcategories:

- random hyperbranched,
- dendigrafts,
- dendrons, and
- dendrimers.

During the last two decades, dendritic polymers (IV), particularly, dendrimers and hyperbranched, have become one of the fastest growing areas of interest in polymer science.

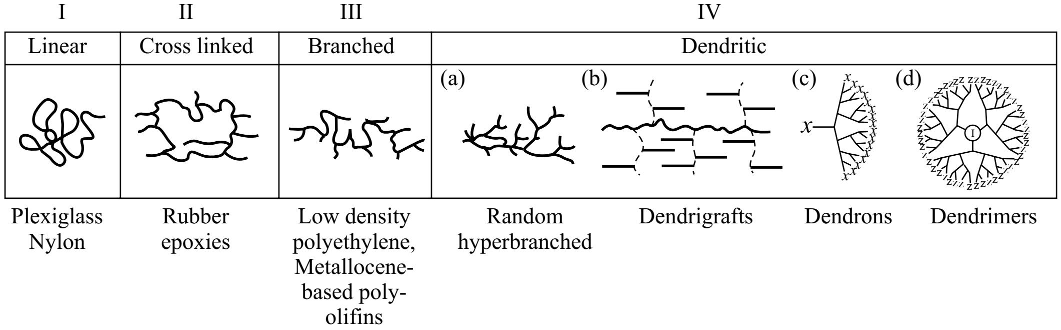


Figure 19.43 *Four major classes of macromolecular architecture.*

[Traditional synthetic polymers: (I) Linear (II) cross linked (III) branched (IV) branched]

Dendritic polymers are belonging to a special class of macromolecules. Due to their unique physical and chemical properties, dendrimers have wide ranges of potential applications. These include adhesives and coatings, chemical sensors, medical diagnostics, drug-delivery systems, high-performance polymers, catalysts, building blocks of supermolecules, separation agents and many more.

Types of dendrimers

- Poly(amidoamine) dendrimers (PAMAM)

  Poly(amidoamine) dendrimers are synthesized by the divergent method starting from ammonia or ethylenediamine initiator core reagents. Products up to generation 109 have been obtained. PAMAM dendrimers are commercially available, usually as methanol solutions. Starburst dendrimers is applied as a trademark name for a subclass of PAMAM dendrimers based on a tris-aminoethylene-imine core. The name refers to the star-like pattern observed when looking at the structure of the high generation dendrimers of this type in two dimensions.
- PAMAMOS Dendrimers

  Radially layered poly (amidoamine-organosilicon) dendrimers (PAMAMOS) are inverted unimolecular micelles that consist of hydrophilic, nucleophilicpolyamidoamine (PAMAM) interiors and hydrophobic organosilicon (OS) exteriors. These dendrimers are exceptionally useful precursors for the preparation of honeycomb-like networks with nanoscopic PAMAM and OS domains.
- **PPI dendrimer**

  PPI-dendrimer (polypropylene imine) describing the propylamine spacer moieties is the oldest known dendrimer-type developed initially. These dendrimers are generally poly-alkyl amines having primary amines as end groups. The dendrimer interior consist of numerous tertiary tris-propylene amines. PPI dendrimers are commercially available up to generation 5 (G5), and has found widespread applications in material science as well as in biology. As an alternative name to PPI, POPAM is sometimes used to describe this class of dendrimers. POPAM stands for polypropylene amine, which closely resembles the PPI abbreviation. In addition, these dendrimers are also sometimes denoted 'DAB-dendrimers' where DAB refers to the core structure, which is usually based on diamino butane.

- **Tecto dendrimer**
  These are composed of a core dendrimer, surrounded by dendrimers of several steps (each type design) to perform a function necessary for a smart therapeutic nanodevice. Different compounds perform varied functions ranging from diseased cell recognition, diagnosis of disease state drug delivery, reporting location to reporting outcomes of therapy.
- **Multilingual dendrimers**
  In these dendrimers, the surface contains multiple copies of a particular functional group.
- **Chiral dendrimers**
  The chirality in these dendrimers is based upon the construction of a constitutionally different but chemically similar branch to chiral core.
- **Hybrid dendrimers linear polymers**
  These are hybrids (block or graft polymers) of dendritic and linear polymers.

Amphiphilic dendrimers

- They are built with two segregated sites of chain end, one half is electron donating and the other half is electron withdrawing.

Micellar dendrimers

These are unimolecular micelles of water soluble hyper branched polyphenylenes.

- **Multiple antigen peptide dendrimers**
  It is a dendron-like molecular construct based upon a poly-lysine skeleton. Lysine with its alkyl amino side-chain serves as a good monomer for the introduction of numerous of branching points. This type of dendrimer was introduced by J. P. Tam in 1988, has predominantly found its use in biological applications, e.g. vaccine and diagnostic research.
- **Frechet-type dendrimers**
  It is a more recent type of dendrimer developed by Hawker and Fréchet based on poly-benzyl ether hyper branched skeleton. These dendrimers usually have carboxylic acid groups as surface groups, serving as a good anchoring point for further surface functionalization, and as polar surface groups to increase the solubility of this hydrophobic dendrimer type in polar solvents or aqueous media.

19.8.1 Properties of Dendrimers

Monodispersity

Dendrimers are synthesized by step-by-step well-organized structures with a very low polydispersity, unlike the traditional linear polymers. It is due to the precise synthesis and purification strategy in the preparing process. High-performance liquid chromatography (HPLC), size exclusion chromatography (SEC), mass spectrometry (MS), gel electrophoresis, and transmission electron microscopy methods characterise the monodisperse property of dendrimers. There exists inherent structural defects during the synthesis of high-generation dendrimers. This could be due to incomplete reactions and steric resistance, which cause the missing of repeating units, intramolecular cyclization, dimmer formation and retro-Michael reaction in dendrimers. However, MS results show that the molecular weight distributions of these high-generation dendrimers still remain very narrow (polydispersity < 1.05).

Nanoscale dimensions and shapes

Dendrimers have nanoscale dimensions due to their well-organized synthesis strategy and size controllable properties. The diameter of dendrimers (PAMAM) family with an ethylenediamine core grows for generation 1–10 increase from 1.1 to 12.4 nm. Also, the shape of dendrimers may vary with their generations. The PAMAM dendrimers of low generation (0–3) with an ethylenediamine core without interior characteristics have ellipsoidal shapes, whereas the PAMAM dendrimers of high-generation (4–10) with well-defined cavities have roughly spherical shapes.

Defined architecture, size and shape control

During the synthesis of dendrimers, they form well-organized three-dimensional structures with highly ordered architectures. For applications such as protein modelling or catalysis, this property is desirable. The control of size of dendrimer is also important in therapeutic applications, as different molecular sizes exhibit different pharmacokinetics. Other dendritic polymers such as dendronized polymers or hybrid linear-dendritic structures can have more potential than pure dendrimers for certain medical applications, but a key requirement for biological applications will be the ability to deliver a pure product; hence hybrid dendritic structures for such applications will generally start with dendrimer construction followed by the hybridization phase. The shape persistence of dendrimers is very important, as it allows the defined placement of functions not only on the dendrimer surface but also inside the dendritic scaffold. This is of crucial importance for several applications, e.g. in sensing. Furthermore, stiff dendritic architectures possess defined pores or voids. This is a prerequisite for defined interactions between the dendrimer and incorporated guest molecules.

Solubility and reactivity

The surface functional groups, dendrimer generation, repeated units, and even the core of dendrimers determine the solubility aspect. Dendrimers were reported to possess perfect solubility in a large number of solvents. Their high solubility in organic solvents leads to rapid dissolution and provides various approaches to characterize their structures. Also, high water-solubility ensures their applications as solubility enhancers for hydrophobic guest molecules. On the other hand, high density of surface functional groups ($-NH_2$, $-COOH$, $-OH$) in PAMAM dendrimers may be expected to conjugate with a series of bioactive molecules. These surface-modified dendrimers with different functions provide us with new nanodevice design strategies. Overall, high solubility and reactivity make dendrimers suitable as a platform in biomedical fields.

Encapsulating guest molecules

Because of their 'globular shape and the presence of internal cavities', it is possible to encapsulate guest molecules in the macromolecule interior. Small molecules like *p*-nitrobenzoic acid can be trapped inside the dendritic box of poly (propylene imine) dendrimer with 64 branches on the periphery. Further, it is allowed to form a shell on the surface of the dendrimer by reacting it with terminal amines of an amino acid (L-phenylalanine). The incorporated guest molecules were stably encapsulated inside the dendrimer.

Polyvalency

The dendrimers allow functionalization fairly easy and permit multiple functionalities for the modified dendrimer molecules. Polyvalency is useful as it provides versatile functionalization; it is also extremely important to produce multiple interactions with biological receptor sites, for example, in the design of antiviral therapeutic agents.

New biological activities of dendrimers are uncovered with unique pharmacokinetics for scaffolded molecules of dendrimer with multiple surface group or groups. Different ligands can be coupled to dendrimers to use them as transfection reagent, e.g. ligands recognizing only the surface of a certain cell type combined with ligands that facilitate the escape from the endosome. Functionalization of the periphery can also result in copolymers with interesting properties, such as viscosity, stability, etc. and dendrimer fillers are already fairly widely used in composites and other materials to modify such properties. Dendrimer properties can be easily tuned by modifying the end groups (e.g. changing the end groups on a same skeleton induces solubility in organic solvents, in CFC or in water).

Loading capacity (molecular container property)

Active dendrimers have materials on their surface and exhibit interesting behaviour. Today, the internal cavities of dendritic structures also can be used to carry and/or store a wide range of metals, organic or inorganic molecules by encapsulation and absorption. The appropriate degree of functionalization results depending on loading capacity. This property makes dendrimers very suitable as drug delivery vehicles and also appropriate for obtaining electrooptic or magnetic devices. It also allows the use of dendrimers to store, for example, nanoparticles of metal and to prevent precipitation, allowing for the creation of dispersions of what some have called 'nanoreactors'. The possibility of loading dyes could lead to novel ways of labelling and has been used to colour polymers with a dendrimer additive (dendrimers can mix and bond better than the raw dye filler).

Biocompatibility/low toxicity

Toxicity levels of dendrimers are generally very low for dendrimers carrying anionic groups on the surface, which are less toxic than those carrying cationic groups. Dendrimers commonly show also negligible or very low immunogenic response when injected or used topically. These properties make them highly suitable for drug delivery and biolabelling. In this sense, high bio-compatibility is crucial both for preventing toxic reaction and for seeking biodegradability options. Dendrimers can be made from biomaterials such with DNA being a popular choice. Dendritic polymers have great potential in various kinds of therapies, especially given their ability to be designed for biological specificity, therefore their biocompatibility and lack of toxicity is important. However, not all dendrimers are biocompatible nor show low toxicities.

19.8.2 Methods of Preparation

1. Divergent method

Divergent dendrimer synthesis is a technique that effectively grows the dendrimer structure from the initiator core to the periphery in a stepwise fashion by iterative addition of monomer units. Specifically, it is initiated by coupling of a monomer unit to a multifunctional initiator core where the dendrimer generation increases by successive addition of the building blocks to the surface of the parent dendrimer. Tomalia and co-workers used this strategy to couple.

PAMAM-NH_2 dendrimers are developed using *N*-(2-aminoethyl) acrylamide monomers to an ammonia core. Each branching unit is synthesized in a two-step sequence starting with exhaustive Michael addition of the acrylate ester to the ammonia core followed by amidation with excess EDA. The first step produces a half-generation, and the addition of the diamine yields the full generation.

2. Convergent method

The deficiencies of the divergent method are taken care of in convergent approach to dendrimer synthesis. Convergent synthesis begins with the dendrimer surface units coupled to additional building blocks to form the branching structure, thus constructing dendrons from the periphery toward the central focal point. Each dendron is then coupled through its focal point to a multifunctional core to produce the complete dendrimer. Unlike divergent synthesis, convergent reactions are simple to purify since the desired dendrons are substantially different from the reaction by-products, thus eliminating the need for highly efficient reactions. While the number of synthetic steps is similar for both convergent and divergent techniques, the convergent approach has fewer nonideal growth events, which leads to improved mono-dispersity of the final dendrimers.

19.8.3 Applications of Dendrimers

The dendrimers have innumerable applications in a broad range of fields, including materials engineering, industrial, pharmaceutical and biomedical applications. Specifically, nanoscale catalysts, novel lithographic materials, rheology modifiers, and targeted drug delivery systems, MRI contrast agents and bioadhesives. Commercial dendrimer-based immunoassay have been synthesised to monitor for heart attack, novel approaches towards the cancer and gene therapies, the discovery of dendrimer-metal nanocomplexes and nanocomposites, better understanding of unique Newtonian dendrimer rheology, the first commercial silicon-containing dendrimers (PAMAMOS), honeycomb-like nanostructured dendrimer-based films, sheets, membranes and coatings for various materials engineering applications, etc.

Some of main contributions to medical health include the following:

It is observed that the dendrimer Newtonian flow behaviour which makes them unique among other known high molecular weight macromolecules and suitable for unprecedented viscosity standards as nanoscopic ‘molecular ball bearings’.

The organosilicon radially layered copolymeric dendrimers was used under the trade name of PAMMOS. These dendrimers are ideal, highly precise globular building blocks for bottom-up preparation of nanoscaled honeycomb-like 3D networks which can be easily formed into elastomeric or plastomeric films, sheets, membranes or coatings on variety of substrates, including glass, metals, plastics, paper, etc.

These unique dendrimer-based networks open the doors for dendrimers to materials science and engineering, offering exceptionally attractive possibilities for applications in nanopatterning (lithography), various types of chemical and biological sensors, decontamination materials, preparation of nanoporous dielectrics, various nanocomplexes and nanocomposites with metals or organometallic compounds, room temperature preparation of quantum dots, etc.

Other applications in medical health are the following

(i) Various routes for dendrimer drug delivery: oral, parenteral, intraocular, nasal.

(ii) Gene therapy, immunodiagnostics: Dendrimers can act as vectors, in gene therapy. PAMAM dendrimers have been tested as genetic material carriers. Numerous reports have been published describing the use of amino-terminated PAMAM or PPI dendrimers as nonviral gene transfer agents, enhancing the transfection of DNA by endocytosis and, ultimately, into the cell nucleus.

(iii) Dendrimer in ocular drug delivery—to enhance pilocarpine bioavailability.

(iv) Dendrimers in pulmonary drug delivery—for Enoxaparin (40 percent increase in relative bioavailability by 2 and 3 generation positively charged PAMAM dendrimers).

(v) Dendrimer in transdermal drug delivery—improvement in solubility and plasma circulation time. PAMAM dendrimer complex with NSAIDs as permeation enhancers.

(vi) Dendrimers for controlled release drug delivery—anticancer drugs like methotrexate, adriamycin.

(vii) Dendrimers in targeted drug delivery—folic acid PAMAM dendrimers modified with carboxymethyl PEG5000 surface chains.

(viii) Dendrimers as nanodrugs: Poly(lysine) dendrimers modified with sulphonated naphthyl groups have been found to be useful as antiviral drugs against the herpes simplex virus can potentially prevent/reduce transmission of HIV and other sexually transmitted diseases (STDs).

(ix) Dendrimers in photodynamic therapy: The photosensitizer 5-aminolevulinic acid has been attached to the surface of dendrimers and studied as an agent for PDT of tumorigenic keratinocytes. Photosensitive dyes have been incorporated into dendrimers and utilized in PDT devices. This cancer treatment involves the administration of a light-activated photosensitizing drug that selectively concentrates in diseased tissue.

19.9 Inorganic–Organic Hybrid Nanoparticles

Polyhedral silsesquioxanes

An emerging class of new materials that have extensive applications are known as 'Hybrid inorganic–organic composites'. Materials are being designed with the good physical properties of ceramics and the excellent choice of functional group chemical reactivity associated with organic chemistry. Silicon containing organic polymer like poly silsesquioxanes, in particular, have generated a great deal of interest because of their potential replacement for and compatibility with currently employed, silicon-based inorganic in the electronics, photonics, and other materials technologies. Network polymers or polyhedral clusters having the generic formula $(RSiO_{1.5})_n$ are obtained by the hydrolytic condensation of trifunctional silanes.

Hence, they are known by the name 'silsesquioxanes'. Each silicon atom is bound to an average of one and a half (sesqui) oxygen atoms and to one hydrocarbon group (ane). Typical functional groups that may be hydrolyzed/condensed include alkoxy- or chlorosilanes, silanols and silanolates.

Synthetic methodologies that combine pH control of hydrolysis/condensation kinetics, surfactant-mediated polymer growth, and molecular templating mechanisms have been employed to control molecular scale regularity as well as external morphology in the resulting inorganic/organic hybrids (from transparent nanocomposites, to mesoporous networks, to highly porous and periodic organosilica crystallites) all of which have the silsesquioxane (or $RSiO_{1.5}$) stoichiometry. These inorganic-organic hybrids offer a unique set of physical, chemical, and size-dependent properties that could not be realized from just ceramics or organic polymers alone. Silsesquioxanes are therefore depicted as bridging the property space between these two component classes of materials. Many of these silsesquioxane hybrid materials also exhibit an enhancement in properties such as solubility, thermal and thermomechanical stability, mechanical toughness, optical transparency, gas permeability, dielectric constant and fire retardancy.

19.10 Nanointermediates

Nanostructured films, dispersions, high surface area materials, and supramolecular assemblies are the high utility intermediates to many products with improved properties such as solar cells and batteries, sensors, catalysts, coatings and drug delivery systems. They have been fabricated using various techniques.

Nanoparticles are obvious building blocks of nanosystems but, require special techniques such as self-assembly to properly align the nanoparticles. Recent developments have led to air resistant, room temperature systems for nanotemplates with features as small as 67 nm. More traditionally, electron-beam systems are used to fabricate devices down to 40 nm.

19.11 Nanocomposite Materials

Polymer nanocomposites consist of a polymeric material (e.g. thermoplastics, thermosets or elastomers) with reinforcement of nanoparticles.

The nanocomposite material is an innovative product having nanofillers dispersed in the matrix. Typically, the structure is a matrix-filler combination where the fillers like particles, fibres, or fragments surrounds and binds together as discrete units in the matrix.

Nanocomposites are materials with a nanoscale structure that improve the macroscopic properties of products. Nanocomposites are clay, polymer or carbon, or a combination of these materials with nanoparticle building blocks.

Nanocomposites have their own role to play for many applications and are synthesized via innovative approaches. The properties of nanocomposite materials depend not only on the properties of their individual parents but also on their morphology and interfacial characteristics.

The fabrication of nanocomposites is optimized for controlled nanosized second phase dispersion, stability and mechanical properties such as adhesion resistance, flexural strength, toughness and hardness which can be enhanced. The possibilities of producing materials with tailored physical and electronic properties at low cost could result in interesting applications ranging from drug delivery to corrosion prevention to electronic/automotive parts to industrial equipment and several others.

Nanocomposites, materials with nanoscale separation of phases can generally be divided into two types:

1. Multilayer structures, and
2. Inorganic/organic composites.

A gas phase deposition or from the self-assembly of monolayers, multilayer structures of nanocomposites can be obtained. Inorganic/organic composites can be formed by sol–gel techniques, bridging between clusters (as in silsesquioxanes), or by coating nanoparticles, in polymer layers. The greatly increased properties of materials can be observed in nanocomposites. It is observed that few ppm level impurities results in the formation of nanoscale aluminide secondary phases in aluminium alloys, increasing their strength and corrosion resistance. Magnetic multilayered materials are one of the most important aspects of nanocomposites as they have led to significant advances in storage media.

The term nanocomposite encompasses a wide range of materials right from three-dimensional metal–matrix composites, two-dimensional lamellar composites and nanowires of single dimension to zero-dimensional core-shells all representing many variations of nanomixed and layered materials. Though various composite materials like fibreglass and reinforced plastics are now in wide use for numerous applications, there has been continued demand for novel composites with desirable properties for many other applications.

Properties

The nanocomposite materials have their own physical, chemical and biological properties which differ from the properties of individual atoms and molecules or bulk matter. It can be remarked, therefore, that by creating nanoparticles, it is possible to control the fundamental properties of materials, such as their melting temperature, magnetic properties, charge capacity and even their colour without changing the materials chemical compositions.

Generally, a very high surface-to-volume and aspect ratios are found in nanoparticles and nanolayers that makes them ideal for use in polymeric materials. Such structures combine the best properties of each component to possess enhanced mechanical and superconducting properties for advanced applications.

The properties of nanocomposite materials depend not only on the properties of their individual parents but also on their morphology and interfacial characteristics. Some nanocomposite materials could be 1000 times tougher than the bulk component.

The zeolites have three-dimensional framework of inorganic components while, clays, metal oxides, metal phosphates, chalcogenides, etc. have two-dimensional framework.

One-dimensional and zero-dimensional materials such as $(Mo_3Se_3–)n$ chains and clusters also exists. Thus, nanocomposites promise new applications in many fields such as mechanically reinforced lightweight components, nonlinear optics, battery cathodes, nanowires, sensors and other systems.

Inorganic layered materials exist in many varieties. They possess well defined, ordered intralamellar space potentially accessible by foreign species. This ability enables them to act as matrices for polymers yielding hybrid nanocomposites. Lamellar nanocomposites represent an extreme case of a composite in which interface interactions between the two phases are maximized.

Nanocomposites could be produced with the broad range of properties by monitoring the polymer-host interactions. Lamellar nanocomposites can be divided into two distinct classes—'intercalated and exfoliated'.

The polymer chains are embedded alternatively with the inorganic layers in a fixed compositional ratio in the intercalated nanocomposites. They have a well-defined number of layers in the intralamellar space. The intercalated nanocomposites are useful for electronic and charge transport properties

In exfoliated nanocomposites, the number of polymer chains between the layers is almost continuously variable and the layers stand >100 Å apart. On the other hand, exfoliated nanocomposites possess superior mechanical properties.

For example, the electronics industry utilizes materials that have high dielectric constants and that are also flexible, easy to process and strong. Finding single component materials possessing all these properties is difficult. The most commonly used ceramic materials with high dielectric constant are found to be brittle and are processed at high temperatures, while polymer materials, which are very easy in processing have low dielectric constants.

Micron-scale ferroelectric ceramic particles are used as the filler in liquid crystal polymer, fluoropolymer (or thermoplastic polymer matrices) to get a nanocomposite material which do not possess ideal processing characteristics and are difficult to form into the thin uniform films used for many microelectronics applications.

19.11.1 The Basic Ingredients of Nanocomposite

Polymer nanocomposites consist of a polymeric material (e.g. thermoplastic, thermosets or elastomers) with reinforcement of nanoparticles. There are different types of nanoparticles that can be added into the polymer matrix to form polymer nanocomposites. A proper selection of a nanoparticle is essential to ensure effective penetration of the polymer or its precursor into the interlayer spacing of the reinforcement. As a result, desired exfoliated or intercalated product can be prepared.

During the preparation of a polymer composite, either, a polymer could be taken as matrix for reinforcement with nanoparticles (nanoclay) or a monomer can be taken, which will be polymerised in situ to give the corresponding polymer-clay nanocomposite.

Most commonly used nanoparticles include the following:

- Montmorillonite organoclays (MMT)
- Carbon nanofibres (CNFs)
- Polyhedral oligomeric silsesquioxane (POSS)
- Carbon nanotubes [multiwall (MWNTs), small-diameter (SDNTs), and single-wall (SWNTs)]
- Nanosilica (N-silica)
- Nanoaluminum oxide (Al_2O_3)
- Nanotitanium oxide (TiO_2)
- Clay and others

Thermosets and thermoplastics used as matrices for making nanocomposites include the following:

- Nylon
- Polyolefin, e.g. polypropylene
- Polystyrene
- Ethylene-vinyl acetate (EVA) copolymer
- Epoxy resins
- Polyurethanes
- Polyimides
- Poly ethylene terephthalate (PET)

There are the following two main challenges in developing nanocomposite materials after the desired polymer has been selected for the purpose:

- The choice of nanoparticles requires an interfacial interaction and/or compatibility with the polymer matrix.
- The processing technique should address proper uniform dispersion and distribution of nanoparticles or nanoparticle aggregates within the polymer matrix.

In addition, the mechanical properties of the nanocomposites are dependent on the significant role of the addition of the amount of nanoparticulate/fibrous to polymer matrix. These are generally added in very small quantities to result in improved properties. This in turn could result in significant weight reductions particularly in military and aerospace applications, greater strength and increased barrier performance for similar material thickness, whereas, the micro-dimensional particles/additives require much higher loading levels to achieve similar performance.

There are the few disadvantages associated with using nanoparticle viz. toughness and impact performance. There is a need for better understanding of formulation/structure/property relationships to platelet exfoliation and dispersion, etc. The improved properties vis-à-vis the disadvantages of the nanoparticles and resultant composites are shown in Table 19.4.

Improved properties	**Disadvantages**
Mechanical properties (tensile strength, stiffness, toughness)	Viscosity increase (limits process ability)
Gas barrier	Dispersion difficulties
Synergistic flame retardant additive	Optical issues
Dimensional stability	Sedimentation
Thermal expansion	Black colour when different carbon containing nanoparticles are used
Thermal conductivity	
Ablation resistance	

Table 19.4

Important characteristics of nanocomposites

19.11.2 Processing of Nanocomposite

Polymer nanocomposites are a class of materials that use nanoscale fillers reinforced into the polymer matrix.

A nanofiller is blended with a polymer to produce a composite with equal or better physical and mechanical properties than their conventionally filled counterparts. Polymer nanocomposites offer the potential for enhanced mechanical properties, barrier properties, thermal properties and flame retardant properties due to the higher surface area available with nanofillers, when compared to conventionally filled materials.

The following methods are considered for the processing of nanocomposites.

Each of the monomers is polymerised first followed by the pelletization process of each of the various polymers. After the individual polymers are pelletized, the formed pellets may be mixed with a nanofiller material in an extruder to form the nanocomposite material. While this process may be efficient for forming nanocomposites, at some instances they appear to be relatively expensive. Three-dimensional metal–matrix composites, two-dimensional lamellar composites and one-dimensional nanowires and zero-dimensional core-shells all represent the various nanomixed and layered materials. These methods of construction combine the best properties of each of the components or give rise to new and unique properties for many advanced applications.

The dispersion of multilayered silicate material into a thermoplastic polymer at a temperature greater than the melting or softening point of the thermoplastic polymer yields a nanocomposite. The thermoplastic polymer is selected from the group consisting of a thermoplastic urethane, a thermoplastic epoxy, polyester, nylon, polycarbonate and their blends.

(a) Carbon nanotube-polymer composite

(i) Carbon nanotube-polymer composites can be prepared by incorporating carbon nanotubes as fillers to polymers matrix. These carbon nanotube-polymer composites have high electrical conductivity and finds application in a conventional cold cathode or a polymer LED. They can also improve electrical conductivity of materials, additionally, with an improved mechanical performance.

CNT incorporated into a polymer matrix can produce composites with a very high strength and high elastic modulus which may lead to the development of ultra-resistant materials for use as reinforcement fibres.

Carbon nanotubes and polymer composites can form foams.

These materials are just at an early stage of investigation and lightweight foams will be produced with improved electrical, mechanical and thermal properties.

(ii) Carbon nanotube-reinforced composites:

The carbon nanotube-reinforced composites could be synthesized using a powder mixing process with a powder-powder blending between carbon nanotubes and ceramic powder or raw metal like aluminium or copper matrix followed by a conventional sintering process.

The relative densities of these nanotube-reinforced composite materials are low (85–95 percent) which accounts for a decrease in mechanical properties. This is the indication of the existence of many fracture sources, such as pores and defects, which could result in low mechanical properties.

The problems arise due to severe agglomeration of carbon nanotubes on the metal powder surface during the use of conventional consolidation processes. The homogeneous dispersion of carbon nanotubes in the metal–matrix prevents the carbon nanotubes agglomeration. This can be achieved by dispersing carbon nanotubes in a predetermined dispersing solvent like water, ethanol, nitric acid solution, toluene, *N*,*N*-dimethylformamide, dichlorocarbene and thionyl chloride. This forms a dispersed solution, which is further subjected to ultrasonic wave. This ultrasonic wave mixed water soluble salts or metal hydrates are dried to remove water vapour, hydrogen and finally calcined to produce a stable carbon nanotube/metal oxide nanocomposite powder.

Metal nanocomposite powders can be used as high-valued abrasive materials or wear-resistant coating materials. The metal nanocomposite powder could further be applied in industrial fields which utilize conventional metal composite materials, such as the aerospace, high-performance machine parts and medical industry, because it has high sintering performance and easily becomes bulky.

(b) Polymer matrix composite materials

The properties of plastic composite materials can be greatly improved by incorporating nanosized fillers. The nanocomposite materials require only a very small amount of nanofiller (less than 5 wt%) to yield marked improvements in mechanical, thermal, electrical and other properties such as gas impermeability, dimensional stability and flame retardance. This is quite different from conventional standard fillers. Furthermore, all of these benefits are available without sacrifices in composite density or light transmission of matrix resins. Nanofillers provide one to three orders of magnitude more interfacial area per particle than conventional mineral filler. Two types of nanofiller that are used in commercial products are nanoclays and nanotubes, such as carbon nanotubes and aluminium silicate nanotubes. Both types of nanofiller materials require surface modification prior to use.

(i) Thermoplastic-based nanocomposites

Thermoplastic materials could be classified as metals, ceramics or polymers. However, the lower densities of polymeric materials offer an added advantage in applications where lighter weight is desired. The addition of thermally and/or electrically conducting fillers helps in developing conducting type nanocomposites.

Thermoplastic nanocomposites are used in a wide array of applications. These includes automotive sector for interior parts and under-the-hood applications, in packaging industry for carbonated beverage bottles, plastic wrap, etc.

Nylon 6 nanocomposites

Mica-type silicates are nanoclays. These nanoclays are functioning as reinforcing fillers for polymers because of their high aspect ratio. They have unique intercalation and exfoliation characteristics as nanocomposites. There are many nanocomposites with the incorporation of organoclays into polymer matrices. An improved method of producing nylon 6–clay nanocomposites using in situ polymerization was carried out. The resulting nylon 6–clay nanocomposites exhibited increased solvent resistance, reduced permeability and increased flame retardant characteristics.

The nanocomposites could also be obtained by direct polymer intercalation, where polymer chains diffuse into the space between the clay galleries. This process could be combined with conventional polymer extrusion to reduce the time to form these hybrids, by shearing clay platelets leading to sample uniformity.

The chemistry of the clay surface and the type of extruder with its screw design can affect the degree of exfoliation and dispersion of layered silicate nanocomposites formed from polyamide 6. Excessive shear intensity or back mixing also causes poor exfoliation and dispersion. The exfoliation and dispersion could be improved by increasing the mean residence time in the extruder. Increased residence time in a low-shearing or mildly shearing environment allows polymer to enter the clay galleries and peel the platelets apart. It is reported that the nonintermeshing, twin-screw extruder yields the best exfoliation and uniformity of dispersion.

(ii) Thermoset-based nanocomposites

Thermoset nanocomposites are complex hybrid materials. A blend of nanoparticles with polymers produces a novel nanostructure with extraordinary properties. Thermoset polymer nanocomposites have received less interest in their scientific development and engineering applications than their thermoplastic counterparts.

Some of these materials could be relatively easy to bring into production. The understanding of characteristics of the inter-phase region and the estimation of technology-structure-property relationships are the current frontier researches in thermoset nanocomposites.

The production of thermoset nanocomposites is restricted to resin nanocomposites such as epoxy resins, unsaturated polyesters, acrylic resins, and so on. Various nanoparticles such as smectite clay, diamond, graphite, alumina and ferroxides have been found to be useful for the preparation of thermosetting polymers. Thermoset nanocomposites results in improved dimensional/thermal stability, flame retardancy and chemical resistance and have potential applications in marine, industrial and construction markets. Such nanocomposite materials are particularly suitable to be used for a large variety of applications.

The materials are eminently processable and can be shaped in conventional shaping steps, such as injection moulding and extrusion processes.

A variety of articles and materials of different shapes can be manufactured from such nanocomposite materials, e.g. fibres, packaging materials and construction materials.

The property comparison between thermoplastic and thermoset resins is given in Table 19.5.

Thermoplastic resin	Thermosetting resin
High MW solid	Low MW liquid or solid
Stable material	Low to medium viscosity, requires cure
Reprocessable, recyclable	Cross-linked, nonprocessable
Amorphous or crystalline	Liquid or solid
Linear or branched polymer	Low MW oligomers
Liquid solvent resistance	Excellent environmental and solvent resistance
Short process cycle	Long process cycle
Neat up to 30% filler	Long or short fibre reinforced
Injection/compression/extrusion	Resin transfer moulding (RTM)/filament winding (FW)/ sheet moulding compound (SMC)/prepreg/pultrusion
Limited structural components	Many structural components
Neat resin + nanoparticles	Neat or fibre reinforced + nanoparticles
Commodity: high-performance areas for automotive, appliance housings, toys	Commodity: advanced materials for construction, marine, aircraft, aerospace

Table 19.5

Comparisons of thermoplastic and thermosetting resin characteristics

(c) Applications of polymer matrix nanocomposite materials

The key nanoelements used in polymer nanocomposite materials are mainly nanoclay and carbon nanotubes. The plastic nanoclay composite materials are used in automotive body panels and under-hood components, and packaging. In packaging, nanocomposites decrease the permeation rate of gases and moisture vapour into plastic by creating a ‘tortuous path’. Due to high stiffness and impact resistance of polymer matrix nanocomposite materials they are used for electrical parts, power-tool housing, appliance components, pallets, and dunnage.

A timing belt cover and food packaging materials are produced based on nylon 6 with nanoclays nanocomposite. Many plastic nanocomposite materials are manufactured for packaging and automotive parts. Nanoclay can also be applied to reinforce fabrics. By adding 1 wt% of nanoclay, mechanical properties of polypropylene fabrics are improved. Nanoclay can potentially also be used in thermoplastic/natural fibre composites to improve their mechanical properties.

19.11.3 Clay–Polymer Nanocomposites

Nanocomposites of exceptional higher mechanical strength and thermal resistance of polymeric materials have gained interest because they are produced based on layered inorganic compounds such as clays. These properties could further be improved by incorporating a certain amount of clay in the polymeric materials. The variables such as type of clay, the choice of clay pretreatment method, the selection of polymer component and the way in which the polymer is incorporated has profound influence on the nature and properties of the final nanocomposite. The purity of the clay and homogeneity of dispersion of clay also affect the properties of nanocomposite.

Polymer and clay are intrinsically nonmiscible due to difference in their polarity. Polymers constitute nonpolar, organic material whereas the clay is more polar and inorganic material. Therefore for successful formation of polymer-clay nanocomposites, clay polarity needs to be altered to make the clay ‘organophilic’.

This could be carried out by using swelling agents like surfactants which increases the interlayer distance of clay structure before it is mixed with a monomeric material, which is then polymerized in the presence of the clay to form nanocomposites.

The clay-based nanocomposite could be produced in the form of an intercalated or exfoliated hybrid structure (Fig. 19.44). In the case of an intercalate, the organic component could be inserted between the layers of the clay such that the inter-layer spacing is expanded, but the layers still bear a well-defined spatial relationship to each other. In an exfoliated structure, the layers of the clay are completely separated and the individual layers are distributed throughout the organic matrix. Complete clay particles could also be dispersed within the polymer matrix as conventional filler.

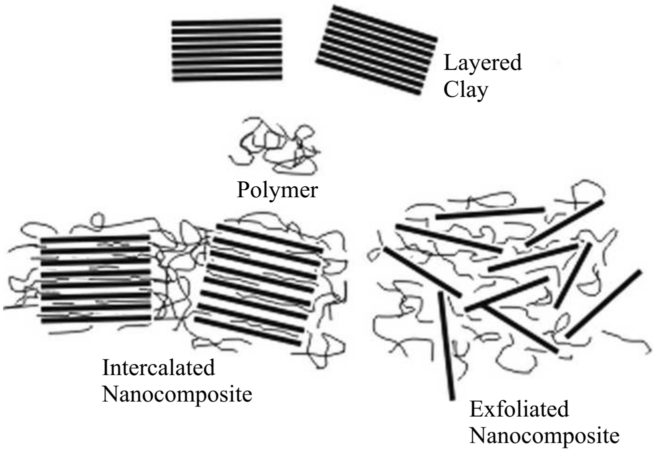


Figure 19.44 *Formation of intercalated and exfoliated nanocomposites from layered silicates and polymers.*

Whether particular organoclay hybrid nanocomposites are synthesized as an intercalated or exfoliated structure depends on the exchange capacity of the clay, polarity of the reaction medium and the chemical nature of the interlayer cations.

(a) Structure of clay material

A particular category of clay mineral material is used for the synthesis of nanoclay plastic composite. This clay has specialized structure with a plate morphology. The modified 'montmorillonite clay' is used for plastic composites. Montmorillonite is a 2-to-1 layered smectite clay mineral with a platy structure. Each layer has two tetrahedral sheets containing an octahedral sheet between them. The platelet thicknesses of each of them are just 1 nm, but surface dimensions are generally 300 to more than 600 nm, resulting in an unusually high aspect ratio. Hundreds or thousands of these layers are stacked together with van der Waals forces to form clay particles.

Figure 19.45 shows platelets shape of nanoclays and Fig. 19.46 shows transmission electronic microscopic image of refined nanoclays. The thickness of these nanoclay platelets is only 1 nm, less than the wavelength of light, so they do not impede light's passage.

(b) Characteristics of clay material

Refined or modified nanoclays are obtained by a two-step process for commercial applications. The modified clay can then be introduced into the resin matrix either during polymerization or by melt compounding. In the first step, it is desired that the hydrophilic surface property of 1 μm natural existing montmorillonite must be changed to hydrophobic for compatibility with organophilic polymers that are generally used in composites.

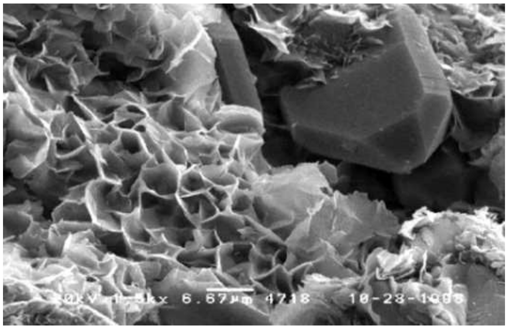


FIGURE 19.45 *A platelet montmorillonite montmorillon.*

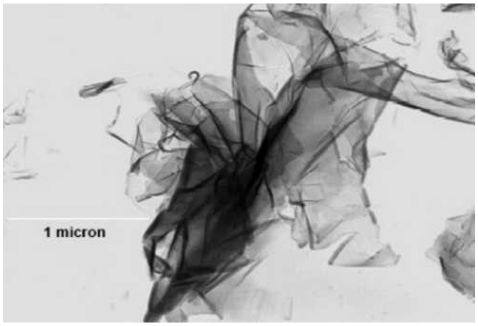


FIGURE 19.46 *TEM of refined structure.*

In the process of surface treatment method, there is exchange of organic ammonium cations with existing inorganic cations on the clay's surface. These modified clays are known as organoclays. Nanoclays normally exist as agglomerated bundles, consisting of thousands of platelets held together by van der Waals force. In the second step, the clay is exfoliated, which is separation of the individual clay platelets from each other (Fig. 19.47). To optimize the gas barrier function and promote good clarity of packaging materials, the platelets must be fully exfoliated in the plastic matrix. However, complete exfoliation is not always good for composite materials. For example, complete exfoliation in polyolefin material decreases particle reinforcement, which may be a problem for automotive application. By combining surface compatibilizing agents and process shear, commercial nanoclays can achieve more than 95 percent of exfoliation and even distribution within polyamide plastics.

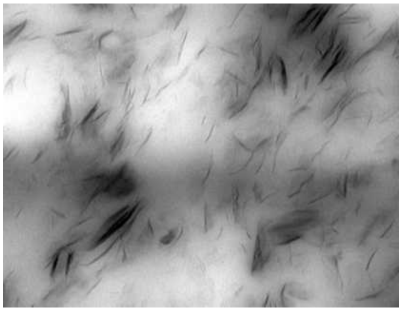


FIGURE 19.47 *Dispersion of exfoliated nanoclays in resin matrix, ethylene propylene diene monomer.*

(c) Fabrication methods

1. In situ polymerization

Montmorillonite is a 2-to-1 layered smectite clay mineral with a platy structure. Insertion of polymer precursors between the clay layers expands the platy structure. The expanded clay layers are then dispersed into the matrix by polymerization. This method can produce well-exfoliated nanocomposites applicable to a wide range of polymer systems, and is particularly useful for thermosetting polymers.

2. Solution-induced intercalation

The dispersion of clays into a polymer solution can be done by allowing the solvent to swell. It is mainly used for water-soluble polymers because of the low cost of water as a solvent as well as its low health and safety risk. In addition, this method can be used for commercial production. However, this method is not applicable to most engineering (i.e. nonwater soluble) polymers because of high solvent cost, and health and safety concerns.

3. The melt process

During the melt process, clays and polymers are intercalated with each other. The efficiency of intercalation is lower than in situ polymerization and often produces a partially exfoliated structure. However, this method can use traditional techniques such as extrusion and injection moulding to produce nanocomposites and is easily adapted to commercial production.

(d) Applications of clay-polymer nanocomposite

1. Barrier for packaging

A nanocomposite material made out of a combination of nylon and nanoclay is used for barrier layers in multilayer polyethylene terephthalate (PET) bottles and films for food packaging. The insertion of nanoclays in composite materials acts as a barrier and create a ‘tortuous path’ for gas molecules to go through the matrix resin. This tortuous path hinders the permeation of gases when they attempt to penetrate a plastic matrix.

The gas barrier property is maintained over time because nanoclays do not degrade in the matrix. These clay nanocomposites are used for oxygen (O_2) and carbon dioxide (CO_2) sensitive products. Figure 19.48 shows a schematic diagram of tortuous path due to the exfoliated nanoclay in matrix resin.

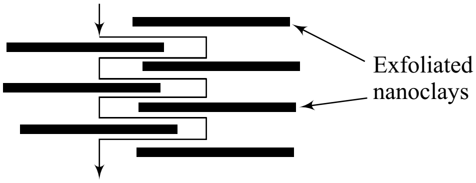


Figure 19.48 *Schematic diagram of tortuous path due to exfoliation of nanoclays.*

The partially hydrolyzed polyacrylamide (HPAM), xanthan gum, guar gum, carboxymethylcellulose (CMC) and hydroxyethylcellulose (HEC) are several types of water soluble polymers. Nylon 6 is the major resin material for gas-impermeable plastic packaging. The focus is on PET bottles, where nanocomposites provide a gas-impermeable barrier to oxygen and carbon dioxide, and mechanical strength. A nanocomposite of nylon 6 with 2 wt% nanoclay has three times the oxygen barrier property than that of neat nylon 6. The Young’s modulus of the material increased because of the addition of 5% of wt of nanoclay and reduces the oxygen transfer rate by five times. The results are shown in Table 19.6.

Material	Young’s Modulus (Mpa)	OTR (CC-mil/100 in 2-days
PA (Nylon) 6	140	2.91
PA (Nylon) 6 + 5%wt% nanoclays	705	0.58

Table 19.6

Changes of Young’s modulus and OTR in polyamide (PA) 6

2. Automotive parts

Nanoclay is used to increase structural strength in plastic polymer nanocomposite materials. Moreover, the synthesis of nanocomposite is of relative low cost compared to carbon nanotubes. Thermoplastic olefin (TPO) and polypropylene (PP)-based nanocomposite materials are used for mainly automotive parts. It is expected that the auto industry will increase usage of nano-PP, mostly to displace existing PP applications. PP nanoclay composite is used in the car seat back, in light truck, and in office furniture. Thermoplastic olefin nanocomposites improve low-temperature impact strength and surface quality. TPO clay nanocomposite materials are used in van (2.5 wt% nanoclay), the body side moulding, and the cargo bed on sport utility truck (Figs. 19.49 and 19.50).

Figure 19.49 *Car side mouldings with nanoclay*

Figure 19.50 *Pick up van bed with composite material.*

3. Flame retardant

Polyolefin, polystyrene (PS) and PP resin matrix are generally used for the preparation of nanoclay composites as flame-retardant. These polymer resins readily form char layers when exposed to heat. This char formation impedes the movement of volatilized polymer from the interior of a plastic matrix to the surface, which decreases the mass loss rate and flammability. The optimal loading amount of nanoclays is 5 wt% to reduce flammability. If 5 wt% of nanoclays added to the resin matrix, then, the peak heat release rates are reduced by 70 percent. When nanoclays are used together with traditional flame retardant, nanocomposites can achieve equivalent effect using significantly less flame-retardant additive.

19.12 Metal–Matrix Nanocomposite Materials

A new class of engineering materials known as 'metal–matrix composites' (MMCs) are tailored to meet specific needs for high impact and wear resistance, high melting point, chemical resistance and high neutron absorption capacity.

The synthetic methods of MMCs are concentrated on aluminium alloys because of their low densities.

Boron carbide (B_4C), silicon carbide (SiC) and boron nitride (BN) ceramic materials can be used to reinforce nanocrystalline aluminium alloys to enhance mechanical properties of metal–matrix nanocomposite. Mechanical milling method is to produce MMCs. Newly developed cryomilling technology is capable of producing a nanocrystalline grain metallic powder. Using this fine nanocrystalline grain powder achieves greater strength in the matrix. Cryomilling, which is mechanical milling at cryogenic temperature, takes advantage of both the extremely low temperature of the liquid nitrogen medium and the attributes associated with conventional milling. This technique is available to produce nanoscale materials in large quantities.

1. Cryomilling process

In the cryomilling process, a mixture of elemental or pre-alloyed powders is ground using shaker mills that generate high-energy compressive forces. In the pulverizing (shaker) mills, a number of hardened steel or tungsten carbide (WC) coated balls are used to ground particle sizes less than 50 μm. The resulting metal–matrix composite powders can be shaped with extrusion or other processes. Metal–matrix composites can be used for the aerospace or automotive industry because of their hardness and light weight. Figure 19.51 shows the mechanism for producing a composite material and a TEM image for a cryomilled aluminium/boron carbide (Al/B_4C) composite powder.

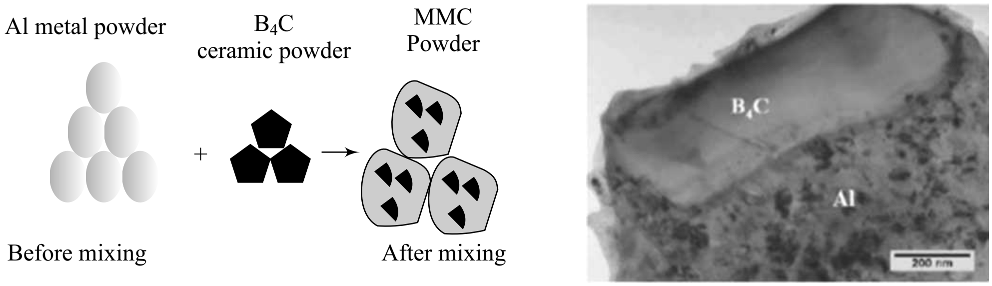


Figure 19.51 *Schematic diagrams showing the formation of composite powder and transmission electron microscopic image for the Al/B4C composite powder cryomilled A.*

In order to synthesize an aluminium MMC material, Al powder with particle size of 40–44 μm and B_4C powder with 1–14 μm particle size are cryomilled. During the milling process, liquid nitrogen is supplied to maintain extremely low temperature. After cryomilling, the matrix grain size of Al is reduced to 24–27 nm, and B_4C particle size is reduced to 0.2–2 μm.

19.12.1 Applications of MMC Nanocomposite

The commercial applications of MMCs are innumerable for MMC nanocomposites.
The applications of nanocomposites are in the following key areas:

- Drug delivery systems
- Anticorrosion barrier coatings
- UV protection gels
- Lubricants and scratch free paints
- New fire retardant materials
- New scratch/abrasion resistant materials
- Superior strength fibres and films

The major applications of MMCs with improved mechanical strength are in numerous automotive and general/industrial applications. These include potential for utilization as mirror housings on various vehicle types, door handles, engine covers and intake manifolds and timing belt covers.

The MMCs find applications as impellers and blades for vacuum cleaners, power tool housings, mower hoods and covers for portable electronic equipment such as mobile phones, pagers, etc.

(a) Food packaging

The MMCs obtained with improved ‘gaseous barrier property’ has been found useful in food packing industries. The data provided from various sources indicate oxygen transmission rates for polyamide-organoclay composites, which are usually less than half of the unmodified polymer. Both the amount of clay incorporated in the polymer and the aspect ratio of the filler contributes to overall barrier performance. The such high ratios and hence the tendencies towards filler incorporation at the nanolevel have quite dramatically enhance gaseous barrier properties.

Passive barrier characteristics are provided by nanoclay particles incorporated via melt processing techniques whilst the active contribution comes from an oxygen-scavenging ingredient. Oxygen transmission results reveal substantial benefits provided by nanoclay incorporation in comparison to the base polymer (rates approximately 15–20 percent of the bulk polymer value, with further benefits provided by the combined active/passive system).

Such excellent barrier characteristics have resulted in considerable interest in nanoclay composites in food packaging applications, both flexible and rigid.

Specific examples include packaging for processed meats, cheese, confectionery, cereals and boil-in-the-bag foods, also extrusion-coating applications in association with paperboard for fruit juice and dairy products, together with coextrusion processes for the manufacture of beer and carbonated drinks bottles. The use of nanocomposite packaging would be expected to enhance considerably the shelf life of many types of food.

(b) Fuel tanks

The ability of nanoclay incorporation to reduce solvent transmission through polymers such as polyamides has been shown. There is a significant reduction in fuel transmission through polyamide–6/66 polymers by incorporation of nanoclay filler. As a result, considerable interest is now being seen in these materials as both fuel tank and fuel line components for cars.

(c) Films

The presence of filler incorporation at nanolevels has also been shown to have significant effects on the transparency and haze characteristics of films. In comparison to conventionally filled polymers, nanoclay incorporation has been shown to significantly enhance transparency and reduce haze.

With polyamide-based composites, this effect has been shown to be due to modifications in the crystallization behaviour brought about by the nanoclay particles. Similarly, nanomodified polymers have been shown, when employed to coat polymeric transparency materials, to enhance both toughness and hardness of these materials without interfering with light transmission characteristics. The ability to resist high velocity impact combined with substantially improved abrasion resistance was also demonstrated.

(d) Environmental protection

The polymeric materials that are exposed to water laden atmospheres can encounter ill effects. Thus, an ability to minimize the extent to which water is absorbed can be a major advantage. The incorporation of nanoclay material to polymer can have a significant reduction of water absorption. Similar effects could also be achieved with polyamide-based nanocomposites. Specifically, increasing aspect ratio diminishes substantially the amount of water absorbed, thus indicating the beneficial effects likely from nanoparticle incorporation compared to microparticle loading.

Hydrophobicity enhancement would clearly promote both improved nanocomposite properties and diminish the extent to which water would be transmitted through to an underlying substrate. Thus, applications in which contact with water or moist environments is likely could clearly benefit from materials incorporating nanoclay particles.

(e) Flammability reduction

It is found that the flammability behaviour could be restricted in polymers such as polypropylene with as little as 2 percent nanoclay loading. In particular heat release rates, as obtained from cone calorimetry experiments, were found to diminish substantially by nanoclay incorporation. Although conventional microparticle filler incorporation, together with the use of flame retardant agents would also minimize flammability behaviour, this is usually accompanied by reductions in various other important properties. With the nanoclay approach, this is usually achieved whilst maintaining or enhancing other properties and characteristics.

Review Questions

1. What is nanoscience?
2. What is nanotechnology?
3. What are the scopes of nanotechnology?
4. What is a nanomaterial?
5. What are the categories of nanomaterials?
6. What are nanotools?
7. What are nanodevices?
8. What are nanostructured materials?
9. What are nanocrystalline materials?
10. What are the applications of nanocrystalline materials?
11. What are carbon-based nanomaterials?
12. What is carbon black?
13. What is amorphous carbon?
14. What is vitreous carbon?
15. What is graphene?
16. Discuss the importance of graphene structure?
17. What is graphite?
18. What are fullerenes?
19. Write a note on fullerene synthesis.
20. What is carbon onion?
21. What are nanoclusters?
22. What are nanostructured materials/particles?
23. What are the types of carbon nanotubes?
24. What is single-wall carbon nanotube (SW-CNT)?

25. What is multi-wall carbon nanotube (MW-CNT)?
26. Discuss the structure of a typical carbon nanotube?
27. Discuss the various methods of synthesis of carbon nanotubes.
28. Discuss top down and bottom up methods of wet chemical synthesis of nanomaterials.
29. Describe sol–gel process of synthesis of nanomaterials.
30. Describe gas phase synthesis of nanomaterials.
31. Describe flame-assisted ultrasonic spray pyrolysis method of synthesis of nanomaterials
32. Describe gas condensation (GPC) synthesis of nanomaterials.
33. Describe chemical vapour condensation (CVC) process of synthesis of nanomaterials.
34. Describe particle precipitation aided CVD process of synthesis of nanomaterials.
35. A brief note on laser ablation process of synthesis of nanomaterials.
36. What is a single-walled carbon nanohorn (SWNH)?
37. What are the types of nanohorn materials available?
38. Discuss the characteristics and properties of nanohorns.
39. Discuss the processes of synthesis of nanohorns.
40. Discuss the manifold applications of nanohorns.
41. What is a nanowire?
42. Describe the synthesis of nanowires.
43. What are the applications of nanowires?
44. What are nanorods?
45. Discuss the applications of nanorods.
46. How nanorods are synthesized by different techniques?
47. What are nanofibres?
48. Discuss different methods of processing carbon nanofibres.
49. What are carbon-based nanofilms?
50. Why is DLC tough and hard? Give reasons.
51. Mention the tribological properties of diamond like carbon (DLC).
52. Discuss the applications of diamond like carbon.
53. Why nanostructured bulk materials are important? Discuss.
54. Write a note on nanoporous carbon.
55. Write a note on nanofoam of carbon.
56. What are carbon aerogels?
57. What are bulk carbon nanocrystals?
58. How is a nanocrystal synthesized using an electrochemical cell?
59. What is a nanocrystalline diamond (NCD)?
60. What are quantum dots (QDs)?
61. How are quantum dots obtained? Mention the fabrication methods.
62. Describe colloidal process for the synthesis of quantum dots.
63. Describe molecular epitaxy method for the synthesis of quantum dots.
64. Discuss the applications of quantum dots.
65. What are dendrimers?
66. Describe the nanoscale fabrication of dendrimer macromolecule.
67. What are the types of dendrimers?
68. Mention the properties of dendrimers
69. How dendrimers are prepared?
70. Give the applications of dendrimers.
71. Write a note on polyhedral sesesquioxanes.
72. What are nanocomposites?
73. What are the types of nanocomposites?
74. Mention the properties of nanocomposites.
75. What are the basic ingredients of nanocomposites?
76. What are the thermoset and thermoplastic material used for nanocomposites?
77. Describe processing of nanocomposite material with a polymer matrix.
78. Write a note on CNT-polymer composite.
79. Write a note on thermoplastic-based nanocomposite.
80. Write a note on thermoset-based nanocomposite.
81. What are clay–polymer nanocomposites?
82. Discuss different fabrication methods of nano-clay–polymer composites.
83. What are the applications of clay–polymer composites?
84. What are metal–matrix nanocomposites?

85. Describe cryomilling process of manufacture of clay–polymer nanocomposites.
86. What are the applications of MMC nanocomposites?
87. (a) What is meant by nanochemistry?
    (b) What are nanorods?
    [Annamalai University, Dec, 2014]
88. (a) Describe the preparation of any two methods of carbon nanotubes.
    [Annamalai University, Dec, 2014]
89. (a) Mention the difference between a nanorod and nanowire.
    (b) Write any two applications of nanotubes.
    [Annamalai University, Dec, 2015]
90. (a) (i) What are the properties that change from bulk form to nanosize form? Explain each with example.
    (ii) Explain chemical vapour deposition technique of synthesis of nanomaterial particles.
    [Annamalai University, Dec, 2015]
91. (a) Discuss the solvothermal and laser ablation methods of synthesis of nanomaterials.
    (b) Compare the properties of molecules, nanoparticles and bulk materials.
    [Annamalai University, Dec, 2015]
92. (a) What are nanorods?
    (b) List out any four properties of nanomaterials. [Annamalai University, May, 2016]
93. (a) Explain how nanomaterials are synthesized by laser ablation and thermolysis.
    (b) What is chemical vapour deposition? Explain thermal CVD and photo CVD.
    [Annamalai University, May, 2016]
94. (a) Write short notes on: (i) Nanoclusters (ii) nanowires.
    (b) Briefly explain any four properties of nanomaterials.
    [Annamalai University, May, 2016]
95. (a) What are carbon nanotubes?
    (b) What is laser ablation?
    [Annamalai University, Dec, 2013]
96. (a) How are carbon nanotubes synthesized? Explain in detail.
97. Distinguish molecules, nanoparticles and bulk materials. [Annamalai University, May, 2013]
98. (a) What are nanowires?
    (b) Distinguish between nanomaterials and bulk materials.
    [Annamalai University, May, 2014]
99. (a) Discuss any four salient properties of nanomaterials.
    (b) Describe any two method of synthesizing nanomaterials.
    [Annamalai University, May, 2014]
100. (a) Discuss the size dependent properties of nanomaterials.
    (b) Explain any six applications of nanomaterials in various fields.
    [Annamalai University, May, 2014]
101. What are carbon nanotubes? Give types with respect to their structure and applications.
    [Pune University, May, 2016]
102. What are carbon nanotubes? Explain their types in detail. [Pune University, April, 2013]
103. Give any one method of preparing carbon nanotubes. State application of carbon nanotubes.
    [Pune University, Dec, 2013]
104. Explain the structure of fullerene. Give any two properties and any two applications of fullerene. [Pune University, Dec, 2015]
105. Write a note on dendrimers.
    [VTU, June/July, 2015]
106. (a) Give the gas condensation and hydrothermal process for the synthesis of nanomaterials.
    (b) Give the synthesis, properties and application of carbon nanotubes.
    [VTU, June/July, 2016]

20

Lubrication

Chapter Outline

Introduction–friction between two surfaces of solid objects. Types of friction: rolling friction, sliding friction, static friction, kinetic friction, fluid friction, lubricated friction, skin friction. Lubrication–lubricants, functions of the lubricants, requirements for satisfactory lubricant performance. Test of lubricants and their significance. Viscosity pour point flash point fire point, autoignition point, demulsibility, neutralization number, precipitation number, oiliness, penetration number. Types of lubrication mechanism:–hydrodynamic or fluid-film lubrication, boundary lubrication, mixed film lubrication (elastohydrodynamic lubrication). Basic types of lubricants–liquid lubricants, synthetic lubricants, semi-solid lubricants. Grease–function of grease, applications of grease, functional properties of grease. Solid lubricants, forms of solid lubricants, requirements to solid lubricants properties. Characterization of solid lubricants, typical applications, types of solid lubricants. Inorganic lubricants with lamellar structure. Oxides. Soft metals. Self-lubricating composite, Emulsions. Gaseous lubricants. Lubricant additives, Wear and Friction Improvers, adsorption or boundary additives, antiwear additives, extreme pressure additives, antioxidants, oxidation inhibitors, contamination of control additives, viscosity improvers. Classification of lubricants by application. Classification of lubricants by additives. Applications of lubricants.

20.1 Introduction

The animal and vegetable products were used by man as lubricants for all purposes and were used in large quantities. These lubricants were rejected because they lack chemical inertness and later, these conventional lubricants were substituted by petroleum products and synthetic materials. Petroleum lubricants are the predominantly occurring hydrocarbon products which are extracted from fluids that occur naturally within the Earth. They are used widely as lubricants because they possess a combination of the following desirable properties:

- available in suitable viscosities,
- low volatility,
- inertness (resistance to deterioration of the lubricant),
- corrosion protection (resistance to deterioration of the sliding surfaces), and
- low cost.

Lubricants obtained from synthetic procedures are generally characterized as oily, neutral liquid materials having some properties similar to petroleum lubricants. The synthetic lubricants have better properties than the conventional vegetable oils lubricants. Synthetic lubricants exhibit greater stability of viscosity with temperature changes, resistance to scuffing and oxidation and fire resistance. Since the properties of synthetics vary considerably, each synthetic lubricant tends to find special applications. There is a wide scope to exploit lubricants of desired qualities for they are extensively needed for working of all types of engines and machineries in daily life.

20.2 Friction between Two Surfaces of Solid Objects

Friction is the resistive force which takes place when two objects or bodies come into the contact with each other or it can be defined as the friction force that opposes the relative motion or tendency of such motion of two surfaces in contact.

20.2.1 Types of Friction

It is noted that different kinds of motion between two objects gives rise to different types of friction between them. Friction is the force that develops at the surfaces of contact of two bodies and impedes (resistive force) their relative motion. This force is also called as '*retarding force*' and it tries to stop the motion of the objects in contact.

A brief discussion of the following types of friction between the two surfaces is discussed here below:

1. Rolling friction

This type of rolling friction occurs when an object rolls over a surface. It acts on the point of the contact of the rolling object and the surface on which it is rolling. It retards the rolling motion of the object. It is due to the combination of various other forces.

The various causes of rolling friction are as follows:

- Deformation of the object
- Deformation of the surface
- The diameter of the wheels surface
- The movement below the surface
- Sliding
- Surface adhesion
- Surface properties
- Material of the object
- Temperature

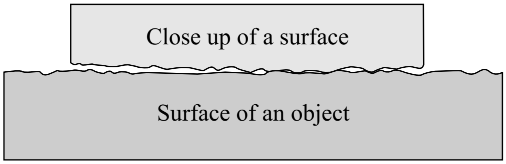


Figure 20.1 *Showing interlocking between object and surface.*

The main reason of rolling friction is the deformation of the object and the surface. A surface which appears to be smooth, on a molecular level is not so, it has many irregularities present on it. Interlocking at molecular level is the reason of friction (Fig. 20.1).

Plowing friction occurs when the wheel is soft and it gets deformed which is an example of another form of rolling friction as shown in Fig. 20.2.

Rolling friction can also occur when an object rolls over a soft surface. An example of rolling of wheels on a surface is as shown (Fig. 20.3).

A person riding a bike through a soft dirt surface is the best example. It is difficult to ride in such a condition because rolling friction comes into action and hence opposes motion. It is very interesting to note that it is easier to drive through soft dirt if the tyre also has less air and this is because the traction is increased which leads to improved driving.

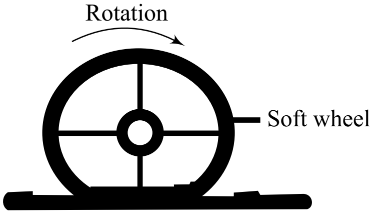


FIGURE 20.2 *A diagram showing plowing friction.*

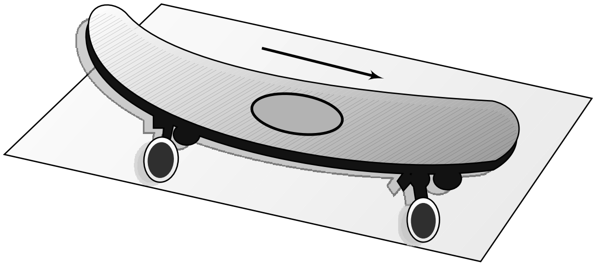


FIGURE 20.3 *A typical rolling of wheels on a surface.*

Temperature also influences rolling friction and it is found to be inversely proportional to friction. The coefficient of rolling friction is

$$f = C_{\text{rf}}W$$

where

f = force of friction
C_{rf} = coefficient of rolling friction
W = weight of the object (wheel) in this case

- Rolling resistance is not directly dependent on the weight of the vehicle as obvious in the above equation.
- When we apply brakes to any moving vehicle, rolling friction changes to kinetic sliding friction.

2. Sliding friction

When two objects are rubbing against each other as shown in Fig. 20.4, there appears sliding friction between the two objects.

Any object moving on a flat desk experiences a friction. This is an example of sliding friction. It is also known as kinetic friction which takes place when an object slides over another.

The direction of the force is such that the relative slipping (Fig. 20.5) is opposed by the retarding force (friction). The aim of sliding friction is to stop an object unlike rolling friction. It acts between the objects already in motion.

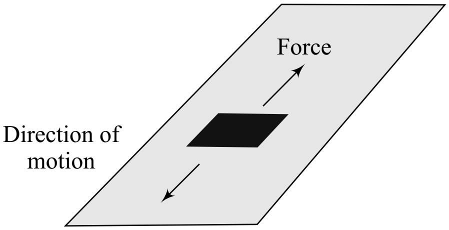


FIGURE 20.4 *Showing sliding friction.*

3. Static friction

Static friction prevents a solid object from moving against a surface (Fig. 20.6). It is the force that keeps an object from sliding off a plane surface, even when the surface is slightly tilted, and that allows you to pick

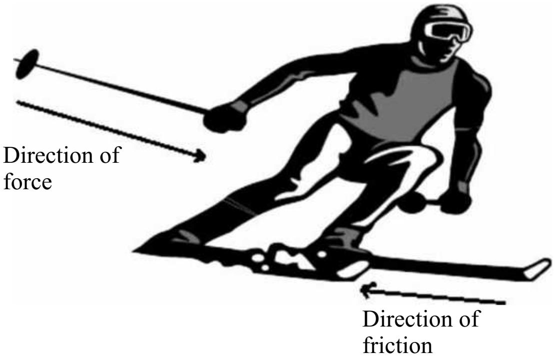


Figure 20.5 *A typical skating or slipping action.*

Figure 20.6 *A suitcase experiences a static friction.*

up an object without the object slipping through your fingers. In order to move the object, the force of *static friction* between the object and the surface on which it is resting is to be overcome. This force depends on the coefficient of static friction (μ_s) between the object and the surface and the normal force (N) of the object. The initial force required to get an object moving is often dominated by static friction.

When two objects are not moving relative to each other, a force opposes or resists the lateral movement of the objects and this force is called 'Static friction'.

It is the opposing force which comes into play when an object does not move over another object, even when the force is applied to make it move.

For static friction,

$$f_s = \mu_s . N \text{ or } \mu_s = \frac{f_s}{N} \text{ and } f_{\max} = \mu_s \cdot N$$

where

f_s = force of friction.
μ_s = coefficient of static friction.
N = normal force pressing the surfaces together

- The coefficient of static friction depends upon the roughness of the surfaces.
- The coefficient of static friction does not depend upon the contact surface area, provided the two surfaces are made of dissimilar enough material so that any cohesive force between the two surfaces is small.

Example It is not possible to push a heavy object even if it is pushed harder and harder and after some time it moves suddenly and once it moves it is easier to push it (Fig. 20.7).

Thus, resistance encountered by a body in static condition while trying to move under the action of an external force is called static friction. Static friction is equal and opposite to the applied force. Limiting friction is the maximum value of static friction and is also called *Traction*.

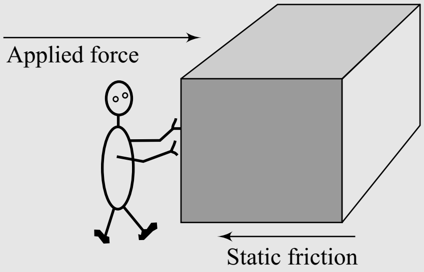


Figure 20.7 *Man trying to overcome static friction force.*

4. Kinetic friction

The appearance of force between two objects that are moving relative to each other is known as kinetic friction.

When two objects are moving relative to each other and rub, kinetic friction occurs as result of chemical bonding between the surfaces rather than interlocking between them.

A solid object sliding off a flat surface and brakes slowing down a wheel are both examples of sliding friction, called *kinetic friction* (Fig. 20.8). Sliding friction acts in the direction opposite to the direction of motion. It prevents an object sliding off a flat surface or wheel from moving as fast as it would without friction. When sliding friction is acting, another force must be present to keep an object moving. The *force is gravity* of the solid object influences an object sliding off a flat surface in this particular example.

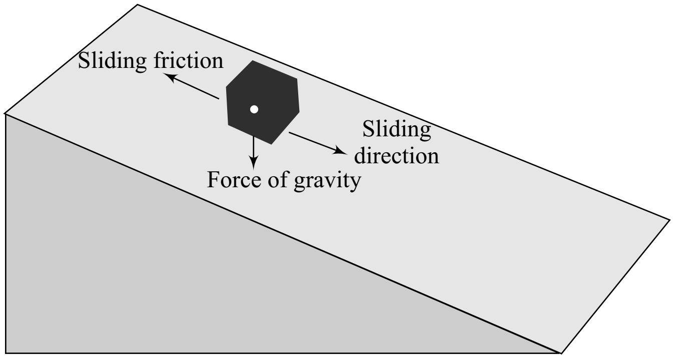


FIGURE 20.8 *A solid object experiencing force of gravity and sliding in the direction is an example of kinetic friction.*

The *force of kinetic friction* depends on the coefficient of kinetic friction between the object and the surface on which it is moving (μ_k) and the normal force (N) of the object. For any pair of objects, the coefficient of kinetic friction is typically denoted as μ_k, and is usually less than the coefficient of static friction (μ_s). This means that it takes more force to start an object sliding than it does to keep the object sliding. The coefficient of kinetic friction (like the static coefficient) depends upon the roughness of the surfaces. However it is generally smaller in value than the static coefficient.

$$\mu_k < \mu_s$$

For kinetic friction,

$$f_k = \mu_k \cdot N \text{ or } \mu_k = \frac{f_k}{N}$$

where

f_k = force of kinetic friction
μ_k = coefficient of static friction
N = normal force pressing the surfaces together

A graph illustrating static and kinetic friction is shown in Fig. 20.9.

5. Fluid friction

Fluid friction is the friction between a solid object as it moves through a liquid or gas medium.

When layers of any viscous liquid move over each other, fluid friction appears. *Fluid friction* or *drag* occurs when objects move through a fluid. Drag acts between the object and the fluid and hinders the motion of the object. The force of drag experienced by the object depends upon the shape of object, material and speed, as well as the fluid's viscosity. Viscosity is a measure of fluid's resistance to flow. It results from the friction that occurs between the fluid's molecules, and it differs depending on the type of fluid.

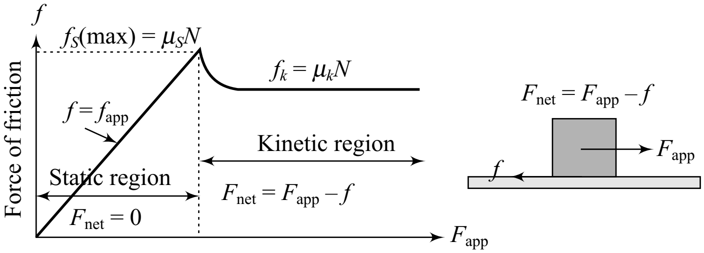


FIGURE 20.9 *Graph showing static and kinetic friction.*

Examples: An airplane's engines help it overcome drag and travel forward, while a fish uses its muscles to overcome drag and swim. It can be observed that while driving a car, it experiences friction due to molecules present in air. It is important to note that air is also a fluid made up of molecules. Fluid friction is responsible for an aerodynamic drag to a large extent.

6. Lubricated friction

A liquid layer between two solid surfaces offers a frictional resistance and is called *lubricated friction* (Fig. 20.10). Lubricated friction is a case of fluid friction where a fluid separates two solid surfaces.

Lubrication is a technique by which a viscous liquid is introduced in between two moving surfaces which aids for motion (smooth) and also prevents wear and tear of the surfaces and hence increasing their life span.

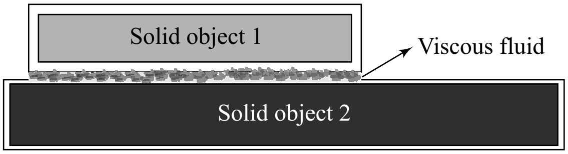


FIGURE 20.10 *Lubricating fluid between two solid objects.*

In most cases, the applied load is carried by pressure generated within the fluid due to the frictional viscous resistance to motion of the lubricating fluid between the surfaces. Adequate lubrication allows smooth continuous operation of equipment, with only mild wear, and without excessive stresses or seizures at bearings. When lubrication breaks down, metal or other components can rub destructively over each other, causing heat and possibly damage or failure.

7. Skin friction

This occurs when solid body is moving over the fluid. The force which resists the movement of a solid body through a liquid is called *Skin friction*. This directly depends on the area of contact of the body with the fluid.

There are various ways to reduce it such as

1. improving the shape of the object
2. reducing the surface area as far as possible.

20.3 Lubrication

The frictional resistance of two solids or metallic surfaces can be minimized by lubricating with a wide variety of substances. The most common lubricants are the oil and grease. Grease is composed of oil and a thickening agent to obtain its consistency, while the oil is what actually lubricates. Lubricating oils can be synthetic, vegetable or mineral-based as well as a combination of these. The application determines which oil, commonly referred to as the base oil, should be used. In extreme conditions, synthetic oils can be beneficial. Where the environment is of concern, vegetable base oils may be utilized.

The process to reduce frictional force between the sliding and rubbing surfaces of metals is called lubrication.

The effect of lubrication appears between the opposing surfaces when any suitable lubricant film is in between them. The applied load is carried by pressure generated within the fluid, and frictional resistance to motion arises entirely from the shearing of the viscous fluid.

Lubrication is a technique by which a viscous liquid is introduced in between two moving surfaces which aids for motion (smooth) and also prevents wear and tear of the surfaces hence increasing their life span.

This type of lubrication is desired when two surfaces of the materials rotate against each other while being completely separated. The liquid lubricants form a very thin film between the moving surfaces and thus avoid the direct metal to metal contact and reduce friction. This condition is known as 'fluid film lubrication'. The resistance to movement of sliding moving parts is only due to the internal resistance between the particles of the lubricants moving over each other. Therefore, lubricant chosen should have the minimum viscosity under working conditions and at the same time, it should remain in place and separate the surfaces.

Delicate machines and light instruments like watches, clocks, guns, sewing machines, scientific instruments, etc. are provided with thick fluid film lubrication. Boundary lubrication may appear when two parts slide or rotate against one another under heavy load, and such conditions may lead to the contact between the moving surfaces and film's breakdown. An oil is used which remains interacting with the solid surface of the metal that will be between the two sliding surfaces. This is achieved because of adsorption of lubricant molecules on the surfaces or by chemical reaction of the lubrication substances with the metal surfaces.

20.4 Lubricants

In all working machineries, the friction between metal to metal parts arise due to the moving surfaces rubbing against each other and experience a resistance that retards their movement. This friction will cause a lot of wear and tear of surfaces of moving parts. Due to the friction, large amount of energy is dissipated in the form of heat thereby the efficiency of machine gets reduced.

Lubricants are the substances used to reduce the frictional force between metal to metal contacts and minimize heat generation, along with certain foreign substances which keep the two rubbing surfaces apart.

Lubrication is simply the use of a material to improve the smoothness, in movement of one surface over another; the material which is used in this way is called a lubricant. Lubricants are usually liquids or semi-liquids, but may be solids or gases or any combination of solids, liquids and gases.

- Solid lubricants are used for railway tract joints, chains, air compressors, open gears, heavy machines, etc. The examples of solid lubricants are wax, talc, mica, molybdenum disulphide graphite, etc. It is used where the machine parts are subjected to slow speed, heavy load and sudden jerks.
- The examples of semi-solid lubricants are grease and vaseline.
- Liquid lubricants are used in delicate and light machines which work at high speed but under low pressure. Mineral oils, vegetable oils and animal oils are the various types of liquid lubricants.

If all other lubricants fail due to extreme temperature, chemical reactive atmosphere or some very particular operating conditions, then polyglycols, silicones, organic amines, imines and amides can be used in such occasions.

20.5 Function of the Lubricants

The very important purpose of the use of lubricants is to reduce the friction between the two solid objects and at the same time, there are many other benefits due to the process of lubrication.

The function of the lubricant is to remove friction between the rubbing surfaces solid bodies. In addition to friction reduction, it also reduces the amount of wear that occurs during operation, reduces operating temperatures, minimizes corrosion of metal surfaces and assists in keeping contaminants out of the system.

Lubricants have many properties that can be mixed and matched to meet your operating needs of machines. For example, there are different chemicals that can be added to allow a machine to efficiently run at extreme temperatures. It is possible to make a lubricant more effective in protecting machine surfaces under extreme pressures. By looking at the demands of the machine, one can properly identify the type of lubricant best suited for its proper function.

Summarising, the important functions of lubricants are as follows:

- It prevents the contact between the moving surfaces and keeps the surfaces separate under all loads, temperatures and speeds and reduces wear, tear and surface deformation of the concerned parts.
- It reduces the energy so that the efficiency of the machine is enhanced.
- It reduces the frictional heat and acts as a cooling fluid removing the heat produced by friction or from external source, prevents the expansion of metals and remains adequately stable in order to guarantee constant behaviour over the forecasted useful life.
- Protects surfaces from the attack of aggressive products formed during operation.
- Sometimes, it acts as a seal preventing the entry of dust and leakage of gases at high pressure.
- It reduces the maintenance and running cost of the machine.
- It minimizes corrosion.
- It also reduces power loss in internal combustion engines.

To accomplish the above functions, a good lubricant should possess the following properties:

- Suitable viscosity
- Oiliness to ensure the adherence of the bearings, for friction loss and wear when the lubrication is in the boundary region and as a protective covering against corrosion.
- High strength to prevent the metal to metal contact and seizure under heavy load.
- Should not react with the lubricating surface.
- A low pour point to allow the flow of lubricant at low temperature to the oil pump.
- No tendency to form deposits by reacting with air, water, fuel or products of combustion
- Cleaning ability
- Nonfoaming characteristic
- Nontoxic
- Nonflammable
- Low cost

A good lubricant exhibits the following characteristics:

- A good lubricant should not undergo any decomposition, oxidation and reduction at high temperature.
- A good lubricant should have higher flash and fire point than the operating temperature.
- A good lubricant should have high oiliness and viscosity index aniline point.
- A good lubricant should not corrode machine parts.

(a) The *gears* are generally subjected to high pressures, so, lubricants must have the following properties:
  - They should possess good oiliness.
  - They should not be removed by centrifugal force from the place of application.
  - They should be highly resistant to oxidation.
  - They should have high load carrying capacity.

(b) Cutting metal tools when used for performing any machining operation such as cutting, sawing, turning, boring, drilling, etc. required the use of certain oil known as cutting oils that must have following properties.
  - The oil must have both lubricating and cooling properties.
  - It should be chemically stable.
  - It should have high conductance to thermal ability.

(c) In steam turbines, lubricating oil is exposed to high temperature and oxidizing conditions and it should have the following properties:
  - It should have emulsification properties.
  - It should have antifoaming properties.
  - It should have corrosion resistance properties.
  - It should have high oxidation properties

20.6 Requirements for Satisfactory Lubricant Performance

Few desirable properties necessary for satisfactory performance of lubricants are as follows:

- The lubricants should have low volatility under operating conditions. Volatility characteristics are essentially inherent in the choice of base oil for a particular type of service and cannot be improved by the use of additive materials.
- The lubricants should have satisfactory flow characteristics in the temperature range of use. Flow characteristics largely depend on the choice of base oil. However, this can be improved by the addition of modifiers. The former improve low-temperature flow properties, while the latter enhance high-temperature viscosity characteristics.
- The lubricants should be stable or able to maintain desirable characteristics for a reasonable period of use. These characteristics depend to some extent on the base materials that enhance base fluid properties in this area. Lubricant stability is affected by the environment in which it operates. Such factors as temperature, oxidation potential and contamination with water, unburned fuel fragments, and corrosive acids limit the useful life of a lubricant. This is the area where additives have made a major contribution in improving the performance characteristics and extending the useful life of lubricants.
- The lubricants should be compatible with other materials in the system. Compatibility of lubricants with seals, bearings, clutch plates, etc. may also be partially associated with the base oil. However, additive chemistry can have a major influence on such characteristics. Additives can be classified as materials that impart new properties to or enhance existing properties of the lubricant or fuel into which they are incorporated.

20.7 Test of Lubricants and Their Significance

The desirable properties of lubricating oils for commercial use are tested and the properties of oils are assessed as follows:

1. Viscosity,
2. Pour point,
3. Flash point,
4. Fire point,
5. Auto-ignition point,
6. Demulsibility,
7. Neutralization number, and
8. Precipitation number.

Standard test methods are used for each of the above. The properties of lubricant oil are briefly explained in the following paragraphs.

20.7.1 Viscosity

The tendency to resist flow in a system is called the viscosity of an oil. A high viscosity liquid flows very slowly. Fluid lubricants have many good properties for application. Of all the properties of fluid lubricants, *viscosity is the most important, since it determines the amount of friction that will be encountered between sliding surfaces and whether a thick enough film can be built up to avoid wear from solid-to-solid contact*.

The viscosity of a lubricant determines the thickness of the layer of oil between metallic surfaces in reciprocal movement. This property is temperature dependent. In variable climates, for example, automobile owners change oil according to prevailing seasons. Oil changes are necessary because heavy oil becomes too thick or sluggish in cold weather and light oil becomes too thin in hot weather. The higher the temperature of an oil the lower its viscosity becomes; lowering the temperature increases the viscosity. In a cold morning, it is the high viscosity or stiffness of the lube oil that makes an automobile engine difficult to start. The viscosity must always be high enough to keep a good oil film between the moving parts. Otherwise, friction will increase, resulting in power loss and rapid wear on the parts. Viscosity customarily is measured by a viscometer that determines the flow rate of the lubricant under standard conditions; the higher the flow rate the lower the viscosity.

It is a measure of the internal resistance of a liquid during its flow. It is expressed in centipoises or centistokes (cSt).

The viscosity of an oil is the time in seconds for a given quantity of an oil to pass through a standard orifice under the specified conditions.

(a) Determination

The two methods are used to determine the viscosity of an oil. They are as follows:

(i) Red wood viscometer
(ii) Saybolt viscometer

The rate is expressed in centipoises, reyns, or seconds Saybolt universal (SSU) depending, respectively, upon whether metric, English or commercial units are used. The time required for 50 mL of the liquid to pass through an orifice of a red wood viscometer is called *as red wood seconds*. The time required for 60 mL of the liquid to pass through the orifice of a saybolt viscometer is called *Saybolt universal seconds*.

(b) Significance

(i) The movement of the machine is restricted due to excessive friction, if the viscosity of lubricating oil is too high.
(ii) If the viscosity of the lubricating oil is too low, the liquid oil film cannot be maintained and good lubricating oil must have moderate viscosity.

   Generally, oils are graded for applications by their viscosities at a certain temperature. The oils are graded by noting the number of seconds required for a given quantity (60 mL) of the oil at the given temperature to flow through a standard orifice. The right grade of oil, therefore, means oil of the proper viscosity.

(c) Viscosity index

The viscosity index (VI) of lubricating oil is a desirable characteristic to be noted which indicates the variations in the viscosity of lubricating oils with changes in temperature.

In other words, *the viscosity index measures the variations in the viscosity with changes in temperature; the higher the level of the viscosity index the lower the variation in viscosity at temperature.*

Consequently, if two lubricants with a viscosity of 40°C are considered, the one with the higher viscosity index will guarantee its usage:

- Because of lower internal friction, better engine starts up at low temperatures.
- Surface separation between the objects is higher due to the density of the layer of lubricants at high temperatures.

Every oil has a viscosity index based on the slope of the temperature-viscosity curve.

Viscosity index is defined as 'The average decreases in viscosity of an oil per degree rise in temperature between 100°F and 210°F'.

The V.I of lubricating oil is low, if the viscosity of the oil decreases rapidly with the increase in temperature.

- Gulf coast oil (consists of naphthenic hydrocarbons) exhibits a larger change in viscosity with increase in temperature and its V.I value is arbitrarily assigned as zero.
- If the viscosity of an oil is slightly affected with the increase in temperature, it has higher V.I.
- Pennsylvanian oil (consists of paraffinic hydrocarbons) exhibits a relatively smaller change in viscosity with increase in temperature and its V.I value is arbitrarily assigned as 100.

(d) Determination

The two standard oils are used to calculate the V.I of a test oil. The test oil is compared at 38°C (100°F) with zero V.I oil (gulf coast) and 100 V.I oil (Pennsylvania oil) both having the same viscosity as the test oil at 99°C (210°F). The V.I of the test oil is given by the following formula

$$\text{V.I} = \frac{(L-U)}{(L-H)} \times 100$$

where U is the viscosity of the test oil at 38°C, L is the viscosity of the low V.I oil at 38°C, H is the viscosity of the high V.I oil at 38°C. The viscosity temperature curve is flatter for the oil of high V.I than the oil of low V.I

A good lubricant should have minimum change in viscosity for a wide range of temperature. A good lubricant should have a high V.I. Also, V.I of a lubricant can be increased by the addition of linear polymers such as polyisobutylene, *n*-hexanol, etc.

20.7.2 Pour Point

The lowest temperature at which the oil barely flows from a container is called the 'pour point' of an oil. It is a rough indication of the lowest temperature at which oil ceases to flow. At a temperature below the pour point, oil solidifies.

The test to determine the pour point of a lubricant is to tilt the jar containing oil for a prescribed period at the lowest temperature at which the oil flows. The pour point of oil indicates the lowest temperature at which the oil may be used in some applications.

(i) Measuring the pour point of petroleum products

The specimen is cooled inside a cooling bath to allow the formation of paraffin wax crystals. At about 9°C above the expected pour point, and for every subsequent 3°C, the test jar is removed and tilted to check for surface movement. When the specimen does not flow when tilted, the jar is held horizontally for 5 seconds. If it does not flow, 3°C is added to the corresponding temperature and the result is the pour point temperature.

(ii) Measuring the pour point of crude oils

Two pour points can be derived which can give an approximate temperature depending on its thermal history. Within this temperature range, the sample may appear liquid or solid. This peculiarity happens because wax crystals form more readily when it has been heated within the past 24 hours and contributes to the lower pour point.

The upper pour point is measured by pouring the test sample directly into a test jar. The sample is then cooled and then inspected for pour point as per the usual pour point method.

The lower pour point is measured by first pouring the sample into a stainless steel pressure vessel. The vessel is then screwed tight and heated above 100°C in an oil bath. After a specified time, the vessel is removed and cooled for a short while. The sample is then poured into a test jar and immediately closed with a cork carrying thermometer. The sample is then cooled and inspected for pour point as per the usual pour point method.

20.7.3 Flash Point

The fire resistance properties of lubricating oils are assessed to indicate their flash points. In order to assess the overall hazard of a material, flash point of lubricating oil is determined and also to notify safety regulations in shipping as 'flammable' and 'combustible' materials.

The temperature at which a lubricant momentarily flashes in the pressure of a test flame is known as its flash point.

The flash point of an oil is the temperature at which enough vapour is given off to flash when a spark is present. Like the pour point factor, the flash point may in some instances become the major consideration in selecting the proper lubricant, especially in lubricating machinery handling highly flammable material. The minimum flash points allowed for navy lube oils are all above 300°F. However, the temperatures of the oils are always far below 300°F under normal operating conditions.

(a) What is an oil's flash point?

The oil is said to be 'flashed' when a flame appears and instantaneously propagates itself over the entire surface. The oil flashes because a flammable mixture results when it is heated sufficiently, causing vapours to emerge and mix with oxygen in the air.

The flash point is the lowest temperature at which an ignition source causes the vapours of the specimen (lubricant) to ignite under specified conditions.

The flash point temperature of an oil corresponds roughly to a vapour pressure of 3–5 mmHg. When a small flame (ignition source) is applied to the oil's surface this vaporous mixture will burn momentarily and then extinguish if the critical temperature has been reached.

(b) Detect and quantify fuel dilution using flash point

The flash points of lubricants are determined to indicate the fire hazard of fuels and oils being stored or transported.

- However, combined with other tests such as viscosity, viscosity index and specific gravity, the flash point can help reveal both the quality of the crude oil from which the lubricant was derived and the quality of the refining process.
- The flash point can also identify whether the base oil was a wide or narrow single cut or whether it represents a blend of two fractions (two base oils of different viscosities mixed together).
- The flash point may give some indication about the volatility and content of the most volatile components of the test oil. The flash point tells nothing, however, about the volatility of the oil as a whole.

20.7.4 Fire Point

The fire point of an oil is the temperature at which the oil will continue to burn when it is ignited. Fire point is the lowest temperature at which a specimen sustains burning for 5 seconds.

20.7.5 Auto-ignition Point

The auto-ignition point of an oil is the temperature at which the flammable vapours given off from the oil will burn. This kind of burning will occur without the application of a spark or flame. For most lubricating oils, this temperature is in the range of 465–815°F.

20.7.6 Demulsibility

The demulsibility, or emulsion characteristic, of an oil is its ability to separate cleanly from any water present, which is an important factor in forced-feed systems.

20.7.7 Neutralization Number

In order to assess the *acid or alkaline content of new oils*, the value could be ascertained through a standard analytical chemical technique (titration).

The neutralization number is a measure of the acid or alkaline content of new oils and an indicator of the degree of oxidation degradation of used oils.

The neutralization number of an oil indicates its acid content and is *defined as the number of milligrams of potassium hydroxide (KOH) required to neutralize 1 gram of the oil (lubricant).* All petroleum products deteriorate (oxidize) in air and heat. Oxidation produces organic acids which, if present in sufficient concentrations, will cause deterioration of:

1. alloy bearings at elevated temperatures,
2. galvanized surfaces, and
3. demulsibility of the oil with respect to fresh and salt water.

The increase in acidity indicates the rate of deterioration of oil and is measured as a part of the work factor test. This test is not applicable to 9250 oil.

20.7.8 Precipitation Number

The precipitation number of an oil is *a measure of the amount of solids classified as asphalts or carbon residues contained in the oil.* The number is reached when a known amount of oil is diluted with naphtha and the precipitate is separated by centrifuging—the volume of separated solids equals the precipitation number. This test detects the presence of foreign materials in used oils. An oil with a high precipitation number may cause trouble in an engine. It could leave deposits or plug up valves and pumps.

20.7.9 Oiliness

Oiliness is also an important aspect that signifies the relative ability to operate under boundary lubrication conditions. The term relates to a *lubricant's tendency to wet and adhere to a surface*. The oiliness property of oil is assessed chiefly through subjective judgment and experience, though there is no formal test for the measurement of oiliness. The most desirable lubricant for the specific use need not necessarily be the oiliest; *e.g.* long-fibre grease, which is low in oiliness as compared with machine oils, is usually preferable for packing rolling bearings.

20.7.10 Penetration Number

The penetration number, applied to grease, is a measure of the film characteristics of the grease. The test consists of dropping a standard cone into the sample of grease being tested. Gradations indicate the depth of penetration: the higher the number the more fluid the grease.

20.8 Types of Lubrication Mechanism

There are three mechanisms to illustrate the lubrication process between the surfaces of two rubbing metals/objects.

1. Fluid-film lubrication
2. Boundary lubrication
3. Elastohydrodynamic lubrication

A lubricant is any substance that reduces friction by creating a slippery film between two surfaces. Lubricants permit one surface to move easily over the other surface. The lubrication of objects to reduce of friction via the use of a fluid appears to be a simple concept; it is actually far more complex. The basic principle of lubrication process can be divided into three basic types or regimes.

20.8.1 Hydrodynamic or Fluid–Film Lubrication

Hydrodynamic lubrication is a system of lubrication in which the shape and relative motion of the sliding surfaces cause the formation of a fluid film having sufficient pressure to separate the surfaces.

The moving metal surfaces are actually separated by lubricating oil film eliminating all metal contacts under hydrodynamic lubrication (Fig. 20.11). This is also known as ‘fluid film lubrication’. The reason for the separation of two surfaces without metal to metal contact is due to the noncompressible nature of this lubricating film, Fig. 20.12(a).The condition in which surfaces are completely separated by a continuous film of lubricating fluid is commonly referred to as ‘hydrodynamic lubrication’. In this regime, the lubricant immediately adjacent to each surface travels at the same speed and direction as that of each surface. As you travel through the width of the film, differential speed graduates and direction is reversed.

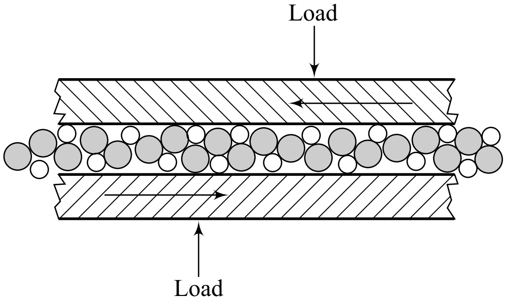
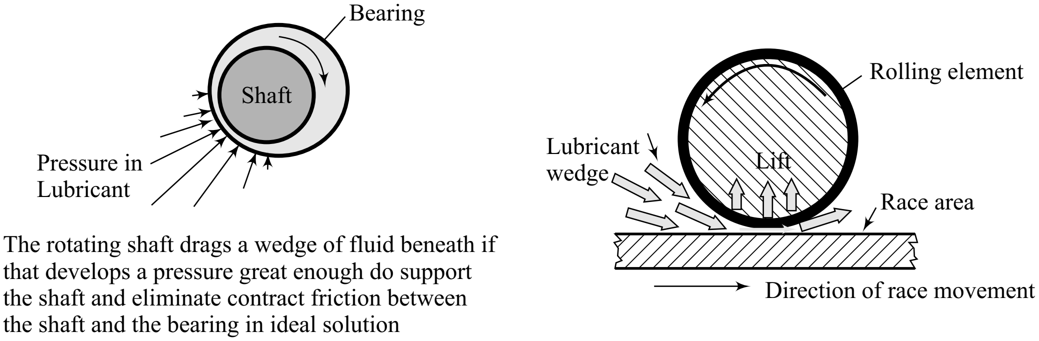


Figure 20.11 *A Hydrodynamic or full fluid film lubrication.*

Hydrodynamic film thickness can be formed by wedging the lubricant through a convergent gap with the tangential surface velocities. Lubricant wedging film action is shown in Fig. 20.12(b). This is similar to a car tyre hydroplaning on a wet road surface.



The rotating shaft drags a wedge of fluid beneath if that develops a pressure great enough do support the shaft and eliminate contract friction between the shaft and the bearing in ideal solution

Figure 20.12 *Different types of hydrodynamic lubrication.*

Although hydrodynamic lubrication is an ideal situation, in many instances it cannot be maintained. There are few factors that affect the hydrodynamic lubrication, such as, lubricant viscosity, rotation speed or RPM, oil supply pressure and component loading. An increase in speed or viscosity increases oil film thickness. An increase in load decreases oil film thickness.

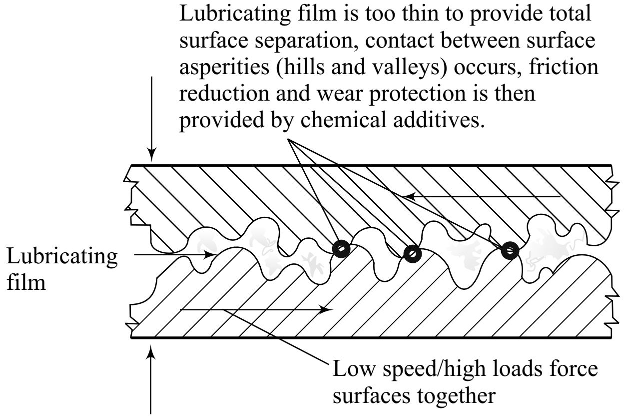
Hydrodynamic lubrication is an excellent method of lubrication since it is possible to achieve coefficients of friction as low as 0.001 ($\mu = 0.001$), and there is no wear between the moving parts. Since viscosity is temperature dependent property, temperature rise due to frictional force has to be taken care of. In such a case, the lubricant is maintained at cooling reservoir to retain the desired viscosity of the fluid. Another way of handling the heat dissipation is to use commercially available additives to decrease the viscosity's temperature dependence.

A special attention has to be paid to the extremes of motion, when using hydrodynamic lubrication: starting and stopping. When the surfaces are at rest with respect to each other, or at very low speeds, the distance of separation is theoretically zero.

20.8.2 Boundary Lubrication

There may be excessive loading, speeds or a change in the fluid's characteristics during the operation of metal parts of machines; as a result, the lubricating film may be very thin in between the metal parts. In such a situation, contact between surface asperities (peaks and valleys) occurs.

Boundary lubrication (thin film lubrication) is a condition in which the lubricant film becomes too thin to provide total separation. Friction reduction and wear protection is then provided via chemical compounds rather than properties of the lubricating fluid (Fig. 20.13). Boundary lubrication often occurs during the start-up and shut down of equipment or when loading becomes excessive. This condition can commonly be observed in certain types of gear sets that need to withstand sliding pressures and shock loadings, such as hypoid gears found in automotive differentials.



Lubricating film is too thin to provide total surface separation, contact between surface asperities (hills and valleys) occurs, friction reduction and wear protection is then provided by chemical additives.

Low speed/high loads force surfaces together

Figure 20.13 *A typical boundary lubrication process.*

In order to overcome this effect, an additive is added to basic mineral oil to create an effective boundary lubricant. Generally, the best *additives are active organic compounds with long chain molecules and active end groups*. These compounds bind tightly and intricately with each other, forming a film that builds up on the surface of the metal itself binding strongly to the metal. This results in a thin film that is very difficult to penetrate. When two surfaces, each covered with a boundary layer, come in contact with each other they tend to slide along their outermost surfaces, with the actual faces of the surfaces rarely making contact with each other. Liquids are rarely good boundary lubricants. The best boundary lubricants are solids with long chains of high inter-chain attraction, low shear resistance so as to slip easily and a high temperature tolerance. The boundary lubricant should also, obviously, be able to maintain a strong attachment to the surfaces under high temperatures and load pressures.

The most common boundary lubricants are probably greases. Greases are so widely used because they have the most desirable properties of a boundary lubricant. They not only shear easily but flow also. They also dissipate heat easily form a protective barrier for the surfaces, preventing dust, dirt and corrosive agents from harming the surfaces.

20.8.3 Mixed Film Lubrication (Elastohydrodynamic Lubrication)

A mixed film lubrication process is obtained by a combination of both hydrodynamic and boundary lubrication. Only occasional asperity (hills and valleys) contact occurs in mixed film lubrication process. Elastohydrodynamic lubrication (EHD or EHL) occurs as pressure or load increases to a level where the viscosity of the lubricant provides a higher shear strength than the metal surface it supports. This regime can occur in roller bearings or gears as the lubricant is carried into the convergent zone approaching a contact area or the intersection of two asperities. As a result, the metal surfaces deform elastically in preference to the highly pressurized lubricant, which increases the contact area and thus increasing the effectiveness of the lubricant.

Consider a shaft of metal resting on a sheet of rubber as shown Fig. 20.15(a).

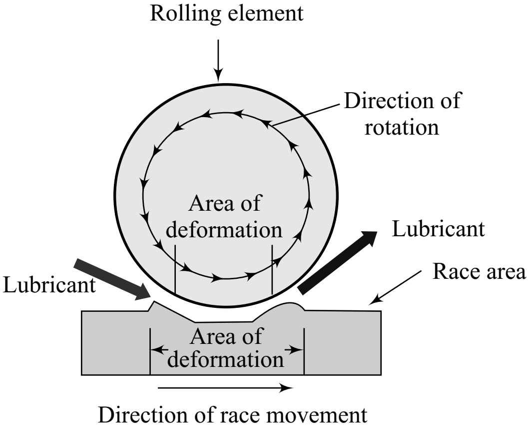


FIGURE 20.14 *Explanation of elastohydrodynamic lubrication.*

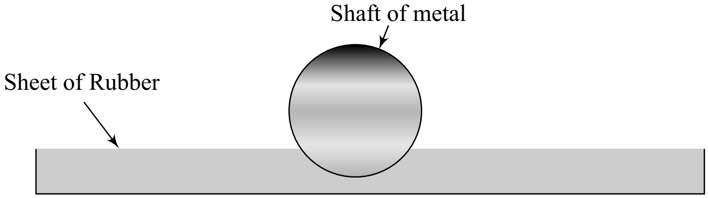


FIGURE 20.15 *(a) A shaft of metal resting on a sheet of rubber.*

With oil lubrication, the pressures of the hydrodynamic film will complicate the scenario. The pressure of the hydrodynamic film will exert pressure on the deformable medium and the shaft lifting it upward. The shape of the rubber will be changed as shown below in Fig. 20.15(b). However, the shaft will be shifted to one side depending on its direction of spin.

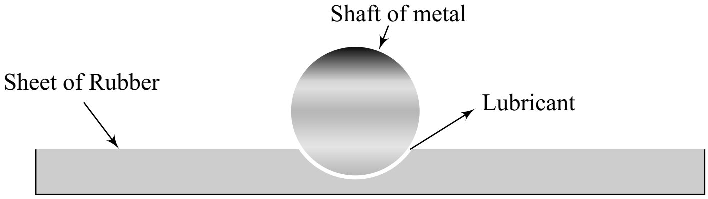


FIGURE 20.15 *(b) A shaft of metal resting on a sheet of rubber with lubricant.*

Let us extend this model to include all deformable medium, even metal gear teeth. Because the metal is very hard it has a very high threshold of elastic deformation, often as high as 100,000 psi in gear teeth. When these two deformable mediums come together, it is theoretically possible to capture a hydrodynamic film between the two faces such that there is never actually any metallic contact occurring. In this case of extreme pressure, the viscosity of the oil film may increase as much as 10,000x behaving virtually like a solid between the two surfaces. This explains the reasons of operating the mechanisms under much harsher conditions. There is a catch though, this type of elastohydrodynamic lubrication works only when the thin film is in the order of 10–40 millionths of an inch. This requires the surfaces to be extremely smooth and carefully aligned. This keeps elastohydrodynamic lubrication from becoming an everyday lubrication answer. It is, however, very applicable and necessary in high pressure situations.

20.9 Basic Types of Lubricants

The process of lubrication is simply to achieve or improve the smoothness between the two moving surfaces and the material which is used in this way is called a lubricant. Lubricants are usually liquids or semi-liquids, but may be solids or gases or any other combination of solids, liquids and gases. Lubricants reduce the friction and the smoothness of movement is improved. This is not, however, always the case, and there may be situations in which it is more important to maintain uniform friction than to obtain the lowest possible friction.

Lubricants are usually expected to reduce wear and often to prevent overheating and corrosion, in addition to reduce or control friction.

Lubricants are usually divided into four basic classes.

1. *Oils*: Liquid lubricants encompass all mineral oils, natural oils, synthetics, emulsions or even process fluids.
2. *Greases*: Greases are oils, which contain a thickening agent to make them semi-solid. It is convenient, however, to include the antiseize pastes and the semi-fluid greases under the same heading.
3. *Dry lubricants*: Many bulky solids, paint like coatings or loose powders in the solid form are used as lubricants.
4. *Gases*: The gas usually used in gas bearings is air, but any gas can be used which will not attack the bearings, or decompose itself.

20.10 Classification of Lubricants

Lubricants are classified on the basis of their physical state as follows:

1. Liquid lubricants

- Vegetable oils: palm oil, castor oil, etc.
- Animal oils: whale oil, tallow oil, etc.
- Mineral oils: petroleum fractions.
- Synthetic lubricant: silicones, polyglycol ethers, etc.
- Blended oils or compounded oils: mineral oils with various additives

2. Semi-solid lubricants

Greases, Vaseline, etc.

3. Solid lubricants

Graphite, molybdenum-disulphide, etc.

4. Emulsions

(a) Oil in water type: Cutting emulsions.
(b) Water in oil type: Cooling liquids.

20.10.1 Liquid Lubricants

(a) Vegetable-based Liquid Lubricants

Liquid lubricants can also be obtained from plant life. Vegetable-based lubricants are manufactured from seeds bearing trees/plants of the following:

- canola
- palm oil
- rapeseed
- sunflower seed

Vegetable lubricants are obtained from soybean, corn, castor, canola, cotton seed and rape seed oils. Vegetable oils are environmentally friendly alternative to mineral oils since they are biodegradable. Lubrication properties of vegetable base oils are identical to those of mineral oils. The main disadvantages of vegetable lubricants are their low oxidation and temperature stabilities.

Though, vegetable oil based lubricants are more environmentally friendly than mineral and synthetic oils, they are rarely used in vehicles or machinery. Natural oils include vegetable oils and animal fats. They are usually excellent boundary lubricants, but they are much less stable than mineral oils, and tend to break down to give sticky deposits. Rapeseed oil is still used by itself or as an additive to mineral oils to give improved boundary lubrication. A recent development is an increase in use of castor oil and palm oil in manufacture of biodegradable and food industry grade lubricants.

(b) Animal lubricants: Lanolin

Animal fats are the source for animal based lubricants. There are two main animal fats as follows:

- Hard fats (stearin) and
- Soft fats (lard).

Greases are mainly manufactured from animal fats. Lanolin is obtained from animal matter as a liquid lubricant. Lanolin, a naturally water repellent substance, present on sheep's wool grease.

Lanolin is a naturally occurring, renewable resource that can be harvested every time a sheep is sheared. Consequently, it is nontoxic, making it an ecologically friendly alternative to mineral and synthetic oils, as well as reducing friction and wear. Lanolin liquid lubricants have anticorrosive properties.

(c) Mineral oil lubricants: fluid lubricant oils

Naturally occurring mineral oils are the source for the production of mineral oil lubricants. Mineral oils (petroleum oils) are the products of refining crude oil. The distillation of crude petroleum oil gives mineral oils that can be processed to create fluids with varying viscosity. The mineral oils used for lubrication were originally just the fractions, obtained by distilling petroleum, which had a suitable viscosity for lubrication. Mineral oils are some of the most common liquids used in machinery. Mineral oils are easily the most widely used lubricating oils. Consequently, they are often the standard with which other oils are compared. Mineral oils generally mean oils obtained from petroleum, although they can also be obtained from similar sources, such as oil shales and tar-stands.

There are three types of mineral oils:

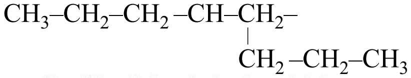
- Paraffinic
- Naphthenic and
- Aromatic.

(i) The majority of the lubricating oil is derived from *paraffin family*, in which the carbon atoms are in straight or branched chains but rings and paraffinic oils are not produced either by hydrocracking or solvent extraction process.

Examples:

$CH_3–CH_2–CH_2–CH_2–$

(Paraffin with straight carbon chain)

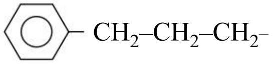


$$\begin{array}{l} CH_3–CH_2–CH_2–CH–CH_2– \\ \qquad\qquad\qquad\quad | \\ \qquad\qquad\qquad\quad CH_2–CH_2–CH_3 \end{array}$$

(Paraffin with branched carbon chain)

Paraffinic oils are relatively viscous and resistant to oxidation. They possess high flash point and high pour point. Paraffinic oils are used for manufacturing engine oils, industrial lubricants and as processing oils in rubber, textile and paper industries.

(ii) *Naphthenic oils* are produced from crude oil distillates. The second most common types are *naphthenes*, in which some of the carbon atoms form rings.

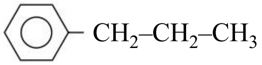


Naphthene

Naphthenic oils possess low viscosity, low flash point, low pour point and low resistance to oxidation. Naphthenic oils are used in moderate temperature applications, mainly for manufacturing transformer oils and metal working fluids.

(iii) *Aromatic oils* are products of refining process in manufacture of paraffinic oils.

Most hydrocarbon molecules of aromatic oils have nonsaturated ring structure.



Aromatics

Aromatic oils are dark and have high flash point. Aromatic oils are used for manufacturing seal compounds, adhesives and plasticizers in rubber and asphalt production.

There is a class of lubricating material known as *asphaltenes* containing other elements, such as oxygen, sulphur, phosphorus and nitrogen.

(d) Motor oils

Motor vehicles demand highly specific properties for motor oils; and these motor oils satisfactorily perform some of the basic functions. All of these functions are considered when designing any motor oil, whether it will be synthetic or mineral-based oil.

- **Permit easy starting**
  Motor vehicle should start the engine easily with the use of proper motor oil. A good motor oil with desired viscosity and other properties is preferable for quick starting of engine, Oil's ability to flow efficiently throughout an engine will be affected by air temperature, therefore, an oil formulated for winter driving must have a low enough pour point for it to flow to all parts of an engine quickly when started and provide an adequate protection once the engine reaches normal operating temperatures.
- **Lubricate and prevent wear**
  Motor oils must also lubricate and prevent wear and tear in all temperature extremes. Oil that does not flow well in cold temperatures will leave parts of the engine with no protection, and oil that burns off

and becomes too fluid will leave little protection in high temperatures. The goal of oil is to provide constant full-film lubrication to an engine's components. This type of lubrication occurs when the moving surfaces are continuously separated by a film of oil. Crankshaft bearings as well as connecting rods, cam-shafts and piston rings normally operate with full-film lubrication. Boundary lubrication occurs when it is impossible to maintain a continuous oil film between moving parts and intermittent metal-to-metal contact results. Additives can greatly reduce the amount of damage that can occur during boundary lubrication. Boundary lubrication conditions always exist during engine starting and often during the operation of a new or rebuilt engine.

- **Reduce friction**
  Motor oils must reduce friction. Automobile manufacturers recommend oils based on SAE grades according to the expected atmospheric conditions. This helps to ensure adequate but not excessive viscosity at normal operating temperatures. Excessive viscosity can make an engine work harder at moving the oil and therefore loses some of its efficiency and create more heat.
- **Prevent rust and corrosion**
  Motor oils must also prevent rust and corrosion. Unburned fuel and soot can mix with water to form sludge and varnish deposits on critical engine parts. Sludge build up may clog oil passages, which reduces oil flow. Varnish build up interferes with proper clearances, restricts oil flow and causes vital engine parts to stick and malfunction. The life of an engine depends on the oil's ability to neutralize the effects of these corrosive substances.
- **Keep engines clean**
  Another important feature of any motor oil related to preventing rust and corrosion is the necessity of keeping engine components clean. Sludge and varnish can be controlled with the proper additives and can be filtered out of vital engine components. In performing its lubrication function, some oil must reach the area of the top piston ring in order to lubricate the rings and cylinder walls. However, when exposed to the heat and flame of burning fuel, part of the oil actually burns off. Modern oils have been chemically engineered to burn as cleanly as possible in order to minimize the harmful deposits left on the walls of the combustion chambers. When excessive deposits build up, this can cause ring sticking and breakage, pinging, engine knock or other combustion irregularities that reduce the efficiency and economy of the engine.
- **Cool vital components**
  Engine oil also acts as a coolant. In fact, the water-antifreeze mixture used in an automobile's cooling system only does about 60 percent of the job cooling only the upper portions of the engine such as the cylinder heads, cylinder walls and the valves. As much as 5–10 percent of engine heat comes from friction produced by closely fitted engine parts such as connecting rods, main bearings, camshafts and piston rings. The crankshaft, the main and connecting rod bearings, the camshaft and its bearings, the timing gears, the pistons and many other components in the lower engine rely on oil to carry the heat load away.
- **Seal**
  The surfaces of the piston rings, ring grooves and cylinder walls are not completely smooth. They feature microscopic hills and valleys that can reduce engine efficiency by allowing combustion pressure to escape into the low pressure area of the crankcase. Motor oils must fill in these hills and valleys on ring surfaces and cylinder walls, allowing maximum combustion pressure.
- **Prevent foaming**
  Antifoam additives in modern oils prevent foaming due to air in the crankcase being whipped into the oil. Foamed oil contains air bubbles that may or may not readily collapse, which inhibit heat transfer and impede lubrication performance due to the high compressibility of air. In order to allow an engine to run efficiently, motor oil must be nonfoaming.

- **Increase fuel economy**
  A final function of motor oil is increasing fuel economy. Various additives such as friction modifiers can allow engines to operate at increased levels of efficiency, resulting in better fuel economy for vehicles.

  All of these functions, to some extent, can be performed by both mineral and synthetic oils. However, synthetic oils do not contain many of the impurities that mineral oils can, and this allows synthetics to perform at higher levels in all categories. AMSOIL has long recognized the benefits that the synthetic oil can offer over a mineral oil. Synthetic fluids exhibit higher tolerances and greater lubricating properties in every aspect of automotive operation, AMSOIL motor oils contain the optimal additive packages to insure that a vehicle can receive the finest in protection and operate at a high level of efficiency.

20.10.2 Synthetic Lubricants

Petroleum lubricants are predominantly hydrocarbon products extracted from fluids that occur naturally within the Earth. They are used widely as lubricants because they possess a combination of the following desirable properties:

- availability in suitable viscosities,
- low volatility,
- inertness (resistance to deterioration of the lubricant),
- corrosion protection (resistance to deterioration of the sliding surfaces) and
- low cost.

Synthetic lubricants generally, having some properties similar to petroleum lubricants and are characterized as oily, neutral liquid materials. These are not usually obtained directly from petroleum. These synthetic lubricants are superior to hydrocarbon products. Synthetics exhibit greater stability of viscosity with temperature changes, resistance to scuffing and oxidation and fire resistance. Since the properties of synthetics vary considerably, each synthetic lubricant tends to find a special application. A few of the more common classes of synthetics and typical uses of each are shown in Table 20.1.

Synthetic lubricant	Typical uses
Di-basic acid esters	instrument oil, jet turbine lubricant, hydraulic fluid
Phosphate esters	fire-resistant hydraulic fluid, low-temperature lubricant
Silicones	damping fluid, low-volatility grease base
Silicate esters	heat transfer fluid, high-temperature hydraulic fluid
Polyglycol ether compounds	synthetic engine oil, hydraulic fluids, forming and drawing compounds
Fluorol compounds	Nonflammable fluid, extreme oxidation-resistant lubricant

Table 20.1

Synthetic lubricants and typical applications

Synthetic lubricant oils are synthesized as desired to suit the engines/machines that have greater operating ranges than mineral oil lubricants. As a result, synthetic oils are better choices than mineral oils in certain circumstances. The best examples of synthetic oils are esters, silicone fluid and fluorocarbons. Silahydrocarbons are synthetic oils composed, synthesized from hydrogen, silicon and carbon. Silahydrocarbons have the potential to provide fluidity and lubrication in space machinery.

Many of the alternative types of oil are synthetic. They are manufactured from various feed stocks by chemical processes. There are different types of synthetic oil and are very different from each other in their performance and properties. A brief discussion of these synthetic oils is given below:

1. Polymeric films

The two known synthetic polymeric films are polytetrafluoroethylene and polychlorofluoroethylene. The major advantage of such film-type lubricants is their resistance to deterioration during exposure to the elements. Thus, $^1/_2$-inch thick plates of polymeric film are used in modern prestressed concrete construction to permit thermal movement of beams resting atop columns. Such expansion and contraction of the structural members is facilitated by the long-lived polymeric film plate.

2. Synthetic hydrocarbons

There are several classes of synthetic hydrocarbon. The most important are branched-chain paraffins which are known as polyalphaolefins (PAO). Polyalphaolefins are the most popular synthetic lubricant. PAO's chemical structure and properties and performance are very similar to some of the most highly refined mineral oils. Polyalphaolefins are one of the most important classes of synthetic oil, yet they resemble mineral oils. Their inherent oxidation resistance is quite good and their boundary lubrication not very good, like those of the highly refined mineral oils.

Polyalphaolefins (synthetic hydrocarbons) are manufactured by polymerization of hydrocarbon molecules (alphaoleins). The process occurs in reaction of ethylene gas in presence of a metallic catalyst.

3. Diesters and polyol esters

Diesters and polyol esters are the most important class of synthetic oils used extensively. They are almost universally used for aircraft gas turbine engine lubrication. Diesters are a class of 'carboxylic esters', containing the carboxylic group –O—C=O. The better thermal stability polyesters are derived from carboxylic esters. Diesters are used as base oils for high-temperature applications in industry, such as hot-rolling oils in steel rolling. They can be used at higher temperatures than mineral oils, have very good lubricating characteristics and are readily available in a variety of viscosities.

Different viscosity grades polyol esters are available in the market and they are now extensively used in aircraft jet engines. Apart from their good high-temperature properties, they also have very good lubricating characteristics.

4. Phosphate esters

The phosphate ester oils are produced by the reaction between alcohols and phosphoric acid.

Due to their outstanding fire resistance, they are extensively used, especially, in high fire-risk situations such as aircraft hydraulic systems, coal mines and hot metal-processing. They are chemically similar to some of the best antiwear additives, and consequently have excellent boundary lubricating behaviour. Their main disadvantages are poor thermal stability and their powerful solvent action on many pains, plastics and rubbers.

5. Silicones

Silicones are a group of inorganic polymers. The silicones have a chain skeleton containing alternate silicon and oxygen atoms and are polymeric substances. The molecules of silicones have a backbone structure containing repeated units of monomers Si=O moieties. Two organic groups are attached to each Si=O moiety: e.g. methyl + methyl $[(CH_3)_2]$, methyl + phenyl $(CH_3 + C_6H_5)$, phenyl + phenyl $[(C_6H_5)_2]$. The most popular silicone is polydimethylsiloxane (PDMS). It is built from monomer, $(CH_3)_2SiO$. PDMS is produced from silicon and methyl chloride. Other examples of silicones are polymethylphenylsiloxane and polydiphenylsiloxane.

Viscosity of silicones depends on the length of the polymer molecules and on the degree of their cross-linking. They are available in a very wide range of viscosity grades. Short noncross-linked molecules make fluid silicone. Long cross-linked molecules result in elastomer silicone. Silicone lubricants (oils and greases) are characterized by broad temperature range (–73°C to 300°C). They are generally used for their good high-temperature stability which enables them to be used over 200°C. They are widely used as base oils for high-temperature greases. The silicones are poor boundary lubricants and this is the main disadvantage especially for steel against steel.

6. Chlorinated biphenyls

The chlorinated biphenyls are available to use as mediocre boundary lubricants; but they are very good fire resistance and chemical inert materials.

7. Polyglycols (PAG)

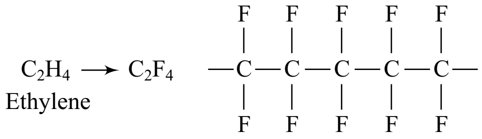
Polyglycols are long-chain polymeric liquids. They are stable to about 200°C. When heated above this temperature, they decompose cleanly without producing any undesirable decomposition products. Polyglycols are produced by oxidation of ethylene and propylene. The oxides are then polymerized resulting in formation of polyglycol. Polyglycols are water soluble. Polyglycols are characterized by very low coefficient of friction. They are also able to withstand high pressures without EP (extreme pressure) additives.

8. Fluorinated ethers and fluorocarbons

These oils have the highest thermal stability and chemical inertness of any oils, and can be used continuously at temperatures over 300°C. Like silicones, they are poor boundary lubricants.

9. Organic lubricants with chain structure of the polymeric molecules

Ethylene on fluorination process yields tetrafluoroethylene which further polymerizes under certain conditions to polytetrafluoroethylene (PTFE). It is often referred to as teflon.



$$\underset{\text{Ethylene}}{C_2H_4} \longrightarrow C_2F_4 \qquad -\underset{\displaystyle F}{\overset{\displaystyle F}{\underset{|}{\overset{|}{C}}}}-\underset{\displaystyle F}{\overset{\displaystyle F}{\underset{|}{\overset{|}{C}}}}-\underset{\displaystyle F}{\overset{\displaystyle F}{\underset{|}{\overset{|}{C}}}}-\underset{\displaystyle F}{\overset{\displaystyle F}{\underset{|}{\overset{|}{C}}}}-\underset{\displaystyle F}{\overset{\displaystyle F}{\underset{|}{\overset{|}{C}}}}-$$

Polytetrafluoroethylene

It is a white solid with slightly waxy appearance, hard to touch but can be easily cut and deformed. It is usually fabricated by sintering or hot compression moulding at above 300°C. It is difficult to form large components accurately, provided that the cutting speed and depth of cut are kept low to avoid thermal expansion. Bonding PTFE to metals has always been difficult, but it is said that treatment with hot alkali gives a surface, which can be bonded. Adhesives are commercially available which are claimed to be suitable. It should also be kept in mind that toxic and irritant vapours can be emitted from PTFE and other fluorinated polymers if the temperature rises during machining.

It is stable in use up to almost 300°C, but it changes its state at 325°C and cannot be used above that temperature. It can also be very low temperatures, even down to –200°C or lower in liquefied gases. There is, however, some evidence that the wear rate increased at very low temperatures. It is also very resistant to oxidation, and can be used for lubrication or sealing in oxygen systems.

The attractive forces between the long straight-chain molecules are low, so that PTFE has fairly poor mechanical strength. For this reason, the long molecules slide easily over each other; thus giving their low friction properties. In view of this stability and low atomic forces, it is not clear why PTFE forms good, smooth, strongly adhering films to metal surfaces; but it does probably more readily that graphite.

PTFE is widely used as an additive in lubricating oils and greases. Due to the low surface energy of PTFE, stable unflocculated dispersions of PTFE in oil or water can be produced. Contrary to the other solid lubricants discussed, PTFE does not have a layered structure. The macro molecules of PTFE slip easily along each other, similar to lamellar structures. PTFE shows one of the smallest coefficients of static and dynamic friction, down to 0.04. Operating temperatures are limited to about 260°C.

Polytetrafluoroethylene (PTFE) and polychlorofluoroethylene are the typical examples of such materials. The molecular structure of the materials consists of long chain molecules parallel to each other. The bonding strength between the molecules is weak therefore they may slide past one other at low shear stresses. The strength of the molecules along the chains is high due to strong bonding between the atoms within a molecule. Such anisotropy of mechanical properties provides good lubrication properties of the materials. Chain structure lubricants are used in the form of coatings (films) applied on the substrates surfaces.

10. Ester oils

Ester oils are produced by reaction of acids and alcohols with water. Ester oils are characterized by very good high temperature and low temperature resistance.

20.10.3 Semi-Solid Lubricants

Semi-fluid lubricants (greases) are produced by emulsifying oils or fats with metallic soap and water at 204°C to 316°C.

One of the best example of mineral oil based grease is vaseline. The properties of grease are determined by a type of oil (mineral, synthetic, vegetable, animal fat), type of soap (lithium, sodium, calcium, etc. salts of long-chained fatty acids) and additives (extra pressure, corrosion protection, antioxidation, etc.). Nonsoap thickeners consist of such inorganic compounds as modified clays or fine silica, or such organic materials as aryl ureas or phthalocyanine pigments. By adding animal and vegetable fats to soap and water, it is possible to create grease lubricants.

Semi-fluid lubricants (greases) are used in variety of applications where fluid oil is not applicable and where thick lubrication film is required such as lubrication of roller bearings in railway car wheels, rolling mill bearings, steam turbines, spindles, jet engine bearings and other various machinery bearings. Large quantities of greases are used as lubricants for surfaces of railway wagon axles. If the melting point of fats and greases is used close to the temperature of the objects around them, the lubricating films get so heated by the two surfaces rubbing against each other (or causing friction) that they melt and turn into oils or semi-solid lubricants. In this state, these semi-solid lubricants are put into bearings by a kind of industrial syringe with a spring piston.

It is desirable to lubricate any machinery parts by grease than lubrication by oil under conditions when (1) less frequent lubricant application is necessary, (2) grease acts as a seal against loss of lubricant and ingress of contaminants, (3) less dripping or splattering of lubricant is called for or (4) less sensitivity to inaccuracies in the mating parts is needed.

20.10.3.1 Grease

Grease comprises mixtures of lubricating oils and thickeners. The thickeners are dispersed in lubricating oils in order to produce a stable and colloidal structure or gel. Thus, a grease consists of oil constrained by minute thickener fibres. There are three components that form lubricating grease (Fig. 20.16). They are *oil, thickener* and *additives*. The base oil and additive package are the major components in grease formulations, and as such, exert considerable influence on the behaviour of the grease. The thickener is often referred to as a sponge that holds the lubricant (base oil plus additives).

Greases always contain three basic active ingredients:

- a base mineral or synthetic oil,
- additives, and
- thickener.

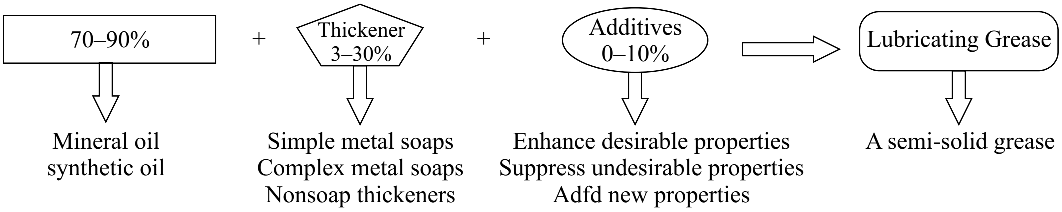


Figure 20.16 *Composition of grease.*

(a) Base Oil

In most greases, the main fluid components are as follows:

- *Mineral oils* are generally used as the base component in grease formulation. About 99 percent of greases are made with mineral oils. Naphthenic oils are the most popular despite their low viscosity index.
- *Synthetic oils* are used for formulation of greases. These greases are expected to operate in extreme conditions. The most commonly used are synthetic esters, phosphate esters, silicones and fluorocarbons. Their most common applications are in high performance aircraft, missiles and in space.
- *Vegetable oils* are also used in making grease intended for the food and pharmaceutical industries, but even in this application their use is quite limited.

These mineral oil-based greases typically provide satisfactory performance in most industrial applications. In temperature extremes (low or high), a grease that utilizes a synthetic base oil will provide better stability.

(b) Thickener

In order to produce a semi-solid structure of lubricant, a thickener material in combination with the selected lubricant is used for the formulation of grease. Soaps are very important in the production of greases. Soap type greases are the most commonly produced. In the chemistry of soap manufacture, it is necessary to heat some fats or oils in the presence of an alkali, e.g. caustic soda (NaOH). Fats and oils can be animal or vegetable, and are produced from cattle, fish, castor bean, coconut, cottonseed, etc. Apart from sodium hydroxide (NaOH) other alkali can be used in the reaction, as for example, lithium, cadmium, aluminium, barium, etc.

The reaction products are soap, glycerol and water. The most commonly used soap type greases are obtained from calcium, lithium, aluminium, and sodium.

In nonsoap-type greases, inorganic, organic and synthetic materials are used as thickeners. Inorganic thickeners are in the form of very fine powders, which have enough porosity and surface area to absorb oil. The most commonly used are the silica and bentonite clays.

Complex greases are made by combining the conventional metallic soap with a complexing agent. The most widely used complex grease is lithium based. These are made with a combination of conventional lithium soap and a low-molecular-weight organic acid as the complexing agent.

Nonsoap thickeners are also gaining popularity in special applications such as high-temperature environments. Bentonite and silica aerogel are two examples of thickeners that do not melt at high temperatures.

(c) Additives

The additives are added to the formulation of greases to primarily enhance the existing desirable properties, suppressing the existing undesirable properties and imparting new properties. The most common additives are oxidation and rust inhibitors, extreme pressure, antiwear and friction-reducing agents. It can play several roles in a lubricating grease. In addition to these additives, boundary lubricants such as molybdenum disulfide

(MoS_2) or graphite may be suspended in the grease to reduce friction and wear without adverse chemical reactions to the metal surfaces during heavy loading and slow speeds.

The additives used in grease formulations are similar to those used in lubricating oils. Some of them modify the soap, others improve the oils characteristics. The most common additives include antioxidants, rust and corrosion inhibitors, tackiness, antiwear and extreme pressure (EP) additives.

The advantages and disadvantages of oils stem from their ability to flow easily. Thus, on the negative side, it is very easy to pour them from a container, to feed them into a bearing by dripping, splashing or pumping and to drain them out of a machine when no longer fit for use. Other advantages are the cooling of a bearing by carrying away heat and cleaning it by removing debris. The behaviour of greases is very similar to that of oils, but the former is used where the advantages of easy flow are outweighed by the disadvantages. Thus grease does not easily leak out of a machine, or container, does not migrate away and will form an effective seal against contaminants.

The advantages and disadvantages of solid lubricants are rather like the extremes for greases, where the lubricant will not flow at all. Similarly, the advantages and disadvantages of gas lubricants are like the extremes of oils, where the flow properties are almost too good.

20.10.3.2 Function of Grease

The function of grease is to remain in contact with and lubricate moving surfaces without leaking out under the force of gravity, centrifugal action or being squeezed out under pressure. Its major practical requirement is that it retains its properties under shear forces at all temperatures it experiences during use.

20.10.3.3 Applications of Grease

Grease and oil are not interchangeable. Grease is used when it is not practical or convenient to use oil. The lubricant choice for a specific application is determined by matching the machinery design and operating conditions with desired lubricant characteristics. Grease is generally used for:

- Machinery that runs intermittently or is in storage for an extended period of time. Because grease remains in place, a lubricating film can instantly form.
- Machinery that is not easily accessible for frequent lubrication. High-quality greases can lubricate isolatedly or relatively inaccessible components for extended periods of time without frequent replenishing. These greases are also used in sealed-for-life applications such as some electrical motors and gearboxes.
- Machinery operating under extreme conditions such as high temperatures and pressures, shock loads or slow speed under heavy load.
- Worn components. Grease maintains thicker films in clearances enlarged by wear and can extend the life of worn parts that were previously lubricated by oil.

20.10.3.4 Functional Properties of Grease

1. Grease functions as a sealant to minimize leakage and to keep out contaminants. Because of its consistency, grease acts as a sealant to prevent lubricant leakage and also to prevent entrance of corrosive contaminants and foreign materials. It also acts to keep deteriorated seals effective.
2. Grease is easier to contain than oil. Oil lubrication can require an expensive system of circulating equipment and complex retention devices. In comparison, grease, by virtue of its rigidity, is easily confined with simplified, less costly retention devices.

3. Grease holds solid lubricants in suspension. Finely ground solid lubricants, such as molybdenum disulphide (MoS_2) and graphite, are mixed with grease in high-temperature service or in extreme high-pressure applications. Grease holds solids in suspension while solids will settle out of oils.
4. Fluid level does not have to be controlled and monitored.

20.10.3.5 Characteristics of Grease

As with oil, grease displays its own set of characteristics that must be considered when being chosen for an application. The characteristics commonly found in product data sheets include the following:

(a) Pumpability

Pumpability is the ability of a grease to be pumped or pushed through a system. More practically, pumpability is the ease with which the pressurized grease can flow through lines, nozzles and fittings of grease-dispensing systems.

(b) Water resistance

This is the ability of grease to withstand the effects of water with no change in its ability to lubricate. The soap/water lather may suspend the oil in the grease, forming an emulsion that can wash away or, to a lesser extent, reduce lubricity by diluting and changing grease consistency and texture.

(c) Consistency

Grease consistency depends on the type and amount of thickener used and the viscosity of its base oil. The grease's consistency is its resistance to deformation by an applied force. The measure of consistency is called penetration. Penetration depends on whether the consistency has been altered by handling or working or not. To measure penetration, a cone of given weight is allowed to sink into a grease for 5 seconds at a standard temperature of 25°C (77°F).

The depth in tenths of millimetre to which the cone sinks into the grease is the penetration. A penetration of 100 would represent a solid grease while penetration of 450 would be semi-fluid.

(d) Dropping point

Dropping point is an indicator of the heat resistance of grease. As grease temperature increases, penetration increases until the grease liquefies and the desired consistency is lost. The dropping point is the temperature at which the grease becomes fluid enough to drip. The dropping point indicates the upper temperature limit at which the grease retains its structure, not the maximum temperature at which a grease may be used.

(e) Oxidation stability

This is the ability of grease to resist a chemical union with oxygen. The reaction of grease with oxygen produces insoluble gum, sludges and lacquer-like deposits that cause sluggish operation, increased wear and reduction of clearances. Prolonged exposure to high temperatures accelerates oxidation in greases.

(f) High-temperature effects

High temperatures harm greases more than they harm oils. Grease, by its nature, cannot dissipate heat by convection like circulating oil. Consequently, without the ability to transfer away heat, excessive temperatures result in accelerated oxidation or even carbonization where grease hardens or forms a crust.

Effective grease lubrication depends on the grease's consistency. High temperatures induce softening and bleeding, causing grease to flow away from needed areas. The mineral oil in grease can flash, burn or evaporate at temperatures greater than 177°C (350°F).

(g) Low-temperature effects

If the temperature of grease is lowered enough, it will become so viscous that it can be classified as hard grease. Pumpability suffers and machinery operation may become impossible due to torque limitations and power requirements. As a guideline, the base oil's pour point is considered as low-temperature limit of grease.

20.10.4 Solid Lubricants

Solid lubricants are solid materials which reduce coefficient of friction and wear of rubbing parts preventing direct contact between their surfaces even under high loads.

20.10.4.1 Forms of Solid Lubricants

In order to reduce friction, the surface areas are covered in the forms of either dispersed particles or surface films:

- Coating (film) of a solid lubricant applied on the part of surface.
- Composite coating consisting of particles of a solid lubricant dispersed throughout a matrix.
- Particles of a solid lubricant dispersed throughout the bulk of the part material (composite material).
- Powder of a solid lubricant delivered to the rubbing area (dry lubrication).
- Additives in lubricating oils or greases.

20.10.4.2 Requirements to Solid Lubricants Properties

- *Shear strength of lubricants should be low in the sliding direction*. This property provides low coefficient of friction due to easy shear movement of the lubricant material.
- *Compression strength should be high in the direction of the load* (perpendicular to the sliding direction). A solid lubricant possessing high compression strength is capable to withstand high loads without sufficient direct contact between the rubbing surfaces.
- *Good adhesion* of the solid lubricant to the substrate surface. This property provides a presence of the solid lubricant on the part surface even at high shear stresses.

Shear strength and compression strength properties of solid lubricants are observed in anisotropic materials like graphite, molybdenum disulphide or boron nitride having lamellar crystal structure.

20.10.4.3 Characterization of Solid Lubricants

Advantages of solid lubricants:

- Ability to work under high loads.
- High thermal stability.
- Diversity of the application forms

Disadvantages of solid lubricants:

- Higher coefficient of friction and wear as compared to hydrodynamic region.
- Low stability of the lubrication film.
- Less convenient system of the lubricant delivery to the friction surfaces. In contrast to solid lubricants, fluid lubricants are continuously supplied, filtered and cooled.

20.10.4.4 Typical Applications

Solid lubricants are useful for conditions where conventional lubricants are inadequate.

- Solid lubricants are used in a sliding or reciprocating motion that requires lubrication to minimize wear; as for example, in gear and chain lubrication. Liquid lubricants will squeeze out while solid lubricants do not escape and prevent for fretting corrosion and galling.

- In the absence of chemically active lubricant additives, polymers and ceramics find applications for a particular surface.
- Graphite and MoS_2 exhibit high temperature and oxidizing atmosphere environments, whereas liquid lubricants typically will not survive. A typical application includes fasteners which are easily tightened and unscrewed after a long stay at high temperatures.

The lamellar structure orient parallel to the sliding surface resulting in high bearing-load combined with a low shear stress. Most applications in metal forming that involve plastic deformation will utilize solid lubricants.

20.10.5 Types of Solid Lubricants

Solid lubricants possess lamellar that prevents direct contact between the sliding surfaces even at high loads. Generally, graphite and molybdenum disulphide (MoS_2) particles are commonly used as solid lubricant. Boron nitride, tungsten disulphide and polytetrafluoroethylene (PTFE) are other solid lubricants. Solid lubricants are also used in the form of dry powder or as constituents of coatings. Solid lubricants are mainly used as additives to oils and greases.

These lubricants are used when liquids or soft solids cannot be used and when the surfaces between which the lubricant will be placed are hard materials. These materials are then mixed with oil or grease to improve their efficiency as lubricants.

In the form of dry powder these materials are effective lubricant additives due to their lamellar structure. The lamellas orient parallel to the surface in the direction of motion as shown in Fig. 20.17.

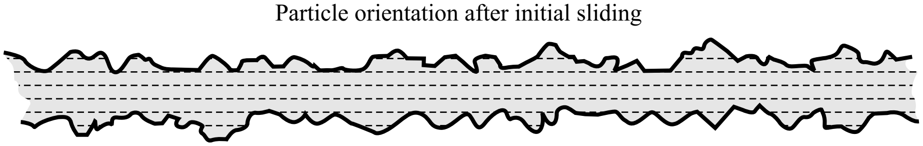


FIGURE 20.17 *Orientation of lamellas parallel to the surface in the direction of motion.*

The lamellar structures of these materials are able to prevent contact even between highly loaded stationary surfaces. In the direction of motion, the lamellas easily shear over each other resulting in a low friction. Large particles best perform on relative rough surfaces at low speed, finer particle on relative smooth surface and higher speeds.

Other components that are useful solid lubricants include boron nitride, polytetrafluoroethylene (PTFE), talc, calcium fluoride, cerium fluoride and tungsten disulphide.

20.10.5.1 Inorganic Lubricants with Lamellar Structure

The crystal lattice of these inorganic lubricant materials has a layered structure consisting of hexagonal rings forming thin parallel planes. Within the plane, each atom is strongly bonded (covalent bonding) to other atoms. The crystal lattice planes are bonded to each other by weak van der Waals forces. This layered structure of inorganic materials allows sliding movement of the parallel planes. Weak bonding between the planes determines low shear strength and lubricating properties of the materials. The most commonly used inorganic solid lubricants with lamellar structure are graphite, molybdenum disulphide (MoS_2) and boron nitride (BN). Other examples of such materials are sulphides, selenides and tellurides (chalcogenides) of molybdenum, tungsten, niobium, tantalum, titanium (e.g.WS_2, WS_2, $MoSe_2$, $TaSe_2$, $TiTe_2$), monochalcoselenides (GaS,GaSe, SnSe), chlorides of cadmium, cobalt, lead, cerium, zirconium (eg. $CdCl_2$, $CoCl_2$, $PbCl_2$, CeF_3, PbI_2) and also some borates (e.g. $Na_2B_4O_7$) and sulfates (Ag_2SO_4).

(a) Graphite

The first solid lubricant to be used in large scale was graphite. It is a grey-black crystalline form of carbon in which the carbon atoms are arranged hexagonally in regular layers (Fig. 20.18).

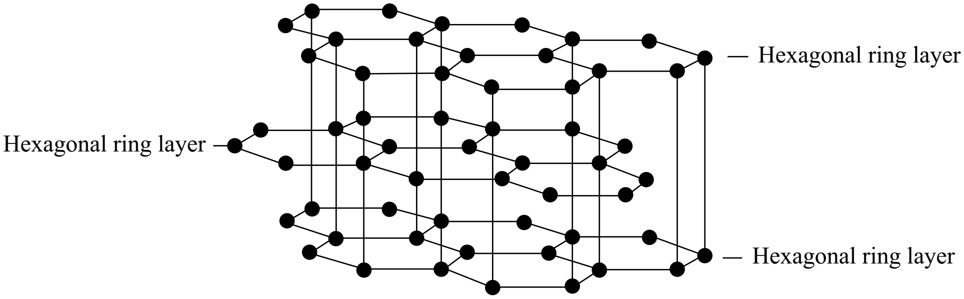


Figure 20.18 *Structure of graphite.*

The bonds between the carbon atoms within a layer are strong covalent bonds in the lamellar hexagonal structure of graphite, as a result, the layers are strong and the crystals strongly resist bending or breaking of the layers. The bonds between the layers are weak van der Waals forces, so that the crystals can be made to split easily between layers, and the layers will slide readily one over the other. The layer-lattice structure of graphite would appear to give good natural low-friction properties. In fact, graphite only gives low friction when it is contaminated or 'intercalated' by water vapour of other condensable vapours. High-quality graphite with water vapour present will give coefficients of friction varying from 0.05 at high contact pressures to 0.15 at low pressures; the friction is smooth and steady. Graphite will adhere readily to surfaces either from a solid block or from dispersions in liquids. It can also be made to adhere more strongly by means of adhesive binders, usually epoxy, phenolic or alkyd, which may also be included in the dispersion.

Graphite is structurally composed of planes of polycyclic carbon atoms that are hexagonal in orientation. The distance of carbon atoms between planes is longer and therefore the bonding is weaker. Graphite is best suited for lubrication in a regular atmosphere. Water vapour is a necessary component for graphite lubrication. The adsorption of water reduces the bonding energy between the hexagonal planes of the graphite to a lower level than the adhesion energy between a substrate and the graphite. Because water vapour is a requirement for lubrication, graphite is not effective in vacuum. In an oxidative atmosphere, graphite is effective at high temperatures up to 450°C continuously and can withstand much higher temperature peaks. The thermal conductivity of graphite is generally low ~1.3 W mK^{-1} at 40°C.

Graphite is characterized by two main groups: natural and synthetic. Synthetic graphite is a high temperature sintered product and is characterized by its high purity of carbon (99.5–99.9 percent). The primary grade synthetic graphite can approach the good lubricity of quality natural graphite.

Natural graphite is derived from mining. The quality of natural graphite varies as a result of the ore quality and post mining processing of the ore. The end product is graphite with a content of carbon (high grade graphite 96–98 percent carbon), sulphur, SiO_2 and Ash. The higher the carbon content and the degree of graphitization (high crystalline) the better the lubricity and resistance to oxidation. For applications where only a minor lubricity is needed and a more thermally insulating coating is required, then amorphous graphite would be chosen (80 percent carbon).

(b) Molybdenum Disulphide

Molybdenum disulphide has been used extensively as solid lubricant. The natural ore containing molybdenum sulphide is molybdenite. Large quantities of crude hexagonal molybdenum disulphides are available in nature.

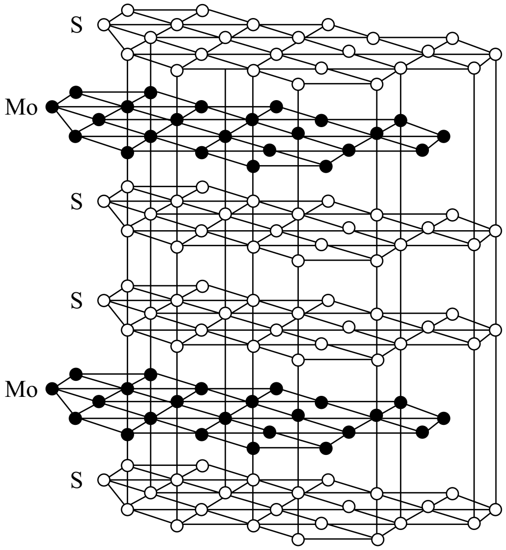


Figure 20.19 *Structure of molybdenum disulphide.*

One molybdenum atom is linked to two sulphur atoms in MoS_2. Like the graphite layer lattice, it has a hexagonal layer-lattice structure, as well as other, nonlubricating, forms. The structure of MoS_2 consists of alternate layers of molybdenum and sulphur atoms (Fig. 20.19).

Molybdenum disulphide differs from graphite mainly in that its low friction is an inherent property and does not depend on the presence of absorbed vapours. Because of this, it can be used satisfactorily in high vacuum, and it has been used for many applications in spacecraft.

It adheres even more strongly than graphite to metal and other surfaces. A useful film can be obtained by simply rubbing or 'burnishing' molybdenum disulphide powder onto a metal surface with cotton wool or cloth.

MoS_2 is present within granite in the form of thin veins and can be mined with high purity suitable for lubricants. MoS_2 a hexagonal crystal structure similar to graphite with the intrinsic property of easy shear. MoS_2 lubrication performance often exceeds that of graphite and is effective in vacuum as well whereas graphite does not. The temperature limitation of MoS_2 at 400°C is restricted by oxidation. The particle size and film thickness are important parameters that should be matched to the surface roughness of the substrate. Large particles may result in excessive wear by abrasion caused by impurities in the MoS_2, small particles may result in accelerated oxidation.

(c) Boron nitride

Boron nitride is an important class of ceramic powder lubricant. It has high temperature resistance of 1200°C service temperature in an oxidizing atmosphere. Further boron has a high thermal conductivity. Interestingly, it is noted that boron nitride exists in two forms of crystal lattice: cubic and hexagonal. Due to its tight diamond-like structure cubic boron nitride is extremely hard and it has poor lubrication properties. It is used in cutting and abrasive tools as a diamond substitute. Hexagonal boron nitride (HBN) is a solid lubricant relating to the class of inorganic lubricants with lamellar structure, which also includes molybdenum disulphide, graphite and some other sulphides, selenides and tellurides (chalcogenides) of molybdenum, tungsten, niobium, tantalum and titanium.

The crystal lattice of hexagonal boron nitride consists of hexagonal rings forming thin parallel planes. Atoms of boron and nitrogen are covalently bonded to other atoms in the plane with the angle 120° between two bonds (each boron atom is bonded to three nitrogen atoms and each nitrogen atom is bonded to three boron atoms) as shown in Fig. 20.20. The planes are bonded to each other by weak van der Waals forces.

The sliding movement of the parallel planes arises due to the layered structure of boron nitride. Weak bonding between the planes provides low shear strength in the direction of the sliding movement but high compression strength in the direction perpendicular to the sliding movement. Friction forces cause the particles of boron nitride to orient in the direction, in which the planes are parallel to the sliding movement. The anisotropy of mechanical properties imparts the combination of low coefficient of friction and high carrying load capacity to boron nitride.

Boron nitride forms a lubrication film strongly adhered to the substrate surface. The lubrication film provides good wear and seizer resistance (compatibility). Similar to molybdenum disulphide, moist atmosphere is not required for lubrication by boron nitride. It demonstrates low friction in dry atmosphere and in vacuum. Coefficient of friction of boron nitride is within the range 0.1–0.7, which is similar to that of graphite and

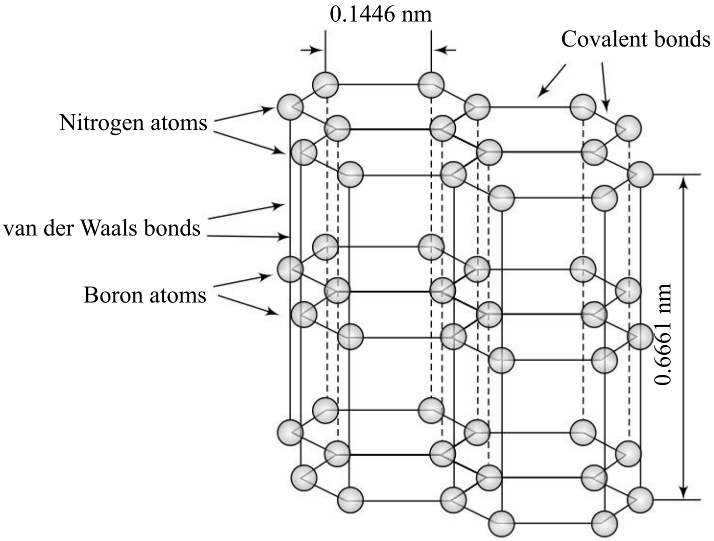


Figure 20.20 *Structure of hexagonal boron nitride.*

molybdenum disulphide. Impurities (e.g. boron oxide) exert adverse effect on the lubrication properties of boron nitride. Boron nitride is chemically inert substance. It is nonreactive to most acids, alkalis, solvents and nonwetted by molten aluminium, magnesium, molten salts and glass. The main advantage of boron nitride as compared to graphite and molybdenum disulphide is its thermal stability. Hexagonal boron nitride retains its lubrication properties up to 2760°C in an inert or reducing environment and up to 870°C in an oxidizing atmosphere. Boron nitride has high thermal conductivity.

Some applications of hexagonal boron nitride:

- Additives in lubricating oils
- Components of polymer based composite antifriction coatings
- Second phase particles of metal based composite antifriction coatings
- Solid lubricant in metal forming
- Release coatings and nonsticking refractory linings in foundry
- Sintered ceramic parts for high temperature applications

20.10.5.2 Oxides

Few oxides of metals are examples of solid state lubricants. They are B_2O_3, MoO_2, ZnO, Re_2O_7, TiO_2, CuO-MoO_2, NiO-Mo_2, PbO-B_2O_3 and CuO-Re_2O_7.

20.10.5.3 Soft Metals

Some soft metals are used as lubricating material due to their inherent low shear strength and high plasticity. Such soft metals are lead (Pb), tin (Sn), bismuth (Bi), indium (In), cadmium (Cd) and silver (Ag). Soft metals are used in pure form. They find application as alloys in the form of coatings (lead based engine bearing overlays), (tin based engine bearing overlays) or the second phase in metal matrix composites (copper based bearing materials, aluminium based bearing materials). Coatings from soft metal lubricants are produced by the methods of electroplating, vapour deposition and thermal spraying. Soft metal composite lubricants are prepared by casting or sintering methods. Soft metals are widely used as solid lubricants in engine bearing materials.

Miscellaneous soft solids are the variety of inorganic solids such as white lead, lime, talc, bentonite, silver iodide and lead monoxide are used as lubricants. Thin films of soft metal on a hard substrate can act as effective lubricants, if the adhesion to the substrate is good. Such metals include lead, tin and indium.

20.10.5.4 Self-lubricating Composites

Solid lubricants are compounded in plastics to form a 'self-lubricating' or 'internally lubricated' thermoplastic composite. Self-lubricating composites are obtained from PTFE, graphite, MoS_2 and also from other antifriction and antiwear additives compounded in polymers.

MoS_2 is compounded in materials for sleeve bearings, elastomer O-rings, carbon brushes, etc. PTFE particles compounded in the plastic form as a PTFE film over the mating surface resulting in a reduction of friction and wear. MoS_2 compounded in nylon reduces wear, friction and stick-slip. Furthermore, it acts as a nucleating agent effecting in a very fine crystalline structure. The primary use of graphite lubricated thermoplastics is in applications operating in aqueous environments.

Application of lubrication methods:

- **Spraying/dipping/brushing:** Solid lubricant, as an additive dispersed in oil water or grease, is used. A dry film lubricant can be sprayed to inaccessible parts after its assembly for lubrication. After the solvent evaporates, the coating cures at room temperature to form a solid lubricant. Pastes are grease like lubricants containing a high percentage of solid lubricants used for assembly and lubrication of highly loaded, slow moving parts. Black pastes generally contain MoS_2. For high temperatures, above 500°C pastes are composed on the basis of metal powders to protect metal parts from oxidation necessary to facilitate disassembly of threaded connections and other assemblies.
- **Free powders:** Dry powder tumbling is an effective application method. The bonding can be improved by prior phosphating the substrate. Use of free powders has its limitations, since adhesion of the solid particles to the substrate is usually insufficient to provide any service life in continuous applications. However, to improve running-in conditions or in metal forming processes, a short duration of the improved slide conditions may suffice.
- **AF-coatings:** Lubricating paints are known as antifriction coatings consisting of fine particles of lubricating pigments, such as moly disulphide, PTFE or graphite, blended with a binder.

  After application and proper curing, these lubricants bond to the metal surface and form a dark grey solid film. Dry film lubricants containing special rust inhibitors protect them from corrosion. AF-coatings are applied where fretting and galling is a problem (such as splines, universal joints and keyed bearings), where operating pressures exceed the load-bearing capacities of ordinary oils and greases, where smooth running in is desired (piston, camshaft), where clean operation is desired (AF-coatings will not collect dirt and debris like greases and oils), where parts may be stored for long periods of time.

  Most long wearing films are of the bonded type but are still restricted to applications where sliding distances are not too long.

20.10.6 Emulsions

- **Water-based lubricants**

  In metal cutting and grinding, water-based lubricants find extensive uses.

  The simplest water-based lubricant is water itself. A more complex, commercial water-based lubricant usually consists of water mixed with some type of base oil. Water-based lubricants are also used as personal lubricants. Because oil-based lubricants can erode latex, water-based lubricants are better choices for safely using condoms.

- **Soaps**
  Soaps are alkali metal salts of fatty acids. Soaps are obtained by reacting oils and fats with strong alkaline solutions. A soap molecule is composed of a long nonpolar hydrocarbon tale which is hydrophobic (repelled by water) and the salt polar end which is hydrophilic (water soluble). The soap molecules attached to the substrate surface provide good adhesion of the soap lubricant and low coefficient of friction.

20.10.7 Gaseous Lubricants

The process of lubrication with a gas is similar in many respects to lubrication with a liquid, since the same principles of fluid-film lubrication are applied. It can be seen that both gases and liquids are viscous fluids and they differ in two important aspects. The viscosity of gases is much lower and the compressibility is much greater than that for liquids. Film thicknesses and load capacities therefore are much lower with a gas such as air. In equipment that handles gases of various kinds, it is often desirable to lubricate the sliding surfaces with gas in order to simplify the apparatus and reduce contamination to and from the lubricant. The list of gases used in this manner is extensive and includes air, steam, industrial gases and liquid-metal vapours.

20.11 Lubricant Additives

The additives for lubricants are chemicals preferably organic or organometallic that are incorporated into oils in few weight percent to improve the lubricating capacity and durability of the oil.

Additives are present in many liquid lubricants. Examples of lubricant additives include:

- antifoaming agents
- antioxidants
- antiwear agents
- corrosion inhibitors
- demulsifying agents
- emulsifying agents
- rust inhibitors
- viscosity improvers

The additives are added to lubricants with specific purposes. They are

- improving the wear and friction characteristics by provision for adsorption and extreme pressure (E.P) lubrication,
- improving the oxidation resistance,
- control of corrosion,
- control of contamination by reaction products, wear particles and other debris,
- reducing excessive decrease of lubricant viscosity at high temperatures,
- enhancing lubricant characteristics by reducing the pour point and inhibiting the generation of foam.

The most common package of additives used in oil formulations contains antiwear and extreme pressure lubrication additives, oxidation inhibitors, detergents, dispersants, viscosity improves, pour point depressants and foam inhibitors.

20.11.1 Wear and Friction Improvers

Among the additives to lubricant, the one which improve wear and friction properties, are probably the most important of all the additives used in oil formulations.

These chemicals are adsorption and extreme pressure additives and they control the lubricating performance of the oil.

These additives can be divided into the following groups:

- adsorption or boundary additives,
- antiwear additives, and
- extreme pressure additives.

20.11.2 Adsorption or Boundary Additives

The additives added to control the adsorption type lubrication are known as adsorption or boundary additives. They are 'friction modifiers' since they are often used to prevent the slipstick phenomena.

Generally, these types of additives are mostly the fatty acids and the esters and amines of the same fatty acids. They usually have the polar group (–OH) at one end of the molecule and react with the contacting surfaces through the mechanism of adsorption. The surface film generated by this mechanism is effective only at relatively low temperatures and loads. The polar group to form a carpet of molecules, which reduces friction and wear, attaches the molecules to the surface.

20.11.3 Antiwear Additives

Specifically synthesized antiwear additives were added in order to protect contacting surfaces at higher temperatures above the range of effectiveness of adsorption or boundary agents. There are several different types of antiwear additives that are currently used in oil formulations. For example, in engine oils the most commonly used antiwear additive is zinc dialkyl dithiophosphate (ZnDDP), in gas turbine oils tricresyl phosphate or other phosphate esters are used. Phosphorous additives are used where antiwear protection at relatively low loads is required.

These additives react with the surfaces through the mechanism of chemisorption, and the protective surface layer produced is much more durable than that generated by the adsorption or boundary agents.

Common examples of these additives are zinc dialkyl dithiophosphate, tricresylphosphate, dilauryl phosphate, diethyl phosphate, dibutylphosphate, tributylphosphate and tripara-cresylphosphate. These additives are used in concentrations of 1–3 percent by weight.

(a) Zinc dialkyl dithiophosphate

Zinc dialkyl dithiophosphate is used as an antioxidant and detergent and later observed to be as an antiwear and mild extreme pressure additive. The term 'antiwear' usually refers to wear reduction at moderate loads. The surface protective films, which are formed as the result of action of ZnDDP, act as a lubricant, reducing wear and friction between two interacting surfaces. The lubricant mechanism of ZnDDP is quite complex as the additive has three interacting active elements that is zinc, phosphorus and sulphur. All of these elements and compounds are involved in surface film formation, and our current understanding of the surface films produced is that they consist of a matrix of zinc polyphosphate with inclusions of iron oxide and iron sulphide. The problem of valve train wear and oil degradation in internal combustion engines was solved by applying ZnDDP.

(b) Tricresylphosphate

Like ZnDDP, it functions by chemisorption to the operating surfaces. It is very effective in reducing wear and friction at temperatures up to about 200°C. Beyond this temperature there is sufficient energy input to the surface for the chemisorbed films to desorb and the compound will then form less effective, much weaker, thick phosphate films with limited load capacity.

Other antiwear additives such as dilauryl phosphate, diethylphosphate, dibutylphosphate, tributylphosphate and triparacresylphosphate are also being used in lubricant formulation.

20.11.4 Extreme Pressure Additives

The extreme pressure additives are introduced to lubricants to react with metal surfaces under extreme conditions of load and velocity, that is slowly moving, heavily loaded gears.

Under these conditions, operating temperatures are high and the metal surfaces are hot. Extreme pressure additives contain usually at least one aggressive nonmetal such as sulphur, antimony, iodine or chlorine. They react with exposed metallic surfaces creating protective and low shear strength surface films that reduce friction and wear. The reaction with the metallic surfaces is a form of mild corrosion, thus the additive concentration is critical.

There are several different types of extreme pressure additives currently added to oils. The most commonly used are dibenzyl disulphide, phosphosulphurized isobutene, trichlorocetane and chlorinated paraffin, sulphurchlorinated sperm oil, sulphurized derivatives of fatty acids and sulpurized sperm oil, cetyl chloride, mercaptobenzothiazole, chlorinated wax, lead naphthenates, chlorinated paraffinic oils and molybdenum disulphide.

- Dibenzyl disulphide is a mild extreme pressure additive, which has sulphur, positioned in a chain between two organic radicals. Examples of this type of additive are butyl phenol disulphide and diphenyl disulphide. The specific type of hydrocarbon radical, e.g. diphenyl, provides a useful control of additive reactivity to minimize corrosion.
- Trichlorocetane and chlorinated paraffin are powerful extreme pressure additives but they are also very corrosive, particularly when contaminated with water. They are applied in extreme situations of severe lubrication problems, e.g. screw cutting.
- Paraffinic mineral oils and waxes can be chlorinated to produce E.P. additives. They are not very popular since the mineral oils are quite variable in their composition and usually a poorly characterized additive results from this procedure. Such additives may have very serious undesirable side effects, e.g. toxicity and corrosiveness.
- Sulphur chlorinated sperm oil is an effective E.P. additive, but is becoming obsolete because of the increasing rarity of harvested sperm whale oil. It is still, however, used in heavy duty truck axles.
- Molybdenum disulphide provides lubrication at high contact stresses. It functions by depositing a solid lubricant layer on the contacting surfaces. It is noncorrosive but is very sensitive to water contamination as water causes the additive to decompose.

20.11.5 Antioxidants

The performance of machinery is always affected due to significant increase in friction and wears. The reason for this cause is due to the oxidation of mineral oils. A gradual increase in the viscosity and acidity of an oil occurs which is due to the oxidation of mineral oils. Oxidation is not the only cause of viscosity increase in lubricating oils. Another cause is diesel soot. Any increase in the elevated oil acidity can cause concentrated corrosion of certain machinery components such as seals and bearings. For example, lead, copper and cadmium are used in the bearing of alloys of an internal combustion engine and they are particularly prone to corrosion.

20.11.6 Oxidation Inhibitors

The onset of oxidation of the oils is controlled by the addition of antioxidant additives to lubricating oils. These are either natural antioxidants or artificially introduced additives that are able to suppress oxidation and any differences in the oxidation resistance of oils largely depend on the presence of these inhibitors. Sulphur and nitrogen of the additive oxidation inhibitors always scavenge the radicals produced by the oxidation process. Sulphur based extreme pressure and antiwear additives are also quite effective as antioxidants.

Widely used antioxidant additives are zinc dialkyl dithiophosphate, metal deactivators, simple hydrocarbons such as phenol derivatives, amines and organic phosphates. Sulphur and phosphorus in elemental form or incorporated into organic compounds are also effective as antioxidants and antiwear additives. They are sometimes added to oils but are likely to cause corrosion problems or may precipitate also lose effectiveness as an additive. Antioxidants are usually added to the oil in very small quantities at a concentration of approximately 1 percent by weight.

Corrosion control activities:

There are two groups of corrosion control activities:

- corrosion inhibitors, and
- rust inhibitors.

Corrosion inhibitors are used to protect the nonferrous surfaces of bearings, seals, etc. against corrosive attack which are treated with various additives, especially those containing reactive elements such as sulphur, phosphorus, iodine, chlorine and oxidation products. It is observed that some of the oxidation products are very acidic. They must be neutralized before they cause any damage to the operating parts of the machinery. One of the major corrosion attacks on nonferrous metallic parts may be due to the combination of corrosive additives, oxidation products, high temperature and very often wear of the parts.

The commonly used additives to control the corrosion of nonferrous metals are benzotriazole, substituted azoles, zinc diethyl dithiophosphate, zinc diethyl dithiocarbamate and trialkyl phosphites.

Rust inhibitors are used to protect the ferrous components against corrosion. The main factors, which contribute to accelerated corrosion attack of ferrous parts, are oxygen dissolved in the oil and water. These can cause an electrolytic attack that may be even more accelerated with increased temperature. Rust inhibitors are usually long chain agents, which attach themselves to the surface, severely reducing the mobility of water. The commonly used additives, which control the corrosion of ferrous metals, are metal sulphonates (i.e. calcium, barium, etc.), amine succinates or other polar organic acids.

20.11.7 Contamination of Control Additives

A new class of additives has been developed for the treatment of internal combustion engines.

Engine oils are regularly exposed to fuel and combustion products which inevitably contribute to their contamination. When sulphur is present in the fuel, sulphurous acid is formed during combustion. Accelerated corrosion occurs on metal parts due to the presence of sulphurous acid is formed in the presence of water.

There can also be many other possible contaminants such as soot from inefficient fuel combustion, wear debris, unburned fuel, breakdown products of the base oil, corrosion products, dust from the atmosphere, organic debris from micro-biological decomposition of the oil, etc. Without proper control of contamination, the oil will lose its lubricating capacity, become corrosive and will be unsuitable for service. Various additives have been developed to control the acidity of the products of sulphurous combustion of dirty fuel and to prevent agglomeration of soot from combustion and wear particles. The agglomeration of particles can be very destructive to engines since it blocks the oil supply pipe-lines or even the filters. Additives, which prevent the development of all these detrimental effects, are known in the literature as ‘detergents’ of ‘dispersants’.

The primary functions of these additives are:

- to neutralize any acids formed during the burning of fuel,
- to prevent lacquer and varnish formation on the operating parts of the engine,
- to prevent the flocculation or agglomeration of particles and carbon deposits which may choke the oil ways.

There are two types of dispersant: a mild dispersant and an over-based or alkaline dispersant.

- Mild dispersants are often composed of simple hydrocarbons or ashless compounds. Mild dispersants are typically low molecular weight polymers of methyl acrylate esters, long chain alcohols or polar vinyl compounds. The function of these additives is to disperse soot (carbon) and wear particles.
- Over-based dispersants are calcium, barium or zinc salts of sulphonic, phenol or salicylic acids. Over-based dispersants means that an excess of alkali is used in the preparation of these additives. The additive is present in the mineral oil as a colloid. The alkaline prepared additive serves to neutralize any acid accumulated in the oil during service. Alkaline dispersants have a disadvantage in that they accelerate oil oxidation and therefore require the addition of an antioxidant to the oil.

20.11.8 Viscosity Improvers

The role of the viscosity improvers is to control the lubricant oils viscosity with temperature. This is an additive to lubricant and is known as viscosity index improver. Viscosity improving additives are usually high molecular weight polymers which are dissolved in the oil and can change the shape from spheroidal to linear as the temperature is increased. This effect is caused by a greater solubility of the polymer in the oil at higher temperatures and partly offsets the decline in base oil viscosity with temperature. The linear or uncoiled molecules cause a larger rise in viscosity in comparison to spheroidal or coiled molecules.

The main problem associated with these additives is that they are easily degraded by excessive shear rates and oxidation. Under high shear rates viscosity improvers can suffer permanent or temporary viscosity loss.

20.12 Classification of Lubricants by Application

Based on the applications of lubricants, they are classified as

- Engine oils
- Gear oils
- Hydraulic oils
- Cutting fluids (coolants)
- Way lubricant
- Compressor oils
- Quenching and heat transfer oils
- Rust protection oils
- Transformer oils (Insulating oil)
- Turbine oils
- Chain lubricants
- Wire rope lubricants

20.13 Classification of Lubricants by Additives

The lubricants are further classified based on the presence of additives,

- Extreme pressure (EP)
- Antiwear (AW)
- Friction modifiers
- Corrosion inhibitors
- Antioxidants
- Dispersant

- Detergents
- Compounded
- Antifoaming agents
- Pour point depressant

20.14 Applications of Lubricants

The main applications of lubricants are primarily used to reduce friction stress between surfaces. They have the following uses:

- As antiwear, antioxidants and antifoaming agents.
- As demulsifying and emulsifying agents.
- As rust and corrosion inhibitors.
- In machinery as engine oils, compressor oils, gear oils and piston oils.
- As hydraulic, brake and gear box fluids.
- Used in the soap and paint industries.

Some specific uses of certain variants of lubricants are as follows:

- Synthetic lubricants are used in turbines, vacuum pumps and semiconductor devices.
- Molybdenum is used as a paint pigment and as a catalyst.
- Liquid lubricants are used in medicines.
- Lubricants are also used as cutting fluids in many industries. Oil, water and oil emulsion are used as cutting fluids. These liquids are used to cool as well as to lubricate surfaces. Emulsions of oil in water are most widely used cutting fluids. Lubricants are also used as cutting fluids in cutting, grinding, trading, and drilling of the metals. Cutting fluids are used in machining operations where friction is very high because of close contact between the workpiece and the tool. This high friction generates a large amount of local heat and the tool is overheated and may even lose its temper and hardness. As a result, liquids, such as lubricating oils, water or water emulsions are used on working parts of the machines. Cutting fluids should have the following qualities:
    - Good lubricating property
    - Low viscosity, so that lubricants can easily fill in the cracks formed on the workpiece

Review Questions

1. What is lubrication?
2. What are the types of frictions?
3. Explain rolling friction.
4. Explain sliding friction.
5. Explain static friction.
6. Explain kinetic friction.
7. Explain fluid friction.
8. Explain lubricated friction.
9. Explain skin friction.
10. What are lubricants?
11. What are the functions of lubricants?
12. What are the requirements for satisfactory lubricant performance?
13. Name the tests for lubrication?
14. Explain viscosity of lubricants.
15. What is viscosity index?
16. What is pour point?
17. What is flash point?
18. What is fire point?
19. What do you mean by oiliness?
20. Give the types of lubrication mechanism.
21. Explain hydrodynamic lubrication.
22. Explain boundary lubrication.
23. Explain elastohydrodynamic lubrication.
24. Which are the basic types of lubricants?
25. How are lubricants classified? Name them.

26. Which are vegetable based liquid lubricants? Name.
27. Name few animal lubricants.
28. What are mineral oil lubricants? Explain.
29. Explain motor oils.
30. What are synthetic lubricants and their uses?
31. Explain polymeric films.
32. Explain synthetic hydrocarbons.
33. Explain the use of diesters and polyol esters.
34. What are phosphate esters?
35. Why are silicones used as lubricant? Give details.
36. What are polyglycols?
37. Explain organic lubricants with chain structure of the polymeric molecules.
38. A note on solid lubricants.
39. Explain the role of grease in lubrication.
40. Mention the application of grease.
41. Which are the characteristic properties of grease?
42. Explain solid lubricants.
43. What are the different forms of solid lubricants?
44. Give the typical applications of solid lubricants.
45. Mention the types of solid lubricants.
46. What are inorganic lubricants with lamellar structure?
47. Write a note on graphite as lubricant.
48. Write a note on molybdenum disulphide as lubricant.
49. Write a note on boron nitride as lubricant.
50. Mention few names of oxides as lubricants.
51. Which are the soft metals used as lubricants?
52. What are self-lubricating composites?
53. Mention the applications of lubrication methods.
54. What do you mean by emulsion lubrication?
55. Name the additives of lubricants.
56. What are wear and friction improvers?
57. Write a note on antiwear additives.
58. What is an antioxidant in lubrication?
59. A note on corrosion control activities of lubrication.
60. Mention the use of viscosity improvers in lubrication.
61. Mention the applications of lubricants.

21

Thermal Analysis

Chapter Outline

Introduction. Thermogravimetric analysis (TGA)—introduction, instrumentation, Sources of error in thermogravimetric analysis, calibration of TGA unit, data analysis of thermograms, factors affecting TG Analysis, factors affecting the TG curves. Applications of thermogravimetric analysis (TGA), quantitative interpretation of the thermogravimatic curves. Interpretation of the TG curves. Introduction to differential thermal analysis (DTA)—The basic principle of DTA, DTA technique, DTA instrumentation, Experimental details, characteristics of DTA curves, DTA data analysis, enthalpy calibration. Instrumentation—source of uniform heating. Factors affecting DTA curves, sources of errors. Interpretation of DTA curves. Differential scanning calorimeter (DSC)—Introduction to DSC, the basic types of DSC systems—principle based on power compensation, principle based on maintaining heat-flux in DSC. Calibration of DSC. Determination of enthalpy change or heat capacity of materials—measurement of enthalpy change, peak area and enthalpy change of sample, heat capacities. Factors affecting DSC curve. Interpretation of DSC curves.

21.1 Introduction

Thermal analysis (TA) is a group of techniques in which changes of *properties of materials* (physical or chemical) of the sample are monitored against time or temperature, while the temperature of the sample is programmed. In practice, thermal analysis gives properties such as enthalpy, thermal capacity, mass changes and the coefficient of heat expansion. Thermal analysis encompasses the following studies:

- Solid state chemistry uses thermal analysis for studying the behaviour of solids.
- Thermal degradation reactions of samples in the solid state.
- Study of phase transitions and phase diagrams of materials.

Thermal analysis includes several different methods. These are distinguished from one another by the property which is measured.

- Thermogravimetric analysis (TGA): mass changes.
- Differential thermal analysis (DTA): temperature difference.
- Differential scanning calorimetry (DSC): heat difference.
- Pressurized TGA (PTGA): mass changes as function of pressure.
- Thermomechanical analysis (TMA): deformations and dimension.
- Dilatometry (DIL): volume.
- Evolved gas analysis (EGA): gaseous decomposition products.

In this chapter, the discussions are confined to only three topics (TGA, DTA and DSC) based on the property that is measured for the materials.

The different thermal techniques of solids and their properties studied are shown in Table 21.1.

Thermogravimetric analysis	TGA	Mass difference	Sample purity Decomposition Dehydration Oxidation
Differential thermal analysis	DTA	Temperature difference	Phase changes Dehydration Decomposition Reactions
Differential scanning calorimetry	DSC	Energy difference	Heat capacity Phase changes Reactions Calorimetry

Table 21.1

Different thermal techniques and their properties studied

21.2 Thermogravimetric Analysis (TGA)

21.2.1 Introduction

A widely used thermal method of sample analysis is ‘thermogravimetric (TG) technique’. It is based on the very simple principle of *monitoring the change in weight of a sample as the temperature is varied.*

The thermogravimetric analysis of the materials is also investigated at a controlled atmosphere (e.g. with O_2 or N_2) which encourage or suppress oxidation reactions, thus controlling to some extent the nature of the thermal events occurring. When materials are heated over a range of temperatures (ambient), many materials undergo weight changes giving characteristic curves known as thermograms. Thermogravimetric (TGA) analysis determines endotherms, exotherms, weight loss on heating, cooling and many more. Materials analysed by the TGA technique include polymers, plastics, composites, laminates, adhesives, food, coatings, pharmaceuticals, organic materials, rubber polymer degradation temperatures, residual solvent levels, absorbed moisture content and the amount of inorganic (noncombustible) filler in polymer or composite material compositions and a host of other materials.

A thermobalance is used to measure the mass change of a sample as a function of temperature or time, under a defined and controlled environment with respect to heating rate, gas atmosphere, flow rate, crucible type, etc.

Thermogravimetry (TGA) is based on the following principles:

1. The amount and rate of change in the weight of samples is observed while the sample is subjected to a programmed heating and atmosphere.
2. The TGA of the sample is studied at a linear temperature programme. It is also carried out when the changes in sample mass with time are followed. Measurements are used primarily to determine the composition of materials and to predict their thermal stability at temperatures up to 1000°C.
3. TGA is inherently quantitative and, therefore, an extremely powerful thermal technique.

The ability to analyse the volatile products during a weight loss is of great value. The technique can characterize materials that exhibit weight loss or gain due to decomposition, oxidation or dehydration, the size of the step in the curve can be used for quantitative analysis.

In thermogravimetry, a continuous graph of mass change against temperature is obtained when a substance is heated at a uniform rate or kept at constant temperature.

The measured weight loss in TGA curve gives information on:

- changes in sample composition
- thermal stability
- kinetic parameters for chemical reactions in the sample.

TGA-applications are as follows:

- Characterization
- Thermal stability
- Material purity
- Determination of humidity
- Examination
- Corrosion studies (e.g. oxidation or reactions with reactive gases)
- Gasification processes
- Kinetic processes

The following processes in TGA reveal the weight gain or loss of samples (Table 21.2)

Process	Weight gain	Weight loss
Adsorption or absorption	√	–
Desorption, drying	–	√
Dehydration	–	√
Sublimation	–	√
Vaporization	–	√
Decomposition	–	√
Solid state Reactions	√	√
Magnetic Transitions	√	√

Table 21.2

Processes in TGA indicating the weight gain or loss of samples

A plot of mass change versus temperature (*T*) is referred to as the thermogravimetric curve (TG curve). For the TG curve, we generally plot mass (*m*) decreasing downwards on the *y*-axis (ordinate) and temperature (*T*) increasing towards the right on the *x*-axis (abscissa) as illustrated in Fig. 21.1.

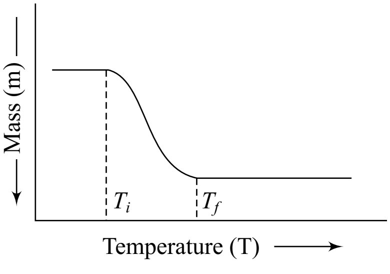


Figure 21.1 *A typical TGA curve.*

The TGA curve (or thermogram) can also be obtained by plotting time (t) in place of temperature (T). TGA curve helps in revealing the extent of purity of analytical samples and in determining the mode of their transformations within the specified range of temperature. In a TGA curve of a single stage decomposition, there are two characteristic temperatures: the initial Ti and the final temperature T_f (Fig. 21.1). T_i is defined as the lowest temperature at which the onset of a mass change can be detected by thermo balance operating under particular conditions and T_f as the final temperature at which the particular decomposition appears to be complete. The difference $T_f - T_i$ is termed as 'reaction interval' for a specific material. In a dynamic thermogravimetry, a sample is subjected to continuous increase in temperature (usually linear) with time whereas in isothermal or static thermogravimetry, the sample is maintained at a constant temperature for a period of time during which any change in mass is noted.

Where thermograms of samples are complex or where changes are subtle, derivative curves (DTG) can be valuable in interpretation. A *derivative weight loss curve* (*DTG*) can be used to indicate the point at which weight loss is most apparent. Figure 21.2 shows an example of a straightforward thermogram of decomposition of calcium oxalate (hydrate) and Fig. 21.3 shows a more complex thermogram illustrating the importance of DTG indicating the temperature at which the weight loss occurs (Fig. 21.3b).

Plot change in mass (*dm*) of the sample with temperature(*dt*). The derivative (*dm*/*dt*) plot of the sample resolve changes more clearly.

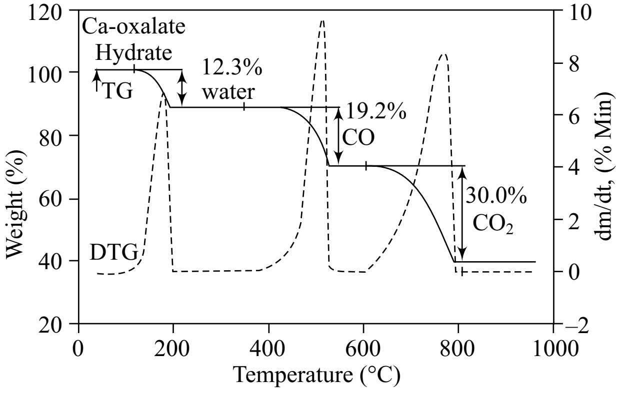


Figure 21.2 *TG curve for calcium oxalate.*

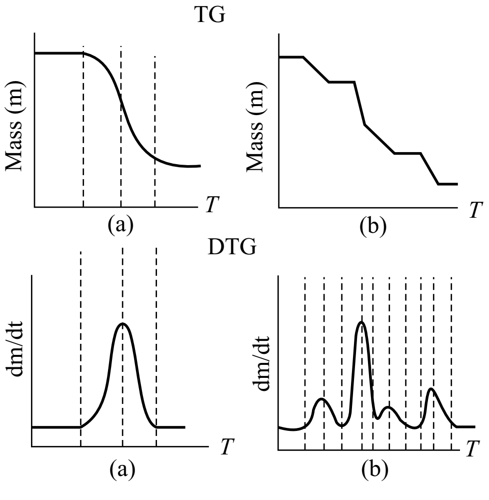
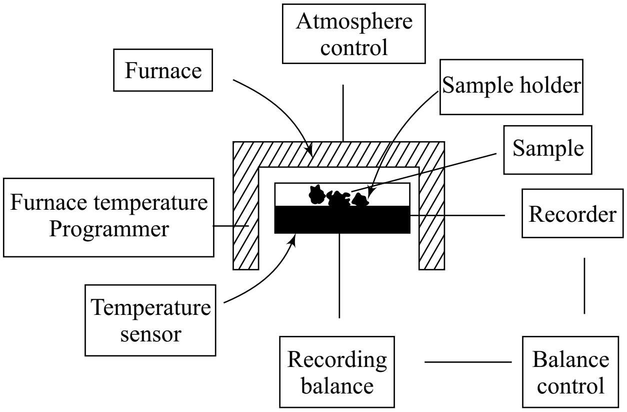


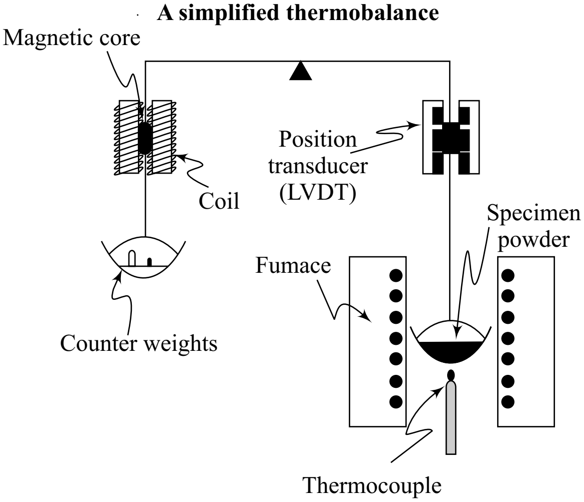
Figure 21.3 *Comparison of TG and DTG curves.*

21.2.2 Instrumentation

The instrument used in thermogravimetry (TG) is called a '*thermobalance*'. It consists of several basic components in order to provide the flexibility necessary for the production of useful analytical data in the form of TGA curve, as shown in Fig. 21.4.



(a) TGA basic principle



(b) Schematic TGA typical construction design.

FIGURE 21.4 *Block diagram of a (TGA) thermobalance balance.*

Basic components of a typical thermobalance are listed as follows:

(i) Balance
(ii) Furnace: heating device
(iii) Unit for temperature measurement and control (Programmer)
(iv) Recorder: automatic recording unit for the mass and temperature changes.

A thermobalance is used to measure the mass change of a sample as a function of temperature or time in a defined and controlled environment with respect to heating rate, gas atmosphere, flow rate, crucible type, etc.

These components may be represented by a simple block diagram in Fig. 21.4.
(a) TGA basic principle (b) Schematic TGA typical construction design.

(i) Furnace

The furnace is controlled to produce a linear heating over the whole working temperature range of the furnace and also provided with a provision to study of materials at any fixed temperatures. A wide temperature range generally between –150 and 1500°C of furnaces is used in different instrument manufacturers depending on the models. The range of furnace basically depends on the types of heating elements that are used.

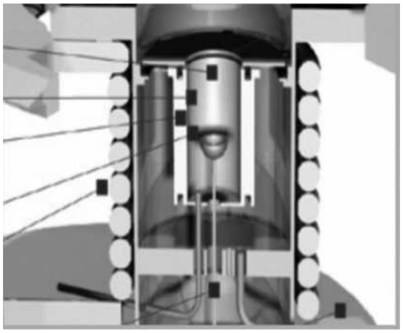
(ii) Temperature measurement and control

Temperature measurement is commonly performed using chromel–alumel thermocouple. They are often used for temperature up to 1100°C whereas Pt/(Pt–10 percent Rh) is employed for temperature up to 1750°C. Temperature may be controlled or varied using a program controller with two thermocouple arrangements: the signal from one actuates the control system whilst the second thermocouple is used to record the temperature.

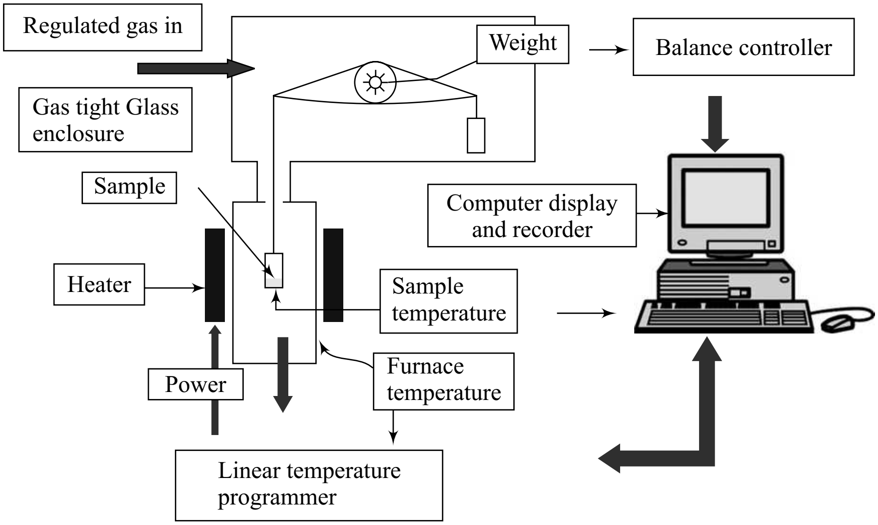
(iii) Recorder

X-Y recorders are commonly used as they plot weight directly against temperature. The instrument may also facilitate microprocessor-controlled operation and digital data acquisition and processing using personal computer with different types of recorders and plotters for better presentation of data.

Figure 21.5 shows a schematic diagram of the TGA unit comprising specific balance and furnace assembly and other accessories as a whole to better understand the working of a thermobalance.



(a) A compact TGA unit



(b) An assembly of TGA unit

FIGURE 21.5 *(a) Compact TGA unit and (b) schematic diagram of an assembly of TGA unit.*

The whole of the balance system is housed in a glass to protect it from dust and provide an inert atmosphere. To maintain an inert atmosphere around the sample, there is a control mechanism to regulate the flow of inert gas to provide an inert atmosphere and water to cool the furnace. The temperature sensor of furnace is linked to the programme to control heating rates, etc. The balance output and thermocouple signal may be fed to recorder to record using an electronic printer device in the TG curve.

The different balances used in TG instruments are having measuring range from 0.0001 mg to 1 g depending on sample containers used.

21.2.3 Sources of Error in Thermogravimetric Analysis

Inconsistencies of results of TGA arise due to few sources of error and they can lead to inaccuracies in the recorded temperature and mass data.

Some of the errors may be corrected by placing the thermobalance at a proper place and handling it with great care. For understanding, we are discussing some common source of errors during operation of a thermobalance.

- It is observed that when a thermally inert crucible is heated empty there is usually an apparent weight change as temperature increases. This is due to the effect of change in 'buoyancy' of the gas in the sample environment with the temperature, the increase in the convection and possible effect of heat from the furnace in the balance itself. Now, in most modern thermobalances, this effect is negligible. However, if necessary, a blank run with an empty crucible can be performed over the appropriate temperature range. The resultant record can be used as a correction curve for subsequent experiment performed under the same condition.
- The mass of the sample is affected due to the condensation of moisture on the sample to be investigated and consequently the shape of TG curve.

This can be avoided by maintaining a dynamic atmosphere around the sample in the furnace so that the entire condensable product may be driven by the flowing inert gases.

- Random fluctuation of balance mechanism
- Reaction between the sample and the container
- Convection effect from furnace
- Turbulence effect from gas flow
- Induction effect from furnace

For good recording of a thermogram of TGA, all these errors have to be minimized.

The design and operation of the furnace are of critical importance in obtaining good quality, reproducible thermograms. Most instruments use electrical resistance heaters, although heating techniques using infrared, laser irradiation or microwave induction heating have been investigated. The latter has a particular attraction as it would enable uniform heating to be employed throughout the sample. The arrangement of the sample pan may be of horizontal, top loading or suspended design. The effects of convection currents within the furnace tube, which can lead to uneven heating, can be minimized by a series of baffles.

Even so, there will still be a thermal gradient within the furnace and careful temperature calibration is needed.

21.2.4 Calibration of TGA Unit

A thermobalance of the TGA unit has to be calibrated before use for analysis. Calibration of thermobalance for the measurement of mass can be done by adding the known mass of the sample container and noting the reading of the chart. Ferromagnetic standards are used for temperature calibration. In a magnetic field, substances show detectable mass changes. The ferromagnetic standards are quite suitable for the temperature range from 242 to 771°C.

One method of calibration which has been used effectively is the *Curie point* method. Ferromagnetic materials lose their magnetism on heating at exactly reproducible temperatures or Curie points. A range of metals or alloys with Curie points between 150 and 1000°C is available. If suitable ferromagnetic calibration standards are placed in the sample pan of the balance and a large permanent magnet is placed below the pan, the sample will experience downwards attraction leading to an apparent increase in weight (Fig. 21.6). At the Curie point, the loss of the ferromagnetism will be reflected by an apparent loss of weight, enabling the temperature experienced by the balance pan to be accurately known. Using a range of standard materials, an accurate calibration curve of the furnace can be produced. However, this does not completely solve the problem of accurate assessment of the temperature actually experienced by the sample. With the exception of microwave induction heating, the heat is absorbed by the exterior of the sample and transferred to the interior by conduction. Thus, a temperature gradient will exist within the sample. Many materials that are the subject of TG investigation will have low thermal conductivities whence this effect will be pronounced. The only remedy is to reduce the sample size to the absolute minimum, e.g. 10 mg, and avoid rapid heating programmes.

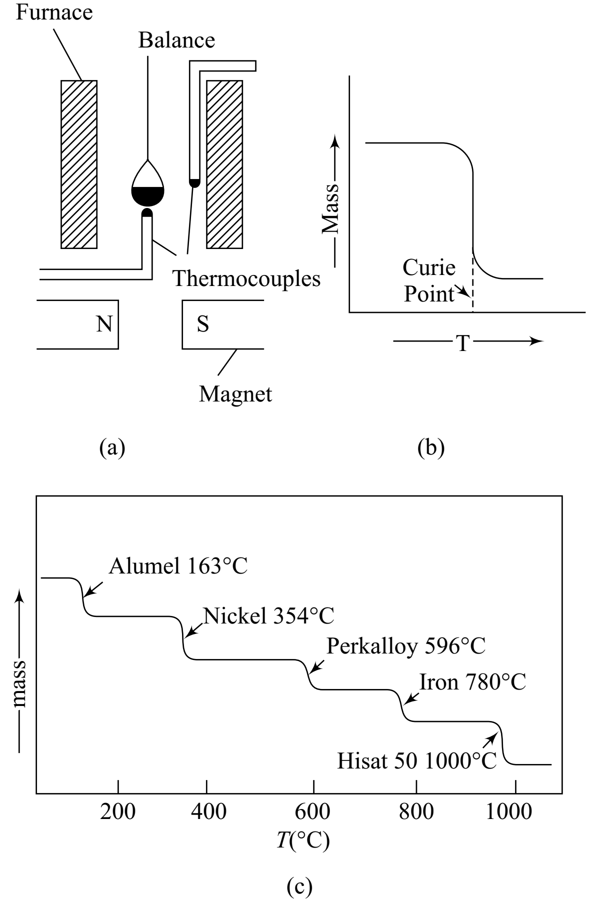


FIGURE 21.6 *Curie point method of temperature calibration.*

21.2.5 Data Analysis of Thermograms

Sample materials at an ambient atmosphere in a furnace give different types of TG curves that are characteristic of each sample.

The thermogram shows that TGA curves are different for different materials depending on the type of thermal reactions occurring in the sample, Fig. 21.7.

A thermogram is a graph of mass of the sample versus temperature (as shown in Fig. 21.8). In thermogram, weight percent of the sample versus temperature can also be reported. In order to find the onset (T_i) and the offset (T_f) reaction temperatures for samples in the TG curve, tangents have to draw to obtain these exact temperatures.

It will be unclear to interpret the thermogram, sometimes. However, simultaneous recording of DTG thermogram along with TGA helps to resolve the weight loss in the thermogram in stages, as shown in Figs. 21.9 and 21.10.

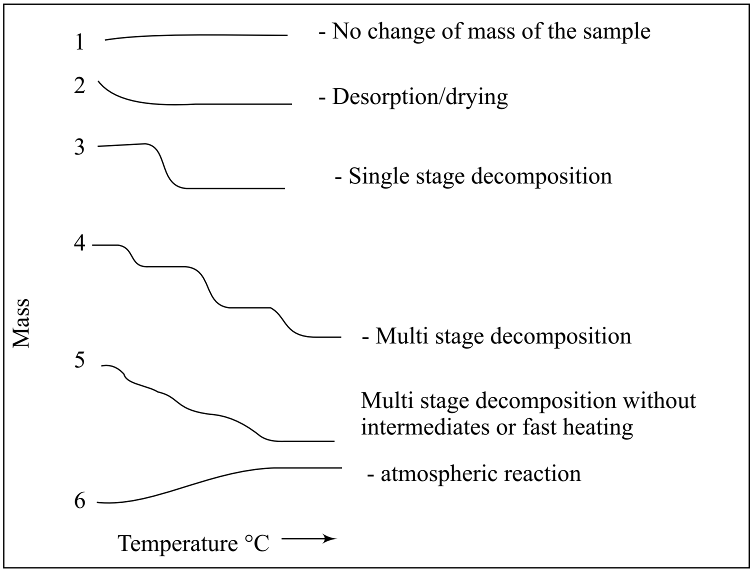


FIGURE 21.7 *Display of different types of thermograms.*

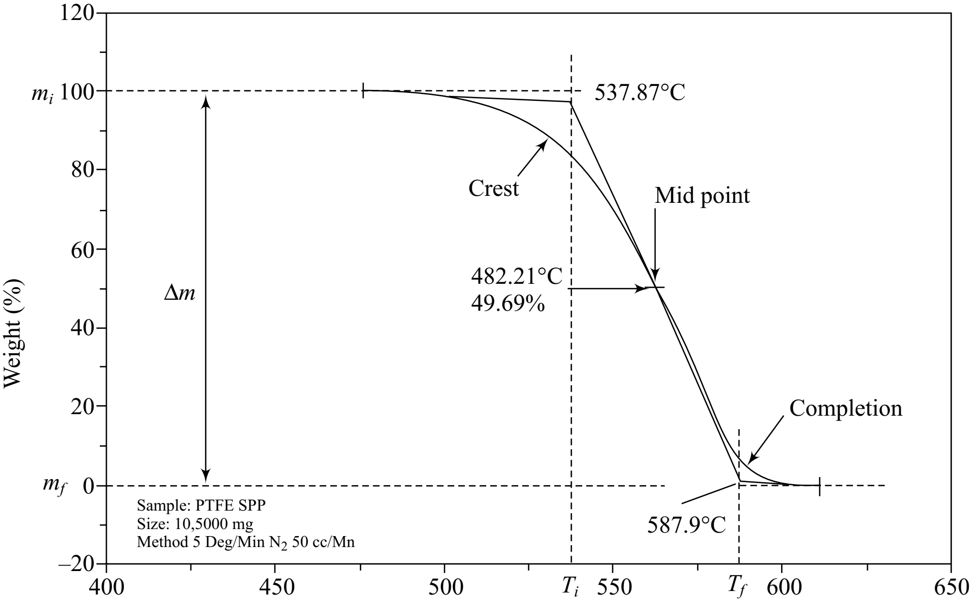


FIGURE 21.8 *A procedure to find onset and offset temperatures in a thermogram. The masses of the sample are marked as m_i, m_f and Δm at the referred temperatures and T_i and T_f depend on the operating variables.*

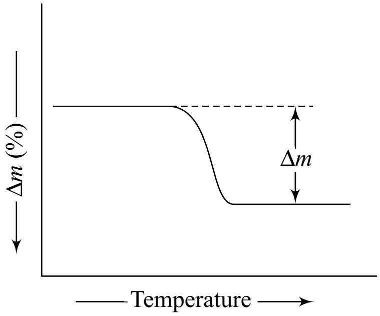


Figure 21.9 *Total change in mass (Δm) of which the sample in thermogram*

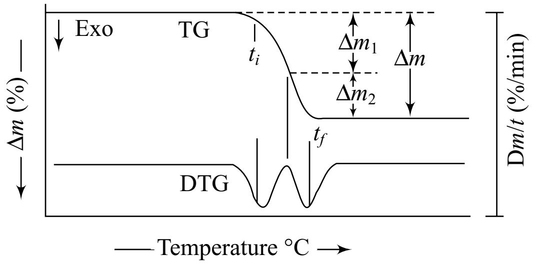


Figure 21.10 *DTG shows two peaks corresponds to $\Delta m = (dm_1 + dm_2)$*

21.2.6 Factors Affecting TG Analysis

Following factors are affecting TG analysis:

- Heating rate and sample size
- Increase in either of which tends to increase the decomposition temperature and to decrease the resolution between successive mass losses
- Particle size and packing of the sample
- Crucible shape
- Gaseous atmosphere
- Nature
- Flow rate

Effect of heating rate

In TG analysis, 5 mg sample is heated at 8 (or 10)°C/minute in air, O_2 in an atmosphere of inert gas/nitrogen to study the effect of heating rate on the sample.

Experiments have to be carried out under consistent conditions to obtain good reproducibility. It is found that the following factors affect the TG curve (Table 21.3).

Heating rate	Increases the temperature at which decomposition occurs
Sample size	
Packing	Affects the progress of the reaction
Crucible shape	
Gas flow rate	
Particle size of sample	

Table 21.3

Factors affect the TG curve

A skilled technician can obtain a resolved TG thermogram by adjusting heating rate (5–8°C/minute) for uniform particle size with a linear programmed heating of the furnace in an ambient atmosphere (O_2, N_2 inert gas).

21.2.7 Factors Affecting the TG Curves

Main factors affecting the shape, precision and accuracy of the thermogram curve of TGA are as follows:

1. Instrumental factors
   - (i) Furnace heating rate
   - (ii) Recording or chart speed
   - (iii) Furnace atmosphere
   - (iv) Geometry of sample holder/location of sensors
   - (v) Sensitivity of recording mechanism
   - (vi) Composition of sample container
2. Sample characteristics
   - (a) Amount of sample
   - (b) Solubility of evolved gases in sample
   - (c) Particle size
   - (d) Heat of reaction
   - (e) Sample packing
   - (f) Nature of sample
   - (g) Thermal conductivity

The reproducibility of the thermogram curve in TGA depends on the precise control of the various parameters as stated above. Moreover, thermograms of the sample have to be recorded under identical conditions of TGA experiment.

(a) Furnace heating rate

The shape of the TG curve is influenced by the heating rate. For example, when calcium carbonate is heated at a heating rate of 3°C/minute in a thermobalance below 600°C, the TG curve does not show any change in mass; however, it is observed that there is mass loss at 250°C at a higher rate of heating.

More specifically, it is observed that the decomposition temperature (T_i) and also the final temperature (T_f) of decomposition/phase transition will decrease with decrease in a heating rate and the TG curve will be shifted to the left. The appearance of an inflection in a TG curve at a fast heating rate may well be resolved into a plateau at a slower heating rate. Therefore, in TGA heating rate the sample should be neither too fast nor too slow to get a well-resolved TGA curve.

(b) Recording or chart speed

In the TG curve, one has to fix the chart speed to such a condition that endo/exothermic peaks obtained are sharp and well-resolved from each other. This has to be set by the experimenter for his advantage to interpret the results. For a slow reaction followed by a rapid one at the lower chart speed, the curve will show less separation in the two steps than the higher chart speed curve. For fast–fast reaction followed by slower one, a similar observation was observed in shorter curve plateaus.

(c) Furnace ambient temperature

The effect of ambient temperature on the TG curve depends on the following:

- (i) the types of the reaction
- (ii) the nature of the decomposition products and
- (iii) type of the atmosphere employed

The effect of the ambient temperature on the TG curve may be illustrated by taking the example of thermal decomposition of a sample of monohydrates of calcium oxalate in dry O_2 and dry N_2 as shown in Fig. 21.11.

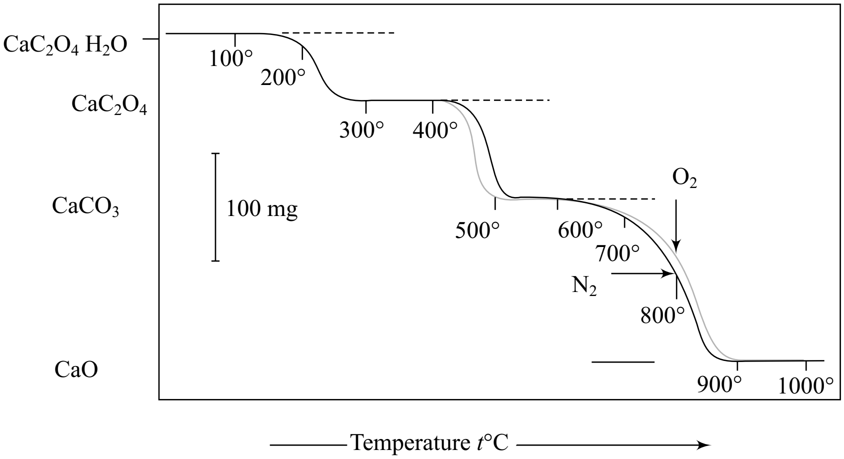


FIGURE 21.11 *TG curve of calcium oxalate in an O_2 and N_2 atmosphere.*

The first step observed in Fig. 21.12 is the dehydration of hydrated oxalic acid at 200°C.

$$CaC_2O_4 \cdot H_2O(s) \rightarrow CaC_2O_4(s) + H_2O(g)$$

There is decrease in weight in the thermogram due to the elimination and flushed out water vapour from the sample holder below 300°C.

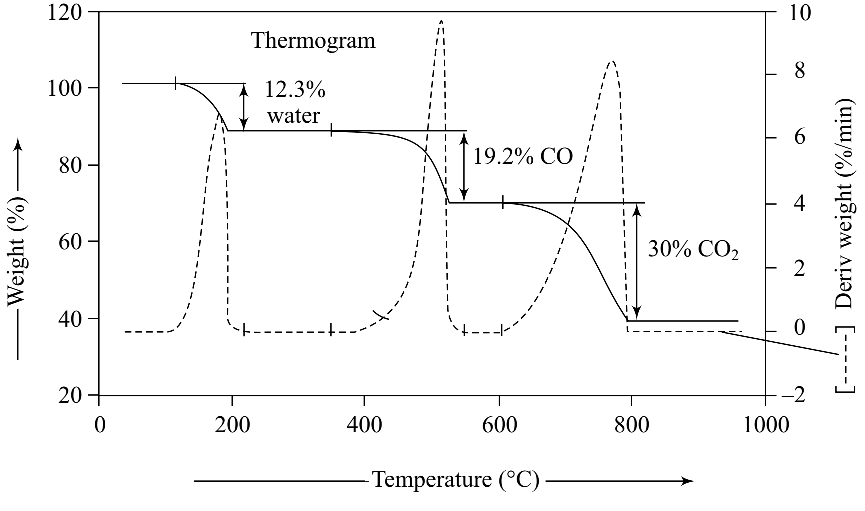


FIGURE 21.12 *TG and DTG curves of decomposition of calcium oxalate.*

The second step of thermal heating is the decomposition of $CaC_2O_4(s)$ at 400°C with the elimination of CO(g) gas, and the reaction is complete at 500°C. A mass loss is observed between 400 and 500°C accounts for the loss of CO.

$$CaC_2O_4(s) \rightarrow CaCO_3(s) + CO(g)$$

The curve diverges in an O_2 atmosphere because the oxygen reacts with evolved CO, yielding CO_2 as a second oxidation reaction which is highly exothermic and increases the temperature of the unreacted sample. The temperature accelerates the decomposition of the compound more rapidly and completely at a lower temperature as shown in the figure in dry O_2.

The third step of the thermogram is the decomposition reaction of $CaCO_3$ with further mass loss between 600 and 800°C.

$$CaCO_3(s) \rightarrow CaO(s) + CO_2(g)$$

This step is not affected by influence of O_2 or N_2 atmosphere. However, there is a slight difference in curves for the two gases as shown in the diagram.

(d) Sample holder

Generally, a tiny platinum crucible is used in TGA thermobalance to hold sample with a capacity of 5 mg and at the same time, minimizes the thermal gradient of temperature across the sample in the holder. The design of such Pt cups with small quantity of sample can minimize the error in recording the weight change of the sample and the onset and offset temperatures in the thermogram very precisely.

(e) Effect of size of sample particles in the container

A uniform and small particle size of sample to be taken to obtain a better well-defined thermogram curve or else it affects the thermogram of TGA.

21.2.8 Applications of Thermogravimetric Analysis (TGA)

TGA is a very useful tool to study the materials because it provides a very wide range of information regarding the nature of reactions. TGA also provides information about stability or instability of materials over a temperature range. Quantitative investigation of samples is useful for science and technology. The scope and applications of thermogravimetric analysis are listed as follows:

- Purity and thermal stability
- Solid-state reactions
- Decomposition of inorganic and organic compounds
- Determining composition of the mixture
- Corrosion of metals in various atmospheres
- Pyrolysis of coal, petroleum and wood.
- Roasting and calcinations of minerals
- Reaction kinetics studies
- Evaluation of gravimetric precipitates
- Oxidative and reductive stability
- Determining moisture, volatile and ash contents
- Desolvation, sublimation, vaporizations, sorption, desorption and chemisorptions

21.2.9 Quantitative Interpretation of Thermogravimatic Curves

Let us enumerate a few important features of quantitative aspects of TGA. The TG curves represent the variation in the mass (*m*) of the sample with the temperature (*T*) or time (*t*). Normally, the plot of mass loss (downwards) on the ordinate (*y*)-axis and mass gain (upwards) is shown in Fig. 21.13. Derivative thermogravimetric (DTG) curves of samples are very useful to interpret when the heating rate of the same is about 5–8°C/minute because the thermogram gives resolved peaks for onset or offset temperatures at a uniform heating rate. A DTG curve

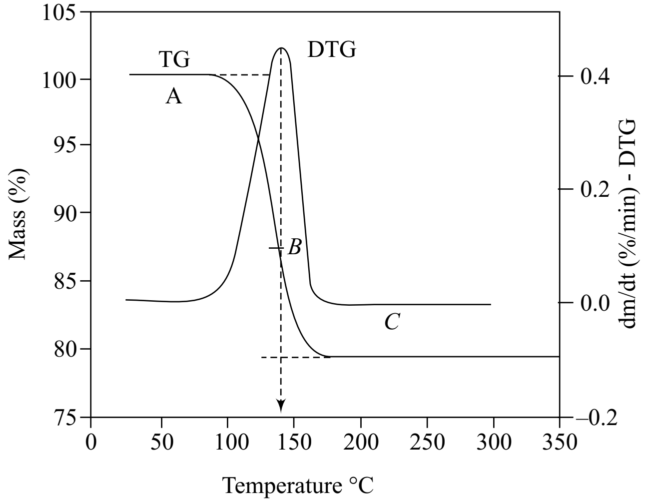


FIGURE 21.13 *A plateau of constant weight (region A) in the TG curve, the mass loss portion (region B) and another plateau of constant mass (region C) and DTG curve is also shown.*

presents the rate of mass change (*dm*/*dt*) as a function of temperature, or time (*t*) against *T* on the abscissa (*x*-axis), as shown in Fig. 21.13, when substance is heated at a uniform rate. In Fig. 21.13, the derivatives of the curve are shown by dotted lines.

21.2.9.1 Interpretation of the TG Curves

1. Thermal decomposition of $CaCO_3$

A characteristic TG curve of a pure compound is obtained in thermogravimetric analysis. TG curves quantitatively account for the mass changes during the course of reaction stoichiometrically with temperature. This approach provides a quantitative analysis of samples, if composition of the sample is known (Fig. 21.14). For example, let us consider the thermal decomposition of $CaCO_3$. The thermogram (TGA curve (a)) obtained by TGA indicates that $CaCO_3$ decomposes in a single step between 650 and 800°C to form stable oxide of calcium (CaO) and liberation of gaseous carbon dioxide (CO_2). The thermal decomposition of $CaCO_3$ is represented as

$$\underset{\text{(MF) } 100.1}{CaCO_3(s)} \rightarrow \underset{56.1}{CaO(s)} + \underset{44}{CO_2(g)}$$

The thermogram of $CaCO_3$ decomposition indicates that the percentage of mass loss by the $CaCO_3$ between 650 and 800°C (from Fig. 21.14) is 44 (100.1–56.1). This weight exactly matches with the weight of CO_2 that is liberated on heating.

The mass change calculation based on stoichiometry of the decomposition reaction confirms the liberated product as CO_2 (44g). The calculated percentage of weight loss due to liberation of CO_2 of $CaCO_3$ is

$$m\% = \frac{M_r(CO_2) \times 100}{M_r(CaCO_3)}$$

$$m\% = \frac{44 \times 100}{100.1} = 44$$

The thermograms of the decomposition of $CaCO_3$ samples are carried out at different heating rates as depicted in Fig. 21.14.

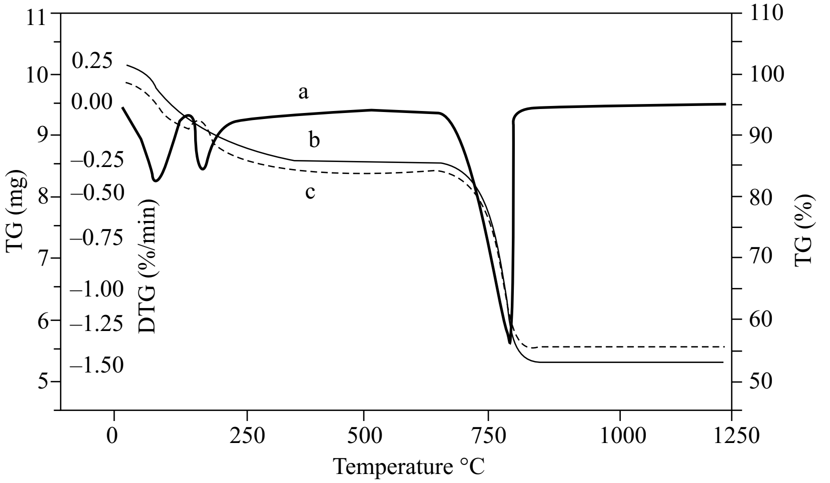


Figure 21.14 *TG and DTG curves of $CaCO_3$ at various heating rates (a = 10°C, b = 8°C, c = 5°C (DTG = rate of change of mass, dm/dt).*

2. Thermal dehydration and decomposition of calcium oxalate

It is possible to ascertain that the sample under TGA investigation is a hydrate or absorption moisture due to exposure to environment. Thermal decomposition of calcium oxalate is analysed using TGA experiment, and the thermogram (Fig. 21.15) is recorded with a programmed linear (10°C/min) heating rate in an ambient atmosphere. The first loss of sample weight at 170.2°C (12 percent of total mass) is due to the dehydration of water. This loss weight of the sample in the TG curve corresponds to the loss of one molecule of water (18.0 g) and hence, the sample is a monohydrate.

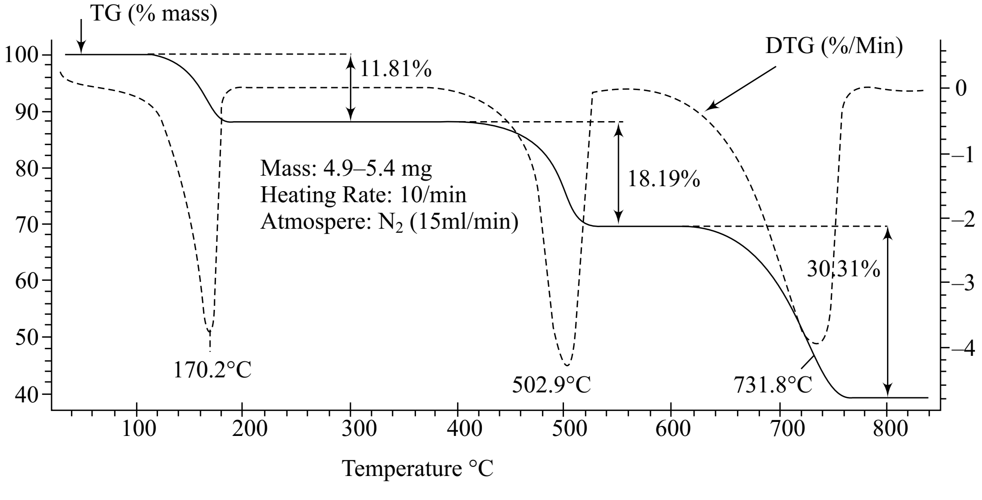


Figure 21.15 *Thermogravimetric curve of decomposition of calcium oxalate hydrate.*

Corresponding to the DTG peak in the thermogram, second weight loss observed at 502.9°C is due to the loss of CO (19 percent) and the third step weight loss of 30.31 percent at 731.8°C is due to the release of CO_2

(30.31 percent) from the sample and the final stable residue is CaO (39.5 percent). This is the fingerprint TG thermogram of pure calcium oxalate mono hydrate.

$$CaC_2O_4 \cdot H_2O \rightarrow H_2O \rightarrow CaC_2O_4$$

$$CaC_2O_4 \rightarrow CaO + CO\uparrow + CO_2\uparrow$$

If the sample contains other oxalates of barium and strontium, then the TG thermogram would indicate additional peaks at different temperatures. The absence of these peaks indicates that the material is a pure calcium oxalate monohydrate.

The stoichiometric equation has been considered for the decomposition of hydrated calcium oxalate at elevated temperatures, and the weight loss for each product formed is calculated. This is in accordance with the experimental results.

3. Thermal decomposition of zinc oxinate

Thermogravimetric (TG) and differential thermogravimetry (DTG) analysis of zinc oxinate was carried out at a programmed uniform heating rate of 8°C/minute in an ambient air atmosphere. The TG curve showed a first weight loss between 125 and 145°C (Fig. 21.16), the second weight loss of the sample is between 300 and 425°C and finally, the major weight loss observed is between 500 and 560°C. The weight loss in the TG curve with an increase of temperature of zinc oxinate is supported by the results of DTG, which indicated their corresponding reaction peaks at 145, 425 and 560°C. The complete removal of water molecules from the sample was at 145°C. A nearly flat region of the TG and DTG curves (Fig. 21.16) showed that the hydrated zinc oxinate, $Zn(C_9H_6ON)_2 2H_2O$, is thermally stable up to 125°C with Zn content at 16.77 percent per mole of zinc oxinate. The observed weight loss at 145°C is 9.34 percent is attributed to the elimination of two water molecules from the hydrated sample [$Zn(C_9H_6ON)_2 2H_2O$]. This weight loss in the TG curve is equivalent to a loss of exactly 36.00 g/mole. As a result of removal of water hydration, the Zn content per mole of oxinate

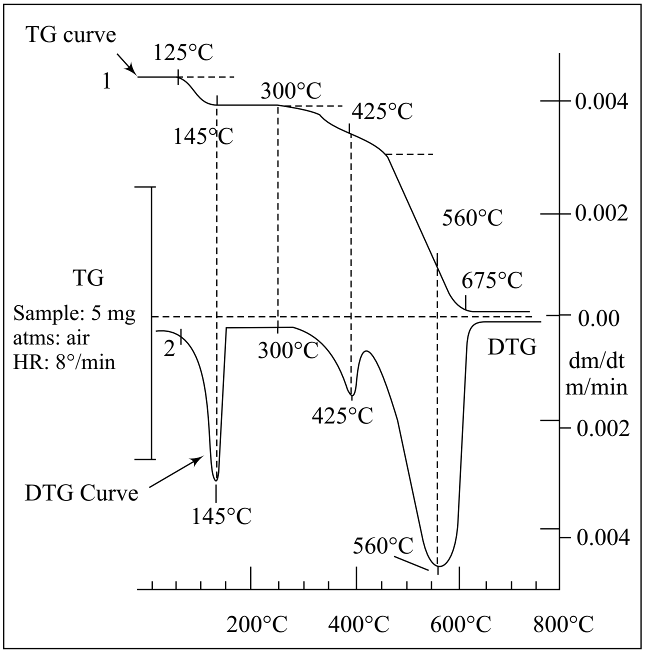


FIGURE 21.16 *TG and DTG results of zinc oxinate between 25 and 800°C in air.*

is increased to 18.48. The anhydrous $Zn(C_9H_6ON)_2$ is stable between temperature 145 and 300°C. Both the TG and DTG curves (Fig. 21.16) indicate that initiation reaction of the thermal oxidative degradation process of oxinate sets in above 300°C. A part of the complex molecule is eliminated (80.0 g/mole) between 300 and 425°C, increasing the percentage of zinc per mole to 21.05 percent. The onset of oxidative thermal decomposition of zinc oxinate (TG and DTG) commenced at slowly at 300°C and quickly between 425 and 625°C. The maximum weight loss per mole of oxinate at 560°C is 52 percent (i.e. 202.6 g/mole between 425 and 560°C) and the Zn content in the residue has increased to 34.94 percent.

Ultimately, the complex oxinate molecule decomposes between 425 and 675°C are shown by TG and DTG curves (Fig. 21.16) and the total weight loss observed between 145 and 675°C is 272.38 g/mole.

The final product obtained above 675°C is ZnO (which is identified and characterized by powder the XRD technique). The percentage of zinc in ZnO (the final product) is 80.40.

The thermal degradation of $Zn(C_9H_6ON)_2 2H_2O$ by TG and DTG data analysis are given in Tables 21.4 and 21.5.

Temperature-range in DTG curve (°C)	DTG peak temperature (°C)	Total weight loss corresponding to DTG peak	Weight loss per mole of oxinate (g)	Percentage weight loss between the DTG peaks	Percentage of total weight loss per mole oxinate
25–145	145	0.5055	36	9.24	—
300–425	425	1.1240	80.0	10.86	20.52
425–560	560	2.8450	202.60	32.00	52.00
560–625	625	4.3304	308.38	27.14	79.14
Total weight loss of sample at 625°C		4.3304	308.38	—	79.14
Weight of the residue, ZnO		1.1416	81.30		20.86
Total		5.4720	389.68	—	100

Note: Weight of $Zn(C_9H_6ON)_2 2H_2O$ taken: 5.4720 mg; mol. weight of $Zn(C_9H_6ON)_2 2H_2O$ = 389.68.

Table 21.4

Thermogravimetric (TG) data of $Zn(C_9H_6ON)_2 2H_2O$

DTG peak temperature (°C)	Weight loss per mole of oxinate (g)	Weight calculated from TG curve (g)/mole	Percentage of Zn in the thermal product Theoretical	Experimental
25–125	—	389.68	16.77	16.77
145	36	353.68	18.48	18.48
425	80.00	309.68	—	21.00
560	202.60	187.08	—	34.94
625	308.38	81.30	80.00	80.0*

Note:* 80 percent of Zn in ZnO.

Table 21.5

TG analysis data of $Zn(C_9H_6ON)_2 2H_2O$

It can be remarked that

(i) Hydrated zinc oxinate phase is stable up to 125°C, while the anhydrous phase is stable between 145 and 300°C.

(ii) A quantitative estimation of Zn as ZnO is possible above 625°C using oxine as a precursor.

4. Thermal dehydration of copper sulphate pentahydrate

Figure 21.17 shows the results of a TGA–DSC measurement of copper sulphate pentahydrate. The different dehydration steps that occur during heating of the material are difficult to distinguish in the TGA-signal, but are clearly visible in the DSC-result. The four water molecules are lost (21.2 percent) below 150°C, but the monohydrate is stable up to 244°C. The loss of water from the monohydrate corresponds to a weight loss of 5.3 percent per mole of hydrate copper sulphate.

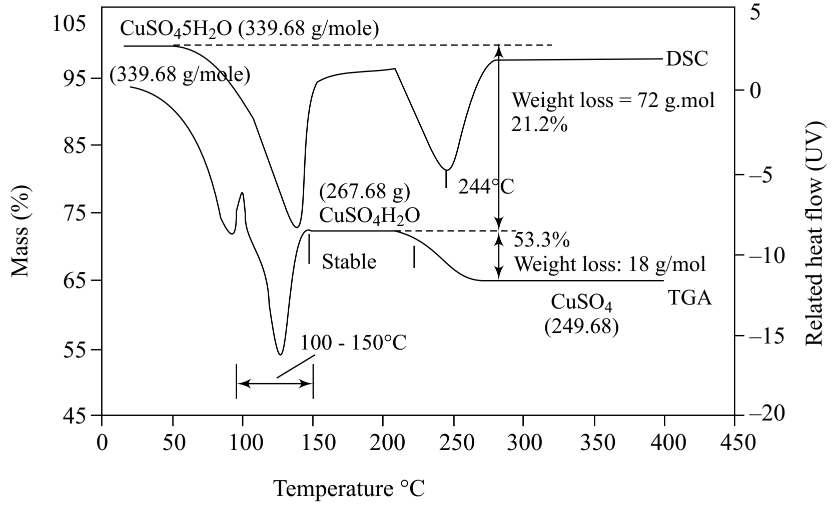


Figure 21.17 *TGA and DSC analysis of copper sulphate pentahydrate.*

From TGA thermogram, it is clear that the weight loss due to two stages of dehydration of water molecules can be illustrated, thus:

$$\underset{339.68\text{ g}}{CuSO_4 5H_2O} \xrightarrow[\substack{\text{weight loss}\\ 21.2\%/\text{mole}}]{-4H_2O(-72\text{ g})} \underset{267.68\text{ g}}{CuSO_4 4H_2O} \xrightarrow[5.3\%/\text{mole}]{-H_2O(-18\text{ g})} \underset{249.68\text{ g}}{CuSO_4}$$

(100 percent) (78.8 percent) (73.50 percent)

Theoretical calculations for loss of weight of H_2O per mole of copper sulphate pentahydrate is

$$\text{Weight loss for } 4H_2O/\text{mole} = \frac{72 \times 100}{339.68} = 21.196 \text{ percent}$$

$$\text{Weight loss for one } H_2O/\text{mole} = \frac{18 \times 100}{339.68} = 5.299 \text{ percent}$$

Weight loss for $5H_2O$/mole = 26.465 percent
Weight of residue, $CuSO_4$ = 73.535 percent = 249.78 g

5. Determination of the amount (in percent) of vinyl acetate in copolymers of vinyl acetate and polyethylene

TGA also provides quantitative information of organic compound decompositions and is particularly useful for studying the behaviour of one of the polymers in a mixture of copolymers. The TGA of vinyl acetate in copolymers of vinyl acetate and polyethylene is carried out to ascertain the stability and decomposition temperature of copolymer. For example, TGA can be used for determining the amount of vinyl acetate in copolymers of vinyl acetate and polyethylene. A typical observation is that vinyl acetate loses a molecule of acetic acid when heated to about 340°C. This observation is taken care of to determine the percentage of vinyl acetate in a mixture of vinyl acetate polyethylene copolymers. The TG curve obtained for several vinyl acetate polyethylene copolymers indicated loss of acetic acid at about 340–390°C (Fig. 21.18) due to the decomposition of vinyl acetate. Each molecule of vinyl acetate loses one molecule of acetic acid. The amount of vinyl acetate in the polymers can be calculated.

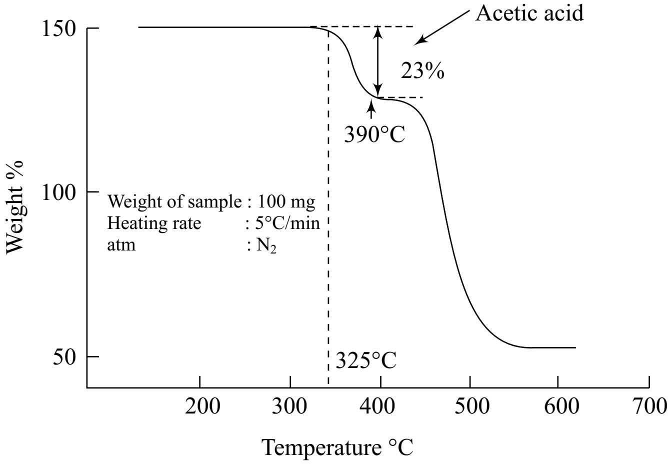


Figure 21.18 *TGA of decomposition of vinyl acetate.*

The TGA investigation of vinyl acetate in copolymers of vinyl acetate and polyethylene is performed by taking 100 mg of sample in a thermobalance and the heating rate of the sample is maintained at 5°C/minute and the furnace programmed for linear heating rate. The atmosphere of the sample is maintained by circulating N_2 gas at a rate of 20 mL/minute. The thermogram curve (Fig. 21.18) indicated that vinyl acetate in copolymers of vinyl acetate and polyethylene is stable up to 325°C. Twenty-three percent weight loss is observed at about 340–390°C/mole of vinyl acetate (Mol Wt: 86.1). The release of acetic acid is from vinyl acetate in copolymers of vinyl acetate and polyethylene. The DTG curve of the sample is very much desired to ascertain the temperature of completion of decomposition of acetic acid from the peak position of DTG.

Percent of acetic acid in copolymers of vinyl acetate and polyethylene

$$= \frac{\text{Weight loss of acetic acid} \times \text{Mol weight of vinyl acetate}}{\text{Mol weight of acetic acid}}$$

$$\text{Acetic acid (percent)} = \frac{23 \times 86.1}{60.1} = 32.95$$

6. Thermogravimetric analysis of rubber

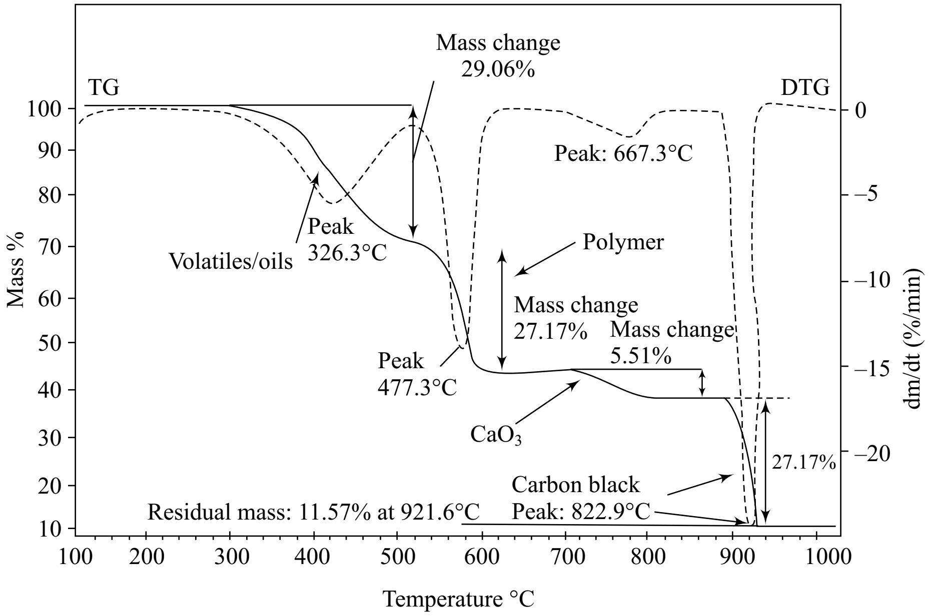


Figure 21.19 *TGA and DTG of rubber analysis.*

	Volatiles	Polymers	CO_2	$CaCO_3$	Carbon black	Residual
	Wt percent	Wt percent	Wt percent	Wt percent	Wt percent	Wt percent
Peaks	326.3°C	477.3°C	677°C	–	822.9°C	>823°C
EPDM compound	29.06	27.17	5.51	12.5	27.17	11.57

Table 21.6

TGA and DTG of rubber analysis

A detailed TG and DTG investigation of the analysis of rubber (EPDM) is carried out. The analytical data of the sample is given in Table 21.6. Based on the shape of the derivative and peak temperature, the polymer could be EPDM (confirmed via FTIR). Most probably, the compound contains calcium carbonate (based on just the weight loss pattern). This could be confirmed via X-ray fluorescence (XRF). Based on this assumption, the sample contains 12.5 percent $CaCO_3$ per mole of EPDM. At 326.3°C, the sample (EPDM) loses 29.06 percent per mole of the sample as volatiles. There is a loss of 27.17 percent of weight per mole of sample due to the decomposition of polymers at 477.3°C. Thereafter, CO_2 is released at 677°C (5.51 percent). The material confirmed after 700°C is $CaCO_3$ (12.5 percent). Carbon black (27.17 percent) is obtained above 823°C. At the end of each TGA measurement at 925°C a certain mass remains (residual mass). The residual mass is the ash content of the sample and contains inorganic components. Because the sample contains 12.5 percent $CaCO_3$, this means that the ash of the sample contains 7 percent CaO. The residual mass without CaO is 4.6 percent. More information about the residual mass can be obtained by performing an elemental analysis with XRF.

7. Analysis of inorganic and organic mixtures

A typical characteristic TG curve is obtained for a pure compound in thermogravimetric analysis. Based on this principle, TG curves can be used to predict relative quantities of the components of a mixture.

Binary mixtures

Consider a mixture of two compounds A and B. Thermogravimetric curves of two compounds A and B have characteristic TG curves which are different from each other as shown in Fig. 21.20.

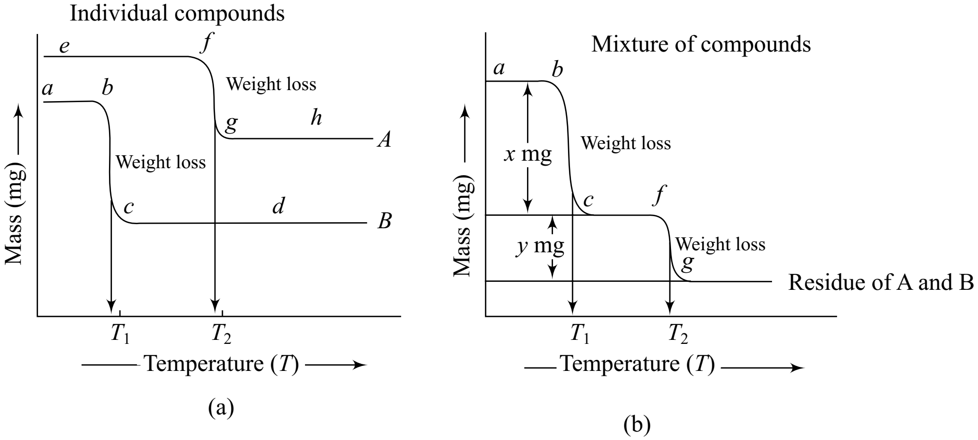


Figure 21.20 *(a) Thermogravimetric curves of two compounds A and B and (b) their mixture.*

The lines of 'ab' and 'ef' are horizontal indicating that the two compounds are stable up to T_1 and T_2, respectively. The residue of the decomposition of two compounds is shown as 'cd' and 'gh', respectively, for two pure compounds A and B.

TG analysis of samples indicated that the pure compounds A and B decompose sharply with single step at T_1 and T_2, respectively. The losses of weight of the compound are 'bc' and 'fg', respectively, in the thermogram. Equimolar quantities of both A and B are taken and homogenized mixtures of these two are taken for TG analysis. The TG thermogram of this mixture is shown in Fig. 21.20(b). Since both the components are stable (flat plateaus) and the thermograms show horizontal line 'ab' till the onset of temperature (T_1) of decomposition compound A and after a weight loss equal to 'bc' of compound A, the onset of decomposition of the second compound B starts at T_2. These decomposition temperatures T_1 and T_2 exactly match with the temperatures of decomposition of pure compounds A and B. The weight loss at T_1 (bc) and T_2(fg) are noted. Let the values be 'x' and 'y' (both in mg), respectively, for compounds A and B in the binary mixture (Fig. 20.20b). By measuring these two quantities x and y from the TG curves of Fig. 21.20, we can determine the relative quantities of x and y in the original binary mixture.

(i) Analysis of a mixture of calcium and magnesium carbonates

A typical TG curve of a mixture of calcium and magnesium carbonates is shown in Fig. 21.21. One can note that a significant mass loss occurs before 210°C. This is due to the moisture present in the mixture. Another mass loss at about 480°C is due to the following reaction:

$$MgCO_3(s) \rightarrow MgO(s) + CO_2(g)$$

Earlier, it was mentioned that $CaCO_3$ decomposes at about 800°C. In Fig. 21.21, mass loss between about 600 and 900°C can be interpreted due to decomposition of $CaCO_3$.

$CaCO_3(s) \rightarrow CaO(s) + CO_2(g)$

The horizontal line 'a–b' represents a mixture of MgO and $CaCO_3$ and 'c–d' represents a residue of the products after heating the above mixture.

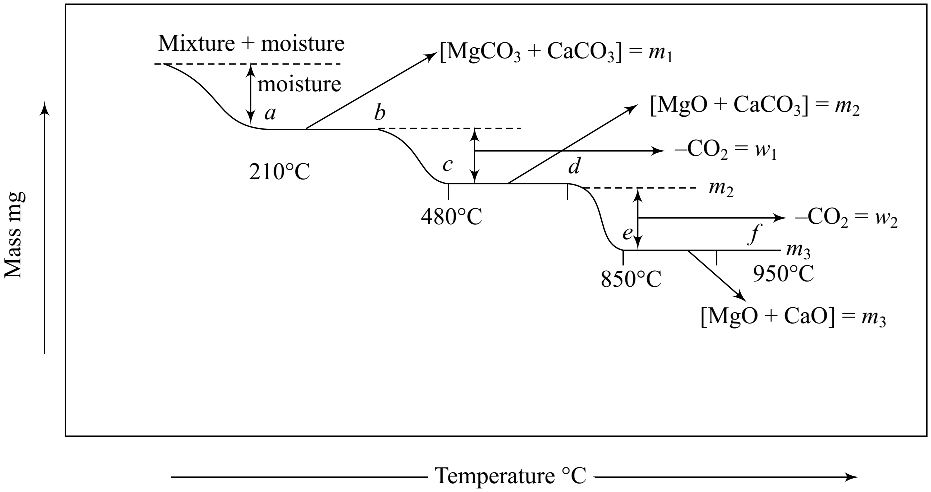


FIGURE 21.21 *TG curve of mixture of calcium and magnesium carbonate.*

In the thermogram curve of a mixture of calcium and magnesium carbonate, horizontal line (*a–b*) denotes the mass of the mixture of calcium and magnesium carbonates as 'm_1' which is stable between *a* and *b*. On heating, the onset of thermal decomposition of magnesium carbonates begins at '*b*' and is complete at 480°C (point *c*). Carbon dioxide (CO_2) is the released product whose equivalent mass is 'w_1'g, leaving MgO along with undecomposed $CaCO_3$ as residue at 480°C. The mass of the residue [MgO + $CaCO_3$] obtained at 480°C is 'm_2'g (Fig. 21.21) which is stable between the temperatures 480 and 800°C. Further heating the above residue, CO_2 is liberated again due to the decomposition of calcium carbonate between 800 and 850°C. Let the mass of CO_2 liberated from calcium carbonate decomposition be 'w_2'g. After release of second mole molecule of CO_2, the mass of the left over residue of the combustion product comprises both MgO and CaO whose mass is recorded as 'm_3'g. The residual mass m_3 g is constant between 850 and 950°C.

By utilizing stoichiometric thermal decomposition equations of both $MgCO_3$ and $CaCO_3$, it is possible to find out the relative masses of MgO and CaO in a binary mixture using TG analysis.

The stoichiometry of the decomposition reaction of $MgCO_3$ is

The MgO is formed by the evolution of CO_2 on the decomposition of $MgCO_3$,

$$\underset{M_w\ 84.0}{MgCO_3(s)} \rightarrow \underset{40}{MgO(s)} + \underset{44}{CO_2\uparrow(g)}$$

From the above equation, one molecule of $MgCO_3$ gives one molecule of CO_2 and one molecule of MgO.

The total mass of a mixture of $MgCO_3$ and $CaCO_3$ is $[M_{MgCO_3} + M_{CaCO_3}] = [84 + 100] = 184.0 \text{ g} = m_1$g.

The total mass of a mixture of MgO and $CaCO_3$ is $[M_{MgO} + M_{CaCO_3}] = [40 + 100] = 140 \text{ g} = m_2$g.

Thus, moles molecules of CO_2 in the given examples = $\dfrac{m_1 - m_2}{M_w(CO_2)}$ and *this is equal to mole molecule of MgO formed.*

Thus, the mass of MgO must be $M_{MgO} = \dfrac{m_1 - m_2}{44} \times M_w(MgO)$

If one molecule of $MgCO_3$(84 g) liberates 44 g of CO_2, then the mass equivalent of MgO produced in the mixture will be $\frac{[184-140]}{44} \times 40 = 40 \text{ g } = w_3\text{g}$

The mass of MgO (m_{MgO}) formed can be calculated

$$m_{MgO} = 0.91\ (m_1 - m_2)$$

$$= 0.91 \times 44 = 40.04 \text{ g} - w_3\text{g (for one mole molecule liberation of } CO_2.)$$

For the second decomposition $CaCO_3$ above 800°C, the stoichiometric relationship is

$$\underset{100}{CaCO_3} \rightarrow \underset{56}{CaO} + \underset{44}{CO_2}\uparrow\ (w_2)$$

Similarly, if one mole molecule of $CaCO_3$(100 g) liberates 44 g of CO_2, then the mass equivalent of CaO produced will be $\frac{[140-96]}{44} \times 56 = 56 \text{ g}$.

The mass of Ca (m_{CaO}) formed can be calculated

$$m_{CaO} = 1.27\ (m_2 - m_3)$$

$$= 1.27 \times 44 = 56 \text{ g} = -\ w_4 \text{ g, (for one mole molecule liberation of } CO_2).$$

We know the mass of residue left (m_3), i.e. 96 g and hence the mass of MgO (w_3) can be calculated.

$$W_3 = m_3 - w_4 = 96 - 56 = 40 \text{ g}$$

8. Drying temperature range for gravimetric precipitates

Thermogravimetric analysis provides information regarding the selection of an exact range of temperatures for drying the precipitated samples and also temperature at which the samples to be ignited to get stable residual masses.

It is a common experience that AgCl sample should be dried at 130–150°C and not beyond at any case. $BaSO_4$ should be ignited at 800–900°C.

$$AgCl \xrightarrow[\text{moisture}]{130-150\,°C} AgCl$$

$$BaSO_4 \xrightarrow[\text{ignited}]{800-900} BaSO_4$$

A TG analysis gives better information for drying the sample after observing a constant mass range between two temperatures of experimentation and the precipitates can be dried in this temperature range of gravimetric analysis. Generally, it is reported that the precipitate of $Al(OH)_3$ should be ignited at 1200–1300°C to get Al_2O_3. But with the help of TG curve, it appears that the precipitate can be ignited at 1030°C to get anhydrous Al_2O_3. The thermogram of zinc oxinate $[Zn(C_9H_6NO)_3]$ indicated that the sample can be dried at 100–120°C to get anhydrous zinc oxinate and not above this temperature range (Fig. 21.16).

TG thermograms of samples help to decide the reaction chemistry of the process of obtaining a precise dry sample. Let us follow the chemistry of obtaining dry sample of beryllium pyrophosphate ($Be_2P_2O_7$) from beryllium salt and $(NH_4)_2HPO_4$ in the presence of excess of ammonia. Beryllium is precipitated as ammonium beryllium phosphate with ammonium hydrogen phosphate in the presence of ammonia.

$$Be^{2+} + (NH_4)_2HPO_4 \rightarrow Be(NH_4)PO_4 \times 6H_2O$$

Its TG curve indicates the elimination of water and ammonia at 30°C, so that there is no possibility of weighing anhydrous beryllium phosphate. The thermogram of $Be(NH_4)PO_46H_2O$ shows a continuous loss in weight from 30 to 600°C and hence, weighing of the sample is undesirable because of instability of the sample. Above 650°C, the mass of the residue ($Be_2P_2O_7$) remains constant. This points out that ammonium beryllium phosphate is unstable and decomposes easily as compared to beryllium hydroxide.

On the other hand, if $BeSO_4$ is treated with excess ammonia to get beryllium hydroxide and the precipitate is washed and dried and ignite the sample. $Be(OH)_2$ is converted to BeO above 900°C.

$$BeSO_4 + NH_4OH \rightarrow Be(OH)_2 + (NH_4)_2SO_4$$

$$Be(OH)_2 \rightarrow BeO + H_2O \text{ above } 900°C$$

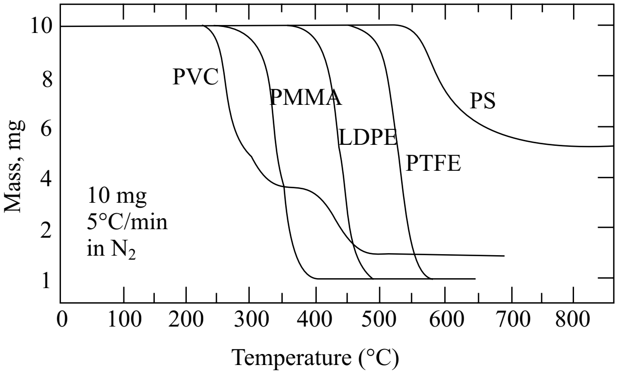
Thus, it is better to precipitate beryllium with ammonium hydrogen phosphate in the presence of ammonia, wash and then ignite it to 640°C to BeP_2O_7. Thus, TG curves help in finding the temperature at which the precipitate should be dried or ignited.

9. TG analysis of some polymers

The thermogravimetric analysis has a very versatile application in polymer analysis. The TG curve provides information about the decomposition temperatures of various polymers, which may be used for identification purposes and also assessment of thermal stability. The basic route by which a polymer degrades can be categorized according to following mechanism:

1. Main chain scission
2. Side chain scission
3. Elimination
4. Depolymerization

Among the above steps, only steps 1–4 lead to change in mass and can be detected in TGA. Figure 20.22 shows thermogravimetric analysis of polyvinyl chloride (PVC), high density polyethylene (HDPE), polymethyl methacrylate (PMMA), polytetrafluoroethylene (PTFE) and polystyrene (PS).



PVC = polyvinyl chloride;
PMMMA = polymethyl methacrylate,
LDPE = low density polyethylene;
PTFF = Poly tetra fluoroethylene;
PS = polystyrene.

Figure 21.22 *TG curves of some polymers.*

TG curves (Fig. 21.22) clearly indicates that polymer (PVC) is the least thermally stable and polymer (PS) is most thermally stable. In polymer (PS) no weight loss is observed at below 500°C and then decomposes abruptly above 600°C. The degradation of the polymer begins at the onset of decomposition temperature in the TG analysis curve. The degradation of PVC starts at lower temperature ~ 150°C and takes place in two steps, while nylon-6, LDPE and PTFE start degrading in single step at higher temperatures 400, 450 and 550°C, respectively. Polymers (PMMA) decomposes more slowly overall than the others as indicated by slopes of TG curves. The TG curve of polymer (PMMA) has less slope than the others. It can be remarked from the analysis of TGA of polymers (Fig. 21.22) that higher is the decomposition temperature of the polymers as observed TGA, greater is the stability of the polymer samples.

21.3 Differential Thermal Analysis (DTA)

21.3.1 Introduction

The measurement of energy changes in materials is investigated by both differential thermal analysis (DTA) and differential scanning calorimetry (DSC) techniques. They are thus the most generally applicable for all thermal analysis methods, since every physical or chemical change involves a change in energy. Differential thermal analysis (DTA) in which a material under investigation is typically subjected to a programmed temperature change; and thermal effects in the material are observed. The term 'differential' indicates the difference in behaviour between the material under study with an inert reference material. Under controlled temperature program in DTA, the temperature difference between a substance and a reference material is measured as a function of temperature. In this manner, the temperature at which any event either absorbs or releases heat can be found. The chemical and physical changes which are not accompanied by the change in energy on heating are not indicated in thermogravimetric (TGA) but such changes are observed in DTA of the samples.

In DTA, the heat changes within a material are monitored by measuring the difference in temperature (ΔT) between the sample and the inert reference. This differential temperature is then plotted against temperature or time to get the DTA curve (Fig. 21.23). Differential thermal analysis is the most widely used and is probably a very suitable thermal analytical method for materials analysis. This allows the determination of phase transition temperatures and the study of order–disorder transitions, decomposition, redox and other chemical reactions.

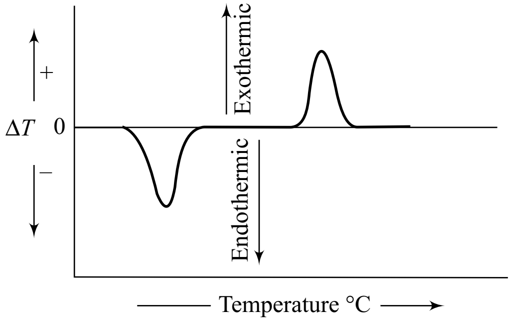


Figure 21.23 *Typical DTA curve.*

20.3.2 The Basic Principle of DTA

Differential thermal analysis is a technique in which the temperature of the substance under investigation is compared with the temperature of a thermally inert material such as α-alumina (α-Al_2O_3) and is recorded with furnace temperature as the substance is heated or cooled at a predetermined uniform rate.

In DTA investigation, the range of temperature is much larger than TG determination.

TGA will not characterize the following processes of materials such as pure fusion reactions, crystalline transition, glass transition and crystallization and solid-state reactions, since there is no mass change during the process. However, these changes are indicated during DTA by showing endothermal or exothermal changes in the curves. DTA is a dynamic method. It is essential that all aspects of the technique be standardized in order to obtain reproducible results. These include pretreatment of specimen, particle size and packing specimen, dilution of the specimen and nature of the inert dilutent.

The principle of method consists *in measuring the change in temperature associated with physical or chemical changes during the gradual heating of the substance.* Thermal changes due to fusion, crystalline structure inversions, boiling, dissociation or decomposition reactions, oxidation and reduction reactions, destruction of crystalline lattice structure and other chemical reactions are generally accompanied by an appreciable rise or fall in temperature. Hence, all these are accounted in DTA. Generally speaking, phase transitions, dehydration, reduction and some decomposition reactions produce endothermic effects whereas crystallization, oxidation and some decomposition reactions produce exothermic effects.

20.3.3 DTA Technique

In the DTA technique, a test sample (S) material under investigation is placed by the side of thermally inert reference material (R) [α-Al_2O_3], in a suitable sample holder (Pt) or block. As both the samples are heated inside the furnace uniformly under identical conditions with a programmed heating rate; the temperature difference (ΔT) between 'S' and 'R' is continuously recorded. The differential temperature 'ΔT' is then plotted against time or against temperature. Chemical, physical, structural and microstructural changes in the sample 'S' lead to the absorption (endothermal) or evolution of heat (exothermal) relative to R. If the response of sample and the reference submitted to an applied heat-treatment programme is not identical, differential temperatures ΔT arise as well. Therefore, DTA can also be used to study thermal properties and phase changes which do not necessarily lead to a change in enthalpy.

The baseline of the DTA curve exhibits discontinuities at transition temperatures and the slope of the curve at any point will depend on the microstructural constitution at that temperature. The area under a DTA peak can be related to the enthalpy change (ΔH) and is not affected by the heat capacity of the sample.

20.3.4 DTA Design

The basic components of a differential thermal analysis (DTA) cell are shown in Fig. 21.24.

The sample and reference materials are placed in two separate Pt cups inside the furnace with a uniform temperature environment. The thermocouples are also placed close to the two Pt cups to record temperatures of sample and the reference. Thermocouples should not be placed in direct contact with the sample to avoid contamination and degradation, even though sensitivity may be compromised. The sample assembly is isolated against electrical interference with the wiring of the oven with an earthed metallic or Pt-coated ceramic sheath.

The furnace should provide a stable, large enough hot-area and a fast response to the temperature-controller commands. Data acquisition and/or the real-time display is handled by a computer system or an analogue device (plotter).

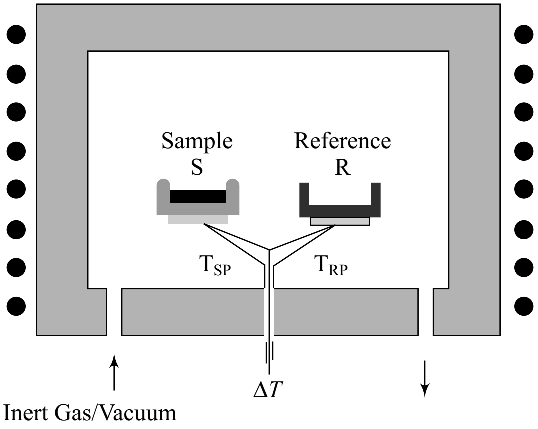


Figure 21.24 *Schematic illustration of a DTA cell.*

21.3.5 Experimental Details

The DTA of samples indicates that the thermal behaviour depends on chemical composition, size, surface-to-volume ratio, heating rate and environment; and these parameters have to be rigidly controlled during the experiment. The packing state of powder samples is important in these thermal decomposition reactions and may lead to significant differences in thermal behaviour between otherwise identical samples. Fine powdered microcrystalline material samples are preferred for structural transition studies. The heat flow rate may sometimes saturate the response capability of the DTA system. It is then advisable to dilute the specimen with a thermally inert material. For the detection of phase transition temperatures, one must also check that the peak temperature T_p does not vary with sample size. The shape of DTA peaks depends on sample weight and on the heating rate employed. Low heating rates (e.g. 5–8°C/min) or small sample weights (5–10 mg) lead to better resolved (sharper) DTA peaks, but the signal-to-noise ratio might however be compromised. Reducing the specimen size or the heating rate is also important in kinetic analysis (isothermal) experiments, where thermal gradients should be minimized.

21.3.6 Characteristics of DTA Curves

An idealized representation of the two major processes observable in DTA is shown in Fig. 21.25, where ΔT is plotted on the y-axis and T on the x-axis. In the DTA curve, peak showing downwards (↓) the baseline is endothermic while the other peak indicating upwards (↑) the baseline is exothermic. Similarly, the temperature of the sample is greater for an exothermic reaction, than that of the reference, for endotherms the sample temperature lags behind that of the reference.

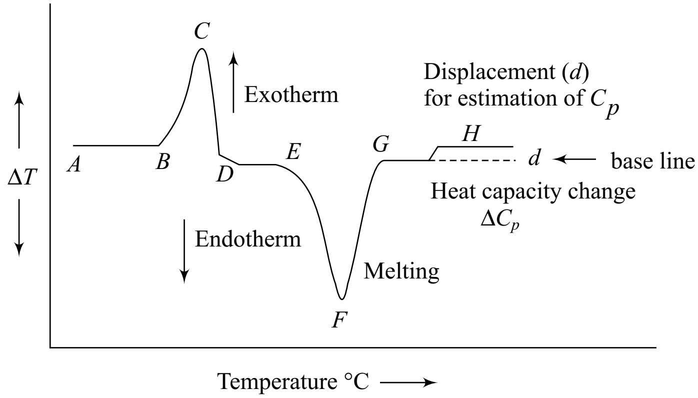


Figure 21.25 *A representation of the DTA curve showing exotherm, endotherm and baseline changes.*

When there is no thermal change in the sample during heating, the temperature of the sample remains similar to that of the reference substance. This is because both are being heated exactly under identical condition, i.e. temperature difference $\Delta T = (T_s - T_r)$ will be zero for no reaction.

However, any thermal change in the sample indicates exothermic or endothermic reaction peak in the DTA curve. A peak appears in the DTA thermogram, if there is a temperature difference 'ΔT' against temperature of furnace or time.

Let us consider the DTA curve in Fig. 21.25. ΔT is zero along the line AB indicating the baseline of the curve. At point marked 'B' where the curve begins to deviate from the baseline corresponds to the onset temperature at which the exothermic reaction starts and gives rise to a peak BCD with a maximum at point 'C'. The peak temperature 'C' corresponds to the maximum rate of heat of evolution. It neither represents the maximum rate of reaction nor the completion of the exothermic process. The usefulness of the method arises from the fact that peak temperature is normally characteristic of the material in the sample. *Area of the peak BCD is proportional to the enthalpy change of reaction.* For endothermic reaction, the peak EFG will be obtained as shown in the idealised curve. This peak shows that the ΔT, i.e. $(T_s - T_r)$ will be negative because heat is absorbed and consequently T_s will be smaller than T_r. Note the levels of baseline s of exotherm curve, AB and DE. Both are at different levels above the x-axis. This is due to the fact that 'heat capacity' of the sample has changed as a result of the exothermic process. Similar explanation can be given for the difference in levels of baseline s of endotherm curve, i.e. DE and GH. DTA curves not only help in the identification of materials but also their peak areas provide *quantitative information regarding mass of sample (m), heat of reactions (enthalpy change, ΔH)* and factors such as sample geometry and thermal conductivity.

The peak area of a DTA curve (either exo/endo) is proportional to the enthalpy change of sample material at that temperature.

Then, peak area (A) can be expressed as follows.

$$\text{Peak area } (A) = \pm \Delta H \cdot M \cdot K$$

where 'K' is the calibration constant which is *temperature dependent*.

Convention is that +ve sign is used for endothermic reaction, ($\Delta H > 0$), e.g. the temperature of the sample will lag behind that of the reference and hence energy has to be supplied to the sample to be at par with the reference.

The negative sign is used for exothermic reaction ($\Delta H < 0$), the temperature of the sample will exceed that of reference, factor K is called calibration constant which is *temperature dependent*. This constant 'K' can be determined by calibrating DTA with some standard. Once we know the value of 'K' at a particular temperature, the peak area can be used for quantitative analysis to determine the mass of sample or energy (enthalpy changes) of a reaction. Besides this, the DTA curve also helps in estimating the 'heat capacity' of a sample.

As Fig. 21.25 shows that there is always a difference in the baselines. The changes in heat capacity (ΔCp) may be determined at a particular temperature by measuring the difference in baselines, i.e. displacement (d) since:

$$\Delta C_p = \frac{1}{\text{Heating rate} \times m} \quad \text{or} \quad \frac{d}{(dT \cdot dt)} \times \frac{1}{m}$$

The area (A) enclosed between the DTA peaks (area BCD), and the baseline is related to the enthalpy change (Q) of the specimen during the phase transition. When the thermocouples are in thermal (but not in physical) contact with the specimen and reference materials, it can be shown that the peak area 'A' is given by

$$A = \frac{Q}{g \cdot K} = \frac{m \cdot q}{g \cdot K}$$

where

m = sample mass (weight)

Q = enthalpy change of the specimen

q = enthalpy change per unit mass

g = (measured) geometrical shape factor

K = thermal conductivity of sample

Errors in estimating the correct DTA peak areas are likely to occur for powder materials and compacted samples which retain some degree of porosity. The gas filling the pores or which eventually evolves from the specimen itself alters the thermal conductivity of the DTA cell environment (compared to calibration experiments), leading to rather large errors in the DTA peak areas.

21.3.7 DTA Data Analysis

It can be remarked that the reason for shift in the baseline of the DTA curve is due to the difference properties like 'heat capacity' and 'thermal conductivity' of the test sample (S) and reference sample (R) which are not identical. This gives rise to a certain displacement between their response in the linear regions of the DTA trace. The DTA peaks are obtained correspond to evolution or absorption of heat following physical or chemical changes in the specimen (S) under analysis.

The detection of phase transition temperatures using DTA is not very accurate. The onset of the DTA peak indicates the start of the phase transition, but temperature shifts from the true values are likely to occur depending on the location of the thermocouples with respect to the specimen, reference or to the heating block. Therefore, the temperature calibration of the DTA instrument using standard materials with the known melting points is necessary.

The curve shows an endothermic (heat-absorbing) peak. If an exothermic (heat-producing) event had occurred, the curve would show a peak in the opposite direction. The area 'A' on the curve is proportional to the heat of the reaction:

$$\Delta H = K \cdot A = K \int \Delta T \cdot dt$$

The constant K comprises many factors, including the thermal properties of the sample, and varies with temperature. Nowadays, the thermocouples are placed below the container, which has the effect of reducing the influence of sample properties on the area of the DTA peak. With such designs, it is easier to determine the variation in K with temperature, and quantitative data are more readily obtained. The area (A) enclosed between the DTA peak and the baseline is related to the enthalpy change (ΔH) of the specimen during the phase transition. Errors in estimating the correct DTA peak areas are likely to occur for powder materials and compacted samples which retain some degree of porosity. The gas filling the pores or which eventually evolves from the specimen itself alters the thermal conductivity of the DTA cell environment (compared to calibration experiments), leading to rather large errors in the DTA peak areas.

21.3.8 Enthalpy Calibration

The energy scale of the DTA instrument must also be calibrated to absolute enthalpy values by measuring peak areas on standard samples over specified temperature ranges. The enthalpy calibration implies measuring at least two different samples for which both heating and cooling experiments must be performed. It is then possible to measure the constant pressure heat capacity (C_P) using the DTA method using the below-mentioned equation:

$$C_p = k \cdot \frac{T_2 - T_1}{m \cdot H}$$

where 'H' is the heating rate employed, k is the enthalpy calibration constant and T_1 and T_2 are the differential temperatures generated during an initial 'empty' run (without sample) and during the actual DTA measurement (with sample), respectively.

21.3.9 Instrumentation

A block diagram of a differential thermal analyser is shown in Fig. 21.26. It consists of the following basic components:

1. Furnace assembly
2. Sample and reference holder with temperature detector

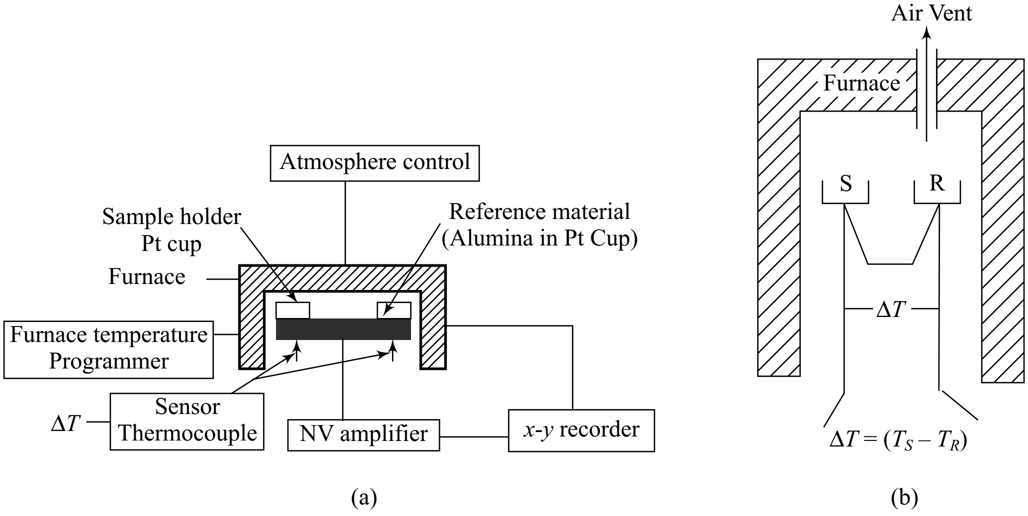


FIGURE 21.26 *Schematic diagram of a differential thermal analyser (a) complete layout (b) furnace part sharing continuous heating of sample and standard specimen holder.*

3. Temperature programmer
4. Amplifier and recorder
5. Atmosphere control equipment for furnace and sample holder

The instrument measures the differential temperature (ΔT) of the sample as a function of temperature or time where the temperature rises at a constant linear rate.

- **Source of uniform heating**
  Nichrome (nickel and chromium alloy) furnace can be used up to 1300°C, platinum and its alloys up to 1750°C and molybdenum (Mo) for higher range up to 2000°C. A special type of higher frequency induction heating may be used for higher temperatures.
- **Temperature regulating system**
  A uniform rate of heating of the furnace is ensured through temperature programmed circuits.
  The specimen holder is designed to accommodate even a small quantity of material and to give maximum thermal effect. It can be of Pt, Ni, stainless steel, Ag and alloy such as Pt–Rh. Certain ceramic materials such as sintered alumina, silica, fire clay, heat resistant glass and even graphite have been recommended as material specimen holders for the sample under investigation and reference material like α-alumina.
- **Measurement of temperature**
  A chromel–alumel thermocouple or rare metal alloys such as Pt-(Pt-10-13 percent Rh) are commonly used as a thermocouple for measuring the temperature. For higher temperature up to 2000°C, W-Mo thermocouple may also be used. Very thin thermocouple is inserted in the sample and reference holder.
- **Temperature recording system**
  Visual galvanometric observations, though inconvenient, can also be used when only a few samples are to be investigated. Nowadays, automatic pen and ink electronic recorder have been found to be more convenient.

21.3.10 Factors Affecting DTA Curves

DTA is a dynamic temperature technique. The sharpness of temperature peaks and their location in the DTA curve should be very precise and well defined for a particular sample. In addition to these behaviours, the extent of deviation on either side of the baseline in the DTA curve is also affected by a large number of instrumental factors.

The main errors in recording the DTA curve are due to the following:

(i) Sample factors and
(ii) Instrumental factors

Improper design and setting of the components of DTA assembly will lead to an error in the DTA curves due to instrumental factors such as improper size and shape of sample holder, cup material, heating rate, the position of thermocouple for reading the temperature of the behaviour of the sample, atmosphere and sensitivity of recording system. Most of these factors are associated with instrumental design. After proper design and setting up the DTA assembly, the exact position of the temperature (endo/exo) peaks obtained in the DTA curve has to be calibrated with reference to standard samples. Reproducibility of the DTA curves for any samples is must and has to be cross-checked frequently for zero error.

Sample characteristics include the amount of sample, particle size, packing density, heat capacity and thermal conductivity, degree of crystallinity, dilutes of diluents, swelling and shrinkage of the sample. In following lines, we will concentrate on some of the important factors in some detail.

- **Amount of sample**: In DTA analysis, the peak area of DTA curve is proportional to the mass of the sample. Preferably, small amounts (say, 5 mg) of powdered and homogenized samples are used for DTA cup/holder.

- **Particle size**: For DTA experiments, samples used are fine homogenized powders. Even for plastics and polymers, small uniform cut sizes are taken for DTA experiment.
- **Sample packing**: Uniform and identical packing density of finely powdered samples is preferable or otherwise it effects the shape of DTA curve.
- **Heating rate**: A skilled worker gets a beautiful resolved DTA peaks because the resolution peak always depends on the heating rate of the sample. It depends on the case study of the investigation and a skilled person adopts either slow heating to obtain resolved peaks or faster heating for desired experimentation. Generally, furnace heating rate of 5–8°C/minute is preferred to get good desired results.
- **Atmosphere around sample**: In order to extract good information, control of the environment of the samples has to be taken care. Depending on the nature of the reaction such as decomposition, redox and others, the environment around the sample under investigation has to be controlled and maintained. A proper flow of the gases such as inert gases, O_2, N_2 are to be made as an ambient atmosphere for DTA experiments. Generally, 15–20 mL/minute of these gases are sent into the sample holder during investigation. At the same time, the sent gas also flushes out the gaseous product and maintains a uniform environment around the sample.

Typical purge gases are air and nitrogen, helium is useful for efficient heat transfer and removal of volatiles. Argon is preferred as an inert purge when examining samples that can react with nitrogen. The experiment can also be carried out under vacuum or under high pressure using instruments of the appropriate design. In Table 21.7, the major factor which can affect the DTA curve is summarized.

Factor	Effect	Suggestions
Heating rate	Change peak size and position	Preferably low heating rate
Location of thermocouple	Erratic temperature reading	Standardize thermocouple position to read standard temperature of sample
Atmosphere	Any gradient change curve	Inert gas atmosphere
Amount of sample	Alter the peak size and position	Minimum sample mass
Particle size of sample	Irreproducible curves with different sizes	Use homogenized sample
Packing density	Irreproducible curves with different packing density	Uniform identical packing density
Sample container	Peak size depends on container	Use platinum cups

Table 21.7

Factors that influence DTA curve

21.3.10.1 Sources of Errors

A few errors appear while recording the DTA curve of the samples even with utmost skill and care, which lead to inaccuracies in the recorded temperatures. Major setback is the improper position of thermocouple near the sample and as a result there is mismatch between the actual temperature and the behaviour of samples. This can be minimized by the repeated calibration for thermocouple position with reference to standard material.

To minimize errors, DTA equipment should be calibrated both for temperature and peak area determination with appropriate standards. For temperature calibration with respect to peak position, standard substances like KNO_3, calcium tungstate, calcium oxalate, etc. are used depending upon the temperature range used for the experimental condition. For area calibration, indium (In) is often employed as a standard.

21.3.11 Interpretation of DTA Curves

A pure compound gives characteristic DTA curves which is the fingerprint of the compound. Energy changes observed in a sample can be correlated using the DTA curve, the manifested thermophysical and chemical changes during the process of controlled uniform rate of heating of the sample. DTA quantitatively accounts for the energy changes in phase transitions, decompositions, redox reactions and other processes. The thermal stability of the samples can be ascertained from DTA investigations. It is a very powerful tool that provides supporting information to TGA and DSC studies.

1. Decomposition of calcium oxalate

Let us consider the example of thermal decomposition of $CaC_2O_4 \cdot H_2O$ in N_2 atmosphere for which DTA curve is shown in Fig. 21.27(a).

The first endothermic peak at 170°C in the DTA curve shows that water is completely eliminated at this temperature from $CaC_2O_4 \cdot H_2O$. The second DTA curve peak is exothermic in nature while the third is again an endothermic peak. The exothermic peaks, Fig. 21.27(a), are due the decomposition of oxalate molecule producing CO and CO_2 molecules at 503 and 731°C, respectively. The TGA results of calcium oxalate (Fig. 21.15) confirm that the nature of reactions occurring in the endotherm at 170°C is dehydration of calcium oxalate. The final residue is the formation of CaO due to decomposition of oxalate molecule with the evolution of carbon dioxide gas.

In the N_2 atmosphere, Fig. 20.27(a), the onset of decarboxylation begins at 503°C with elimination of CO molecule and ultimately, removal of CO_2 at 731°C. Final stable product of reaction is the formation of solid residue, CaO.

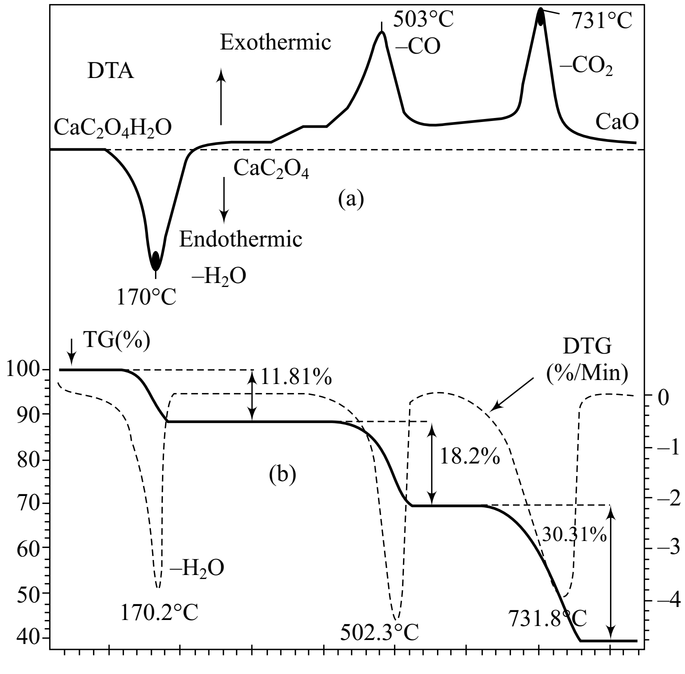


Figure 21.27 *DTA curve of decomposition of $CaC_2O_4 \cdot H_2O$ in nitrogen.*

Simultaneously, DTG of the decomposition of calcium oxalate is also recorded to corroborate the above results. The DTG peaks are obtained at 170.2, 502.3 and 731.8°C, respectively (Fig. 21.27b) which confirms the above observation of DTA studies. The stoichiometric calculations of the molecular mass of the eliminated products like H_2O, CO and CO_2 are performed which exactly matches with DTA results.

A typically designed instrument operates in the range from room temperature to 1000°C.

2. Thermal decomposition of zinc oxinate in air

The DTA of zinc oxinate $[Zn(C_9H_6ON)_2 \cdot 2H_2O]$ is recorded in an atmosphere of air between 25 and 800°C. There is no fusion/volatilization peaks due to the presence of traces of oxine in the DTA curve. This indicates the absence of oxine as impurity in zinc oxinate. The first endothermic peak at 145°C (Fig. 21.28) is observed in the DTA curve due to the dehydration process. It is evident that the chelate complex of zinc obtained under the conditions of preparation is quite stable up to about 125°C. Two other exothermic peaks are also noticed at 425 and 560°C in the DTA curve. These peaks, especially, the one at 560°C is huge indicating possibly the decomposition of the complex molecule, which is highly exothermic. The initial exothermic peak begins at about 300°C and ends up at 425°C, which may be due to thermal elimination of a part of the complex molecule. A huge exothermic reaction due to oxidative thermal decomposition begins at about 500°C and is complete at 625°C with a maximum peak positioned at 560°C. At 625°C, the final residue obtained is ZnO. The XRD pattern characteristic of pure ZnO is obtained (ASTM Card file No 5-0664). The results of DTA curve are corroborated by TG and DTG of the sample over the range of temperatures (Fig. 21.28).

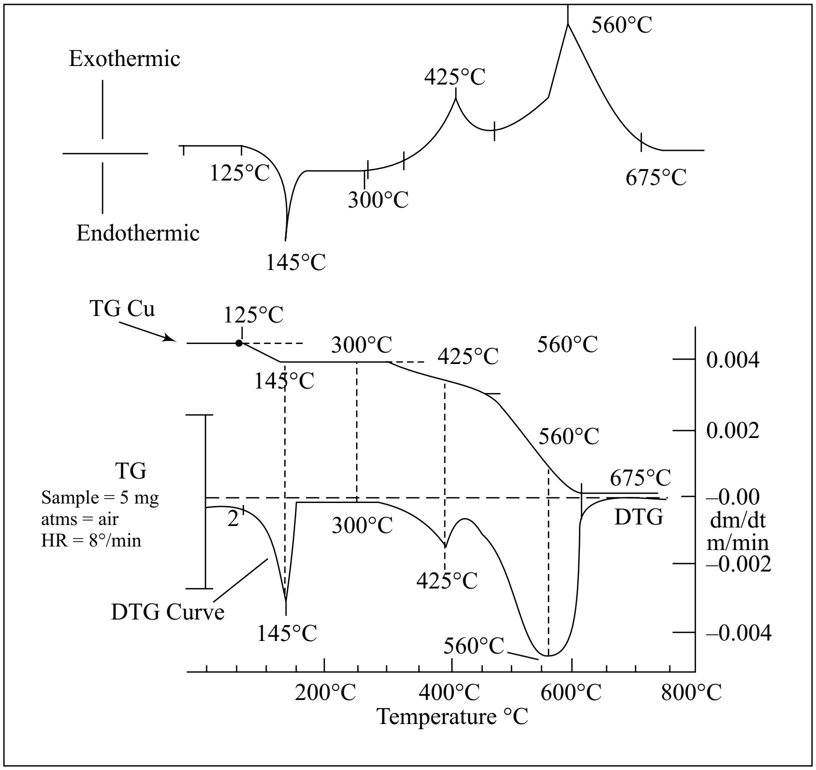


FIGURE 21.28 *DTA, TGA and DTG analysis of zinc oxinate.*

3. Phase transition studies by DTA technique

Sample materials exhibit phase transition from low temperature to high temperature structure at a particular temperature by absorbing a certain amount of energy (Δ*H*) from the environment at programmed heating.

In other words, the two forms of samples of the same material differ in structure and enthalpy content(ΔH). Among these, a few structural transformations are thermally reversible process.

(a) Phase transition investigation of copper pyrovanadate ($Cu_2V_2O_7$)

Palanna O.G and Rao N.S have reported for the first time new phase transitions in copper pyro (and ortho) vanadates using the DTA technique and which was further corroborated by high temperature XRD investigation of both samples.

The DTA of $Cu_2V_2O_7$ phase obtained (Fig. 21.29) by a conventional solid-state technique by heating in air an intimate mixture of 2CuO:V_2O_5 (both of high purity grade) at 600°C for 24 hours in a Pt crucible which was then melted in air. This phase exhibits an endothermic peak at 475°C and a knee at 500°C indicating phase transformations in the $Cu_2V_2O_7$ phase (heating curve). On cooling the molten pyrovandate ($Cu_2V_2O_7$) sample in Pt cup of DTA unit in static air, two distinct exothermic peaks are observed at 485 and 435°C, respectively. The thermo cycle for the above phases is found to be reversible.

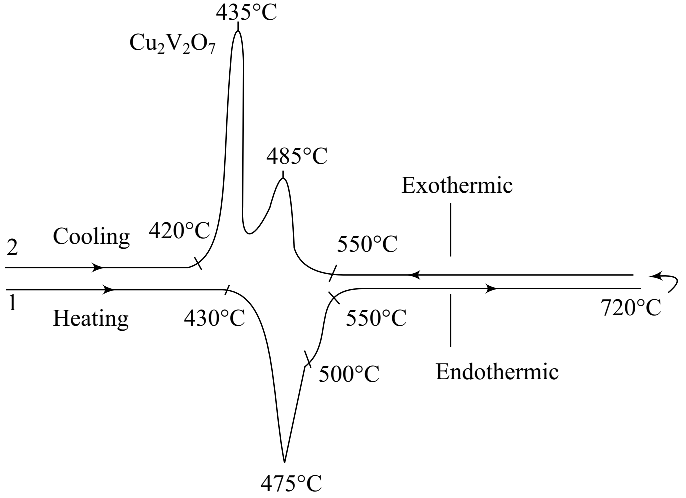


FIGURE 21.29 *DTA of $Cu_2V_2O_7$ in a static air atmosphere (Palanna and Rao).*

The phase transition of $Cu_2V_2O_7$ can be represented as

$$\alpha\text{-}Cu_2V_2O_7 \leftrightarrow \beta\text{-}Cu_2V_2O_7 \leftrightarrow \gamma\text{-}Cu_2V_2O_7$$

The results are reproducible for $Cu_2V_2O_7$. The above results are confirmed by high temperature XRD studies.

(b) Phase transition studies of copper orthovanadate ($Cu_3V_2O_8$)

The $Cu_3V_2O_8$ sample was synthesized by the conventional solid-state ceramic technique and melted. The sample was annealed in an atmospheric air. The results of the differential thermal analysis (DTA) of the $Cu_3V_2O_8$ phase are presented in Fig. 21.30. On heating the copper orthovanadate phase in an atmosphere of static air, the DTA (curve 1) showed an endothermic peak at 510°C and knee at 545°C. On cooling (curve 2) the above phase, exothermic peaks were observed at 532 and 435°C.

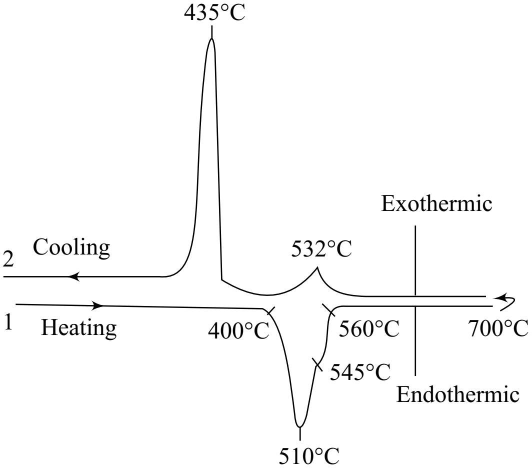


FIGURE 21.30 *DTA curves for $Cu_3V_2O_8$ in static air (Palanna and Rao).*

Therefore, it appears that $Cu_3V_2O_8$ sample prepared as above undergoes the phase transitions upon increasing the temperature. The results are confirmed by the high temperature XRD of the $Cu_3V_2O_8$ sample.

The phase transitions in the $Cu_3V_2O_8$ sample is summarized as

$$\underset{\text{(Phase 1)}}{\alpha\text{-}Cu_3V_2O_8} \xrightleftharpoons[435^\circ C]{510^\circ C} \underset{\text{(Phase II)}}{\beta\text{-}Cu_3V_2O_8} \xrightleftharpoons[532^\circ C]{545^\circ C} \underset{\text{(Phase III)}}{\gamma\text{-}Cu_3V_2O_8}$$

4. Composition and heat capacity of the sample

DTA curves are useful for determining the composition of the sample. The composition of the sample is determined from the position and shapes of the peaks of DTA curve. Equation (1) is used to relate peak area with the heat of the reaction and amount of sample used for analysis. The calibration of instrument with known standards is used to annul the calibration constant, *K*. However, the value of the calibration factor, *K*, may be eliminated from the quantitative calibration by comparison of a peak area of unknown sample with the known sample under identical condition.

For example *if A_1 is the peak area of known mass (m_1) and A_2 is the peak area of unknown sample having mass equal to m_2,* then using Eq. (1) we can write.

$$\frac{m_1}{m_2} = \frac{A_1}{A_2} \quad \text{or} \quad m_1 = m_2 \cdot \left[\frac{A_1}{A_2}\right] \qquad \ldots(1)$$

Similarly, the heat of reaction of a unknown sample can also be calculated by comparing a sample of known heat of reaction. Note that the calibration factor, *K*, in Eq. (1) is temperature dependent in DTA situation and therefore both the known and unknown samples should be run at same temperature.

It is evident from DTA studies that the initial and final baselines of peaks of DTA curves generally do not coincide. Therefore, finding out the area under the peak may be rather difficult. A method has been suggested for determining the peak area which is illustrated in Fig. 21.31. Both baselines are extended to a perpendicular line drawn from the maximum of the curve and the area under the two halves of the curve are determined and added to give the total area.

The molar enthalpy change of reactions, ΔH_m, can be calculated from the enthalpy of reaction obtained DTA experiment by using the formula (2):

$$\Delta H_m = \Delta H_r \cdot \frac{M_r}{m} \qquad \ldots(2)$$

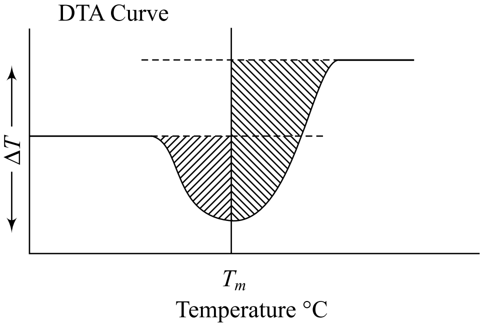


Figure 21.31 *Illustration depicting the determination of DTA peak areas.*

where, ΔH_m = molar enthalpy of reaction, ΔH_r = enthalpy of reaction, M_r = related molar mass of the compound, *m* = mass of substance used for analysis.

Determination of heat capacity

It is possible to determine the heat capacity (C_p) of the material from DTA curve using Eq. (3). The DTA curves for empty container and sample with container is recorded under same conditions.

The absolute change in temperature (ΔT) has been measured and substituted in the equation for calculations:

$$C_p = K \cdot \frac{T_2 - T_1}{mH} \qquad \ldots(3)$$

where C_p is the heat capacity at temperature T, T_1 and T_2 are differential temperatures generated when the instrument is first run without any sample and then with the sample whose heat capacity is to be determined. '*H*' is the heating rate (dT/dt) and K is calibration factor of the DTA unit under identical condition which is obtained against the standard substance of the known enthalpy change.

5. Construction of phase diagram

The melting points of an eutectic mixture of different samples with variable composition can be obtained from DTA run of the sample. These melting points can be used to construct a phase diagram. The critical temperature of organic compounds can be determined by DTA if a sealed sample holder is used. While observing the cooling curve of DTA, a discontinuity is observed at the critical temperature T_c. At Curie point temperature, there is a sudden change of property at this point can also be determined by the DTA technique. The specific heat increases gradually upon the Curie point, e.g. 357°C for nickel, observed simply by plotting dQ/dT or dT against temperature.

6. DTA thermograms (or peaks) are the fingerprints of pure samples

Both qualitative and quantitative analyses of the samples are possible after ascertaining the peak area or peak height for a wide range of materials. DTA also provides scope for the study of thermal characteristics, stability, degradation and reaction kinetics on small samples, over a wide range of temperatures ranging from 25 to 1000°C.

If there is energy changes observed for peaks in DTA curves of the samples in the temperature range used; then, the detection of any minerals in a sample is possible. Since most of the minerals undergo such changes, choice of an appropriate temperature range should have priority. Difficulties for interpretation of the DTA curve may arise due to the presence of organic matter. Oxidation of organic materials, if present, masks the thermal effect of a substance. Consequently, the analysis of such cases should preferentially be carried out in an inert atmosphere or in vacuum. Organic material sometimes be removable by a suitable solvent or may be oxidized by treating with H_2O_2. The DTA peak at <200°C for samples shows the elimination water/moisture. A broad exothermic peak observed between 200 and 600°C may be due to the presence of organic material in the clay/ore.

In practice, however, difficulties do arise in a case where peak areas for two distinct samples of same mineral give peaks at two different temperatures. It is, therefore, necessary for qualitative work to know the peak area given by pure mineral identical to that in the sample under investigation. This, however, is not possible. Another limitation in the quantitative work in the occurrence of overlapping of peaks, e.g., when kaolinite and illite are present in the same sample almost completely overlapping takes place. A representative DTA curve of some minerals is shown in Fig. 21.32. It may be observed that none of these minerals show an exothermic peak in the range 200–600°C corresponding to organic material as mentioned above.

7. Differential thermal analysis of polymeric materials

The polymeric materials are characterized through DTA technique, which is used for the identification of thermophysical, thermochemical, thermomechanical and thermoelastic changes or transitions. DTA provides important parameters for polymer processing and its end use. The DTA also gives useful information about quantitative aspect, degree of fusion, crystallinity, phase equilibrium, heat of polymerization, degree of curing, etc. A typical differential thermal analysis (DTA) of a polymer is shown in Fig. 21.33, which indicated the following transitions: glass transition, crystallization, melting and oxidation, represented as T_g, T_c, T_m and T_d, respectively in the DTA curve.

For analysis of polymeric mixture of samples, DTA technique is used qualitatively and quantitatively. The individual pure polymers exhibit their own characteristic DTA peaks. Figure 21.34 shows the differential thermal analysis of a polymeric mixture comprising seven components. Each endothermic peak recorded in

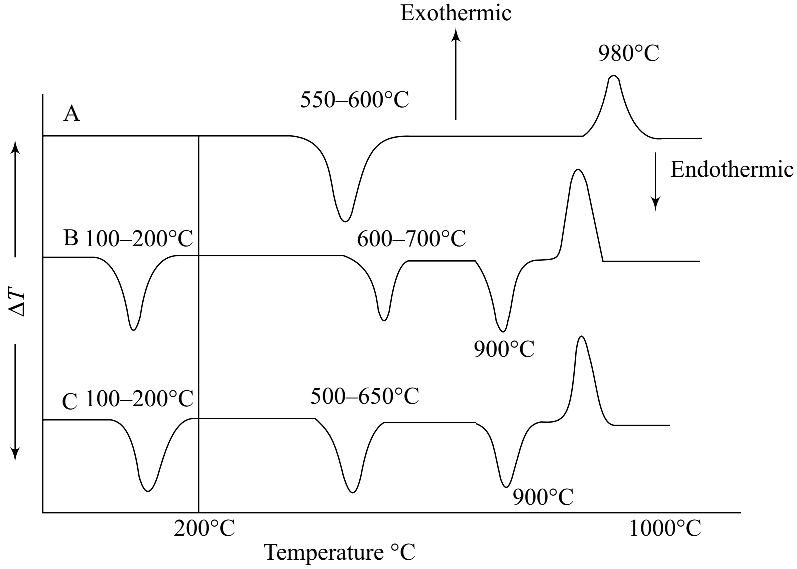


FIGURE 21.32 *DTA curves for (A): kaolinite, (B) mortmonillonite and (C), illite.*

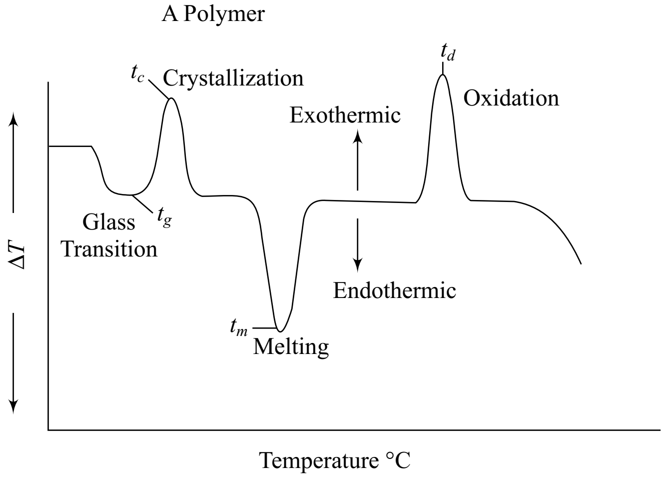


FIGURE 21.33 *DTA curve of a typical polymeric sample.*

the DTA thermogram (Fig. 21.34) indicates the melting temperatures of the following polymers: polytetrafloroethylene (PTFE), high0pressure (high density) polypropylene (LPPE), low0density polypropylene (LPPE), polypropylene (PP), polypropylene POM, nylon 6 and nylon 66.

The DTA curve (Fig. 21.34) shows characteristic peaks of all the typical pure polymers and, therefore, indicates the presence of individual polymers in the analysed sample. *The area under the peak is related to the heat of reaction and related to the quantity of material present in the mixture*. Similarly, the DTA curve of ethylene propylene block copolymer indicates two peaks for ethylene and propylene. The comparison of the peak areas of the two polymers gives 51 percent ethylene and 49 percent propylene, respectively, present in the analysed block copolymer.

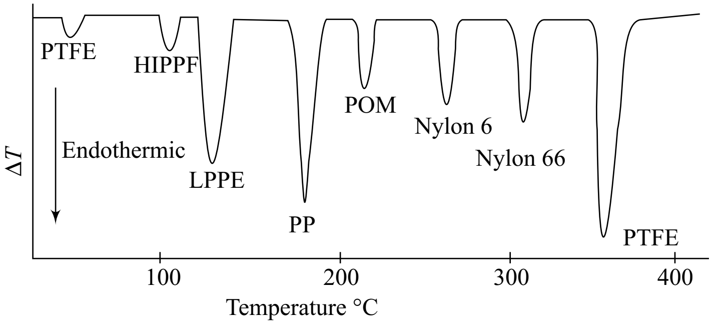


Figure 21.34 *DTA curve of a typical polymeric mixture.*

8. Glass transition (T_g)

DTA is useful for the determination of glass transition (T_g) of a polymer, a second-order transition caused by relaxation of chain segment in the amorphous portion of a polymer. As such this transition is not associated with latent heat but rather a sudden change in specific heat to bring a liquid into the glassy state. The necessary condition is to cool the polymer rapidly to approximately two-third of its melting point.

$$\frac{T_g}{T_m} = 0.66$$

The glass transition of a polymer can be obtained using number average molecular mass M ranging from 8×10^3 to 3×10^6 from expression:

$$T_g = \frac{[96.5 + / -1.0) - (2.8 + / -)] \times 10^3}{M}$$

From the DTA curve of glass transition temperature of polymer, it can be stated that

(a) the slope of the curve ΔT surface temperature should be sigmoidal.
(b) the maximum value of ΔT at the glass transition temperature should be linearly dependent on the heating rate.
(c) the inflection point T_g should increase with the heating rate. The T_g depends on heating rate, volume fraction and molar mass.

9. Measurement of crystallinity

A common application of DTA is the measurement of the mass fraction of crystalline material in semi-crystalline polymers. The method is based upon the measurement of the polymer's heat of fusion, ΔH_{f}, and the plausible assumption that this quantity is proportional to the crystalline content. If by some process of extrapolation, the heat of fusion, $\Delta H_{\mathrm{f}*}$, of a hypothetical crystalline sample is known, then the mass fraction of a crystallinity is

$$\text{Mass fraction} = \frac{\Delta H_f}{\Delta H_f^*}$$

Thus, *crystallinity of a polymer sample (X) can be determined by measuring the total energy absorbed (ΔH) by the sample per gram for a specific temperature interval minus the amount of energy which would be absorbed by 1 g of totally amorphous material [ΔH_f (a)] for the same temperature interval, and then dividing by the heat of fusion of 1 g of a perfectly crystalline sample [ΔH_f (c)].* The crystallinity can be expressed by the following equation.

$$X = \Delta H - \frac{\Delta H_f(a)}{\Delta H_f(c)}$$

Another method for the determination of polymer crystallinity is based on the ability of the instrument to cool a molten sample rapidly to a specific temperature, where crystallization is allowed to occur isothermally. A number of such crystallization curves may be obtained at different temperatures. The difference in crystallinity may be caused due to branching, nucleation and molecular effects.

10. Degree of polymerization

The area covered under the observed DTA peak of a sample of polymer is directly related to heat of polymerization and is expressed in terms of per mole or per gram. Consider the polymerization reaction process of formation of trialkylcyanurate and triallylisocyanurate by the DTA technique. Two parts Al_2O_3, one part monomer and 0.1 part catalyst were mixed as a paste in 50 percent tricresyl phosphate and the mixture was heated at a rate of 8°C/minute to obtain a polymer.

Typical values of heat of polymerization process as estimated from the peak area for two polymers are given in Table 21.8.

Materials	Heat of polymerization (ΔH_1)	(ΔH_2)
Trialkyl cyanurate	35	107
Triallylisocyanurate	56	

Table 21.8

Estimated heat of polymerization

Further, the area (size) of peak appears in the DTA curve had a great value in assessing the degree of curing. This is carried out by the residual cure remaining in a polymer system after various treatments. This approach has been applied for estimation of degree of curing in an unsaturated polyester–styrene copolymer cured at an ambient temperature. The variation in the size of the curve represents the percentage of curing in 2 hours 63 percent, 3 hours 68.6 percent, 4 hours 74.3 percent, 5 hours 77.0 percent and 6 hours 78.2 percent. Similarly, the relative change in the heat of reaction measured by DTA also gives information regarding the role of catalyst, degree of crystallinity and decomposition of polymer samples.

21.4 Differential Scanning Calorimetry (DSC)

21.4.1 Introduction

DSC is one of the analytical techniques belonging to a group called thermal analysis (TA). Differential scanning calorimetry (DSC) is used for examination of all physical processes taking place during heating and cooling of sample materials. Determination of the thermal behaviour is a required part for physical characterization of material samples.

A calorimeter measures the heat into or out of a sample. A differential calorimeter measures the heat of sample relative to a reference. A differential scanning calorimeter does all of the above and heats the sample with a linear temperature control. Endothermic heat flows into the sample. Exothermic heat flows out of the sample. The measurements provide quantitative and qualitative information about physical and chemical changes that involve endothermic or exothermic processes or changes in heat capacity.

Differential scanning calorimetry (DSC) is a technique for measuring the energy necessary to establish zero temperature difference between a test sample 'S' and an inert reference material 'R', while the two samples are subjected to an identical temperature programme.

Differential scanning calorimetry (DSC) is a thermoanalytical technique in which the difference in the amount of heat required to increase the temperature of a sample and a reference are measured as a function of temperature. Both the sample and reference are maintained at very nearly the same temperature throughout the experiment. Generally, the temperature program for a DSC analysis is designed such that the sample holder temperature increases linearly as a function of time. The reference sample should have a well-defined heat capacity over the range of temperatures to be scanned.

The basic principle underlying this technique is that when the sample undergoes a physical transformation such as phase transitions, more (or less) heat will need to flow to the sample than the reference to maintain both at the same temperature. Whether more or less heat must flow to the sample depends on whether the process is exothermic or endothermic. For example, as a solid sample melts to a liquid, more heat flows into the sample to increase its temperature at the same rate as the reference. This is due to the absorption of heat by the sample as it undergoes the endothermic phase transition from solid to liquid. Likewise, as the sample undergoes exothermic processes (such as crystallization) less heat is required to raise the sample temperature. By observing the difference in heat flow between the sample and the reference, differential scanning calorimeters are able to measure the amount of heat absorbed or released during such transitions.

DSC may also be used to observe more subtle phase changes, such as glass transitions. DSC is widely used in industrial settings as a quality control instrument due to its applicability in evaluating sample purity and for studying polymer curing. The result of a DSC experiment is a heating or cooling thermogram of a sample.

The main application of DSC is in studying phase transitions such as melting, glass transitions or exothermic decompositions. These transitions involve energy changes or heat capacity changes that can be detected by DSC with great sensitivity.

The applications of DSCs are the following:

- Phase transition including glass transition, Curie point
- Melting points
- Crystallization times and temperatures
- Heats of melting and crystallization
- Percent crystallinity
- Oxidative stabilities
- Compositional analysis
- Heat capacity
- Purities
- Thermal stabilities
- Polymorphism
- Heat set temperatures
- Kinetic studies
- Material fingerprinting

It is one of the sensitive tools among thermal analysis techniques that monitor heat effects associated with phase transitions and chemical reactions as a function of temperature. In a DSC, *the difference in heat flow to the sample and a reference at the same temperature is recorded as a function of temperature*. The reference is an inert material such as alumina (α-Al_2O_3). The temperature of both the sample and the reference are increased at a constant rate. Since the DSC is at constant pressure, *heat flow is equivalent to enthalpy changes*:

$$\left[\frac{dq}{dt}\right]_p = \frac{dH}{dt} \qquad ...(4)$$

Here, *dH*/*dt* is the 'heat flow' measured in mcal/second. The 'heat flow' difference between the sample and the reference is

$$\Delta\frac{dH}{dt}=\left[\frac{dH}{dt}\right]_{\text{sample}}-\left[\frac{dH}{dt}\right]_{\text{Reference}} \qquad ...(5)$$

and can be either positive or negative. In an endothermic process, such as most phase transitions, heat is absorbed and, therefore, heat flow to the sample is higher than that to the reference. Hence $\Delta dH/dt$ is positive. Other endothermic processes include helix-coil transitions in DNA, protein denaturation, dehydrations, reduction reactions and some decomposition reactions. In an exothermic process, such as crystallization, some cross-linking processes, oxidation reactions and some decomposition reactions, the opposite is true and $\Delta dH/dt$ is negative.

The DSC curve of the sample may be obtained by placing separate heating devices in the sample and reference chambers. The heaters are programmed to ensure that the temperatures of both sample and reference advance at exactly the same rate. It follows that when endotherms or exotherms occur in the sample, the power to the heater will need to be varied in order to maintain $\Delta T = 0$. Thus by monitoring the difference in power supplied to the heaters (ΔE) the thermal changes in the sample may be followed. This basically measures the amount of energy absorbed or evolved in a particular process, and hence gives calorimetric measurements directly. It is worth noting that the measurement of ΔE is effectively a direct measurement of the energy change in the sample. This makes DSC particularly appropriate as a technique for the measurement of ΔH values which can be derived from the areas of the peaks obtained.

The result of a DSC experiment is a heating or cooling curve (Fig. 20.35). This DSC curve can be used to calculate enthalpies of transition processes. This is done by integrating the peak corresponding to a given transition. It can be shown that the enthalpy of transition can be expressed using the following equation:

$$\Delta H = KA$$

where ΔH is the enthalpy of transition, 'K' is the calorimetric constant and 'A' is the area under the curve of peak. The calorimetric constant will vary from instrument to instrument and can be determined by analysing a well-characterized sample with the known enthalpies of transition.

A DSC curve is obtained by plotting *dH*/*dt* (mJ/s) as a function of time or temperature. A typical DSC thermogram in which the heat flow is measured and plotted against temperature or time is shown in Fig. 21.35.

If the DSC curve is above the base (zero) line, then it is called exothermic peak (–heat flows out of the sample) and the curve is below baseline, it is called endothermic peak (–heat flows into a sample).

The area 'A' under the peak is directly proportional to the heat evolved or absorbed by the reaction and the height of the curve is directly proportional to the rate of reaction. Therefore, Eq. (5) is equally valid for DSC scheme also. The only difference is the calibration factor '*K*' which is independent of temperature in the case of DSC. This is the major advantage of DSC over DTA.

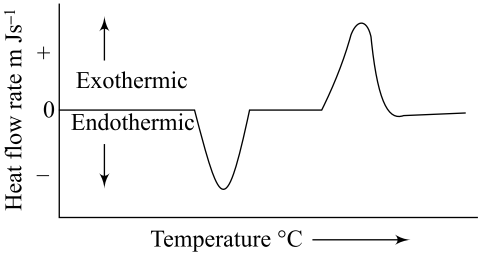


Figure 21.35 *A typical DSc curve.*

21.4.2 The Basic Types of DSC Systems

Differential scanning calorimeter works on two basic types of principles.

- based on *power compensation* and
- heat flux (flow) method

In 'power compensation method', a smaller secondary heater is attached to equalize the generated energy difference between sample and reference materials. While in 'heat flow' technique, heat flux passing through sample and reference is evaluated and its difference is related to energy consumed or released in the thermal reactions**.**

(a) Principle based on power compensation

- Temperature difference is maintained zero, i.e. $\Delta T = 0$, by supplying heat into the sample or reference according to heat emission or absorption.
- Electrical power is proportional to heat change in the sample.

That is $P = I^2 \cdot \mathrm{R}$

The sample (S) and reference (R) temperatures are controlled independently using separate heaters (1) and (2) (Fig. 21.36). The temperature difference between the sample and reference is maintained to zero by varying the power input to the two furnaces. This energy is then a measure of the enthalpy or heat capacity changes in the test sample 'S' (relative to the reference 'R').

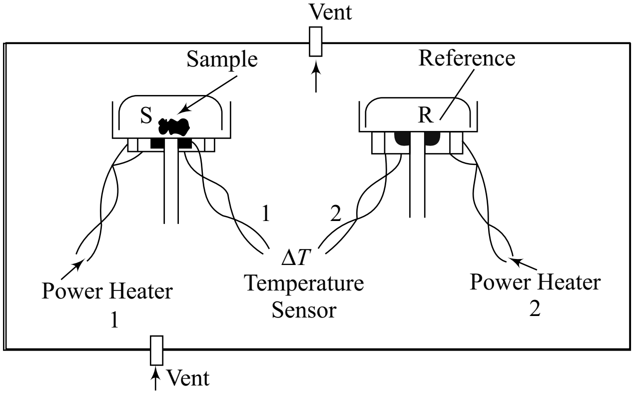


Figure 21.36 *Power compensation DSC.*

The working of DSC can be enumerated as follows:

- Pt cup is used to take a small quantity of sample. Platinum metal makes a good contact between sample, cup and heat flux plate.
- Symmetrical heating of the furnaces containing S and R is achieved by constructing the furnace from a metal of high thermal conductivity, e.g. silver.
- The two furnaces are kept on a large temperature-controlled heat sink.
- Sample and reference holders are at the middle of the furnaces.
- Sample and reference material are heated by *separate independent heaters (1 and 2)*.
- The temperatures of sample and reference are monitored continuously using *Pt which* are imbedded in the furnaces.
- There is provision for gas flow through the cell to sweep away volatiles and maintain the required atmosphere and allow heat transfer.
- Control of the furnace, signal acquisition and data storage and analysis are handled by a computer.
- Two control circuits are used to obtain differential thermograms.
  (i) One for *average temperature control*
  (ii) One for *ΔT control*
- S and R, temperature signals are fed into a *differential amplifier* via a *comparator circuit* that determines which is greater.

- The *amplifier output* then adjusts the power input to the two furnaces in such a way that their temperatures are kept identical.
- Throughout the experiment, S and R are isothermal.
- A signal proportional to difference in power input to the S and R furnaces is transmitted to the data acquisition system.
- The *power differences* are plotted as a function of the sample temperature and its unit is milliwatts (mW).

The differential scanning calorimeter consists of a sample holder and a reference holder as shown in Fig. 21.37.

Platinum cups are used for both sample and reference to allow high temperature range operation. Underneath each Pt holder, a resistance heater and a temperature sensor is placed. Currents are applied to the two heaters to increase the temperature at the specified rate. The difference in the power to the two holders, necessary to maintain the holders at the same temperature ($\Delta T = 0$), is used to calculate dH/dt. A schematic diagram of a DSC is shown in Fig. 21.38.

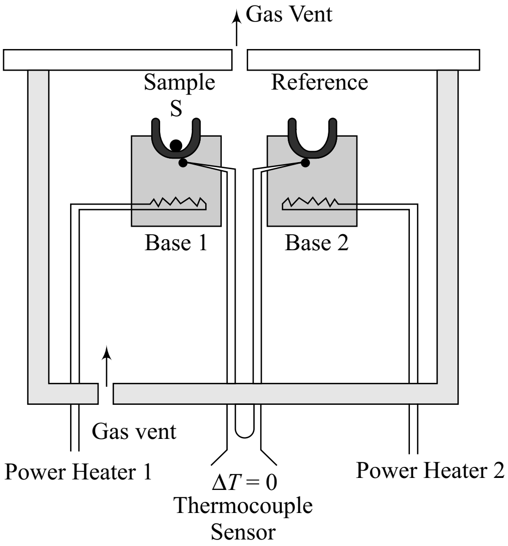


FIGURE 21.37 *Power compensated DSC assembly.*

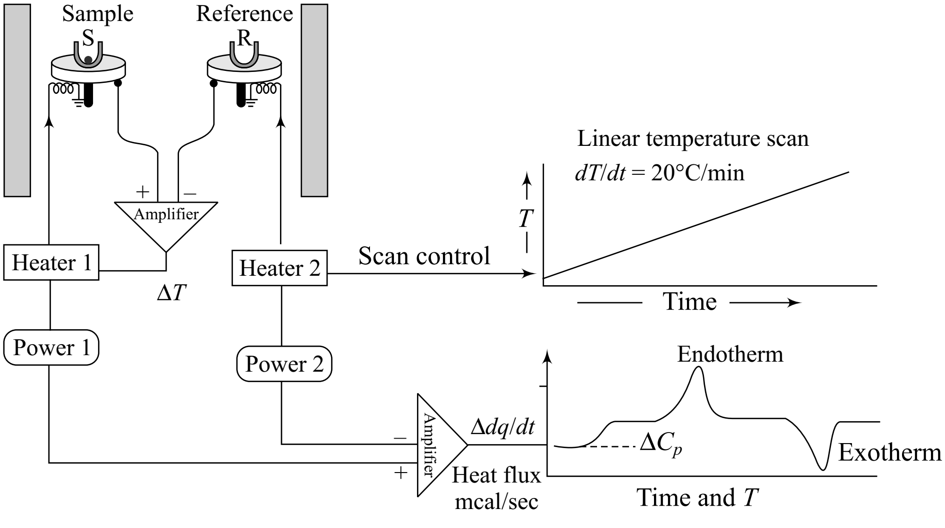


FIGURE 21.38 *Power compensated DSC with accessories for recording.*

The sample is maintained at an atmosphere of dry nitrogen/oxygen during the investigation. The nitrogen atmosphere also eliminates air oxidation of the samples at high temperatures. The sample is sealed into a Pt (cup)/pan. The reference is usually an empty (Pt) pan and cover. The maximum amount of sample taken is about 1–10 mg of material. Set the furnace for a linear temperature scan rate. The amplifiers are shown as triangles in the above figure which determine the difference in the two input signals. The sample heater power is adjusted to keep the sample and reference at the same temperature during the scan.

The heat capacity of the sample is calculated from the shift in the baseline at the starting transient. Glass transitions cause a baseline shift. Crystallization is a typical exothermic process and melting a typical endothermic process, and ΔH_{tr} is calculated from the area under the peaks. The peaks with positive and negative $\Delta dH/dt$ may be recorded during the heating of a sample and each peak corresponds to a 'heat effect' associated with a specific process, such as crystallization or melting (Fig. 21.39).

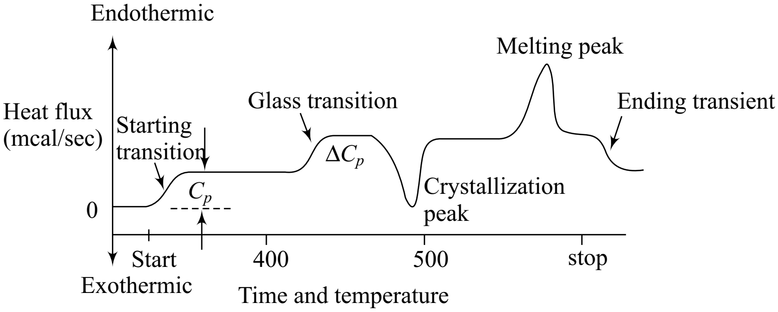


FIGURE 21.39 *A typical DSC scan of a sample polymer.*

In order to get information from DSC, note the temperature at which a certain process such as melting, transition, decomposition, etc. occurs. The 'peak temperature' is associated with the temperature at which maximum reaction rate occurs. In polymers, the glass transition temperature, '*Tg*', is an important aspect to note. This is the temperature at which amorphous (noncrystalline) polymers are converted from a brittle, glasslike form to a rubbery, flexible form.

At glass transition temperature, the specific heat, the coefficient of thermal expansion, the free volume and the dielectric constant (in the case of a polar polymer) all change rapidly for polymers. Since the mechanical behaviour of polymers changes markedly at the glass transition temperature, it is an important characteristic of every polymer. In the DSC experiment, '*Tg*' is manifested by a drastic change in the baseline, indicating a change in the heat capacity of the polymer (Fig. 21.40). No enthalpy is associated with such transition; therefore, the effect in a DSC curve is slight and is observable only if the instrument is sensitive enough.

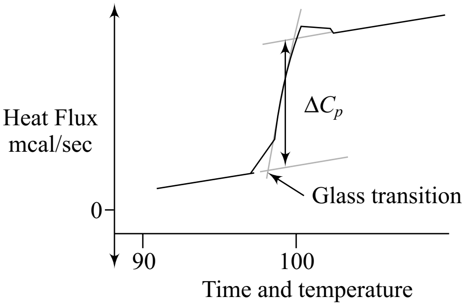


FIGURE 21.40 *A typical glass transition process.*

The second direct information obtainable from DSC curves is the enthalpy (ΔH) associated with certain processes.

If there are sloping baselines before and after the glass transition, extrapolate the baselines forwards and backwards (as shown by dotted lines) and take the baseline shift when the transition is about 63 percent complete (as shown by arrows).

(b) Principle based on maintaining heatflux in DSC

Unlike the power compensation method, the test sample 'S' and reference material 'R' (usually an empty Pt cup) are enclosed in the same furnace together with a metallic block with high thermal conductivity that ensures a good heatflow path between 'S' and 'R'. The enthalpy or heat capacity changes in the specimen 'S' lead to temperature differences relative to 'R'. This results in a certain heatflow between 'S' and 'R'; however, small compared to those in DTA, because of the good thermal contact between 'S' and 'R'. The temperature difference ΔT between 'S' and 'R' is recorded and further related to the enthalpy change in the specimen using calibration experiments. The heat-flux DSC system is thus a slightly modified DTA system: the only important difference is the good heatflow path between the specimen and reference crucibles.

Construction of heat-flux DSC cell

The main assembly of a typical heat-flux DSC cell is enclosed in a heating silver block which dissipates heat to the specimens (S and R). A constantan discs are placed on the Ag block (Fig. 21.41). The constantan disc has two cups on which the 'S' (sample) and R (reference) pans are placed. A chromel disc and connecting wire are attached to the underside of each platform. These chromel-constantan thermocouples are used to determine the differential temperatures of interest. Alumel wires are also attached to the chromel discs to provide chromel–alumel junctions which measure the sample and reference temperatures separately. Another thermocouple is embedded in the Ag block and serves as temperature controller for the programmed heating/cooling cycle.

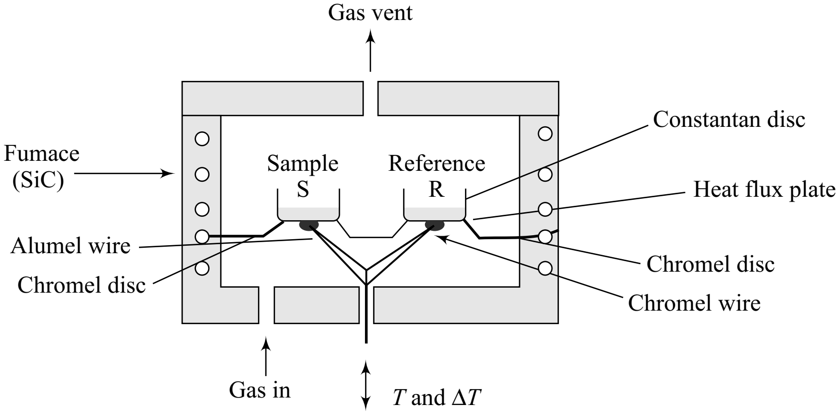


Figure 21.41 *Main components of heat–flux DSC cell.*

In heat-flux DSC instruments, the difference in energy required to maintain both 'S' and 'R' at the same temperature is a measure of the energy changes in the sample 'S' (relative to the inert reference 'R'). The thermocouples are usually not embedded in neither 'S' or 'R' materials. The temperature difference ΔT that develops between 'S' and 'R' is proportional to the heat flow between the two. In order to detect such small temperature differences, it is essential to ensure that both 'S' and 'R' are exposed to the same temperature programme.

The measurements are usually performed under vacuum or inert-gas flow. The flow rate of inert gas (40 mL/minute) is maintained constant throughout the experiment.

In brief, the working of the heat flux assembly is as follows:

- S and R are heated by a programmed *single heater.*
- *Differential heat flow* into the 'S' and 'R' cup/pans is monitored by chromel disc/constantan thermocouple.
- *Differential heat flow* into 'S' and R pans is directly proportional to the difference in output of the two thermocouple junctions.
- Sample temperature is estimated by chromel/alumel junction under the sample disk.

The atmosphere around the heaters are maintained with the following conditions:

- Typical purge gases are air/N_2
- Helium is useful for efficient heat transfer and removal of volatiles.

- Argon is preferred as an inert purge when examining samples that can react with nitrogen.
- The experiment can also be carried under a vacuum or under high pressure.

In the DSC method, 10 mg of uniform sized sample particles are placed in a sample container for analysis. A wide range of heating rate (0.5–20°C/min) can be used. The DSC instruments are very sensitive and detect heat evolution or absorption at a rate less than one milli-calories per second. Electrical signals are amplified and recorded similar to TGA and DTA. Under uniform thermal process in DSC heat is either liberated or absorbed.

For the endothermic process, ΔH is positive and the sample heating device is energized and a positive signal is obtained; when ΔH is negative the reference heating device is energized and a negative signal is obtained. An idealized representation of the three major processes observable in DSC is given in Fig. 21.42. The peak area in DSC is proportional to the amount of sample, the heat of reaction and similar to DTA peak area can be expressed by following equation.

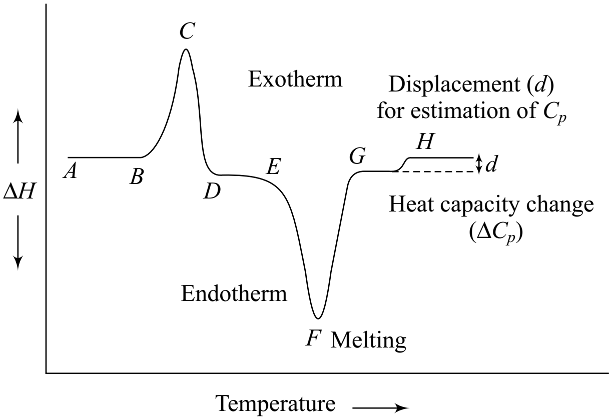


Figure 21.42 *An idealized representation of the three processes observed in DSC.*

21.4.3 Calibration of DSC

- *Temperature calibration* is carried out by running standard materials, usually very pure metals with accurately known melting points (Tables 21.9 and 21.10).
- *Energy calibration* may be carried out by using either known heats of fusion for metals, commonly indium or known heat capacities.

Standard	Melting point t (°C)	Heat of fusion (J/g)
Indium	156.6	28.42
Tin	231.9	
Lead	327.5	
Zinc	419.5	
K_2SO_4	585.0	
$K_2Cr_2O_7$	670.5	

Table 21.9

Commonly used standards and reference materials

Substance	Transition	Transition temperature t (°C)
Cyclopentane	Crystal	–151.16
Cyclopentane	Crystal	–135.06
Cyclohexane	Crystal	–87.06
Cyclohexane	Melt	6.56

(*Contd.*)

(*Contd.*)

Substance	Transition	Transition temperature *t* (°C)
n-heptane	Melt	–90.56
n-octane	Melt	–56.76
n-decane	Melt	–29.66
n-dodecane	Melt	–9.65
n-octa decane	Melt	28.24

Table 21.10

Commonly used standards and reference compounds

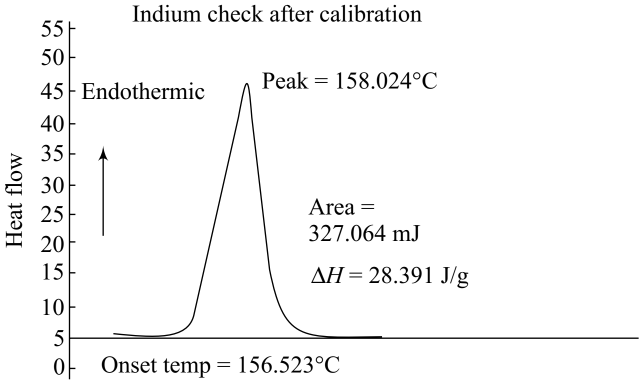


FIGURE 21.43 *Calibration peak with indium.*

21.4.4 Determination of Enthalpy Change or Heat Capacity of Materials

Differential scanning calorimeter provides an accurate method of measuring heat capacities and enthalpy changes of materials. For the evaluation of the total enthalpy change of a process, the integral under the DSC peak (above the baseline) is considered

$$\int\left(\frac{dH}{dt}\right)_{\text{sample}} \cdot dt = \Delta H_{\text{sample}} \qquad \text{...(6)}$$

Assuming that the heat capacity of the reference is constant over the temperature range covered by the peak, ΔH_{refer} will cancel out because the integral above the baseline is taken. Therefore, Eq. (5) is also valid when the integral is taken from the DSC plot of $\Delta dH/dt$. Heat capacities and changes in heat capacity can be determined from the shift in the baseline of the DSC thermogram. The 'heat capacity' is defined as

The temperature scan rate is

$$\text{Scan reat} = \frac{dT}{dt}$$

Using the chain rule:

$$C_p = \left(\frac{dH}{dT}\right) = \frac{dH}{dt}\frac{dt}{dT} \qquad \text{...(7)}$$

where *dH*/*dt* is the shift in the baseline of the thermogram (Fig. 20.39) and the last derivative is just the inverse of the scan rate. For differential measurements, we determine the difference in the heat capacity of the sample and the reference:

$$\Delta C_p = C_p(\text{sample}) - C_p(\text{Reference}) \qquad ...(8)$$

$$\Delta C_p = \Delta\left(\frac{dH}{dT}\right) = \Delta\frac{dH}{dt}\frac{dt}{dT} \qquad ...(9)$$

The units of the heat flow are mcal/second and the temperature scan rate is usually expressed as °C/min. So to be consistent with units you must multiply by 60 second/minute:

$$\Delta C_p = \left(\frac{\text{mcal}}{\text{second}}\right)\left(\frac{\text{minute}}{t°\text{C}}\right)\left(\frac{60\text{ sec}}{\text{minute}}\right)$$

The basic principle underlying this technique is that *when the sample undergoes a physical transformation such as phase transitions, more (or less) heat will need to flow to it than the reference to maintain both at the same temperature.*

Commercial instruments provide a recorder output of the constant pressure heat capacity, as a function of temperature.

$$C_p = \left(\frac{dH}{dT}\right)_p$$

The area under such a curve between any two temperature limits (T_1 and T_2) yields an enthalpy change:

$$\Delta H = \int_{T_1}^{T_2}\left(\frac{dH}{dt}\right)_p = \int_{T_1}^{T_2}(C_p \cdot dT) \qquad ...(10)$$

21.4.5 Measurement of Enthalpy Change

To determine the enthalpy change of a process, the power compensated method is used in DSC measurements. In this type of DSC assembly, there are two small sample holders: one for the sample and the other to hold a reference material. Energy is supplied to each sample holder independently, and the temperatures of each are monitored. Energy in the form of heat is absorbed or evolved by the sample which is compensated by increasing or decreasing the electrical energy delivered to the sample holder heater. The temperatures are monitored by comparing the respective signals from platinum resistance thermometer (thermocouples) in each of the holders. This adjustment of electrical input to the heating unit provides a varying electrical signal which is opposite but equivalent to the varying thermal behaviour of the sample. For example, when a phase transition point is reached from lower temperatures, the sample will absorb additional energy. Then additional electrical power must be supplied to the heater of this sample so that the reference and sample holders will be at the same temperature. When the transition is completed, the temperature of the sample will increase, and additional heat needs no longer be supplied. The varying electrical signal thus provides a record of the thermal behaviour of the sample.

The experimental C_P–T curve displays a broader peak, and *the enthalpy change ΔH for the transition is taken as the area under the peak* (Fig. 21.44). The nominal transition temperature is often taken as the point at which the curve first rises from the baseline; alternatively, if the peak is very sharp, the temperature at the maximum can be used.

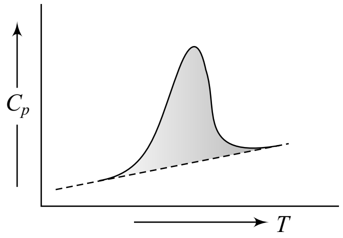


Figure 21.44 *C_p versus temperature plot in DSC.*

(a) Peak area and enthalpy change of sample

$$\text{Peak area } (A) = \pm\Delta H \cdot m \cdot K \qquad \ldots(11)$$

where 'ΔH' represents the sample enthalpy change and 'm' is the mass of sample and 'K' is a constant called *calibration factor*. Unlike DTA, it is independent of temperature. Using Eq. (11), we can determine enthalpy change (ΔH) of a sample directly from peak area (A), if we know the value of K. If area (A) is measured in cm^2 and the unit of ΔH is J/g, the unit of K will be cm^2/J

$$K = \frac{\text{cm}^2}{\text{g.J./g}} = \text{cm}^2/\text{J}$$

In SI units, a change in enthalpy, ΔH is kJ/mol. It is also expressed as kcal/mol.

We can also determine the enthalpy change by comparing the ΔH of the sample with the known ΔH of the standard. That is

$$\Delta H_s = \frac{A_k \cdot m_k \cdot \Delta H_s}{A_s \cdot m_s} \qquad \ldots(12)$$

where ΔH_s is the enthalpy change for sample, ΔH_k is the enthalpy change for the known standard, m_s and m_k are masses of sample and known standard, respectively, and A_s and A_k represent/s the area of peaks of sample and standard materials, respectively.

(b) Heat capacities

The DSC technique is not only sensitive for the determination of ΔH but also very sensitive for the determination of *heat capacities* (C_p). when a sample is subjected to a heating programme (DSC), *the rate of heat flow into the sample is proportional to its heat capacity*. This may be detected by the displacement of the baseline, as illustrated in Fig. 20.42. The value of C_p may be determined at a particular temperature by measuring this displacement (d):

$$c_p = \frac{d}{\text{Heating rate} \times m} \qquad \ldots(13)$$

$$C_p = \frac{\left(\dfrac{dH}{dt}\right)}{\left(\dfrac{dT}{dt}\right)} \times \frac{1}{m} \qquad \ldots(14)$$

Using Eq. (14), we can deduce the unit of C_p.

DSC curve displacement (d) will be measured in m J/s. If heating rate is in °C/s and 'm' is expressed in g, then,

$$C_p = \frac{\text{mJ/s}}{\text{g}} \times \frac{1}{°\text{C/s}} = \text{mJ/g/C}$$

Though in SI units, it is expressed as J/mol/K.

C_p can also be expressed as mcal g/C(conversion factor for J and cal. is 1 calorie = 4.2 J).

In practice, we normally measure the baseline shift, 'd', by reference to a baseline obtained for empty sample and reference pans. To further minimize experimental error, we usually determine the heat capacity of the sample by comparing with the known heat capacity of the standard.

$$K' \, (\text{H}_2 - \text{H}_1) = mC_p \frac{dT}{dt} \quad \text{or } K'd = mC_p \frac{dT}{dt} \qquad \ldots(15)$$

where H_1 and H_2 are differential heat generated when the instruments are first run without any sample at all and then with the test sample in position [in DSC curve ($H_2 - H_1$) is expressed as displacement, 'd']K' is calibration factor, it can be determined by calibration against standard substance. However, K' from Eq. (15) can be eliminated, if a material with a known heat capacity is used to calibrate the instrument.

Once of the commonly used standard is α-aluminium oxide (Al_2O_3) or synthesized sapphire for which specific heat has been determined to five significant figures in the temperature range 0–1200 K. After the baseline and sample program, a third program is run with a weighed sapphire structure. At any temperature T, following equation applies

$$K'd = mC_p \frac{dT}{dt} \qquad \text{...(16)}$$

$$K'd' = m'C'_p \frac{dT}{dt} \qquad \text{...(17)}$$

where d and d' are ordinate deflections (displacements) due to the sample and the standard, respectively, $m'C'_p$ are the mass and heat capacity of the standard. Dividing the Eq. (16) by Eq. (17) we get

$$\frac{d}{d'} = \frac{m}{m'}\frac{C_p}{C'_p} \quad \text{or} \quad \frac{C_p}{C'_p} = \frac{dm'}{d'm} \qquad \text{...(18)}$$

Thus, the calibration requires only the comparison of the two displacement values at the same temperature. We can easily calculate value of C_p on putting the rest values in Eq. (18). The basic components of DSC are quite similar except the differential energy measuring system.

21.4.6 Factors Affecting DSC Curve

The factors affecting DSC thermogram may be due to instrumentation or nature of sample.

The main factors that affect the shape of the curve, precision and accuracy are dependent on the following factors:

1. Instrumental factors
    - Furnace heating rate
    - Recording or chart speed
    - Furnace atmosphere
    - Geometry of sample holder/location of sensors
    - Sensitivity of recording mechanism.
    - Sample container
2. Characteristics of samples
    - Amount of sample
    - Solubility of evolved gases in sample
    - Particle size
    - Heat of reaction
    - Sample packing
    - Nature of sample
    - Thermal conductivity

21.4.7 Interpretation of DSC Curves

1. Fingerprint of a pure compound

A DSC thermogram is the fingerprint of the pure compound. A typical pattern of DSC indicates the transition temperature as well as the amount of heat required for the transition process. Therefore, the DSC curve can be used to ascertain the thermal behaviour of samples. The endothermic or exothermic peak obtained for the sample deviates from the baseline. In the DSC curve, the peak above the baseline (Fig. 21.45) is exothermic while that below the baseline is endothermic. *Area under the DSC curve is proportional to the amount of heat evolved for a process*. The area under the DSC curves is used for the stoichiometric ratio of analysed compounds (quantitative interpretation). The thermal stability of a material can also be assessed through DSC experiment of the sample. Information obtained from DSC can be used to select material for certain end use application and for predict product performance and improve product quality.

DSC curves of polymeric mixture is illustrated pointing out the probable transitions (Fig. 21.45) and change in behaviour of a polymer sample.

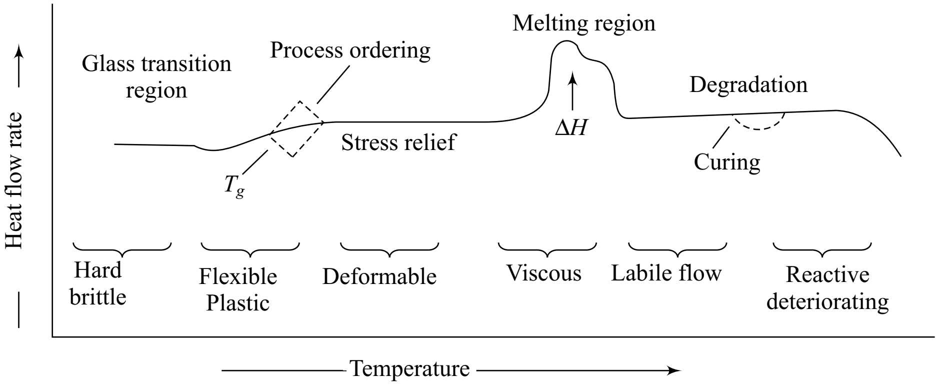


FIGURE 21.45 *Change in behaviour of polymeric materials in DSC.*

The DSC technique is more sensitive than DTA and the technique has been adopted to ascertain the presence of polymorphism, degree of crystallinity, curing fraction, etc. of samples. Figure 21.45 shows the peaks for the glass transition, ordering, melting and decomposition of individual polymers. The ratio of areas under the curves divided by the respective enthalpy change provides the ratio of individual monomers in an analysed copolymer sample.

The heat of reaction (ΔH_r) obtained in DSC can be further used to calculate the molar enthalpy of reactions by using following formula:

$$\Delta H_m = \Delta H_r \cdot \left(\frac{M_r}{m}\right) \qquad ...(19)$$

where ΔH_m = molar enthalpy of reaction
ΔM_r = relative molar mass of analysed compound
m = mass of substance used for analysis.

2. Epoxy mixtures

Two samples with different epoxy resin to hardener ratios are obtained. The epoxies were maintained at room temperature for several weeks prior to the DSC measurements. Figure 21.46 shows DSC curves of room-temperature cured, low-shrinkage epoxy samples.

The black curve shows the epoxy mixed in accordance with the manufacturer's (five parts of epoxy resin with one part of epoxy hardener) instructions. The dotted DSC curve shows the result for a four-part epoxy resin to one-part epoxy hardener ratio. The endothermic peak temperature of the prescribed (5:1) is 53.64°C and that of the 4:1 ratio mixture is higher at 55.89°C. The total endothermic peaks (energies) are also different. Therefore, DSC can be used as a tool for quality control of epoxy mixture ratios.

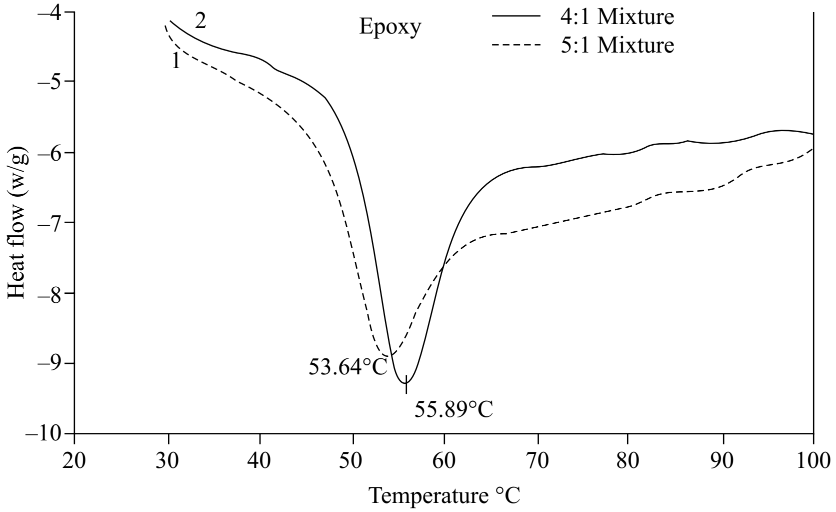


Figure 21.46 *A DSC scan of epoxy mixture.*

3. Phase transitions

When every solid-state material is heated continuously, phase transitions occur at different temperatures. For example, when ice is heated at standard atmospheric pressure, it will undergo a phase change from solid to liquid (melt) at a transition temperature (melting point) of 0°C, and when the resulting water is heated, it will eventually undergo another phase transition from liquid to gas (vaporize) at a transition temperature (boiling point) of 100°C.

Similarly, structural transitions can also be detected in solid-state materials through DSC measurements. Examples cited ($Cu_2V_2O_7$, $Cu_3V_2O_8$, etc.) in DTA topic are the best. The energy per unit mass used in bringing about such a phase change is known as the 'latent heat'.

A DSC measures the heat absorbed or released by a material as a function of temperature or time (isothermally). This then allows the determination of important thermal characteristics such as transition temperatures, for example, melting and boiling points as well as other transition temperatures such as glass transition temperature (T_g) and crystallization temperature (T_c). It also allows for the determination of quantitative measurements such as the specific heat capacity, heat of fusion, heat of vaporization, heat of polymerization, heat of crystallization and heat of reaction. Practically, all DSC instruments allow such measurements to be done under various atmospheres.

(a) Phase transitions in liquid crystals

When energy is supplied in the form of heat to most solids, they transform to the molten state with increased degree of freedom and also orient themselves randomly. In other words, the constituent molecules lose both positional and orientational order in the liquid state.

Many liquid crystals show properties neither of solids nor liquids. Such liquid crystals exhibit phase transitions before melting to liquid states. One form of liquid crystal gets transformed to the other form of liquid crystal with varying degrees of orientational and translational order.

For example, *p*-azoxyanisole (PAA) (Fig. 21.47) transforms from the solid state to the liquid crystal state (solid) at 391 K, and then there is a transition to the normal liquid state at 408 K.

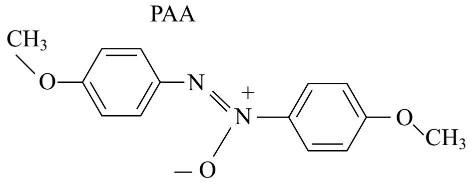


Figure 21.47 *Structure of p-azoxyanisole.*

By determining the enthalpy change associated with each of the transitions, it should be able to deduce whether the liquid crystal phase resembles more closely that of the solid or the liquid. Several kinds of liquid crystal phases are possible, depending on the liquid crystal system. A good system to study transitions is a liquid crystal phase, cholesteryl myristate (Fig. 21.48).

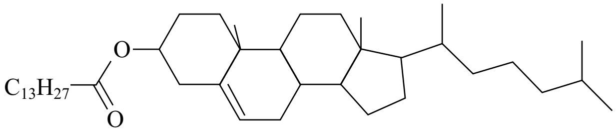


Figure 21.48 *Structure of cholesteryl myristate.*

One can note the following phase transitions in cholesteryl myristate liquid crystal. This phase exhibits a solid-to-sematic-A transition at 344 K, a sematic-A-to-chiral-nematic transition at 352 K and then the transition to the normal liquid at 358 K.

(b) Structural transitions in ammonium chloride

Structural phase transitions are quite common in solids. Innumerable solids exhibit such structural transformations at higher temperatures. One of the best examples of structural transformation observed is ammonium chloride, which is an ionic material that can exist in several different crystal structures. Phase I, which is stable above 457.6 K, is a cubic crystal which resembles common NaCl. Below the transition temperature of 457.6 K, phase II is stable and also has cubic structure, which is identical to the structure of CsCl. Transitions between such different crystal structures are quite common and are generally of the sharp type also called ‘first order’. However, they may take a longer time to occur, and so there is some danger of superheating or supercooling. One should be able to get fairly good values of the transition temperature and ΔH for the change in state.

$NH_4Cl(s)$ (II) → $NH_4Cl(s)$(I) $T = 457.6$ K, $P = 1$ atm

[Cubic – NaCl type] [Cubic – CsCl type]

(c) Structural transformation temperatures of a shape memory alloy

On heating Cu–Al–Ni alloy at a rate of 5°C/minute in an atmosphere of N_2, an endothermic peak is observed in the DSC thermogram (Fig. 21.49) at 153.99°C. Cu–Al–Ni alloy shape memory material undergoes a mar-

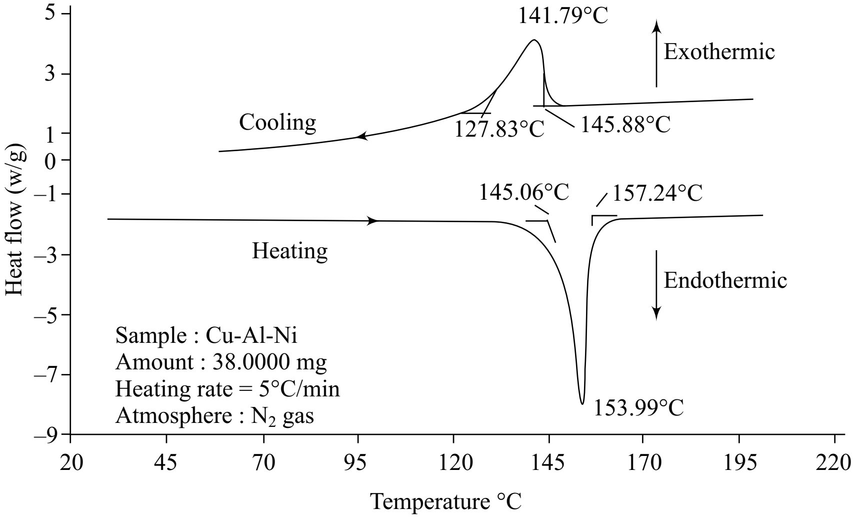


Figure 21.49 *DSC thermogram of Cu–Al–Ni alloy.*

tensite to austenite transformation at 153.99°C. Upon controlled cooling, a reversible austensite to martensite transition could be observed in the DSC curve at 141.79°C. This reversible transformation is an exothermic process. The temperatures of these transitions are a function of the alloy composition. There is a hysteresis in the transformation on heating and cooling which is notable in the DSC thermogram. The onset temperatures, 145.06°C on heating and 145.88°C on cooling are close.

(d) Phase transitions of polymer (PEEK, polyetheretherketone)

Generally, the specimen sample is continuously heated (or cooled) with a steady heating/cooling rate in dynamic DSC/DTA investigations. Figure 21.50 illustrates a typical DSC thermogram of a well-known polymer (PEEK, polyetheretherketone) showing the phase transition during the first heating cycle. An endothermal peak around 140°C corresponds to the glass-transition of PEEK and may be taken as proof for the initial glassy-state of the sample. Upon further heating, crystallization occurs with an exothermic peak (heat released) at 169°C, followed by melting of the sample (endothermic—heat absorption) at 339°C.

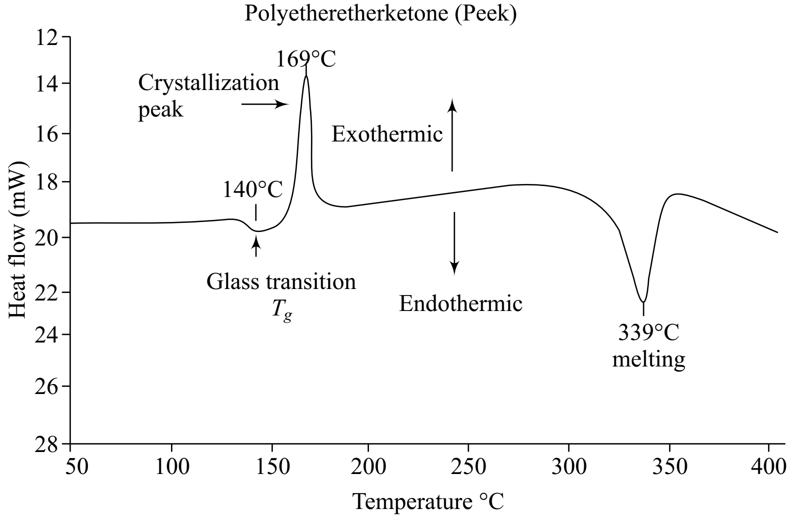


Figure 21.50 *A DSC thermogram of PEEK.*

4. Glass transition of silicon rubber

Glass transitions can occur in amorphous materials. They are referred to as second-order transitions because the glass transition has no latent heat associated with it. The heat flow versus temperature for a glass transition of silicon rubber is shown in Fig. 21.51.

The molecules are more mobile because of the increased heat capacity of silicon rubber and so the material is now in a state where its atoms/molecules have more energy. In order to keep the heat rate constant, the sample pan heater needed to increase heat flow. The transition at t_g for silicon rubber is endothermic. The material is still amorphous but has undergone a transition from a lower energy state to a higher energy state. The temperature at which the material undergoes a glass transition is known as the glass transition temperature (T_g), and the T_g is assigned to the midpoint temperature of the range of temperature during the transition (Fig. 21.51).

A simple example of a glass transition can be seen in plastics. When a plastic or rubber material is placed in a freezer, it becomes very brittle and can easily be shattered, the plastic or rubber is then in a 'glass' state. When it is warmed, the plastic or rigid rubber becomes very pliable or goes into a 'rubbery' state. The glass transition temperature of polyethylene terephthalate (PET), the plastic materials by which various beverage containers are made, has a T_g of about 70°C and the T_g for glass used as containers is about 550°C.

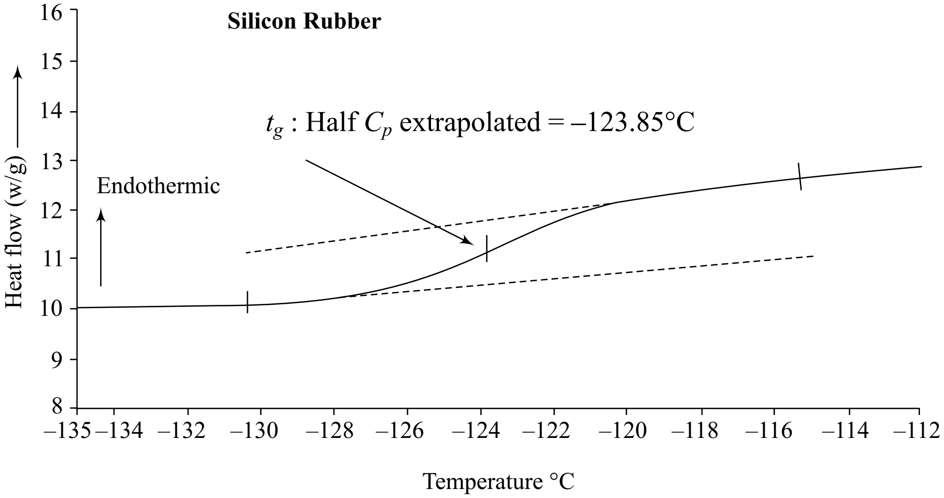


FIGURE 21.51 *Glass transition temperature (T_g) of silicon rubber.*

5. Crystallization

The crystalline nature of a sample material refers to well-'ordered state'. The crystallization process of material which takes place with intake of latent heat of transition is a first-order transition. The process is exothermic (Fig. 21.39), that is thermal energy is released during the process. Therefore, in order to maintain a constant heating rate, the sample heater supplies less energy than the reference heater. Crystallization is a time process dependent on nucleation and growth.

The crystallization temperature (T_c) is observed as a peak in the curve in the DSC thermogram. The peak area of the crystallization process can be assessed. Using the heating rate and the mass of the sample, the latent heat of crystallization (ΔH_{tr}) can be calculated. At the crystallization temperature (T_c), the material has gained enough energy so that its atoms/molecules could rearrange into a more stable and ordered state. Generally, materials tend towards the lowest energy state, and therefore, the lowest energy state is possible in a perfect-ordered crystalline state. A crystalline material can also undergo solid–solid phase transitions.

For example: Caesium nitrate undergoes a solid–solid phase transition at 154°C.

6. Melting

A solid-state material melts to a disordered state to give a liquid. If multiple phases are present, DSC thermogram shows multiple melting transitions. Melting of a sample could be due to the following:

- Order to more disordered state
- Always a change in volume (positive or negative)
- Always a change in heat capacity

Crystallization (T_c) and melting (T_m) are thermodynamically reversible occur either as exothermic or endothermic process. Figure 21.39 depicts an endothermic process for melting of a sample. During such a process of melting, the sample heater supplies more energy per unit time as compared to the reference heater so that the furnace heating rate DSC remains constant. The position of the peak in the curve is the melting temperature (T_m) and the area under the curve may be used to find the latent heat of fusion (melting).

A thermogravimetric analyser (TGA) is used to complement a DSC and measures mass changes of a sample as a function of temperature or time. The measurements also can be performed under various atmospheres just like the DSC. The TGA, thus, allows, among others, the determination of thermal and oxidative stability and optimum processing conditions. The TGA can be used in conjunction with other thermal equipment such as the DSC. For example, if one sees an endothermic or exothermic reaction using a DSC, the TGA measurements will help determine whether the change is physical or chemical in nature. The reaction is chemical in nature if a mass change is associated with it and physical if no mass change occurs.

Figure 20.52 shows a simultaneous DSC and TGA thermal profile of barium yttrium oxides showing loss of adhered and hydrated water in the sample followed by decomposition into component crystalline and amorphous materials. The latter was confirmed by X-ray diffraction measurements.

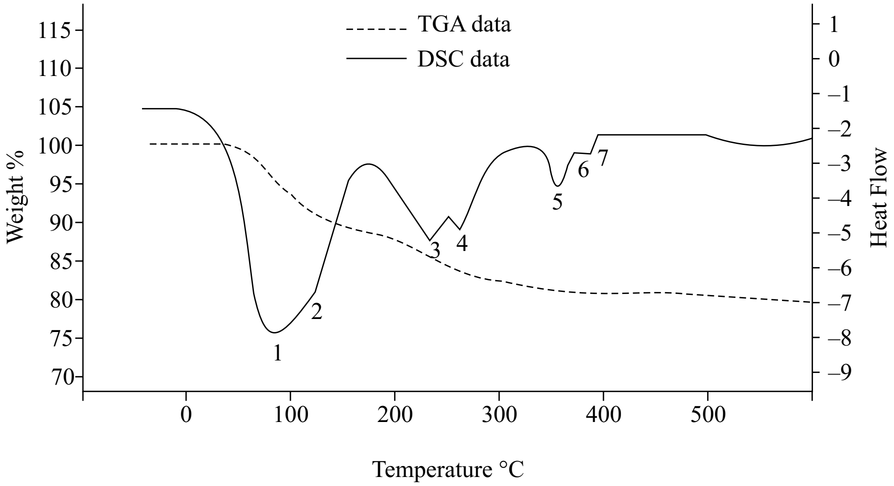


Figure 21.52 *Thermal profile of barium yttrium oxide showing TGA (weight changes) and DSC (heat flow).*

7. Industrial raw material inspection

A material can be identified through DSC investigation. The known DSC profiles may be used to identify few similar materials. For example, by running samples and finding the value for T_g and T_m, the material can be identified and qualified by running a sample of an unknown content and comparing the DSC profile to that of a known material. In the example shown in Fig. 21.53, DSC is used in quality assurance and control, where

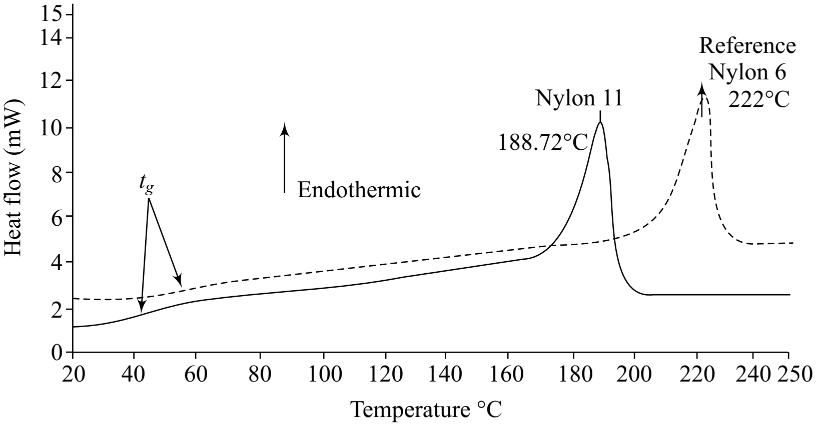


Figure 21.53 *Industrial raw material inspection.*

'incoming Nylon's melting point is compared to the quality standard for melting temperature' (nylon 6). In this case, it turns out that the incoming nylon turns out to be nylon 11 instead of the required nylon 6 which has a higher melting point.

8. Magnetic transition

Unlike DTA technique, magnetic transition in the materials can be detected by a DSC run of the sample.

The magnetic transitions are of second-order with $\Delta H = 0$ but with a maximum heat capacity at the transition temperature. Technique is sensitive to detect these minor variations. A typical DSC curve for a magnetic transition is shown in Fig. 21.54. Similarly, a glass transition can be detected easily with DSC.

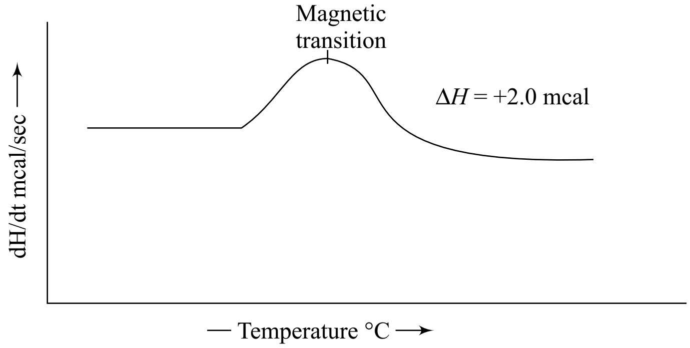


Figure 21.54 *A typical magnetic transition observed in DSC.*

9. Determination of 'T_g' of polymeric material (HDPE)

A polymeric material such as high-density polyethylene (HDPE) can be subjected to DSC investigation to show phase transitions. The DSC thermogram of the sample showed three-phase transition temperatures: glass transition temperature (T_g), crystallization temperature (T_c) and the melting temperature (T_m). From Fig. 21.55, it is possible to determine 'T_g' of HDPE material as 83.94°C at half C_p extrapolated value.

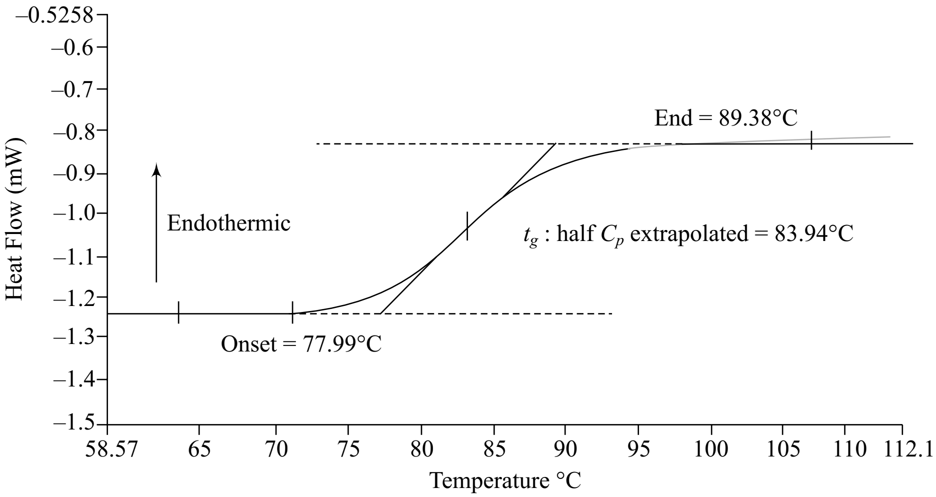


Figure 21.55 *Glass transition temperature of HDPE.*

Polymer characterization

- 'T_g' may be used to identify polymers.
- The amount or effectiveness of a plasticizer may be judged by how much it reduces T_g or affects the shape of the transition.
- Examination of the transitions in polymer blends gives information as to their compatibility.
- Curing reactions result in an increase in T_g and measurements can be used to monitor the extent of cure.

Figure 21.56 illustrates how T_g values of materials are affected by the variant conditions and composition of samples.

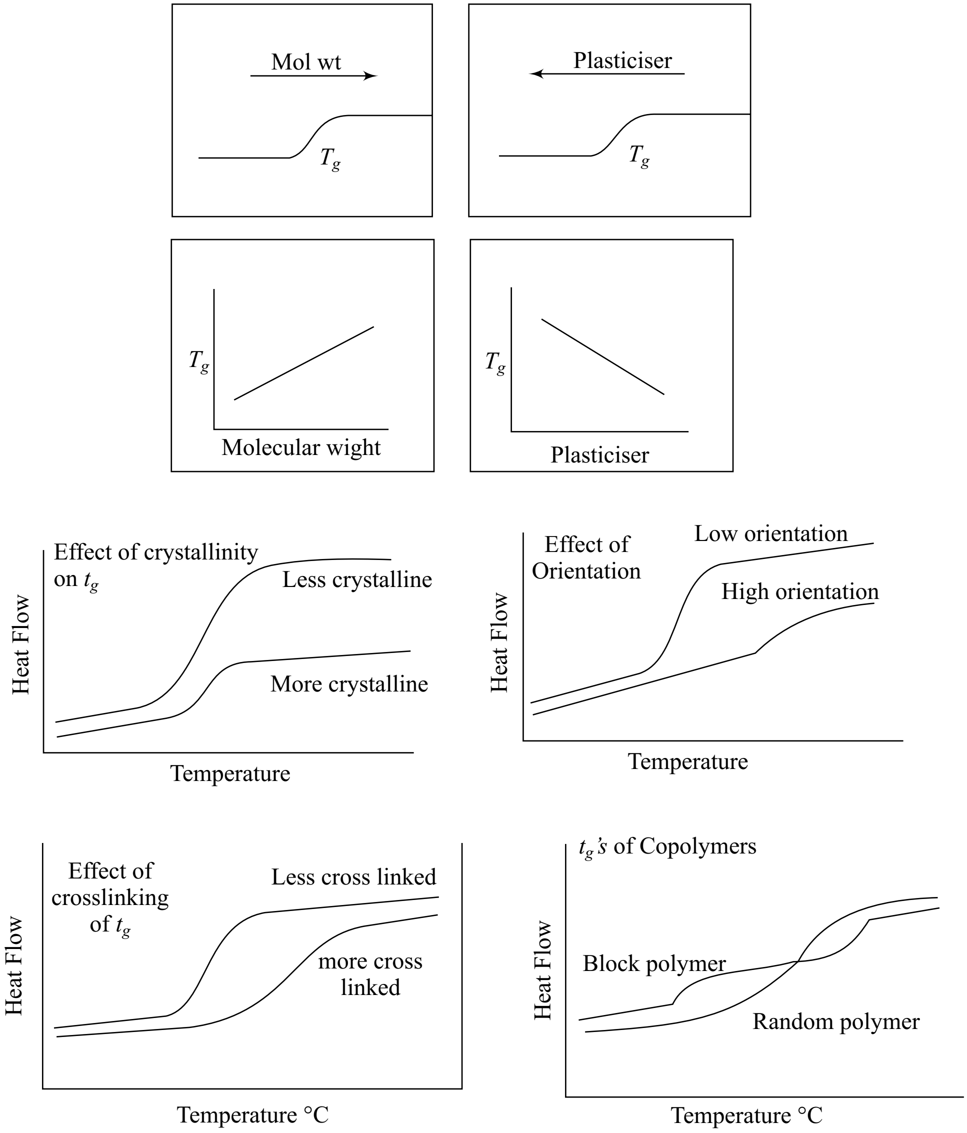


Figure 21.56 *Variation of T_g under different conditions for samples.*

Polymer characterization

- DSC thermograms of the blend of (A + B) samples show two distinct T_g (Fig. 21.57).
- Therefore, the components of the above blend of samples are immiscible in each other.
- T_g also varies with chain length for a related group of polymers.

Additional features occurring in the glass transition region, often a superimposed endothermic peak, are related to the ageing undergone by the material in the glassy state, and can sometimes obscure the transition, making precise temperature measurement difficult.

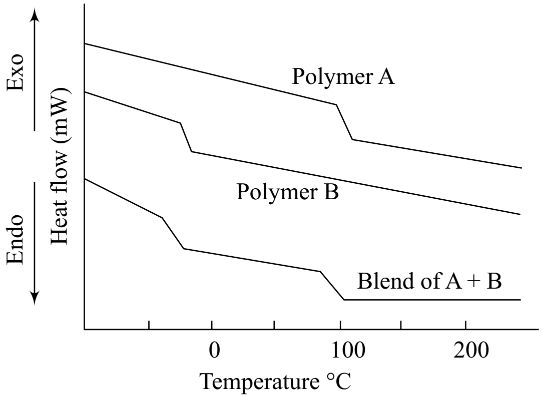


Figure 21.57 *Polymer characterization of polymers.*

10. Analysis of explosives

One of the important components of high explosives is ammonium perchlorate. The stability of this material is of concern and critical to their safe handling. In order to ascertain its stability, the decomposition of ammonium perchlorate was investigated by DSC (Fig. 21.58). Literature values for E_a = 37 – 260 kJ/mol with different mechanisms proposed. The DSC work clarified the mechanism and identified the activation energy as 115 kJ/mol.

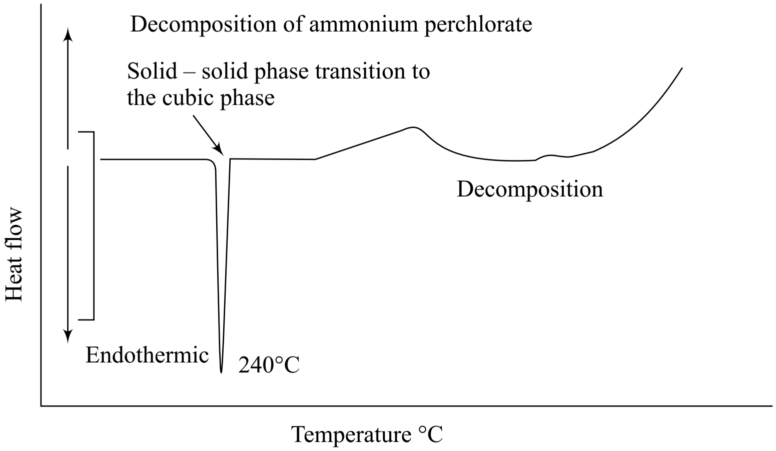


Figure 20.58 *DSC of decomposition of ammonium perchlorate.*

11. Analysis of lubricants

The stability of lubricating oils at elevated temperatures and pressures has to be ascertained before it is being used in the automotive industry. Its utility as a lubricant in motors at high-operating temperature and under an oxygen atmosphere is desirable factors. It is observable that at some point, the oil begins to oxidize and then quickly decomposes exothermically. It is observed that the synthetic oil has a much longer oxidation induction time (OIT) than does the mineral oil (Fig. 21.59).

12. Thermophysical properties of polymers and biological systems

Differential scanning calorimetry (DSC) is widely used to characterize the *thermophysical* properties of polymers. DSC can measure important *thermoplastic* properties including:

- Melting temperature
- Heat of melting

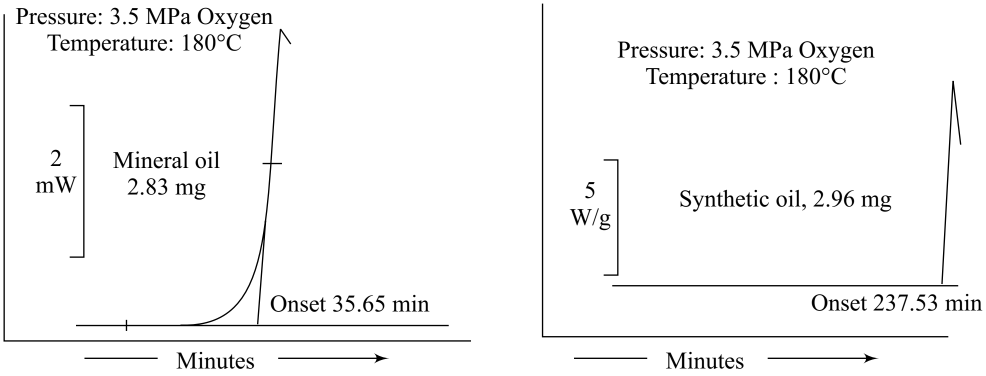


Figure 21.59 *Oxidation of induction of oils.*

- Percent crystallinity
- T_g or softening
- Crystallization
- Presence of recyclates/regrinds
- Plasticizers
- Polymer blends (presence, composition and compatibility)

The DSC studies can also be extended to the study of biological molecules.

1. Differential scanning calorimetry (DSC) is unsurpassed for understanding the stability of biological systems.
2. DSC directly measures heat changes that occur in biomolecules during controlled increase or decrease in temperature, making it possible to study materials in their native state.
3. DSC measures the enthalpy (ΔH) of unfolding due to heat denaturation. A biomolecule in solution is in equilibrium between the native (folded) conformation and its denatured (unfolded) state. The higher the thermal transition midpoint (T_m), when 50% of the biomolecules are unfolded, the more stable the molecule.
4. DSC is also used to determine the change in heat capacity (ΔC_p) of denaturation (Fig. 21.60).

DSC measurements can elucidate the factors that contribute to the *folding and stability of native biomolecules*, including hydrophobic interactions, hydrogen bonding, conformational entropy and the physical environment.

Applications include

- Protein stability and folding
- Liquid biopharmaceutical formulations
- Process development
- Protein engineering
- Rank order binding
- Antibody domain studies
- Characterization of membranes, lipids, nucleic acids and micelle systems
- Assessment of the effects of structural change on a molecule's stability
- Measurement of ultra-tight molecular interactions (up to 10^{20} M^{-1})
- Assessment of biocomparability during manufacturing

The DSC is preferential to ascertain the following:

- Ideal for stability and folding studies.
- Provides insights into mechanisms of unfolding and refolding.

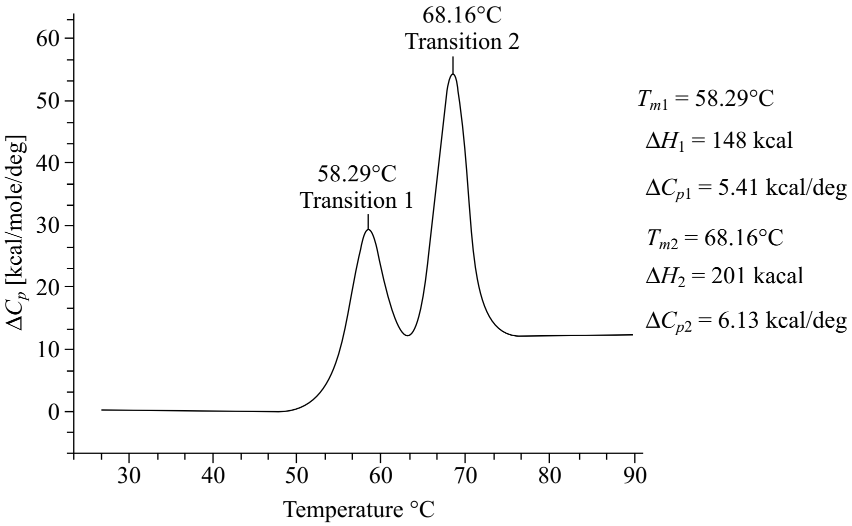


Figure 21.60 *Determination of the change in heat capacity (ΔC_p) during denaturation.*

- Monitors reversibility of thermal processes.
- Study molecules in their native state without labelling. Can be used with solutions that interfere with optical methods including turbid or collared solutions or particulate suspensions.
- Monitors conformational energetics of proteins and biopolymers.
- Continuously measures excess heat capacity.

13. Measures the stability and unfolding of a protein, lipid or nucleic acid

The evaluation of the stability and unfolding of a protein, lipid or nucleic acid can be determined through a powerful analytical technique, differential scanning calorimeter (DSC).

In DSC, the biomolecule is heated at a constant rate and there is a detectable heat change associated with thermal denaturation.

The DSC experiment can provide the following information simultaneously

- Transition midpoint: T_m
- Enthalpy (ΔH) and heat capacity change (ΔC_p) associated with unfolding.

A biomolecule in aqueous solution is in equilibrium between the native (folded) conformation and its denatured (unfolded) conformation. The native (folded) state is always stable which depends on the magnitude of the Gibbs free energy (ΔG) of the system and also due to thermodynamic relationships between enthalpy (ΔH) and entropy (ΔS) changes of the system. A positive ΔG indicates that the native state is more stable than the denatured state—the more positive the ΔG, the greater the stability. For a protein to unfold, stabilizing forces need to be broken. Conformational entropy overcomes stabilizing forces allowing the protein to unfold at temperatures where entropy becomes dominant.

ΔH of unfolding due to heat of denaturation (unfolded configuration) can be measured by the DSC experiment. The transition midpoint T_m is the temperature where 50% of the protein is in its native conformation and the other 50% is denatured. The higher the T_m, the more stable the molecule. During the same experiment, DSC also measures the change in heat capacity (ΔC_p) for denaturation. Heat capacity changes associated with protein unfolding are primarily due to changes in hydration of side chains that were buried in the native state, but become solvent exposed in the denatured state.

Many factors are responsible for the folding and stability of native biopolymers, including hydrophobic interactions, hydrogen bonding, conformational entropy and the physical environment (pH, buffer, ionic strength, excipients, etc.).

14. Determination of crystallinity of a polymer by DSC measurement

Study of the thermal transitions such as glass transition (T_g), melting (T_{m}) and crystallization transition (T_{c}) of polymers are investigated by a differential scanning calorimetric technique.

(a) Heat capacity

In the DSC measurement, the sample and the reference are heated uniformly using two separate heaters maintaining a particular heating rate in an inert atmosphere, and *when the sample undergoes a physical transformation such as phase transitions, more (or less) heat will need to flow to it than the reference to maintain both at the same temperature.*

The DSC thermogram was obtained by plotting the heat flow or absorbed by the sample against temperature as shown in Fig. 21.61.

The measurement of heat flow into (or out of) a sample at a given temperature will help us to understand the basic concepts. The *heat flow* to the sample (absorbed) is in units of heat (q) supplied per unit time (t). The *heating rate* is temperature increase (ΔT) per unit time, 't'.

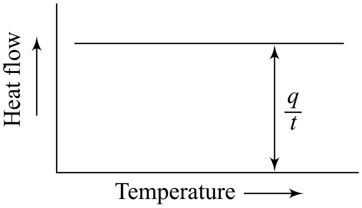


FIGURE 21.61 *Heat flow or absorbed by the sample against temperature.*

$$\text{Heat flow} = \frac{\text{Heat}}{\text{time}} = \frac{q}{t}$$

$$\text{Heating Rate} = \frac{\text{Temperature increase}}{t} = \frac{\Delta T}{t}$$

If heat flow and the heating rate of the sample are known by DSC experiment, then it is possible to calculate the heat capacity (C_p) of a sample. Therefore

$$\text{Heat capacity } (C_p) \text{ of a sample} = \frac{\text{Heat flow}}{\text{Heating rate}} = \frac{\left(\frac{q}{t}\right)}{\left(\frac{\Delta T}{t}\right)} = \frac{q}{\Delta T} = C_p$$

(b) Calculation of crystallinity of a polymer sample

DSC results indicate the percentage of crystallinity and amorphous nature of polymers in the sample. Many polymers contain both amorphous and crystalline materials. If the latent heat of melting, ΔH_{m} of a material sample is known, percentage of crystallinity of the sample can be calculated. The DSc thermogram of the sample containing both amorphous and crystalline materials is recorded. The 'peak area' of melting (endothermic) of the polymer is measured from the DSC curve. The thermogram of the sample is a plot of 'heat flow per gram of material' versus temperature. 'Heat flow' is the amount of heat absorbed per second. The area of the endothermic peak is given by units of heat × temperature × time^{-1} × mass^{-1}.

$$\text{Peak area} = \frac{\text{Heat} \times \text{Temperature}}{\text{Time} \times \text{mass}} = \frac{\text{JK}}{\text{sec} \times \text{g}}$$

Or joules × Kelvin/(seconds) × (grams).

From DSC results, divide the area of the peak by the heating rate during the experiment for the sample. The heating rate is in units of K/second.

On simplification of expression, we have

$$\frac{\text{Peak area}}{\text{Heating rate}} = \frac{\dfrac{\text{JK}}{\text{Sec}\cdot\text{g}}}{\dfrac{\text{K}}{\text{Sec}}} = \frac{\text{J}}{\text{g}} = \text{J/g} \;\; \text{(No. of joules/g of sample)}$$

If mass (g) of the sample material is known, the total heat absorbed during melting of the sample is given as

Total heat absorbed by the sample = $\frac{J}{g} \times g = \text{J}$

Similarly, the total heat absorbed during crystallization of the sample of the polymer is also calculated.

Let us call the total heat absorbed during melting as $(H_\text{m})_\text{total}$ and that for the total heat absorbed for crystallization of material as $(H_\text{c})_\text{total}$

The difference between the total of the two enthalpy changes of processes is represented as

$$\text{Heat absorbed by a portion of crystalline polymer sample} = H_{m(\text{total})} - H_{c(\text{total})} = H'$$

This value H' obtained is divided by the specific heat of melting, H_c^* (the amount of heat absorbed by one gram of a polymer) to get the percentage of crystallinity of a polymer sample.

The units are: H' is in joules, and the specific heat of melting (H_c^*) is in joule/g.

The total amount, m_c (in grams) of crystalline polymer below T_c is

$$\frac{H'}{H_m^*} = m_c \; \frac{\text{J}}{\dfrac{\text{J}}{\text{g}}} = \text{g}$$

Total amount (in grams) of crystalline polymer = $\frac{H'}{H_m^*} = m_c$

$$= \frac{\text{J}}{\dfrac{\text{J}}{\text{g}}} = \text{g}$$

To get the fraction of the sample as crystalline in the total sample, (m_c) divide by the total amount of crystalline polymer calculated, m_total.

Crystalline fraction in the total polymer sample = $\frac{m_c}{m_\text{total}}$

Therefore,

$$\text{Percentage (\%) of crystallinity} = \text{crystalline fraction} \times 100$$

Review Questions

1. What are the different techniques of thermal analysis of materials?
2. What is the basic principle of thermogravimetric (TGA) analysis?
3. Explain thermogravimetry (TGA) analysis.
4. Explain the mass loss with increase of temperature of a material with a help of a TG curve.
5. Why is DTG technique useful along with TG analysis?
6. Explain the instrumentation technique of TGA.
7. What are the sources of error in TG analysis?
8. How do you calibrate TGA unit?
9. What are the factors affecting TG analysis?
10. Mention the factors affecting the T_g curves.

11. Mention the applications of TGA.
12. Write a note on quantitative interpretation of TGA of the decomposition of $CaCO_3$.
13. Give the quantitative interpretation of TGA of the decomposition of calcium oxalate.
14. Write a note on the thermal decomposition of nickel oxinate.
15. Write a note on thermal dehydration of copper sulphate pentahydrate by TGA.
16. Write a note on the TG analysis of rubber.
17. Write a note on TG analysis of inorganic and organic mixtures.
18. Give TG analysis of a mixture of calcium and magnesium carbonates.
19. Write a note on the TG analysis of polymers.
20. What is the basic principle of differential thermal analysis (DTA)?
21. Explain DTA technique.
22. Give DTA design and experimental details.
23. Write a note on the characteristics of DTA curves and explain.
24. How is differential thermal analyser calibrated for enthalpy measurement?
25. Explain the instrumentation details of DTA set-up.
26. What are the factors affecting DTA curves?
27. Mention the sources of error in DTA measurement.
28. Interpret the decomposition of calcium oxalate in DTA curve.
29. Interpret the decomposition of calcium oxinate sample.
30. Interpret the phase transition studies of copper pyrovanadate sample.
31. Interpret the phase transition studies of copper othovanadate sample.
32. How do you study the composition and heat capacity of the sample from the position and shapes of the peaks of DTA curve.
33. DTA thermograms (or peaks) are the finger prints of pure sample. Justify.
34. Discuss the DTA of polymeric materials.
35. How do you study the glass transition (T_g) temperature of a sample by DTA technique.
36. What is the basic principle of differential scanning calorimeter?
37. What are the applications of DSC?
38. Illustrate with figure the DSC curve of heating and cooling of a sample.
39. What are the basic types of methods adopted for DSC studies?
40. Explain the DSC method based on the principle power compensation.
41. Explain the DSC method based on the principle of maintaining heat flux.
42. How do you calibrate DSC instrument? Give reasons.
43. How do you determine the enthalpy change or heat capacity of materials?
44. How do you determine the peak area and enthalpy change of the sample?
45. What are the factors affecting the DSC curve?
46. How do you justify that DSC is a fingerprint of a pure compound?
47. Write a note on DSC studies of epoxy mixtures.
48. Explain DSC curves of copper pyrovanadate and copper orthovanadate.
49. How can phase transitions in liquid crystals be studied by DSC method?
50. How can structural transformation studies be done by DSC method?
51. How do you explain the phase transition of silicon rubber in DSC?
52. How do you observe magnetic transition in DSC curve? Explain.
53. How can T_g of polymeric material be observed through DSC curve? Describe the DSC diagram of the sample.
54. How do you execute DSC analysis for explosive materials?
55. How are analysis of lubricants done through DSC curve? Explain.
56. How do you characterize the thermophysical properties of polymer by DSC.
57. How do you evaluate the stability and unfolding of a protein, lipid or nucleic acid by DSC studies?
58. Describe how the heat capacity and crystallinity of a polymer sample determined.

22

Dynamic Electroanalytical Methods

Chapter Outline

Introduction to dynamic electroanalytical methods. Fundamentals of bulk electrolysis, Ohmic potential: IR drop, polarization. Bulk electrolysis and Faraday's laws. Electrogravimetry—types of electrogravimetric methods, constant-current electrogravimetry, constant potential electrogravimetry. Physical characteristics of metal deposits. Electrodes, applications of electrogravimetry, determination of copper by electrogravimetry, the electrogravimetric, determination of copper in alloy, analysis of copper in a mixture by electrogravimetry. Coulometry: introduction, types of coulometers, electrochemical coulometers. Types of coulometric methods, constant current coulometry—coulometric titration (or amperostatic coulometry), maintaining current efficiency, determination of end point of electrolysis, construction, cell assembly, external generation of titrant, detection of end points in coulometric titrations. Coulometric titrations—acid–base titration, precipitation and complex formation titrations, Redox coulometric titrations, Karl-Fischer reaction, determination of film thickness of metallic coating, coulometric determination of arsenic by electrogenerated bromine and dead-stop end-point detection, comparison of coulometric and volumetric titration, applications. Advantages of the coulometric method. Controlled potential coulometric analysis—(potentiostatic coulometry)—instrumentation, applications, advantages, disadvantage.

22.1 Introduction

Electroanalytical methods are a class of techniques in analytical chemistry which study an analyte by measuring the potential (volts) and/or current (amperes) in electrochemical cell containing the analyte solution under investigation. Electroanalytical techniques can be generally classified into the following types depending on the specified conditions of experimentation. Furthermore, these methods can be broken down into several categories (Fig. 22.1), as described below, depending on which aspects of the cell are controlled and measured.

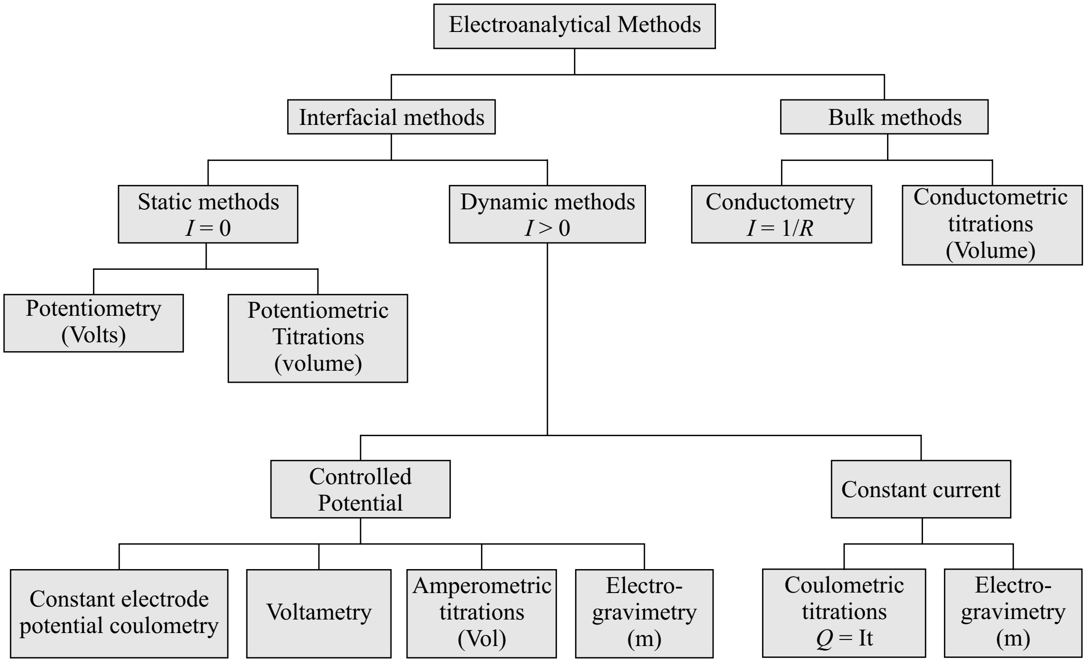


Figure 22.1 *A classification of dynamic electroanalytical methods.*

In electrochemical techniques, a clear distinction is made between 'bulk techniques' and 'interfacial techniques'. In 'bulk techniques', a property of the solution is measured in the electrochemical cell; while in the 'interfacial techniques', a potential or current is measured which depends on the species present at the interface between an electrode and the analyte solution. The electroanalytical interfacial methods are carried at static ($I = 0$) and dynamic ($I > 0$) conditions to get information of analyte solutions. In static potentiometry ($I = 0$), the difference in electrode potentials in volts is measured and while in potentiometric titrations volume (ml) of titrant needed at neutralization points is noted; and as such, on principle, these are two different electroanalytical methods. It is of interest to note that dynamic ($I > 0$) electroanalytical studies are also carried out at constant potential (volts) and at constant current (I) conditions for analyte.

In a potentiometric method of analysis, the determination of an analyte's concentration is carried out by measuring the potential of an electrochemical cell under 'static' conditions. 'Dynamic techniques', in which current passes through the electrochemical cell, are important for electrochemical methods of analysis. In this chapter, the subject of electroanalytical techniques is confined only to two topics, that is, 'electrogravimetry' (at constant current and constant potential) and 'coulometry' (at constant current and constant potential). Before studying the concept and applications of 'electrogravimetry' and 'coulometry', let us be familiar with few basics of electrochemical aspects of analytes under different conditions of potentials and currents.

22.1.1 Fundamentals of Bulk Electrolysis

The electrochemical cells are classified as a '*galvanic cell*' and an '*electrolytic cell*'. A galvanic or voltaic cell is a device which converts chemical energy into electrical energy. The galvanic cell usually consists of two electrolytic solutions in which two electrodes of different materials are dipped in their respective ionic salt solutions and they connected by a salt bridge. A Daniel cell is an example of a galvanic cell (Fig. 22.2).

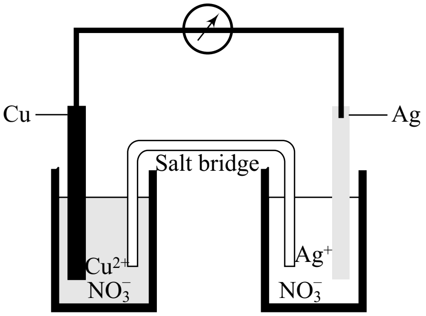


FIGURE 22.2 *An electrochemical cell.*

Consider the following electrochemical cell:
The half-cell reactions of electrodes are

$$Cu^{2+} + 2e \rightarrow Cu(s) E° = +0.34 \text{ V}$$

and

$$Ag^+ + e \rightarrow Ag(s)\ E° = +0.80 \text{ V}$$

The cell reaction is spontaneous:

Anode (oxidation): $Cu(s) \rightarrow Cu^{2+} + \cancel{2e}$; $E° = –0.34$ V

Cathode (reduction): $2Ag^+ + \cancel{2e} \rightarrow 2Ag(s)$; $E° = +0.80$ V

Cell reaction:

$$Cu(s) + 2Ag^+ \rightarrow Cu^{2+} + 2Ag(s);\ E° = 0.80 – (+0.34) = +0.46 \text{ V}$$

Electrons flow spontaneously from the negative copper anode to the positive silver cathode generating electricity. The copper anode oxidizes and silver is deposited on the silver cathode. This electrochemical cell is an example of a *galvanic* or *voltaic cell.*

When electrical energy is supplied from an external source, the current flows through the cell containing two electrodes immersed in their respective solutions. This is known as '*electrolytic cell*'. In potentiometry, reduction takes place at the cathode and is assigned a positive sign and an anode (oxidation) assigned a negative sign in a galvanic cell. However, in an *electrolytic cell*, the cathode is negatively charged while the anode is positively charge. The capacity to do electrical work by a cell is called the cell potential (E) and is expressed in volt.

Consider a cell of the type where copper is deposited at the cathode and oxygen is evolved at the anode.

$$\underset{\text{(Cathode)}}{Cu \mid Cu^{2+}} \parallel \underset{\text{(Anode)}}{2H^+, \tfrac{1}{2}O_2 \mid Pt}$$

$$\begin{array}{lll}
\text{Cathode} & Cu^{++} + 2e \rightarrow Cu° & E_{\text{cathode}} = +0.337 \text{ V} \\
\text{Anode} & H_2O \rightarrow 2H^+ + \tfrac{1}{2}O_2 & E_{\text{anode}} = –1.229 \text{ V} \\
\hline
 & Cu^{++} + H_2O \rightarrow 2H^+ + \tfrac{1}{2}O_2 + Cu° & E_{\text{cell}} = E_c – E_a = 0.337 – (+1.229) = – 0.892 \text{ V.}
\end{array}$$

The potential of the electrochemical cell, (E_{cell}), is the difference between the electrode potential of the cathode and the electrode potential of the anode. That is

$$E_{\text{cell}} = E_{\text{cathode}} – E_{\text{anode}}$$

where E_{cathode} and E_{anode} are the half-cell potentials of the cathode and anode, respectively.

Consider the electrolytic cell shown in Fig. 22.3.

A voltage E_{applied} is applied to the cell in such a way that a *current flows through the cell*. When $E_{\text{applied}} > E_{\text{cell}}$, there will be a flow of current in the circuit. When there is *current*, the potential of the cell is less than the thermodynamic potential because one of the following phenomena is operating:

- *IR* drop
- concentration polarization and
- kinetic (chemical) polarization

(i) Ohmic potential: IR drop

In electrochemical cells, the current flowing through the cell encounters resistance (R), like the metallic conductors. Ohm's law describes the effect of this resistance to conduction of any conductors. For electrolytic conductors, cell develops Ohmic potential ($-IR$) which is described as the product of current (I) and the resistance (R) or it is known as the *IR drop of the cell*.

It can be remarked that when no current flows ($I = 0$) through the cell, it is observed that external applied potential is equal to the emf of the electrolytic cell ($E_{\text{applied}} = E_{\text{cell}}$). When the applied potential is gradually increased, a small current appears in the circuit. This current through the cell encounters resistance (R) resulting in a potential drop of $-IR$ volts. In other words, *the applied potential must be greater than the theoretical cell potential by $-IR$ volts*. Thus, in the presence of a current, a cell potential must be modified by the addition of the term $-IR$.

$$E_{\text{applied}} = E_{\text{cell}} - IR \qquad \ldots(1)$$

$$E_{\text{applied}} = (E_{\text{cathode}} - E_{\text{anode}}) - IR \qquad \ldots(2)$$

where E_{cathode} and $E_{\text{anode}} = E_{\text{cell}}$ are electrode potentials obtained from the Nernst equation.

The above equation can be rearranged to give

$$I = \frac{-E_{\text{applied}}}{R} + \frac{(E_{\text{cathode}} - E_{\text{anode}})}{R}$$

$$I = \frac{E_{\text{cell}} - E_{\text{applied}}}{R} \qquad \ldots(3)$$

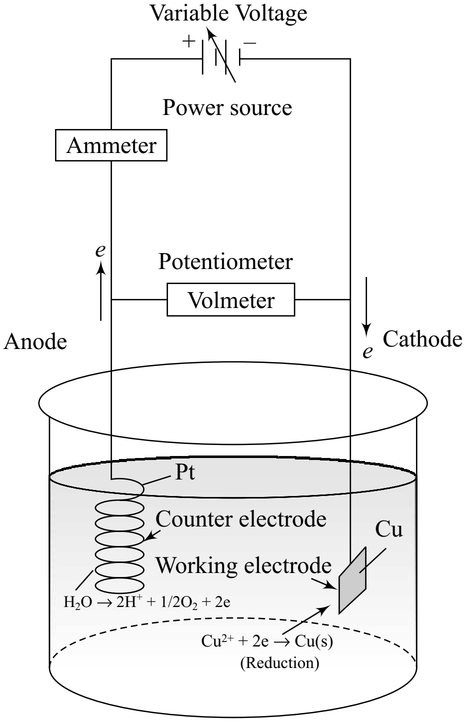


FIGURE 22.3 *An electrolytic cell.*

For short period of time and small current, E_{cathode} and E_{anode} remain relatively constant during electrolysis. The cell behaviour can be represented by the reaction.

$$I = \frac{-E_{\text{applied}}}{R} + k \qquad \ldots(4)$$

where k is a constant.

As shown in Fig. 22.4, a plot of current as a function of applied potential in an electrolytic cell should be a straight line with a slope equal to the negative reciprocal of the resistance $\left(-\frac{1}{R}\right)$. The plot is indeed linear with small currents as in Fig. 22.4. As the applied voltage increases, the current deviates significantly from linearity. Galvanic cells also behave in a similar way Fig. 22.5(a) and (b).

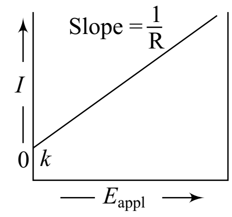


FIGURE 22.4 *A plot of current versus potential.*

Under the above conditions, cells are said to be '*polarized*' when they exhibit nonlinear relationship. The degree of polarization at the respective electrodes is expressed as '*overvoltage*' or '*overpotential*'. Polarization in an electrolytic cell demands the application of a potential greater than the theoretical value to give a current of the expected magnitude. Thus, an overpotential of –0.04 V is required to obtain a current of 0.06 A in the electrolytic cell as illustrated in

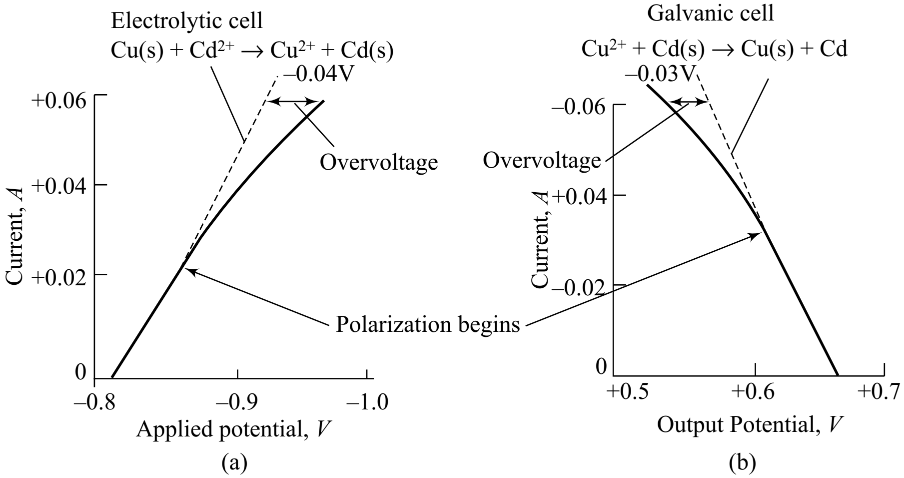


FIGURE 22.5 *Current/voltage curves for (a) an electrolytic and (b) a galvanic cell.*

Fig. 22.5(a). However in a galvanic cell, the output cell potential decreases by about 0.03 V, i.e. the overvoltage is 0.03 V as illustrated in Fig. 21.5(b). The overvoltage is negative in each of the above cases. Hence, for an electrolytic cell affected by overvoltage, Eq. (1) then becomes,

$$E_{\text{applied}} = E_{\text{cell}} - IR - \text{overvoltage} \qquad \ldots(5)$$

In simple terms, *the overvoltage is the potential difference between the theoretical cell potential determined by Eq. (1) and the actual cell potential at a given level of current.* On the other hand, polarization is the departure of the electrode potential from its theoretical value on passage of current.

(ii) Polarization

The polarization at an electrode of an electrolytic cell is illustrated in Fig. 22.4(a). On further increasing E_{applied}, the current becomes independent of E_{applied}. This gives rise to *limiting current* as shown in Fig. 22.6. After this point, *the electrode is said to be completely polarized, since its potential can be changed widely without affecting the current.*

Polarization at an electrode can be conveniently divided into two categories:

- Concentration polarization and
- Kinetic (chemical) polarization.

Polarization is an electrode phenomenon that may affect either or both of the electrodes in an electrolytic cell.

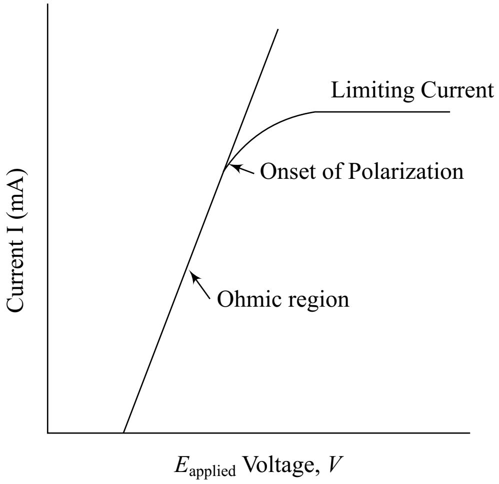


FIGURE 22.6 *Current potential curve for electrolyte illustrating Ohmic region (linear portion of the curve) and limiting current situation.*

(a) Concentration polarization

As a result of passage of current in an electrolytic cell, the ions are discharged and deposited on an electrode during the process of electrolysis, and there is depletion of ions in the immediate vicinity of the electrode. This imbalance of ions around vicinity layer of the electrode offers resistance to the passage of current. This type of *resistance due to changes in* *concentration of the electrolyte around the electrode is known as 'concentration polarization'*. Concentration polarization occurs during electrolysis when the ions do not arrive at the cathode or the product species do not leave the anode fast enough to maintain the desired current.

Generally, electrolytic reactants from the bulk solution are transported to an electrode surface by three mechanisms: (1) diffusion, (2) migration and (3) convection. Concentration polarization arises when the effects of diffusion, migration and convection are insufficient to transport the ions at a rate that produces a current of desired magnitude.

If a situation created with a depletion of ions in the vicinity of the electrode even after proper agitation of bulk solution, the cathode will experience more and more negative potential as the applied potential is increased. As a result, a back emf is set up and if the current has to flow, the applied emf must be sufficiently above the equilibrium potential to overcome the polarization. Hence, concentration polarization requires applied potentials that are larger than the theoretical, i.e. calculated from Eq. (1), to maintain a given current in an electrolytic cell. For this reason, vigorous stirring and heating are important in electrodeposition to minimize the concentration polarization.

(b) Kinetic polarization (chemical polarization)

It is a common experience to note that kinetic polarization arises due to the chemical nature of the electrode position of any metal on the surface of the electrode material in an electrolytic process. This phenomenon also could be seen due to the coating of the electrode with a layer of gas such as hydrogen or oxygen in an electrolytic cell. This leads to *kinetic or chemical polarization at the electrode* in an electrolytic process. Kinetic polarization may also be caused by the formation of films on the electrode. For example, an aluminium anode is liable to be passivated by a coat of alumina on it. Kinetic polarization is most pronounced in electrode processes that yield gaseous products. This is known as gas overpotential.

Gas overpotential depends on several factors as follows

- Electrode material
- Current density, temperature
- pH of solution, etc.

Overpotential for many metals differ. The overvoltage at a given current density decreases if the electrode surface is roughened. This is due to an increase in effective surface area and the consequent decrease in actual current density. An increase in current density invariably increases the overpotential to a limiting value. As the temperature increases, the overvoltage decreases. pH of the solution also has an effect on overpotential. Hydrogen overvoltage is greatest in metals such as cadmium, lead, tin, zinc and mercury.

One of the essential requirements of electrode position is that the deposit must be compact, fine and adherent to the electrode. Simultaneous evolution of gas results in a spongy and feathery deposit. When high current densities are applied, the applied potential exceeds the hydrogen overvoltage and gasification starts. Gasification can be avoided by controlling the potential and by using suitable buffers. Complex ions yield smoother deposits. For example, silver when deposited from a $[Ag(CN)_2]^-$ complex yields a satisfactory deposit.

Importance of overvoltage:

- The lead (Pb) metal is deposited on the cathode instead of hydrogen being evolved during charging in a lead accumulator due to the high overvoltage of lead.
- It is possible to deposit electrolytically the metals, which have a more negative potential than hydrogen such as zinc, cadmium, tin, etc. from an acid solution because of high overvoltage of hydrogen on such metals.
- Overvoltage plays an important role in the industrial production of chlorine and sodium hydroxide by the electrolysis of brine–sodium chloride solution.
- When electrodes of high overvoltage are used in electrolytic reduction, it would mean to increase the reducing power. Electrodes with high overvoltage like lead instead of platinum are used in organic electrolytic reductions to obtain the desired products.

22.1.2 Bulk Electrolysis and Faraday's Laws

(a) Faraday's first law

'*The mass of a substance liberated at the electrodes during electrolysis is directly proportional to the quantity of electrical charge (Q) that passed through the electrolyte.*'

If 'm' is the mass of a substance deposited by a current of 'I' amperes in 't' seconds, then according to the first law of electrolysis

$$m \propto Q \text{ or } m \propto I \times t$$

or $$m = Z \times I \times t \qquad \text{...(6)}$$

where Z is the proportionally constant.

Thus, the first law relates *the quantity of current passed and the extent of chemical change that took place.*

(b) Faraday's Second law of electrolysis

'*The mass of different substances evolved/deposited by the passage of same quantity of electricity is proportional to their chemical equivalence.*'

$$m \propto M_r/n = \text{equivalent weight of substance}$$

where 'm' is the mass of the substance liberated or deposited, M_r is the molar mass of the substance liberated/deposited and 'n' is the valency number of ions of the substance (electrons transferred per ion).

If m_1 and m_2 are the mass of two different substances liberated by the passage of same amount of electricity and M_1 and M_2 are their respective molar masses and n_1 and n_2 are their respective valence number of ions, then

$$\frac{m_1 n_1}{M_1} = \frac{m_2 n_2}{M_2} \qquad \text{...(7)}$$

For example, let us consider the passage of electricity through solutions of $CuSO_4$ and $AgNO_3$ connected in series, so that the same amount of electricity is passed through them, then from the second law, we get

$$\frac{\text{mass of copper deposited} \times 2}{\text{Molar mass of copper}} = \frac{\text{mass of silver deposited} \times 1}{\text{molar mass of silver}}$$

(c) Units of electricity

(i) 'Coulomb' is the unit of electricity.

The quantity of charge that is transported in one second by a constant current of 1A is known as one coulomb (C).

The product of 'current strength' or charge and 'time' for which the current is passed is equal to the quantity of electricity or charge (Q).

$$Q = \text{current} \times \text{time}$$

$$= I \times t \text{ or } Q = f_0^t \cdot I\,dt \qquad \text{...(8)}$$

The quantity of electricity (Q) in coulombs passed through an electrolytic cell when 2 A flows for 10 seconds are 20 coulombs.

$$Q = \text{ampere } (I) \times \text{seconds } (t) = 2 \times 10 = 20 \text{ C.}$$

(ii) *The quantity of electricity carried by one mole of electrons is called Faraday (F).*

The charge of an electron is 1.6022×10^{-19}. Hence, the charge on one mole of electrons would be

$$1.6022 \times 10^{-19} \times 6.022 \times 10^{23} \text{ C} = 96485 \text{ C.}$$

One mole of electrons (6.022×10^{23}) is equal to 96,485 coulombs of charge, and is called a *faraday*.

Now consider a reaction

$$M^{n+} + ne \rightarrow M \qquad \ldots(9)$$

The quantity of electricity (Q) required for the deposition of 1 mole of an element will be equal to *nF*, where '*n*' is the number of moles of electrons required for the reaction. Let '*M*' be the atomic mass of one mole atom of an element.

$$w = \frac{QM}{nF}$$

where w = weight of substance oxidized or reduced

M = Molecular weight of the substance

Q = number of coulombs passing through the cell.

No of moles (n) of electrons:

$$\text{Moles } (n) \text{ of } e = \frac{I \times t}{F} = \frac{\text{Coulombs}}{\text{Coloumbs/mole}}$$

Example If $I = 0.05$ A and time, $t = 5$ minutes, how much silver electrode is oxidized?

No. of moles of electron flowing through the cell

$$\text{No. of moles } (n) \text{ of } e = \frac{I \times t}{F} = \frac{0.05\left(\frac{C}{\text{sec}}\right) \cdot 5 \times 60\left(\text{min} \times \frac{\text{Sec}}{\text{min}}\right)}{96{,}500 \text{ mol}^{-1}}$$

$$= 1.55 \times 10^{-4} \text{ mole electrons}$$

22.2 Electrogravimetry

The two dynamic electroanalytical techniques are based on Faraday's statements; they are electrogravimetry and coulometry. Electrogravimetric and conventional gravimetric analyses are more or less similar. The material is deposited on an electrode by the application of a '*potential*' instead of chemical precipitation from a solution. Electrogravimetry involves the reduction of a solution of a metal salt under controlled potential in an electrolytic cell till the current falls near zero; and the metal is deposited onto working electrode (either the cathode or the anode).

In electrogravimetry, a metal is deposited quantitatively on an electrode by an electrolytic reaction and the amount of the metal is determined by weighing the electrode before and after electrolysis.

22.2.1 Types of Electrogravimetric Methods

The deposition and quantitative determination of metals by electrogravimetric technique has been widely used. In principle, for most applications, the metal is deposited on a weighed platinum cathode and the increase in mass is determined.

There are two types of electrogravimetric methods as follows:

(i) Constant current electrolysis and

(ii) Constant potential electrolysis.

(a) Constant current electrogravimetry

In this method of constant current electrogravimetry, *a quantitative electrodeposition is carried out at the electrode (WE) by maintaining the current constant*. During the process of electrodeposition, a periodic increase of applied potential for an analyte solution is required. The set-up for constant current electrolysis consists of a suitable electrolytic cell and a direct current variable DC source (Fig. 22.7). A variable power source used in the assembly is a 6–12 V storage battery. An ammeter and a voltmeter are used to indicate the current (A) and applied voltage (V), respectively. The resistor (R) in the circuit controls the applied voltage (E_{applied}) to the cell. A cylindrical platinum gauze is used as a cathode of the electrolytic cell. The metal is deposited from an analyte solution onto an electrode by the application of a potential from a variable DC power supply. The electrode reaction of interest occurs at the electrode is called 'working electrode' (WE).

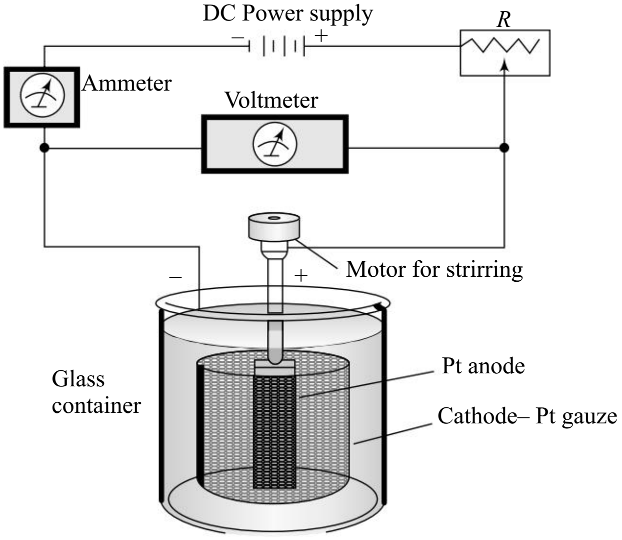


Figure 22.7 *Apparatus for constant current electrolysis.*

In electrogravimetry, a metal to be determined is quantitatively electrodeposited onto an electrode (WE) and, from the gain in mass of the electrode, the amount of metal in the sample is calculated, assuming no interferants are also electroplated onto the electrode.

(b) Constant potential electrogravimetry

Utilizing *controlled-potential electrolysis assembly*, it is possible to separate two elements whose deposition potentials differ sufficiently (by a few tenths of a volt). The potential of the *cathode is controlled* during the constant potential electrolysis so that it never becomes sufficiently negative to allow the deposition of the next element. It may be observed that due to concentration polarization, the potential of the cathode becomes negative and that code position of the other species begins before the analyte is completely deposited. The advantage of using a three-electrode system is that the assembly avoids a large negative drift in the cathode potential.

A 'potentiostat' is used to control and maintain the working electrode (cathode) at a constant potential during the process of electrolysis of an analyte solution. The current in the reference electrode (RE) is essentially zero ($I = 0$) at all times during the process.

The controlled-potential electrolysis assembly is shown in Fig. 22.8. In the assembly unit for electrolysis, there are two independent electrode circuits that share a common electrode, *the working electrode (cathode) at which the analyte is deposited.*

- The *electrolysis circuit* consists of a potentiostat which permits continuous variation in the potential applied across the working electrode (WE), a counter electrode (CE)—generally platinum, and a digital milliammeter.

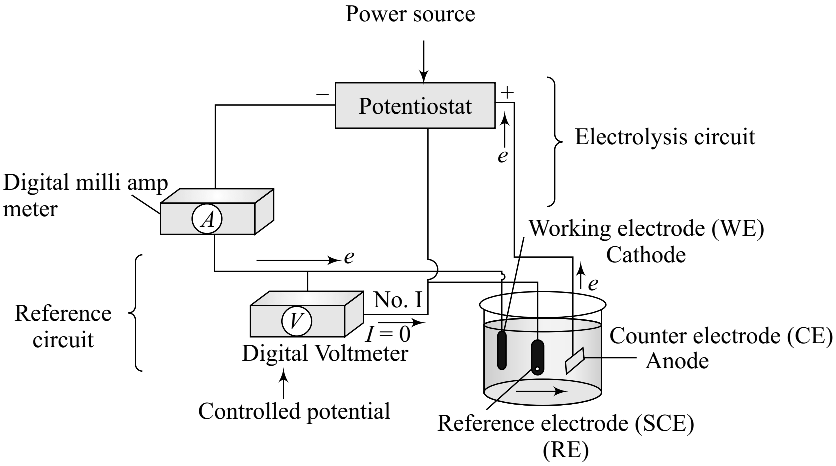


FIGURE 22.8 *An assembly for controlled-potential electrolysis.*

The *reference circuit* consists of a reference electrode (SCE), a high resistance digital voltmeter and the working electrode (WE). The purpose of the reference circuit is to *monitor continuously the potential between the working electrode and the reference electrode.*

The high impedance reference circuit monitors the E_{cell}, while the electrolysis circuit supplies essentially all the current (I) for the deposition of metal at the working electrode (WE). The current in the reference electrode circuit is essentially zero ($I = 0$) at all times. The feedback circuitry of potentiostat varies E_{applied} to maintain constant E_{cathode}.

The potentiostat controls the electrolysis potential for selectivity of electrolysis and to hold the *potential between the cathode (WE) and a reference electrode (SCE) constant*. When there is no flow of current ($I = 0$) through the reference electrode of the assembly, this (RE) is not subjected to IR drop, concentration polarization or overvoltage. The equilibrium of reference electrode (SCE) is not disturbed so that E_{RE} *remains constant*. When the potential *reaches a point* at which codeposition of an interfering species is about to begin, the potential across the working (WE) and counter electrode (CE) is decreased using potentiostat (Fig. 22.7). Since the potential of the *counter electrode (CE) remains constant* during the change, the *cathode potential becomes smaller*, thus preventing codeposition. To avoid codeposition of metals with higher reduction potentials, E_{cell} should be held constant by varying E_{applied} as IR and Π change. This is achieved with a three-electrode arrangement in which the working electrode potential (cathode) is continuously monitored and this signal is used via a feedback circuit to maintain constant E_{cell}.

The changes in cell potential (E_{cell}) and current (I) during a controlled-cathode potential electrolysis of copper is shown in Fig. 22.9. It can be noted that the applied cell potential (E_{applied}) *decreases* continuously throughout the electrolysis of copper. This *decrease, in turn, diminishes*

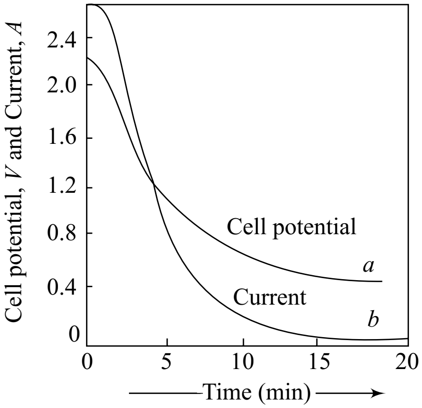


FIGURE 22.9 *Change in cell potential (a) and current (b), during a controlled-cathode potential deposition of copper. The cathode is maintained at –0.36 V versus SCE throughout the experiment.*

the current. When the current in the circuit approaches zero, electrolysis of copper is complete. A constant attention is desired throughout the experimentation.

Cathode potential change during the electrolysis of Cu (II) ion in acid medium at constant applied potential:

The electrolysis of Cu(II) ion in an acid medium is carried out at *constant applied potential*. Initially, the applied potential (E_{applied}) to the cell (at point, A) is about –0.25 V, which leads to a current of about 1.5 A, as shown in Fig. 22.10(a). The electrolytic deposition of copper at the cathode (WE) is then completed at this applied potential (about –0.25 V). As the process proceeds, it is noted that *IR* decreases continuously, Fig. 22.10(b). This decrease is due to *concentration polarization* at the cathode, which limits the rate at which copper ions are brought to the electrode surface and thus limits *the current. The decrease in IR must be compensated by an increase in the cathode potential* (E_{cathode}) *since the cell potential* (E_{cell}) *is constant.* The increase in cathode potential is slowed down at point B by the reduction of hydrogen ions, Fig.22.9(b). Since the solution contains a large excess of acid, the current is no longer limited by concentration polarization, and codeposition of copper and hydrogen goes on simultaneously (beyond B and C) until all the copper ions are deposited. Under these conditions, the *cathode is said to be depolarized* by hydrogen ions.

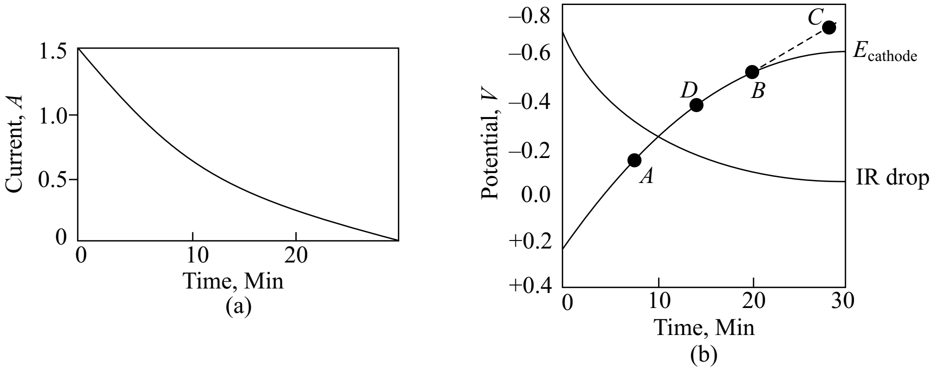


Figure 22.10 *(a) Current and (b) IR drop and cathode-potential change during the electrolytic deposition of copper at a constant applied cell potential.*

22.2.2 Physical Characteristics of Metal Deposits

It is noted that the best form of deposits of metals obtained by electrolysis has the following properties: *pure, strongly adherent, dense and smooth so that it can be washed, dried and weighed without any loss*. Spongy, powdery or flaky deposits are likely to be less pure and less adherent.

The principal factors that influence the physical characteristics of deposits are as follows:

- Current density,
- Temperature, and
- *The presence of complexing agents.*

If the current densities are less than 0.1 A cm^{-2}, best deposits may be obtained by electrolytic process. The quality of deposits can be improved generally by stirring the analyte. The time of deposition is reduced when the solution is stirred vigorously or if the electrode is rotated at a uniform speed. Stirring of analyte solution lowers the concentration overpotential and enables a higher current density without any adverse effect. Avoid evolution of gas at an electrode during deposition, since continuous evolution of bubbles on the electrode surface disturbs the orderly growth of the crystal structure of a metal deposit. Porous and spongy deposits are obtained if there is continuous evolution of gases.

The physical form of the deposit of metal depends on the chemical nature of the ion in solution.

For example, the electrolysis of Cu^{+2} ions in nitric acid solution yields a bright and adherent deposit of copper on the electrode (WE). In contrast, a coarse tree-like deposit of silver is got under similar conditions. However, a suitable, good and adherent deposit of Ag can be obtained using solution of $-Ag[CN]_2$ complex during the electrodeposit process. It is desirable to use complex ions in analyte solution, since 'throwing power'—a property of this complex solution, by virtue of which a relatively uniform deposit of metal is obtained on irregular surfaces. Even, the cyanide and ammonia complexes often are found to provide the best deposits. Any increase in temperature of analyte solution favours diffusion. But the hydrogen overpotential is decreased and the stability of many complex ions is decreased. In practice, constant current electrolysis is limited to the separation of an easily reduced cation from cations that are more difficult to be reduced than hydrogen. Thus, Cu(II) ion can be deposited from an acid solution.

Let us consider implications of the electrodeposition of Cu^{2+} and Pb^{2+} ions from a solution containing these two ions. Lead (II) begins to deposit at point 'A' on the cathode potential curve, Fig. 22.10(b). Hence, lead (II) ion would codeposit well before copper deposition was complete and would interfere. Codeposition of hydrogen during electrolysis often leads to spongy and flaky deposits that do not adhere to the electrode. Such a situation is unsatisfactory for analytical purposes. The process of electrolysis demands the use of depolarizers which improves the efficiency of the process.

A depolarizer is a substance which gets reduced at the cathode without gasification or gets oxidised at the anode before oxygen evolution and stabilises the potential of the working electrode by minimizing concentration polarization.

Thus, copper deposited from a nitric acid solution is smoother and more adherent because nitrate ions act as cathodic depolarizers and prevent evolution of hydrogen. Generally, reducing agents like hydrazine hydrochloride will act as anodic depolarizers. The cathodic depolarizer, nitrate ion gets reduced at a potential less negative than at point C, Fig. 22.10(b) as

$$NO_3^- + 10H^+ + 8e \leftrightarrow NH_4^+ + 3H_2$$

Nitrate ions have long been used in the constant current deposition of copper and lead, wherein copper is deposited at the cathode and lead is deposited as lead dioxide at the anode. Nitrate ions aid in the quantitative deposition of cupric ions at the cathode and prevent the formation of metallic lead at the cathode. Table 22.1 illustrates the typical applications of constant current electrolysis.

Analyte	Weighed as	Cathode	Anode	Conditions
Ag^+	Ag	Pt	Pt	Alkaline cyanide solution
Cd^+	Cd	Cu on Pt	Pt	Alkaline cyanide solution
Cu^+	Cu	Pt	Pt	H_2SO_4/HNO_3 solution
Mn^{2+}	MnO_2 (on anode)	Pt	Pt dish	HCOOH/HCOONa
Ni^{2+}	Ni	Cu on Pt	Pt	Ammoniacal solution
Pb^{2+}	PbO_2 (on anode)	Pt	Pt	Strong HNO_3 solution
Zn^{2+}	Zn	Cu on Pt	Pt	Acetic citrate solution
Br^-	AgBr (on anode)	Pt	Ag	—

Table 22.1

Typical applications of constant current electrolysis

22.2.3 Electrodes

The inert platinum gauze electrodes are used in the electrolytic cell. The electrode can be ignited to remove any organic matter or gases and thus, eliminates the possible harmful effect on the physical properties of the deposit, if they are present adhering to the electrode. Certain metals cannot be deposited directly onto platinum; and metals like bismuth and zinc cause permanent damage to the electrode. Hence, a protective coating of copper is always deposited on a platinum electrode.

Mercury cathode

The mercury cathode is used preferentially because they interact and remove easily reducing metals in an electrolytic process. Figure 22.11 shows a mercury cathode. For example, copper, nickel, cobalt, silver and cadmium are readily separated at this electrode from ions in solution.

The metals that get deposited at the electrode dissolve in mercury with little hydrogen evolution because mercury has a high hydrogen overvoltage and hence no hydrogen evolution even at high applied potentials.

Advantages of mercury cathode are as follows:

1. It forms an amalgam with a number of metals
2. It has a high hydrogen overvoltage and hence no hydrogen evolution even at high potentials.

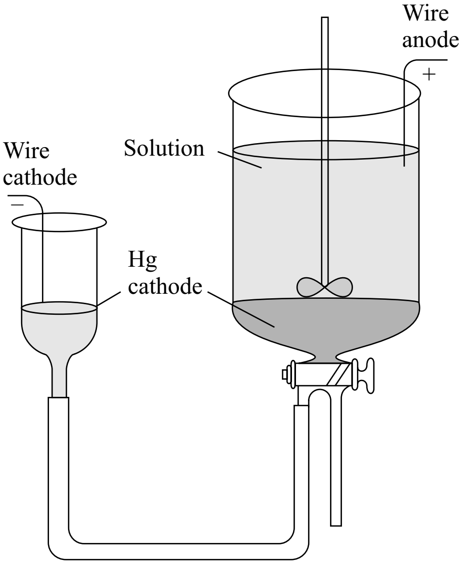


FIGURE 22.11 *A mercury cathode for the electrolytic removal of metal ions from solution.*

22.2.4 Applications of Electrogravimetry

Constant current and constant potential electrogravimetric techniques have wide range of application in quantitative identification of metals and in selective separation. Controlled-cathode potential is very useful in the separation of quantitative determination of metallic species.

The few examples of the applications of electrogravimetry are as follows:

(i) Successive deposition of the metals

Electrogravimetry is capable of very high precision and accuracy for determination of a metal. The technique involves deposition of the specific metallic element on a previously weighed cathode (Pt) by an electrolytic method. The dry electrode is weighed again after the deposition is complete. The difference in weight gives the weight of the deposited metal.

It can be noted with great success that a mixture of copper, bismuth, lead, cadmium, zinc and tin can be determined by the successive deposition of metals on a platinum cathode. First, copper can be deposited on the electrode by maintaining a constant cathode potential at –0.2 V (vs. SCE). After weighing the cathode, the copper-plated cathode is used to deposit bismuth at a constant potential of –0.4 V and weighed dry. Then, lead is deposited over the same cathode by maintaining a constant cathodic potential of –0.6 V and the dry electrode with the contents are weighed. A neutral solution containing tartarate ions is used to deposit copper, bismuth and lead onto the electrode during electrolysis, while tartarate ions form complex with tin (IV) and prevent its deposition at the cathode. When lead deposition is complete and determined, the remaining solution is made alkaline with ammonia. Cadmium and zinc are deposited successively at a constant potential of –0.2 and –1.5 V, respectively. Finally, the solution is acidified to decompose the tin–tartarate complex and tin is deposited at a potential of –0.65 V using a fresh cathode. Tables 22.2 and 22.3 list the common metals that can be determined by a constant current electrolysis method and constant potential electrolysis method, respectively.

Analyte	Weighed as	Cathode	Anode	Conditions
Ag^+	Ag	Pt	Pt	Alkaline cyanide solution
Cd^+	Cd	Cu on Pt	Pt	Alkaline cyanide solution
Cu^+	Cu	Pt	Pt	H_2SO_4/HNO_3 solution
Mn^{2+}	MnO_2 (on anode)	Pt	Pt dish	HCOOH/HCOONa
Ni^{2+}	Ni	Cu on Pt	Pt	Ammoniacal solution
Pb^{2+}	PbO_2 (on anode)	Pt	Pt	Strong HNO_3 solution
Zn^{2+}	Zn	Cu on Pt	Pt	Acetic citrate solution
Br^-	AgBr (on anode)	Pt	Ag	—

Table 22.2

The common metals that can be determined by a constant current electrolysis method

Metal	Potential vs. SCE	Electrolyte	Other elements that can be present
Ag	+0.10	Acetic/acetate buffer	Cu and heavy metals
Cu	–0.30	Tartarate + hydrazine + Cl^-	Bi, Sb, Pb, Sn, Ni, Cd, Zn,
Bi	–0.40	Tartarate + hydrazine + Cl^-	Pn, Sb, Cd, Sn

Table 22.3

The common metals that can be determined by a constant potential electrolysis method

(ii) Simultaneous depositions

A mixture of copper and lead can be determined by the deposition of copper in the presence of nitrate ions at the cathode (WE) at the electrolytic cell. Lead is deposited as lead dioxide at the anode (CE).

(iii) Preparative organic chemistry

If an organic compound can undergo a series of reductions (or oxidations) each at a *definite potential*, it is then possible to reduce the starting material to the desired product by controlling the potential of the cathode during reduction. The type of electrolytic reduction is much more economical than chemical reductions, where side reactions produce undesired products.

22.2.5 Determination of Copper by Electrogravimetry

A metal ion is quantitatively electroplated as a metal onto a preweighed working electrode (WE) in electrogravimetry. The reduction process takes place at the cathode (negative electrode) and copper is deposited. From the gain in mass of the electrode (cathode), the amount of metal in the sample can be calculated. The conditions of electrogravimetry are such that no interferants codeposit onto the electrode.

A good quantitative analysis of copper can be achieved by maintaining suitable conditions and 100% copper is deposited onto the electrode. Electrogravimetry is capable of very high precision and accuracy for determination of a metal. The measurements in the technique are limited to the clean and dry cathode electrode before and after the experiment. The weighing is readily capable of very high precision and accuracy.

There are two basic methods: controlled-potential and controlled-current electrogravimetry which can be adopted for the determination of copper.

(a) *In controlled-potential electrogravimetry*, a known *constant potential* is applied to the electrode for a sufficiently long enough time to plate out 100% of the analyte. The current (I) decreases as the metal is plated out until only the small *charging* or *background current* flows. This provides a measure of selectivity if there are two or more metals that can plate out; the metal ion that is more easily reduced can often be plated out quantitatively without any of the other metal also plating out. This procedure generally provides a nice smooth coating, but requires more sophisticated and expensive instrumentation and quite a long analysis time.

(b) *In controlled-current electrogravimetry*, a fairly *large voltage* (E_{applied}) is applied to the working electrode (WE) in order to force a constant relatively *large current* to flow through the cell. The current is often the order of milliamps (mA). If the concentration of metal ion in the electrolysis solution is insufficient to consume *all* the current, then other reactions such as hydrogen ion reduction must occur in order to consume all the current that is forced to pass through the cell. Obviously, if any interferents are present, they will also plate out; but the method is significantly faster than using a controlled-potential and the required instrumentation is simpler and cheaper. In either electrolysis variation, the solution is stirred or agitated in some manner in order to speed up the analysis.

A dilute acidified solution of copper ion is taken in the cell. A preweighed Pt electrode is taken in the electrolytic cell and copper is to be deposited on it.

$$Cu^{2+}(aq) + 2e \rightarrow Cu(s)$$

The cupric ion does not require very negative potentials to be plated out quantitatively.

Thus, most common metals do not behave as interferents in analyte solution and they do not deposit onto an electrode because of their reduction potential values compared to copper. The E_o for Cu^{2+} reduction is only +0.34 V, as opposed to the E_os for Pb, Sn(II), Ni, Cd and Zn which are –0.13, –0.14, –0.25, –0.40 and –0.76 V, respectively. In general, the more positive the E_o for a redox couple, the easier a species can be reduced. Of the more common metals that can plate out under the conditions used in this experiment, only Ag(I), Hg(I) and Hg(II) with E_os = +0.80, +0.80 and +0.85 V, respectively, would interfere.

Procedure

Preparation of electrodes

1. The larger gauzy electrode is used as a cathode of the electrolytic cell and the anode (being smaller) will rotate during electrolysis to stir the solution.
2. Platinum electrodes are cleaned using concentrated nitric acid for about few minutes.
   This will dissolve any Cu that has already been deposited on the cathode as well as removing traces of grease and other materials from the surface. The electrode (cathode) is washed and rinsed well with deionized water.
3. Wet cathode (larger outer electrode) is washed with little quantity of alcohol.
4. The electrode (cathode) is dried in an oven at about 105°C for 5 minutes. The electrode is removed from the desiccators after cooling and weigh accurately.

Electrolysis

1. Dilute the unknown copper solution carefully up to the mark of a volumetric flask and mix well.
2. Pipette out 25.00 mL of the unknown copper solution each to two beakers and add about 50 mL of deionized water, 2 mL of concentrated sulphuric acid and 1 mL of concentrated nitric acid to each beakers.
3. Connect a set of electrodes to the assembly of an electrolytic cell.
4. Lower the electrode assembly into the beaker so that the top of the cathode is about ¾ inch below the rim of the beaker. Start the stirring motor. Add sufficient deionized water to raise the level of the solution to about half an inch from the top of the cathode. Cover the beaker with a split watch glass.
5. Use 1 to 2 A of current from a DC power source and voltage is maintained below 4 V.
6. Repeat the procedure for the second sample.
7. Using this voltage, electrolyze the copper solution until the blue-green colour of copper ion has disappeared (about 15–30 minutes). Add 1 mL of urea solution and continue the electrolysis for an additional 5 minutes.
8. Add sufficient deionized water to cover the cathode completely. Make this addition with a wash bottle and use this opportunity to wash drops of solution from the surface of the split watch glass and the sides of the beaker into the bulk of the solution.
9. Continue the electrolysis for an additional 15 minutes at 0.5 A current. If no additional copper is deposited on the fresh cathode surface, the deposition is complete.
10. To stop the electrolysis, turn off the stirrer but do not interrupt the current at this time.
    Slowly raise the electrode assembly with one hand while washing the exposed portion of the cathode with a stream of deionized water from a wash bottle. As soon as the cathode is completely out of the solution, turn off the current and replace the beaker with a beaker of deionized water. Raise the beaker of deionized water to cover the electrodes and gently swirl. Wash the electrodes with a second portion of deionized water and carefully disconnect the cathode.
11. Rinse the cathode with alcohol (wash bottle) and place it in a beaker in an oven at 105°C *for no more than 5 min* to dry. Do not heat longer because the fresh copper surface oxidizes easily, the electrode will gain mass, and your results will be high.
12. Cool the electrode to room temperature in a desiccator and weigh accurately. It is important to rinse, dry, cool and weigh the plated electrode in the same manner as was the cleaned electrode previously, because the *difference* in mass will represent the total mass of copper in the 25.00 mL aliquot of unknown copper solution.

22.2.6 The Electrogravimetric Determination of Copper in Alloy

The electrogravimetric method of analysis of copper in a sample of an alloy is based on the deposition of metallic copper on a copper cathode. The hydrous oxide of tin ($SnO_2 \cdot H_2O$) formed during addition of nitric acid should be removed by filtration. Lead dioxide is deposited quantitatively at the anode from a solution with a high nitrate ion concentration, whereas copper is only partially deposited on the cathode under these conditions. High nitrate ion concentration is not a desirable factor for quantitative deposition of copper on the cathode. It is, therefore, necessary to eliminate the excess nitrate. Removal is accomplished by the addition of urea:

$$6NO_3^- + 6H^+ + 5(NH_2)_2CO \rightarrow 8N_2(g)\uparrow + 5CO_2(g)\uparrow + 13H_2O$$

In such a condition, copper deposits quantitatively from the solution after the nitrate ion concentration has been minimized.

Procedure

Accurately weigh a 1 g sample into 250 mL beaker. Cover the beaker with a watch glass. Cautiously add about 30 mL of 6 M HNO_3. Digest for at least 30 minutes; add more acid if necessary. Evaporate to about 5 mL but never to dryness.

SnO_2 removal

If oxide of tin is present in solution, then it is to be removed first from the solution. To the sample, add 5 mL of 3 M HNO_3, 25 mL of water and one quarter of a tablet of filter paper pulp; digest without boiling for about 45 minutes. Filter off the SnO_2H_2O, using a fine-porosity filter paper; collect the filtrate in a big electrolysis beaker. Use many small washes with hot 0.3 M HNO_3 to remove the last traces of copper; test for completeness with a few drops of NH_3(aq). The final volume of filtrate and washings should be between 100 and 125 mL; either add water or evaporate to attain this volume.

Electrolysis

With the current switch off, attach the cathode to the negative terminal and the anode to the positive terminal of the electrolysis apparatus. Briefly turn on the stirring motor to be sure the electrodes do not touch. Add 10 mL of 3M H_2SO_4 followed by 5 g of urea to each beaker. Cover the beaker with split watch glasses and commence the electrolysis. Maintain a current of 2A until the solutions are colourless. To test for completeness of the electrolysis, remove one drop of the solution with a medicine dropper, and mix it with a few drops of NH_3(aq) in a small test tube. If the mixture turns blue, rinse the contents of the tube back into the electrolysis vessel and continue the electrolysis for an additional 10 minutes. Repeat the test until no blue $Cu(NH_3)_4^{2+}$ is produced. When electrolysis is complete, discontinue stirring but leave the current on. Rinse the electrodes thoroughly with water as they are removed from the solution. After rinsing is complete, turn off the electrolysis apparatus, disconnect the electrodes and dip them in acetone. Dry the cathode for about 3 minutes at 110°C. Allow the electrode to cool in air and then weigh it.

The percentage of copper in the alloy.

$$\%\text{Cu} = \frac{m_{\text{cu}}}{m_{\text{sample}}} \times 100$$

Notes

1. Chloride ion must be totally excluded from this determination because it attacks the platinum anode during electrolysis. This reaction not only is destructive but also causes positive errors in the analysis by codepositing platinum with copper on the cathode.
2. If desired, the tin content can be determined gravimetrically by ignition of the $SnO_2 \times H_2O$ to SnO_2.
3. It is important to maintain a potential between the electrodes until they have been removed from the solution and washed. Some copper may redissolve if this precaution is not observed.

22.2.7 Analysis of Copper in a Mixture by Electrogravimetry

Copper in a copper-nickel mixture solution is analysed by electrodepositing only copper onto a previously weighed inert platinum mesh electrode. The increase in weight of the electrode is then related to the amount of metal present in the unknown solution of the mixture containing copper and nickel. Cu(II) ion is easier to reduce than Ni(II) ion at a platinum cathode in an electrolytic cell. When a negative voltage is applied on to the cathode electrode, the copper will be selectively deposited. After most of the copper has been removed from the solution by this process, Ni(II) would then begin to deposit, provided the required condition is met. However, this can be avoided by performing the electrolysis in a sulphuric acid-nitric acid solution. Both hydrogen ion and nitrate are less easily reduced than Cu(II) but more easily reduced than Ni(II). When most of the copper is deposited on the cathode electrode, the following reaction begins to carry the current, and Ni(II) is not able to deposit.

$$NO_3^- + 10H^+ + 8e^- = NH_4^+ + 3H_2O$$

It is desirable to have a secondary reaction to occur which does not cause gas bubbles to form at the cathode (e.g. the reduction of hydrogen ion). Care must be taken to avoid the formation gas bubble which results in a metal deposit with very poor qualities. The reduction of nitrate avoids this difficulty. A platinum mesh electrode is used for deposition of the metal because its large surface area yields low current density (A cm^{-2}) and a more adherent deposit results. A platinum wire anode is centred inside the cylindrical cathode. It is positioned carefully to produce a uniform deposit on all parts of the cathode.

Procedure

Use a Pt mesh electrode. It is important to treat the platinum electrodes with care, for they are quite expensive and easily deformed.

1. The unknown is supplied in a 50 mL beaker. Transfer the contents of the beaker quantitatively to a 100 mL volumetric flask, dilute to the mark with water and mix well.
2. Clean the platinum mesh electrode by immersing it in warm concentrated HNO_3 for about 1 min. If the platinum mesh is left in the HNO_3 solution for more than 2 minutes, some platinum can dissolve, leading to error. Rinse the electrode with distilled water and then with a small portion of acetone. Place the electrode in a clean beaker and dry in the oven at 110°C for 3–5 min. Remove from the oven, allow the electrode to cool in air for 5 min, and then weigh to the nearest 0.01 mg.
3. Place the Pt mesh electrode in the electrolysis assembly, and adjust its position so that it does not touch the inner Pt wire anode
4. Pipet a 10 mL aliquot of unknown into a 300 mL tall beaker. Add 5 mL of 2M HNO_3, 10 mL of 3M H_2SO_4 and enough deionized water to make the final volume about 150 mL. Add a stirring bar to the solution.
5. Position the beaker so that the electrodes are centred in it and the top of the Pt mesh electrode is about 2 mm above the solution level. Stir the solution at a moderate rate. Cover the solution with the plastic cover provided.
6. Turn on the power supply and slowly *increase the applied voltage* to the cell until the current meter indicates 1 A. Allow the electrolysis to proceed for 30 minutes. Make sure that the current maintains at 1 A throughout the experiment. There should be a stream of gas bubbles coming from the anode at all times. If this is not the case, the anode and the cathode may be touching and Cu will no longer be depositing. Ensure that the electrodes do not touch each other, especially when placing the plastic cover on the tall beaker.
7. After 30 minutes, rinse the plastic cover and sides of the beaker with a stream of water, until the top of the Pt mesh electrode is immersed. Continue the electrolysis for another 10 minutes.
8. If no visible copper deposit can be seen on the newly immersed portion of the electrode, the electrolysis is complete. Without turning off the power supply, remove the plastic cover and stirring motor, and slowly lift the electrodes from the beaker while washing the Pt mesh electrode with a stream of water. Immediately wash the electrode with water, and then with acetone. Dry the electrode in the oven for 5 minutes in a small, clean beaker, then allow to cool and weigh.
9. Repeat steps 2 through 8 twice more.

Report the weight of copper (in grams) in your total sample.

22.3 Coulometry

22.3.1 Introduction

Coulometry is based on an exhaustive electrolysis of the analyte. By exhaustive we mean that the analyte is completely oxidized or reduced at the working electrode (WE) or that it reacts completely with a reagent generated at the working electrode and to measure the quantity of electrical charge (electrons) involved for the process.

Coulometry is an electroanalytical technique for measuring an unknown concentration of an analyte in solution by completely converting the analyte from one oxidation state to another during an electrolysis reaction by measuring the amount of electricity (in coulombs) consumed or produced.

It is known that the coulometric procedures are as follows:

- Rapid
- Do not require the product to be weighable solid
- As accurate as conventional gravimetric and volumetric procedures
- Readily automated.

In a reaction, if electrons are transferred from one molecule/ion to another, then these reactions are called as 'redox reactions'. Coulometric technique is applicable to such redox reactions. The basic principle is that the reaction is controlled by controlling constant potential/or current; and the amount of electricity (i.e. the number of electrons) needed to complete the reaction is the main measurement of the technique. Coulometry is an absolute measurement similar to gravimetry or titration. In coulometry, *current is easily measured* rather than volumes or masses. The coulometric technique is very reliable and accurate. The important advantage of coulometric methods involves the possibility of using unstable titrants which are impossible to use in conventional volumetric methods. This is because these titrants are generated in situ and are used as soon as they are formed.

The end point can be detected using visual indicators or some electrochemical methods like potentiometry or biamperometry. In all cases, fast reaction between titrants and analytes are required. Instrumentation involved in coulometric titrations is very simple and can be conveniently assembled from basic components. The amperostatic or potentiostatic conditions for coulometric studies are to be controlled and maintained for accurate current measurements.

22.3.2 Coulometers

A coulometer is a device used for measuring the quantity of electricity required to bring about a chemical change of the analyte.

Coulometry uses a *constant current source* to deliver a measured amount of charge. In the coulometric method, ammeter (which measures the current) is replaced with a coulometer (which measures the quantity of electricity). Each coulometer is a second electrolytic cell, in which an anode and cathode may be in the same compartment or in separate compartments, supposed to proceed at 100% current efficiency. It is put in series with the electrolytic cell, so that the same quantity of electricity passes through both the cells. Various types of coulometers are as follows:

- Gasometric coulometers: hydrogen-oxygen coulometer
- Gravimetric coulometers
- Silver coulometers
- Copper coulometers
- Titrimetric coulometers: iodine coulometers
- Colorimeter coulometers and others

(i) Gasometric coulometers (hydrogen-oxygen coulometers)

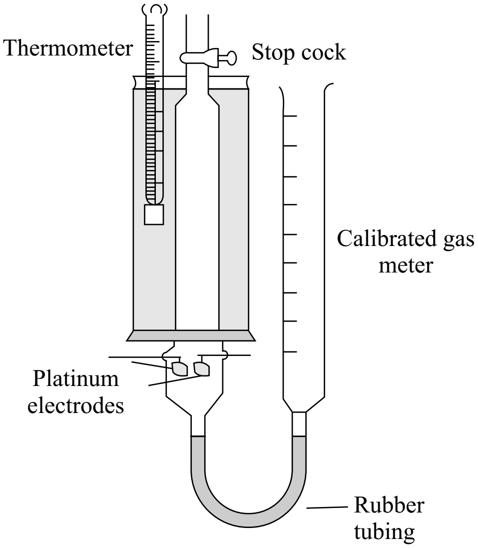


Fig. 22.12 *Hydrogen oxygen coulometer.*

A calibrated gas burette is connected to the electrolysis cell (tube) by means of a pressure rubber tube and which can be moved vertically to adjust the pressure of the collected gases to atmospheric pressure before measuring the volumes of gases. A glass tube of about 50 cm long and a diameter of 2 cm is connected to the gas burette through rubber tubing. The electrodes are two small Pt sheets attached to connecting Pt wires as shown in Fig. 22.12. A 0.5 M solution of potassium sulphate is used as the electrolyte.

On electrolysis the following reactions occur at the anode and cathode

Anode: $4OH^- \rightarrow 2H_2O + O_2 + 4e$

Cathode: $4H^+ + 4e \rightarrow 2H_2$

Overall reaction: $2H_2O \rightarrow 2H_2 + O_2$

Two moles of hydrogen and one mole of oxygen are produced by consumption of 4 Faradays (liberated 4e) which is equivalent to 0.1741 cm^3 at NTP. of mixed gas per coulomb. Such coulometers show an accuracy of about 0.1% at a current density of 0.1 A cm^{-2}, falling to –4% at 0.01 A cm^{-2} and the loss of efficiency increases as the current density decreases. At low current strengths below 50 mA, they cause negative errors. The relationship between gas production and current density has been examined.

It has been reported that the loss of efficiency is due to the formation of hydrogen peroxide at the anode, which aggravates the situation by being reduced at the cathode, so that the cathodic current efficiency also suffers.

The other coulometer is the hydrogen–nitrogen coulometer. The electrolyte used is 0.1 M hydrazine sulphate. The electrode reactions are similar except nitrogen gas is evolved at the anode.

Anode: $N_2H_5^+ \rightarrow N_2 + 5H^+ + 4e$

Cathode: $4H^+ + 4e \rightarrow 2H_2$

Overall Reaction: $N_2H_5^+ \rightarrow N_2 + 2H_2 + H^+$

Three moles of gas are produced by consumption of 4 Faradays and the yield is theoretically 0.1741 cm^3 at NTP of mixed gas per coulomb.

Precautions to be adopted with oxygen–hydrogen coulometer are as follows:

1. The gas burette should be enclosed with water circulation and the temperature variation to be within ±0.1°C.
2. An electrolyte saturated with oxygen and hydrogen mixture is taken.
3. The gas burette must be lowered periodically to hold the gas pressure close to atmospheric pressure so that solubility equilibrium is maintained during experimentation.

(ii) Silver coulometer

It is used as gravimetric coulometer. In silver coulometer, the amount of metal deposited at the cathode or the amount of metal stripped from an anode is determined.

The advantage of a silver coulometer (Fig. 22.13) is that it be used most satisfactorily either in the cathodic deposition mode or in the anodic stripping mode in perchloric acid media. A silver coulometer consists of a platinum cup or silver cup which acts as a cathode (–ve) and contains a solution of 1 M silver nitrate as the electrolyte. The anode of the cell is a rod of pure silver enclosed in a porous pot. The current density at the anode is maintained at 0.2 A cm^{-2}. After electrolysis, the cathode is dried weighed. The increase in mass of the cathode gives the amount of silver deposited. From the mass of the silver deposited, the coulomb involved in the reaction can be calculated.

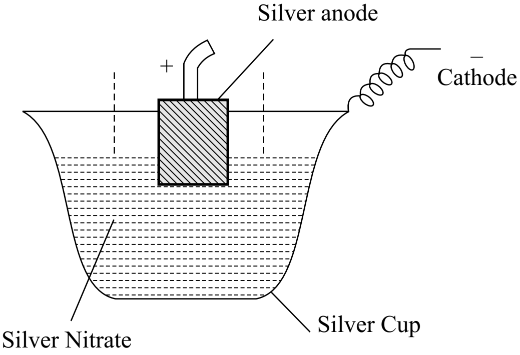


Figure 22.13 *Silver coulometer.*

(iii) Iodine coulometer

Iodine coulometer can also be used as a titration coulometer. The anodically generated iodine in the cell is titrated with thiosulphate or arsenic (III) solution. This has been used in the determination of the Faraday constant.

(iv) Colorimetric coulometers

The principle of analysis is based on the interaction of metal ions released by anodic stripping or cathodic reduction during electrolysis with a reagent to develop colour and measurement of absorbance with a spectrophotometer. For example, the formation of a coloured species for a cobalt ion with nitroso-R-salt and measurement of absorbance with a spectrophotometer. A cathodically reduced metal ion produced may be made to interact with a reagent and subsequent absorbance may be measured. For example, iron(III) is reduced to iron(II) cathodically in a coulometric cell and subsequent measurement of absorbance with 1,10-phenanthroline reagent.

(v) Radioactivity coulometers

A microcoulometer based on counting deposits of ^{110}Ag has been described. The unknown silver solution is spiked with a known amount of radioactive silver (^{110}Ag) and then deposition is made on a platinum wire in a cell in series with one containing only radioactive silver. Equal amounts of silver are plated on each cathode. The ratio of the radioactivities gives the amounts of silver in the unknown solution. The cell containing the radioactive silver is the radioactivity coulometer.

The method is useful for small quantities and plating need not be taken to quantitative completion, as it involves the measurement of the ratio of the radioactivities in the two cells.

The above types of coulometers can be put in series in an electrolytic reaction and can be used as a current measuring device.

(vi) Electronic coulometer

Electronic coulometer is based on the application of the operational amplifier in the 'integrator' type circuit. The current passed through the resistor R_1 makes a potential drop which is integrated by operational amplifier on the capacitor plates. The current can be not constant, the higher current, the larger potential drop. In such scheme V_{out} is proportional of the passed charge (i.t). Sensitivity of the coulometer can be changed by choosing of the appropriate value of R_1.

22.3.3 Electrochemical Coulometers

There are two common types of coulometers to measure the total electricity (Q) based on electrochemical processes:

- Copper coulometer
- Mercury coulometer

(a) Copper coulometer

Copper coulometer is also commonly used for general applications.

The copper coulometer (Fig. 22.14) consists of two identical copper electrodes immersed into the slightly acidic pH-buffered solution of the copper sulphate. On passing the current through the electrolytic cell leads to the anodic dissolution of the metal on anode and simultaneous deposition of the copper-ions on the cathode. The reaction takes place with 100% efficiency over a wide range of the current densities. The total quantity of electricity passing through the cell can be easily calculated by mass changes of any of the electrodes:

$$Q = \frac{\Delta m \times 2 \times F}{63.546}$$

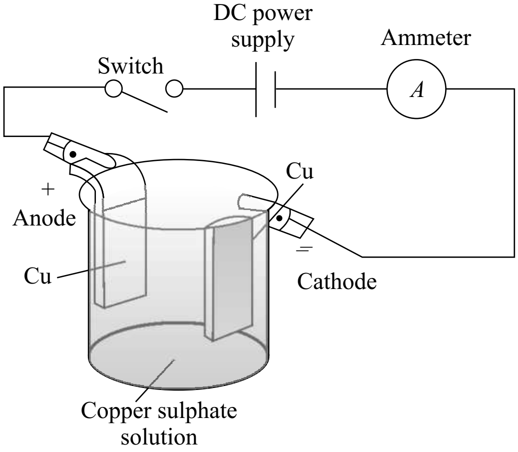


Figure 22.14 *A copper coulometer.*

where Q is the quantity of electricity (coulombs), Δm is the mass changes (g) and F is the Faraday constant (96485).

(b) Mercury coulometer

The working of the mercury coulometer is based on the electrochemical redox processes at the cell reaction

$$Hg^{2+} + e \rightarrow Hg$$

These redox processes have 100% efficiency with the wide range of the current densities.

The change in the mass of the mercury electrode gives information about the quantity of electricity used up for the process. Mass of the electrode can be increased during cathodic deposition of the mercury ions or decreased during the anodic dissolution of the metal.

$$Q = \frac{\Delta m \times 2 \times F}{200.59}$$

where Q is the quantity of electricity (coulombs), Δm is the mass changes (g) and F is the Faraday constant (96485).

Construction

The coulometer construction is based on mass measurements. A typical construction is shown in Fig. 22.15. It consists of two reservoirs connected between itself by thin graduated capillary.

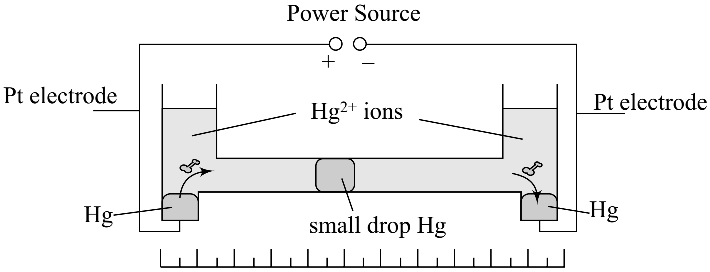
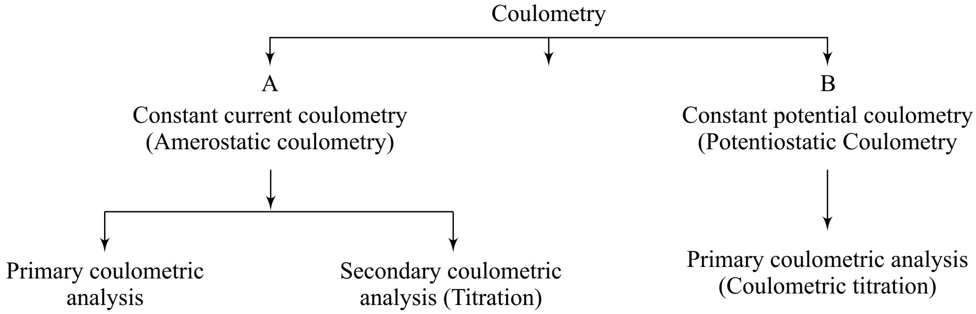


Figure 22.15 *A mercury coulometer.*

All systems contain the solution of the mercury (II)-ions. An electrode is dipped in each reservoir. Another small drop of mercury is placed into the capillary. The current facilitates the dissolution of the droplet of metallic mercury on the one border of the drop in capillary and deposition on another border of the same drop. This drop starts to move. Because of the 100% efficiency of the deposition/dissolution of the mercury under the current influence, mass or volume of this small drop will be a constant and its movement will be linearly correlated with the passed charge. If you change direction of the current, the drop starts move in the opposite direction. Sensitivity of this type of coulometers depends on the diameter of the capillary.

22.3.4 Types of Coulometric Methods

Coulometric methods can be classified into the following categories:



There are two types of coulometry: controlled-current coulometry, in which a constant current is passed through the electrochemical cell, and controlled-potential coulometry, in which a constant potential is applied to the electrochemical cell.

- One category, *coulometric titration or amperostatic coulometry*, the current (measured in amperes) in the electrolytic cell is kept constant. There are two techniques:
  - (i) *Primary constant current coulometry*
    In this technique, the element to be estimated undergoes direct reaction at the electrode with 100% efficiency.
  - (ii) *Secondary constant current coulometry*
    'Coulometric titration' is a frequently used technique and has wider applications. In a coulometric titration, one of the reactants, titrant is quantitatively produced (generated) at an electrode which then stoichiometrically reacts with the analyte ion to be estimated.
- The other, called *potentiostatic coulometry*, the potential of the working electrode is kept constant during the reaction in an electrolytic cell.

It may be noted that during an electrolytic process, the total charge, Q (in coulombs) passing through the electrolytic cell is proportional to the absolute amount of analyte by Faraday's law

$$Q = nF \cdot N_A \qquad ...(1)$$

where n is the number of electrons per mole of analyte, 'F' is Faraday's constant (96487 C mol^{-1}), and 'N_A' is the moles of analyte. 'A' coulomb is equivalent to an A sec; thus, when passing a constant current, 'I', the total charge for the process is

$$Q = i \cdot t_\mathrm{e} \qquad ...(2)$$

where 't_e' is the electrolysis time. If the current varies with time, as it does in controlled-potential coulometry, then the total charge is

$$Q = \int_0^t I \cdot dt \qquad ...(3)$$

In coulometry, current as a function of time is monitored and Eq. (2) or Eq. (3) is used to calculate Q. Knowing the total charge, Eq.(1) is used to determine the moles of analyte. To obtain an accurate value for 'N_A', all the current must be used to oxidize or reduce the analyte. In other words, coulometry requires 100% current efficiency.

22.3.4.1 Constant Current Coulometry

(a) Coulometric titration (or amperostatic coulometry)

The basic principles of coulometric titrations are as follows:

- Based upon the constant-current electrolytic generation of a titrant which reacts quantitatively with the substance to be determined.
- Carried out with a constant-current source called an amperostat. Magnitude of the constant current is analogous to the 'concentration of a standard titrant' solution.
- Time required to complete the titration is equivalent to the 'volume of titrant' solution.

As one of the successful techniques in coulometry is to use a *constant current* in place of a constant potential, which produces the current versus time profile shown in Fig. 22.16. Controlled-current coulometry has two advantages over controlled-potential coulometry.

- First, the analysis time is shorter for the process because the current does not decrease over time. A typical analysis time for controlled-current coulometry is less than 10 minutes, compared to approximately 30–60 minutes for controlled-potential coulometry.
- Second, because the total charge is simply the product of current and time [Eq. (2)], there is no need to integrate the current–time curve in Fig. 22.16.

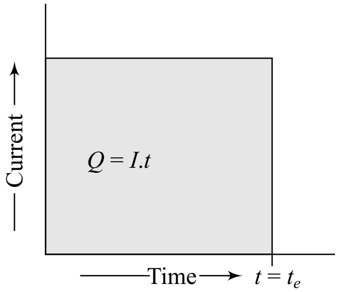


Figure 22.16 *Current versus time for a controlled-current coulometric analysis. The measured current is shown by the black line. The integrated area under the curve, shown in grey, is the total charge (Q).*

There are two important experimental problems when using a constant current.

- First, during electrolysis the analyte's concentration and, therefore, the current due to the its oxidation or reduction continuously decreases as shown in Fig. 22.17.

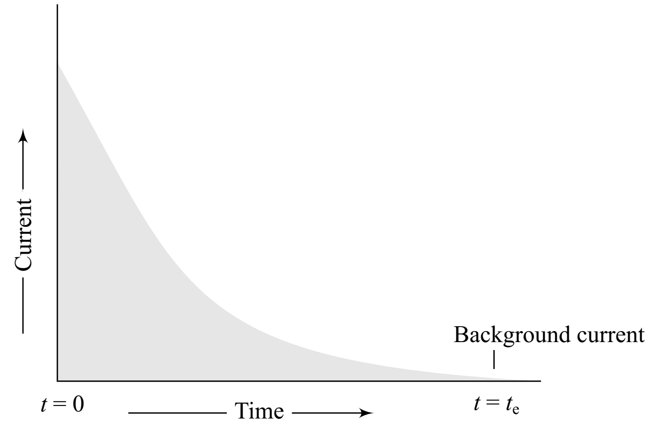


Figure 22.17 *Exponential decrease of current with time.*

In order to maintain a constant current, the system demands the potential to change until another oxidation reaction or reduction reaction occurs at the working electrode. If the conditions are not satisfactory, the secondary reaction occurs at the electrode and decreases the current efficiency to less than 100%.

- The second problem is to find a method to identify the completion of the electrolysis process. In controlled-potential coulometric analysis, when the current reaches zero or when it reaches a constant background or residual current, indicates that the electrolysis is complete as shown in Fig. 22.17. In a controlled-current coulometric analysis, however, current continues to flow even when the analyte's electrolysis is complete. A suitable method for determining the reaction's endpoint, t_e, is needed.

(b) Maintaining current efficiency

All electrolytic processes are expected to be carried out at 100% efficiency. It is noted that any change working electrode's potential may result in a current efficiency <100%. Let us consider the coulometric analysis for Fe^{2+} based on its oxidation to Fe^{3+} at a Pt working electrode in 1M H_2SO_4.

$$Fe^{2+}{}_{(aq)} \rightleftharpoons Fe^{3+}(aq) + e^-.$$

Initially, it is observed that the potential of the working electrode remains nearly constant at a level near its initial value. As the concentration of Fe^{2+} decreases in a cell, the working electrode's potential shifts towards more positive values until the oxidation of H_2O begins. As a result, the anode potential to increase to the point where the decomposition of water becomes a competing process:

$$2H_2O(l) \rightleftharpoons O_2(g) + 4H^+(aq) + 4e^-$$

Because a portion of the total current comes from the oxidation of H_2O, the current efficiency for the analysis is less than 100% and we cannot use $Q = I \cdot t$ to determine the amount of Fe^{2+} in the sample.

The drifting of the potential of working electrode in a cell cannot be prevented until another species undergoes oxidation. However, 100% current efficiency can be maintained, if any product of that oxidation reacts both rapidly and quantitatively with the remaining Fe^{2+}.

In order to avoid the above problem, an excess of Ce^{3+} is added to the analytical solution. In such a case, when the potential of the working electrode shifts to a more positive potential, Ce^{3+} eventually begins to oxidize to Ce^{4+}.

$$Ce^{3+}(aq) \rightleftharpoons Ce^{4+}(aq) + e^-$$

This Ce^{4+} formed at the working electrode rapidly mixes with the solution where it reacts with any available Fe^{2+}.

$$Ce^{4+}(aq) + Fe^{2+}(aq) \rightleftharpoons Ce^{3+}(aq) + Fe^{3+}(aq)$$

The overall reaction shows that the net reaction is the oxidation of Fe^{2+} to Fe^{3+}

$$Fe^{2+}(aq) \rightleftharpoons Fe^{3+}(aq) + e^-$$

thus, maintaining a current efficiency of 100%. A species, such as Ce^{3+}, which is used to maintain 100% current efficiency, is called a 'mediator'. In such a coulometric titration, *a titrant that is electrolytically generated reacts stoichiometrically with another and proceeds with l00% efficiency*.

(c) Determination of end point of electrolysis

Adding a mediator solves the problem of maintaining 100% current efficiency, but it does not solve the problem of determining when the analyte's electrolysis is complete. Example, when the oxidation of Fe^{2+} is complete current continues to flow from the oxidation of Ce^{3+}, and, eventually, the oxidation of H_2O. It does not give any information regarding the end point of electrolysis or the presence of no more Fe^{2+} in solution.

For our purposes, it is convenient to treat a controlled-current coulometric analysis as a reaction between the analyte, Fe^{2+}, and the mediator, Ce^{3+}. This reaction is identical to a redox titration; where the end points for a redox titration are indicated by visual indicators. For example, ferroin provides a useful visual end point for the Ce^{3+} mediated coulometric analysis for Fe^{2+}, changing colour from red to blue when the electrolysis of Fe^{2+} is complete. Instrumental techniques can be used to signal the end of a controlled-current coulometric analysis:

- potentiometric methods
- amperometric methods
- spectrophotometric methods

Constant current coulometric analysis is used to accurately quantify the concentration of a species in analytical reaction. In this case, the applied current (I) is equivalent to a 'titrant'. *Constant current is applied to the unknown solution until all of the unknown species is either oxidized or reduced to a new state, at which point the potential of the working electrode shifts dramatically.*

This potential shift indicates the end point. The only measurement required in this type of constant current coulometry is the 'time' it takes to complete the electrolysis. The product of this 'time' and the 'current' is then used to determine the *total amount of electricity* ($Q = I \cdot t$) used. The end point of the titration can be determined analytically by using an indicator that is placed in the sample and signals when the system reaches equilibrium. Alternatively, the end point can be determined from data provided by potentiometric, amperometric or conductance measurements. The magnitude of the current (I, amps) and the duration of the current (t, seconds) can be used to determine the moles of the unknown species in solution. When the volume of the solution is known, then the molarity of the unknown species can be determined.

Since concentration polarization is inevitable in coulometric titrations, most of the reaction must occur distant from the working electrode. Otherwise, the potential will need to constantly increase as the reaction progresses to maintain the production of products. Therefore, coulometric titration reactions usually have other ions in the sample that react directly with the electrode and then react with the sample molecules. Coulometric titrations can also be done by back-titration.

(d) Construction

An amperostat is used to maintain a constant current in coulometric titrations. It reacts to changes in the resistance of the cell by altering its output potential. Both amperostats and potentiostats can be built relatively cheaply using operational amplifiers.

The equipment of the assembly includes

- a source of constant current: an amperostat.
- a titration vessel,
- an electric timer and
- a device for monitoring current (digital Ammeter).

Figure 22.18 depicts the principal components of a typical coulometric titrator. A source of constant current (amperostat) and a switch that simultaneously initiates the current and starts an electronic timer are required. The potential drop across the standard resistor, R_{std}, is used for current measurement or a sensitive ammeter is used.

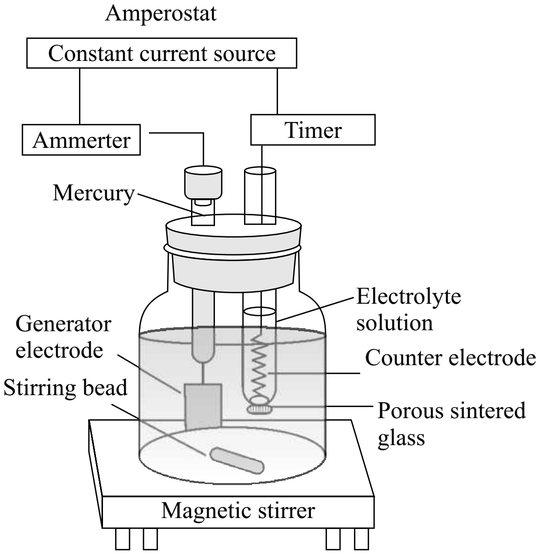


FIGURE 22.18 *A typical coulometric titration cell.*

(e) Cell assembly

The cell assembly consists of a 'generator electrode' at which the titrant/reagent is generated and a 'counter electrode' to complete the circuit as in Fig. 22.19. The generator electrode is a platinum foil or gauze with a large surface area. The products formed at the second electrode normally interfere with the electrochemical reaction. For example, the anodic generation of oxidising agents ($Ce^{3+} \rightarrow Ce^{4+}$) is often accompanied by the evolution of hydrogen from the cathode. This hydrogen reacts with the oxidising agent and leads to a positive error. Hence, the electrode is isolated from the test solution by a sintered glass disk. Besides, this type of arrangement helps to maintain 100% efficiency.

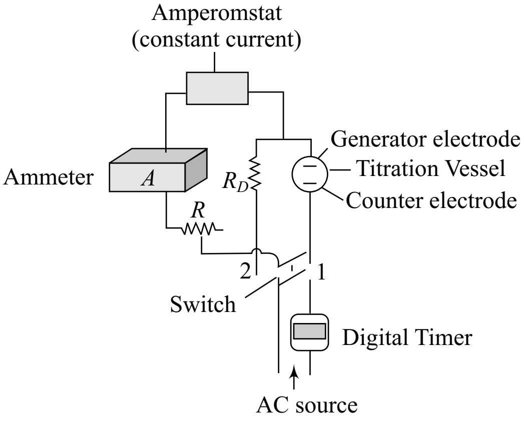


FIGURE 22.19 *A constant current coulometric titrator.*

(f) External generation of titrant

Since internal generation of titrant sometimes interferes with the titration, it becomes necessary to generate the titrant externally and deliver to the titration cell. An assembly of the type shown in Fig. 22.20 is generally used. For example, hydrogen and hydroxyl ions are produced by the electrolysis of sodium sulphate solution.

$$2H_2O + 2e \rightarrow H_2 + 2OH^- \text{ (cathode reaction)}$$

$$H_2O \rightarrow \frac{1}{2} O_2 + 2H^+ + 2e \text{ (anode reaction)}$$

The cell consists of two electrodes of platinum gauze separated by a sintered glass disk. The anode compartment produces H+ ions, while the cathode compartment produces OH^-ions. The apparatus is so arranged that flow of electrolyte continues even after the current is discontinued, thus flushing the residual reagent into the titration vessel.

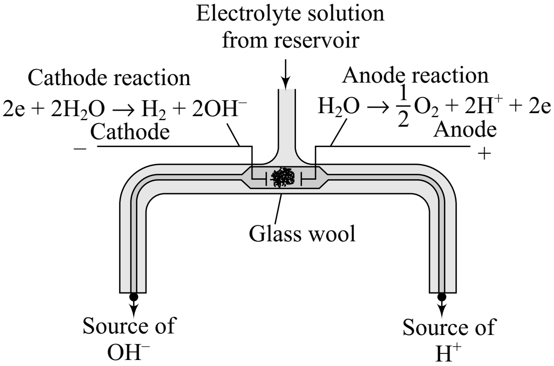


FIGURE 22.20 *A cell for the external coulometric generation of acid and base.*

This type of external generation apparatus has also been used for the generation of other reagents such as iodine by the oxidation of iodide at the anode, generation of titanous ions, etc. The external generation technique has been successfully applied to titrations involving hydrogen, hydroxyl, iodine, etc. and for the generation of reagents, which are quite unstable, such as chlorine, bromine, titanium (III) ions, etc. Thus, coulometric titrations are more advantageous than the conventional volumetric titrations.

It should be noted that external generation of reagents is possible if only the titrant be generated with 100% current efficiency.

(g) Detection of end points in coulometric titrations

Indicators are used to detect the end points of coulometric titrations and are quite satisfactory.

For example,

- In titrations involving iodine, starch indicator is used.
- Methyl orange indicator is used in titrations for hydrazine.
- Hydroxylamine or thiocyanate against electrolytically generated bromine.

However, coulometric end point can be more accurately determined by instrumental methods of detection of end points such as potentiometric, conductometric, amperometric or photometric methods.

(i) Potentiometric endpoint detection has been found to be suitable in acidimetric and redox titrations. This method of detection of end point has the advantage that the titrations can be stopped exactly at the end point. Besides, the observed emf (change in potential) does not depend on factors like geometry of the cell, conditions of stirring, etc. Thus, these methods of detection of endpoints are more advantageous.

(ii) Amperometric detection of end point is useful in redox and precipitation titrations and is still more advantageous than the potentiometric method of detection of end point since the choice of the indicator electrode for detecting the end point does not depend on the availability of reversible indicator electrode. Even if one of the ions is electroactive, the end point can be detected easily.

(iii) Spectrophotometric end point detection has been developed and in recent years, this is of great importance in the titration of olefins with electrolytically generated bromine.

Some of the special advantages of the method of detection of end points are as follows:

(a) The exact end point can be determined by extrapolation of the linear portion of the titration curves before and after the end point.

(b) By proper choice of λ_{max} and molar absorptivity, the sensitivity and selectivity of the methods can be improved.

22.3.4.1.1 Coulometric Titrations

Coulometric titrations have been developed for all types of volumetric titrations. Some important applications of coulometric titrations are given below:

(i) Acid–base titration

The hydroxide ions generated at a cathode is neutralized with a solution of strong acids or weak acids in constant current coulometer

$$2H_2O + 2e \rightarrow 2OH^- + H_2(g) \qquad \text{...(Cathode reaction)}$$

The cells shown in Fig. 21.20(a) and (b) can be used to generate hydroxide ions.

The end points of the neutralization reactions are detected by the use of indicator. However, potentiometric end points can also be employed depending upon the availability of instruments. A real advantage of such coulometric titrations is that interference from carbonate ions can be avoided by the generation of hydroxide ions, free from carbonate ions. However, it is necessary to deaerate the solution with an inert gas before beginning the analysis to avoid the interference of carbonate ions.

Titrations of strong/weak bases are possible with hydrogen ions generated at a platinum anode

$$H_2O \rightarrow \tfrac{1}{2}O_2(g) + 2H^+ + 2e \qquad \text{...(Anode reaction)}$$

Here, the cathode must be isolated from the solution to prevent the interference of hydroxide ions produced at that electrode.

(ii) Precipitation and complex formation titrations

The silver ions (Ag^+) are generated at the anode for coulometric titrations. The generator electrode consisting of a heavy silver wire can be constructed and a cell of the type shown in Fig. 22.19 can be used. End points can be detected with adsorption indicators or by potentiometric method. Similar titrations based upon the generation of mercury (I) ions at a mercury anode have been described.

Titration of several cations is performed as complexometric titrations by generating ethylene diaminetetraacetate ions (HY^{3-}) at a mercury cathode of the electrolytic cell. Reduction of the amine mercury (II) EDTA chelate at a mercury cathode was used to generate HY^{3-} ions.

$$HgNH_3Y^{2-} + NH_4^+ + 2e \leftrightarrow Hg + 2NH_3 + HY^{3-}$$

Since mercury chelate is more stable than the corresponding complexes with calcium, zinc, lead or copper; titrations of these cations are possible by this method.

(iii) Redox coulometric titrations

The redox titrations are carried out for unusual oxidation states of certain ions such as Ag^{2+}, Mn^{3+} and U^{4+} which can be generated at the electrode of electrolytic cell and can be titrated with the oxidising reagents such as iodine, bromine, chlorine, Ce^{4+}. Table 22.5 lists some of the applications of these reagents in coulometric titrations.

Coulometric redox titrations gain importance because of in-situ generation of unusual/unstable reagents such as bromine, dipositive silver ions, tripositive manganese ions, etc. These reagents cannot be prepared under normal circumstances of conventional volumetric analysis. Electrogenerated bromine has been proved to be very useful for the determination of phenol, aniline, mustard gas, As (III), Sb(III), etc. A bromine solution is quite unstable and cannot be used as a titrant in conventional volumetry and hence one has to use a Winkler's solution (bromate-bromide mixture) to carry out reactions involving bromine. Tables 22.4 and 22.5 provide summary of applications of coulometric titrations.

Species determined	Generator electrode reaction	Secondary analytical reaction
Acids	$2H_2O + 2e \rightarrow 2OH^- + H_2$	$OH^- + H^+ \rightarrow H_2O$
Bases	$H_2O \leftrightarrow 2H^+ + ½O_2 + 2e$	$H^+ + OH^- \rightarrow H_2O$
Halides, mercaptans	$Ag \rightarrow Ag^+ + e$	$Ag^+ + Cl^- \rightarrow AgCl(s)$, etc.
Mercaptans	$Ag \rightarrow Ag^+ + e$	$Ag^+ + RSH \rightarrow AgRS(s) + H^+$
Zn^{2+}	$Fe(CN)^{3-} + e \rightarrow Fe(CN)^{4-}$	$3Zn^{2+} + 2K^+ + 2Fe(CN)^{4-} \rightarrow$
Ca^{2+}, Cu^{2+}, Zn^{2+} and Pb^{2+}	$HgNH_3Y^{2=}NH_4^+ \rightarrow Hg_{(l)} + 2NH_3 + HY^{3-}$	$HY^{3-} + Ca^{2+} \rightarrow CaY^{2-} + H^-$, etc.

Table 22.4

Applications of coulometric titrations involving neutralization, precipitation and complex formation reactions

Reagent	Generator electrode reaction	Substance determined
Br_2	$2Br^- \rightarrow Br_2 + 2e$	As(III), Sb(III), U(IV), Tl(I), I^- SCN^-, NH_3, N_2H_4, NH_2OH, phenol, aniline, mustard gas
Cl_2	$2Cl^- \rightarrow Cl_2 + 2e$	As(III), I^-
I_2	$2I^- \rightarrow I_2 + 2e$	As(III), Sb(III), $S_2O_3^{2-}$, S ascorbic acid
Ce^{4+}	$Ce^{3+} \rightarrow Ce^{4+} + e$	Fe(II),Ti(III), U(IV), As(III), I^- $Fe(CN)_6^{4-}$
Mn^{3+}	$Mn^{2+} \rightarrow Mn^{3+} + e$	$H_2C_2O_4$, Fe(II), As(III)
Ag^{2+}	$Ag^+ \rightarrow Ag^{2+} + e$	Ce(III), V(IV), $H_2C_2O_4$, As(III)
Fe^{2+}	$Fe^{3+} + e \rightarrow Fe^{3+}$	Cr(VI), Mn(VII), V(V), Ce(IV)
Ti^{3+}	$TiO^{2+} + 2H^+ + e \rightarrow Ti^{3+} + H_2O$	Fe(III), V(V), Ce(IV), U(VI)

Table 22.5

Applications of coulometric titrations involving redox titrations

22.3.4.1.2 Karl Fischer Reaction

The amount of water in a sample can be determined using Karl Fisher reaction uses coulometric titration. The concentration of water in any substance can be determined of the order mgL^{-1}. The procedure is applicable to water content determination in substances like butter, sugar, cheese, paper and petroleum. The working electrode for this analysis reaction is an iodine electrode, while the reference electrode is a platinum electrode. The reaction involves converting solid iodine into hydrogen iodide in the presence of sulphur dioxide and water. Methanol is most often used as the solvent, but monoether glycol and diethylene glycol also work. Pyridene, which is a base, is often used to counteract the formation of sulphuric acid, although the use of imidazole and diethanolamine for this role is becoming more common. All reagents must be anhydrous for the analysis to be quantitative.

In the presence of methanol and pyridine, a molecule of water reacts with a molecule of iodine as given by the stoichiometric equation

$$[C_6H_5NH]\,SO_3CH_3 + I_2 + H_2O + 2C_6H_5N[C_6H_5NH]SO_4CH_3 + 2[C_6H_5NH]I$$

It is observed that any moisture content of atmosphere can affect the titration results. Therefore, the system is usually isolated with drying tubes or placed in an inert gas container. Always dry solvents to be used to carry out the titration.

The usual procedure is to adopt a back or direct titration method to determine the water in the sample. In the direct method, just enough of the reagents will be added to completely use up all of the water. At this point in the titration, the current approaches zero. It is then possible to relate the amount of reagents used to the amount of water in the system via stoichiometry. The back-titration method is similar, but involves the addition of an excess of the reagent. This excess is then consumed by adding a known amount of a standard solution with the known water content. The result reflects the water content of the sample and the standard solution. Since the amount of water in the standard solution is known, the difference reflects the water content of the sample.

22.3.4.1.3 Determination of Film Thickness of Metallic Coating

An approach for the determination of the thickness of metallic coatings has slightly a different principle. In this method, the quantity of electricity needed to dissolve a well-defined area of the coating is measured. The film thickness 'Δ' is proportional to the constant current 'I', the molecular weight 'M' of the metal, the density 'ρ' of the metal and the surface area 'A':

$$\Delta = \frac{I \cdot M}{A \cdot \rho}$$

If the type of metal used for the coating is known then the amount of current necessary to remove the coating is proportional to the thickness of the coating.

Usually, the electrodes for dissolving a well-defined area is a platinum electrode and an electrode that relates to the reaction. A tin electrode is used for tin coating on a copper wire, and a sodium chloride-zinc sulphate electrode is used to determine the zinc film on a piece of steel. Special cells have been created to adhere to the surface of the metal to measure its thickness. These are basically columns with the internal electrodes with magnets or weights to attach to the surface. The results obtained by this coulometric method are similar to those achieved by other chemical and metallurgic techniques.

22.3.4.1.4 Coulometric Determination of Arsenic by Electrogenerated Bromine and Dead-Stop End-Point Detection

In-situ generation of titrants which are, otherwise, unstable is an important feature of the coulometric titrations. Coulometric titrations are widely used for determination of many such important analytes.

The determination of arsenic can be carried out by electrogenerated bromine in a coulometric experiment and the end point is detected through the use of an auxiliary dead-stop end-point detection circuit. The cell used for the purpose has four electrodes. Two small identical Pt electrodes are used for end point detection of reaction. A large Pt electrode is used as the generator electrode and an isolated auxiliary electrode which is the cathode. The indicator electrodes of the detection circuit pass current only when Br_2 and Br^- are present in excess.

The electrodes reactions are as follows:

(a) At generator electrode

$$2\,Br^- \rightarrow Br_2 + 2\,e$$

(b) At auxiliary electrode

$$2\,H^+ + 2\,e \rightarrow H_2$$

or

$$2\,H_2O + 2\,e \rightarrow H_2 + 2\,OH^-$$

(c) Redox reaction

$$H_2AsO_3^- + Br_2 + H_2O \rightarrow AsO_4^- + 2\,Br^- + 2\,H^+$$

Chemicals and Reagents

1. Solid KBr
2. 5 M NaOH solution
3. Arsenic unknown (predried and desiccated)
4. 5 M HCl solution
5. Predried and desiccated arsenic trioxide (analytical grade).

Apparatus

Coulometric titrator or assembly of the coulometric circuit.

Procedure

Preparation of Arsenic solution

1. Dissolve an accurately weighed sample of analytical grade arsenic trioxide (0.08–0.1 g) into a 500 mL volumetric flask. This can be achieved by dropwise addition of 5 M NaOH solution and swirling. When the sample is dissolved, add about 200 mL of distilled water followed by 5 M HCl equivalent to the amount of the 5 M NaOH added. Adjust the final HCl concentration in the flask to 0.2 M HCl.
2. Repeat for the unknown arsenic sample but take about 0.1–0.15 g of the unknown (accurately weighed).
3. Dissolve about 6 g of KBr in about 100 mL of distilled water in a 250 mL beaker into which a stirring bar was dropped.
4. Set up the circuit described in the attached figure and adjust the location of the electrodes. The two small platinum electrodes are connected to the detection circuit while the other large platinum electrode is the generator electrode.
5. Start the stirrer at a suitable rate which should be constant throughout the experiment.
6. Electrolyse for few seconds by closing the generator circuit through the switch. The galvanometer should read a current of about 5 mA. This is a necessary step for the oxidation of available oxidizable species. Switch off the generator circuit.
7. When the indicator galvanometer deflects, add dilute arsenite solution dropwise to consume any bromine in solution.
8. Transfer 10 mL of the standard arsenic solution and switch on the generator circuit and record the time versus indicator galvanometer reading.

9. Repeat steps 7 and 8 three times for the standard and three more times for the unknown arsenic solution. You do not have to discard the solution each run.
10. From results obtained in steps 8 and 9, plot current versus time, and find the percentage of As_2O_3 in the unknown sample.

22.3.4.1.5 Comparison of Coulometric and Volumetric Titration

(a) Both have observable end point

- Current (e^- generation) serves the same function as a standard titrant solution
- Time serves the same function as volume delivered
- Amount of analyte determined by combining capacity
- Reactions must be rapid, essentially complete and free of side reactions

(b) Advantages of coulometry

- Both time and current easy to measure to a high accuracy
- Stable titrant
- Easier and more accurate for small quantities of reagent: small volumes of dilute solutions
- Used for precipitation, complex formation oxidation/reduction or neutralization reactions
- Readily automated

(c) Sources of error

- Variation of current during electrolysis
- Departure from 100% current efficiency
- Error in measurement of current
- Error in measurement of time
- Titration error (difference in equivalence point and end point).

22.3.4.1.6 Applications

In electroanalytical chemistry, coulometric tritrations have innumerable applications.

- Using electrochemically generated hydroxide ions at the cathode of the cell, it is possible to carry out neutralization of weak as well as strong acids. Similarly, bases can be neutralized using anodically generated H_3O^+.
- Ag^+ ions are electrogenerated at the electrode and are allowed to react with halides resulting in precipitation of the silver halide. Such precipitation reactions can be achieved in coulometry.
- It is convenient to use electrogenerated oxidants in coulometric studies for redox titrations. Most of the oxidants, such as Cl_2, Br_2 and I_2 are electrogenerated which are used for oxidation of many reducing agents.

The high accuracy and the speed of analysis is an added advantage for coulometric experiments and are most preferred technique, which are widely used.

- Mercury, silver chloride or silver bromide electrodes are good counter electrodes for this reaction.
- The coulometric technique has one real advantage over the volumetric technique, which is that the interference of carbonate ions (CO_3^{2-}) is easily removed by bubbling a carbon dioxide-deficient gas through the sample. This removes carbon dioxide in the sample, which would otherwise form carbonate ions and interfere with the analysis.

Other applications coulometric titrations are as follows:

- Coulometric titrations can be done for ions such as halides, Zn^{2+} and mercaptans by complex formation or precipitation in the presence of anodically generated silver ions.
- The oxidation of cerium from lower oxidation (Ce^{3+}) to higher oxidation state (Ce^{4+}) can be achieved in the presence of 1N H_2SO_4. This can be used to characterize several systems including titanium,

iron and uranium. Polarized platinum electrode and a nonpolarized reference electrode, such as a lead amalgam—lead sulphate electrode are used in such a colorimetric cell.

- The reduction of Fe^{3+} to Fe^{2+} in acidic solutions can be used to determine the amount of permanganate (MnO_4^-), chromate (CrO_4^{2-}) or dichromate ($Cr_2O_7^{2-}$) ions in a sample. Coulometric titrations can be used for these systems. In the coulometric titration technique, the ratio of iron ions will change when the sample is added and the endpoint occurs when the ratio returns to its initial amount.
- The amount of water in a sample can be determined in the Karl Fischer reaction.
- The thickness of a metal film can also be determined.

22.3.4.1.7 Advantages of Coulometric Methods

1. In the coulometric technique, the currents (*I*) can be controlled and the time (*t*) of flow current till the completion of electrolysis can be measured with utmost precision and hence, the measurements are more accurate and precise than classical methods.
2. Preparation, standardization and storage of standard solutions are not necessary in coulometric titrations. This procedure is useful especially for preparation of unstable reagents such as dipositive silver ions, tripositive manganese ions, bromine, chlorine, titanium (III), etc. which can be generated in situ, since the conventional analysis cannot handle these reagents.
3. Titrations which cannot be performed by conventional methods can be easily performed coulometrically. Titrations of high hazardous materials, titrations involving unstable or difficulty prepared titrants such as bromine, tin (II), titanium (III), chromium (II), silver (II), etc. and titrations in molten salts can be performed.
4. It is very convenient to perform coulometric titrations where titrants are generated at the working electrode. For example, electrolytic generation of iodine used to estimate hydroquinone, ascorbic acid, antimony (III), etc. Electrolytically generated bromine used to estimate organic substances has been found to be the most satisfactory intermediate for the estimation of several organic compounds.
5. Constant current coulometry has widely been used than the controlled-potential coulometry since the former is faster and requires simpler instrumentation and less expensive.
6. It is easy to handle small quantities of reagents by coulometric titrations. By proper choice of current, microquantities of substance can be analysed with greater accuracy and ease.
7. A single constant current source can be used to generate precipitating, complexing and redox reagents.
8. In chloride titrator silver (I) ion is generated coulometrically.
9. In water titrators, Karl Fischer reagent is generated electrolytically to determine a trace level of concentrations of water content/moisture.

22.3.4.2 Controlled-Potential Coulometric Analysis (Potentiostatic Coulometry)

The controlled-potential techniques are based on controlling potential and measuring the resulting current for a complete electrolytic process. In potentiostatic coulometry, the applied potential (E_{appl}) is the driving force for the redox reaction and the potential is always measured versus a reference electrode (RE). Potential of the working electrode (WE) is maintained at a constant level relative to the reference electrode (RE) throughout the electrolysis. The working electrode (WE) is kept at a *constant potential* which allows for the analyte's reduction or oxidation without simultaneously reducing or oxidizing other species in the solution. The current (*I*) which flows through the circuit is measured. As the substrate is consumed with time, the current (*I*) also decreases, approaching zero (negligible), when the conversion is complete (Fig. 22.17).

The current (*I*) flowing through the cell for '*t*' seconds is proportional to the analyte concentration. Modern instruments, however, use electronic integration to monitor charge as a function of time. The total charge can be read directly from a digital readout or from a plot of charge versus time or means of determining the total charge passed during electrolysis to monitor the current as a function of time and also, determine the area (Fig. 22.17) under the curve (area = $Q = \int_0^t I \cdot dt$).

In a Potentiostatic coulometry, the potential at the working electrode is not equal to the potential of the entire cell. It is related by the equation

$$E = E_0 - \frac{RT}{nF} \ln Q$$

where 'R' is the Molar gas constant, 'n' is the stoichiometric number of electrons, 'F' is the Faraday constant E_0 is the cell potential and 'Q' is the instantaneous ratio of the concentrations of the component in the system. This relationship explains the *exponential decrease in current observed in a system at constant potential.* Here, the current is initially high but decreases exponentially with time and approaches zero at a time required to attain a reaction position that is quantitatively complete.

The entire time of electrolysis of the analyte is to be minimized at 100% efficiency of the process. From the following aspects of potentiostatic coulometry, minimization of the electrolysis time can be achieved giving due importance to the following aspects.

- An exhaustive electrolysis, therefore, may require a longer time.
- The current at 't' is $I = I_0 \cdot e^{-kt}$
  where, 'I_0'is the initial current, 'K' is a constant that is directly proportional to
  (i) area of the working electrode
  (ii) rate of stirring, and inversely proportional to the volume of solution.

For an exhaustive electrolysis in which the analyte is oxidized or reduced, the current at the end of the analysis, 't_e', may be approximated is nearly zero, i.e. ($I \le (10^{-4})I_0$.

Since $\quad I = I_0 e^{-kt}$

$$t_e = \frac{1}{k}\ln(1 \times 10^{-4}) = \frac{9.21}{k}$$

Thus, increasing 'k' leads to a shorter analysis time.

For this reason, controlled-potential coulometry is carried out in

- small-volume electrochemical cells
- using electrodes with large surface areas
- with high stirring rates.

The proper maintenance of constant potential is desirable so that the desired oxidation or reduction reaction goes to completion without interference from redox reactions involving other components of the sample matrix.

For example, reduction of copper ions to copper occurs when the working electrode's potential is more negative than +0.342 V.

$$Cu^{2+}(aq) + 2e \rightarrow Cu(s)$$

To maintain a 100% current efficiency, the potential of the electrode must be selected so that the reduction of H^+ to H_2 does not contribute significantly to the total charge passed at the electrode.

(a) Instrumentation

The electrolytic cell of potentiostatic coulometry consists of a three-electrode system. The entire assembly comprises a potentiostat, an electrolysis cell, a chemical coulometer or an electronic integrating device (for determining the number of coulombs) are placed in series with the working electrode (WE), reference electrode (RE) and counter electrode (CE). The chemical coulometer such as hydrogen–oxygen coulometer/silver coulometer can be used to measure the number of coulombs (Q) of electricity for the completion of the process.

A schematic diagram of a potentiostatic coulometry is shown in Fig. 22.21(a). The control of potential could be varied at the working electrode by applying a *variable voltage source at the noninverting input* of the operational amplifier. A necessary high current to the circuit is sent through a booster amplifier which is attached in series with a recording unit and an integrator device. Potentiostat is an electronic instrument that measures and controls the voltage difference between a working electrode (WE) and a reference electrode (RE) and maintains a constant potential. Integrating device (or ammeter) measures the current (*I*) flow between the working (WE) and counter electrodes (CE).

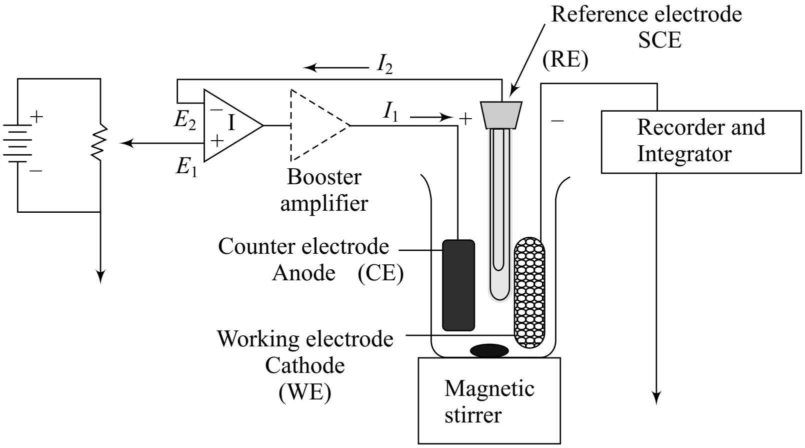


FIGURE 22.21(a) *A schematic diagram of a potentiostatic coulometry.*

Consider the circuit shown in the diagram of Fig. 22.21(b).

The two resistances (R_s and R_u) as marked in Fig. 22.21(b) of a potentiostatic coulometric circuit correspond to resistances in two parts of the electrochemical cell, where 'R_s' is the cell resistance between the counter electrode (CE) and the tip *P* of the reference electrode (RE); and 'R_u' is the so-called *uncompensated cell resistance* which is the cell resistance between *P* and the working electrode (WE).

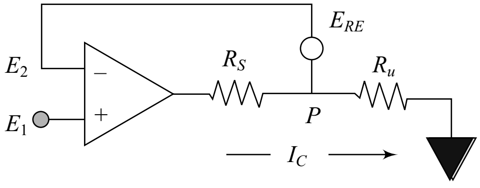


FIGURE 22.21(b) *Circuit diagram of a potentiostatic coulometric cell.*

Because of very high resistance of the inputs to the operational amplifier, there is no current in the feedback loop at the inverting input.

The function of the operational amplifier is to maintain E_1 is equal to E_2 in the noninverting configuration. In order to achieve this condition, the *cell current 'I_C'* is supplied by the operational. If we consider the path between the inverting input and the circuit common at the output, we see that

$$E_2 = E_1 = E_{\mathrm{RE}} + I_C R_u = E_{\mathrm{RE}} + E_C$$

where 'E_C' is the 'cathode potential' and is equal to the potential difference between *P* and the working electrode (WE). Since E_1 and E_{RE} are constant, $I_C R_u$ must also be constant. During electrolysis, if R_U and R_S changes, the operational amplifier output voltage changes in such a way so as to maintain $E_C = I_C R_u$ at a constant level. If R_u decreases, the operational amplifier output voltage *increases* to maintain E_C constant. If R_u increases as a result of an increase in the cell resistance or due to concentration polarization, the output voltage of the operational amplifier decreases, which leads to a decrease in I_C.

The cell assembly (Fig. 22.21c) consists of a platinum gauze working electrode (WE), a platinum wire counter electrode (CE) and a saturated calomel electrode (SHE) as the reference electrode. The counter electrode (CE) is separated from the test solution by a porous tube containing the same supporting electrolyte.

Figure 22.22 illustrates two types of cells that are used for potentiostatic coulometry depending on the type of samples to be used.

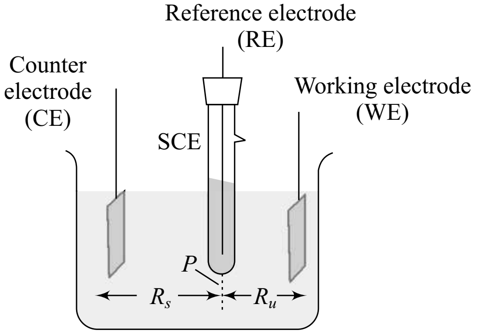


FIGURE 22.21(c) *A cell for potentiostatic coulometer.*

1. Generally, the coulometric cell consists of a platinum-gauze working electrode, Pt wire counter electrode and a saturated calomel electrode [Fig. 22.22(a)]. The counter electrode is separated from the analyte solution by a salt bridge that usually contains the same electrolyte as the solution being analysed. The bridge prevents the reaction products formed at the counter electrode from diffusing into the analyte solution and interfering.
2. A mercury pool-type cathode, Fig. 22.22(b), is particularly useful for separating easily reduced elements as a preliminary step in an analysis. Mercury pool-type cell is also useful for coulometric determination of several metallic cations that form metals that are soluble in mercury and coulometric determination of certain types of organic compounds. This cathode is useful for separating the easily reducible elements during an analysis. For example, copper, nickel, cobalt, silver and cadmium are easily separated from ions such as aluminium, titanium, alkali metals and phosphates. The precipitated elements dissolve in mercury and form amalgams.

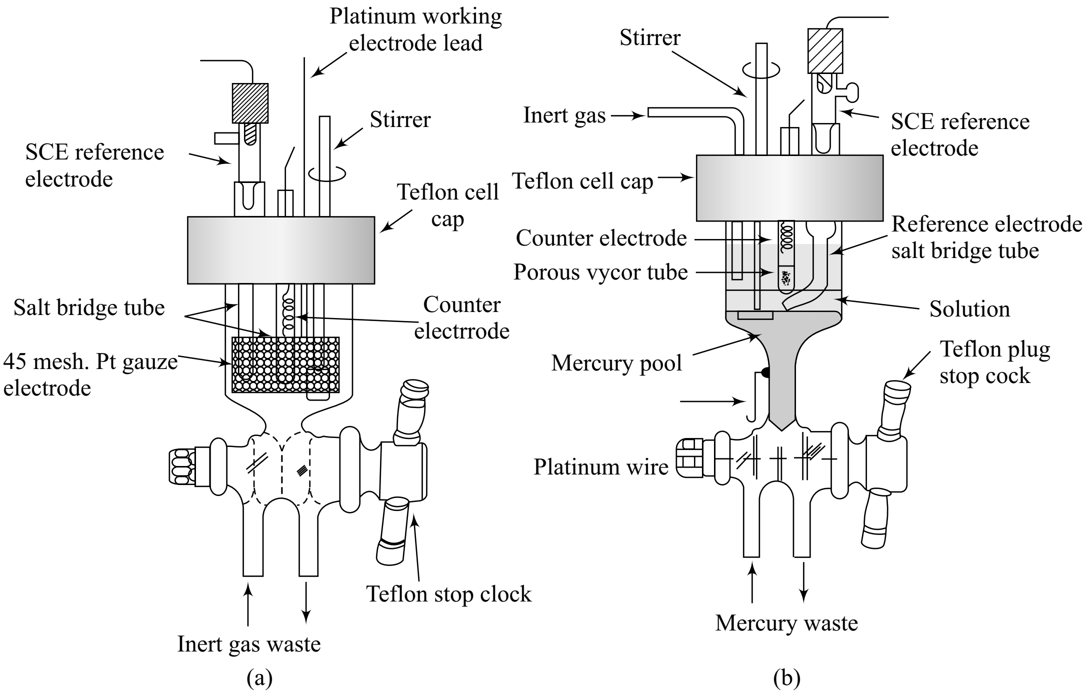


FIGURE 22.22 *Electrolysis cells for potentiostatic coulometry (a) platinum gauze and mercury pool-type.*

Recently, ‘electronic integrators’ are being used in potentiostatic coulometry to determine the number of coulombs required to complete electrolysis.

Another important aspect for controlled-potential electrolysis is efficient stirring of solution, since all analyte species must be swept up to the electrode surface, so that the electrochemical reaction is completed.

In this technique, a ratio of electrode surface to solution volume is to be determined and this ratio plays a role in determining the ‘time’ needed for complete electrolysis, since the electrical current (*I*) will be directly proportional to electrode area, ‘*A*’, and also directly proportional to the concentration, ‘*c*’, of the analyte.

$$I \,\alpha\, c \times A$$

$Q = I \times t$, where Q is the coulomb of electricity consumed for complete electrolysis.

$$Q = I \times t \,\alpha\, c \times A \times t$$

22.3.4.2.1 Applications

(i) Potentiostatic coulometry techniques have been used to determine over 50 elements.

(ii) They are used very frequently for the determination of uranium and plutonium since potentiostatic coulometry techniques do not interfere much with the sample.

(iii) Another useful application is the determination of the amount of oxygen in a sample by using a cadmium electrode and a porous silver electrode. This system for determining oxygen is convenient because it does not require an external power supply or a potentiostat to control the reaction since it does not reach a potential that is high enough to oxidize other substances.

(iv) In addition, potentiostatic techniques can be used to separate materials, to measure the diffusion current and to determine and sometimes synthesize organic compounds.

(v) Inorganic analysis: Controlled-potential coulometric methods have widespread use in the determination of several metal ions. As many as 55 elements of the periodic table can be determined by the cathodic reduction of metal ions to metallic state. Most of the metals (about two dozen element) can form amalgams with mercury, and hence controlled-potential coulometry with mercury cathode is usually preferred.

(vi) Analysis of radioactive materials: The technique is widely adopted for the determination of uranium and plutoninum and thus finds extensive use in the nuclear energy field. Reduction of UO_2^{+2} to U^{4+} can be carried out in H_2SO_4 medium with a mercury pool cathode (–0.6 V vs. SCE). Samples containing 7–75 mg of uranium have been analysed with an accuracy of ±0.1%.

(vii) Micro analysis: Controlled-potential coulometry is more popular than the electrogravimetric methods since it avoids the final step of weighing the product. The tedious process of drying and weighing the electrode after each electrolysis is avoided. This technique is especially useful for the determination of small amounts of analyte (0.01–1 mg) with an accuracy of (±0.5%).

(viii) Multistep controlled-potential electrolysis: Determination of several metal ions in the same solution is possible with controlled-potential electrolysis using a mercury pool cathode. A sample solution containing several metal ions such as Cu^{2+}, Bi^{3+}, Pb^{2+} and Zn^{2+} can be analysed by controlled-potential analysis. When the cathode potential is controlled at about +0.008 vs. SCE, Cu^{2+} is reduced to Cu(s). When the current decays to zero, then the potential can be controlled so that Bi^{3+} can be reduced to Bi. Subsequently by controlling the potentials, Pb^{2+} can be determined as Pb and Zn^{2+} as Zn. A sample of brass/bronze which contains Cu^{2+}, Zn^{2+}, and Pb^{2+} can be analysed by this technique.

(ix) Continuous monitoring of gas streams: Determination of trace level concentration of oxygen in a gas stream can be done by controlled-potential analysis. The cell consists of a porous silver cathode and a cadmium anode (Fig. 22.23).

The reactions are

$$O_2(g) + 2H_2O + 4e^- \rightarrow 4\ OH^- \text{ (Cathode)}$$

$$Cd(s) + 2OH^- \rightarrow Cd(OH)_2(s) + 2e \text{ (Anode)}$$

The porous silver cathode serves to break up the incoming gas into small bubbles, wherein the reduction of oxygen takes place quantitatively within the pores. The hydroxyl ions formed during reduction reacts with the cadmium anode and forms a $Cd(OH)_2(s)$ product.

A special feature of this set-up is that a galvanic cell is formed and hence no external power supply is needed. There is no need of a potentiostat to control the potential of the working electrode. The current produced is passed through a standard resistor and the potential drop is recorded. The oxygen concentration is proportional to this potential and a digital display can indicate the oxygen concentration directly. The set-up can be used for the determination of oxygen in any gas stream and can detect oxygen concentration from 1 ppm to 1%.

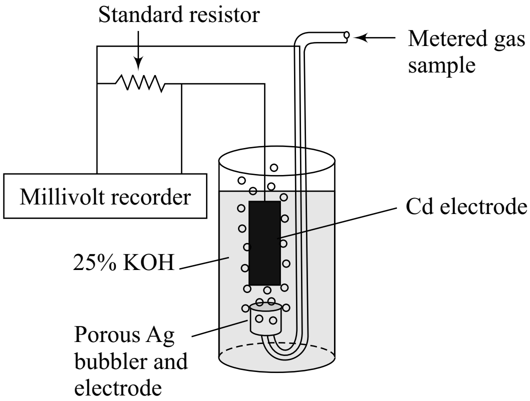


Figure 22.23 *An instrument for continuous recording the oxygen content of a gas stream.*

(x) Electrolytic determination of organic compounds: Controlled-potential coulometry offers a new step for the electrolytic determination of organic compounds. Trichloroacetic acid and picric acid are quantitatively reduced at a mercury cathode. Coulometric methods permit the analysis of these compounds with an accuracy of 0.1%.

$$Cl_3CCOO^- + H^+ + 2e \rightarrow Cl_2HCCOO^- + Cl^-$$

(xi) Electrolytic synthesis of new organic compounds: Synthesis of new species and novel chemical compounds are possible. No chemical reagents are required since electron itself is the reagent for carrying out these reactions. No contamination of the products takes place.

(xii) Determination of *n*-values of the reaction: Controlled-potential coulometry can be used to determine '*n*' values of the reactions. Determination of '*n*' values offers a route to deduce the kinetics and mechanism of the overall reactions. Picric acid is reduced at a mercury pool cathode, in which '*n*' value of the reaction was found to be 18. The reaction product is identified to be triaminophenol (Fig. 22.24).

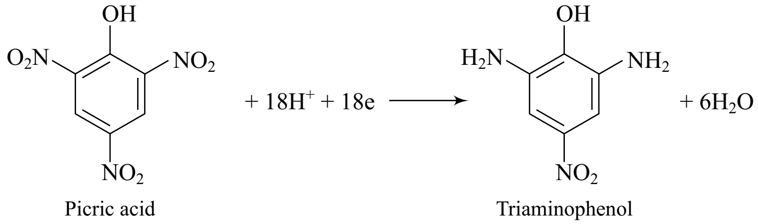


Figure 22.24 *Controlled-potential coulometric reduction of picric acid to triaminophenol.*

22.3.4.2.2 Advantages

- It is more specific than amperostatic coulometry
  - avoids redox of species that may interfere with amperostatic coulometry
  - can be used for over 55 elements without major interference

22.3.4.2.3 Disadvantages

- It does take longer than amperostatic titration
    - current (i) decreases with time
    - conversion becomes slower as less analyte around to oxidize or reduce

Review Questions

1. Describe fundamentals of bulk electrolysis in an electrochemical cell with an example.
2. Give detailed discussion of Ohmic potential (IR drop) in an electrochemical cell.
3. What is polarization of an electrode in an electrochemical cell?
4. Write a note on concentration polarization.
5. Write a note on kinetic polarization.
6. Mention the advantages of overvoltage in a cell.
7. State Faraday's laws.
8. What is electrogravimetry? Discuss.
9. What are the types of electrogravimetric methods adopted?
10. Describe constant current electrogravimetry with a neat figure.
11. Describe constant potential electrogravimetry with a neat figure.
12. Compare the two methods of electrogravimetry and discuss the merits and demerits.
13. A note on mercury electrode as a cathode in electrogravimetry.
14. Discuss the applications of electrogravimetry.
15. Explain the determination of copper by electrogravimetry.
16. How is copper determined in alloy by electrogravimetric method? Give details.
17. Describe how analysis of copper in a mixture is determined by electrogravimetry.
18. What is coulometry? Discuss.
19. What are the types of coulometers in coulometry?
20. Describe in detail Gasometric (hydrogen–oxygen) coulometer.
21. What is silver coulometer?
22. What is iodine coulometer?
23. What is electronic coulometer?
24. Describe with neat figure working of copper coulometer in an electrochemical coulometer.
25. Mention the types of coulometric methods.
26. Discuss in detail amperostatic (constant current) coulometry and mention the basic principles involved in the method of determination.
27. Mention how the determination of end point of electrolysis in constant current coulometric analysis is done.
28. Mention different methods of detection of end points in coulometric titrations.
29. What is coulometric titration? Explain coulometric acid-base titrations.
30. Write a note on coulometric precipitation and complex formation titrations.
31. Write a note on redox coulometric titrations.
32. Determine the amount of water in a sample using Karl Fischer reaction in coulometric titration.
33. Explain the principle and method for the determination of film thickness of metallic coating by the coulometric technique.
34. Describe coulometric determination of arsenic by electrogenerated bromine by dead-stop endpoint detection.
35. Compare coulometric titrations with that of volumetric titrations.
36. Give the applications of coulometric titrations.
37. What are the advantages of coulometric methods?
38. Mention the basic principle and operation of controlled coulometric analysis.
39. Explain the instrumentation and working principle of potentiostatic coulometry.
40. What are the applications of potentiostatic coulometry.
41. Mention the advantages and disadvantages of potentiostatic coulometry.

23

Supercapacitors

Chapter Outline

Introduction. Types of supercapacitors. Electrochemical double-layer capacitors, activated carbons, carbon aerogels, carbon nanotubes. Pseudocapacitors—conducting polymers, metal oxides. Hybrid supercapacitors—composite, asymmetric, battery-type. Advantages of supercapacitors. Limitations. Applications.

23.1 Introduction

In response to the changing global landscape, energy has become a primary focus of the major world powers and scientific community. As the concern grows over fossil fuel usage, in terms of global warming and resource depletion, there will be a progressive swing to renewable energy. This will necessitate the development of improved methods for storing electricity when it is available and retrieving when it is needed. There has been great interest in developing and refining more efficient energy storage devices. One of such devices is the supercapacitor that has matured significantly over the last decade and emerged with the potential to facilitate major advances in energy storage. In the field of energy storage, two main parameters are fundamental for any storage devices: the 'energy density' and the 'power density'. The first parameter defines the amount of energy that can be stored in a given volume or weight. The power density defines the way this energy can be stored into the device. If this parameter is high, the time for loading and unloading the amount of needed energy is reduced. The ideal storage device should satisfy both high energy density and high power density. Supercapacitors are governed by the same fundamental equations as conventional capacitors, but utilize higher surface area electrodes and thinner dielectrics to achieve greater capacitances. This allows for energy densities greater than those of conventional capacitors and power densities greater than those of batteries. As a result, supercapacitors may become an attractive power solution for an increasing number of applications such as power storage devices in traction, space flight technology, power electronics and other related fields.

The supercapacitor technology has emerged with potential to enable major advances in energy storage; and in the future supercapacitors, therefore, represent a new breed of technology that occupies a niche amongst other energy storage devices that were previously vacant. It is also stressed that there is substantial scope for technology development in this newly emerging area, where materials science and polymer technology will have a pivotal role in conjunction with electrochemistry.

Many renewable resources such as wind and solar power are intermittent, i.e. they are not available all the time. Storing energy from the renewable source allows supply to more closely match demand. For example, a storage system attached to a wind turbine could store energy captured around the clock, whenever the wind blows; and then dispatch that energy into the consumer market. Such energy storage enables solar electricity to be used both day and night.

Supercapacitor [also known as electric/electrochemical double-layer capacitor (EDLC)] is a unique electrical storage device, which can store much more energy than conventional capacitors and offer much higher power density than batteries. EDLCs fill up the gap between the batteries and the conventional capacitor (Fig. 23.1), allowing applications for various power and energy requirements, i.e. backup power sources for electronic devices, load-levelling, engine start or acceleration for hybrid vehicles and electricity storage generated from solar or wind energy.

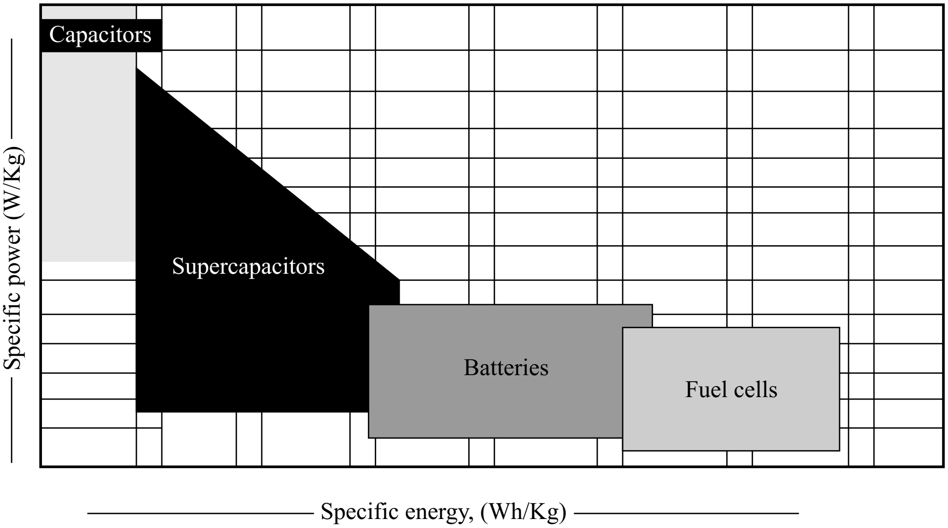


FIGURE 23.1 *Comparison of different electrochemical devices in terms of their specific power and energies.*

Electric double-layer capacitor (EDLC) works on the principle of double-layer capacitance at the electrode–electrolyte interface where electric charges are accumulated on the electrode surfaces and ions of opposite charge are arranged on the electrolyte side. Electrical energy can be stored directly in an electrostatic way as negative and positive electric charges on the plates of an EDLC by a process termed as nonfaradaic electrical energy storage. Generally, these supercapacitors have the following characteristics:

- Greater density of stored energy,
- High power discharge; rapid charge and discharge,
- Long useful life,
- Start quickly, accelerate quickly, climbing powerfully,
- The capacitance is 30 times as the lead acid batteries, it is the most important characteristic of the electrical automobile.

There are two main types of double-layer capacitors as classified by the charge storage mechanism:

(a) The capacitance associated with charging and discharging of the electrical double-layer at the electrode–electrolyte interface and are hence called electrical double-layer capacitors (EDLCs),

(b) The pseudocapacitance with electrosorption or surface redox reactions which are referred to as pseudocapacitors.

A new type hybrid supercapacitor is also available in combination with the above two supercapacitors (Fig. 23.2).

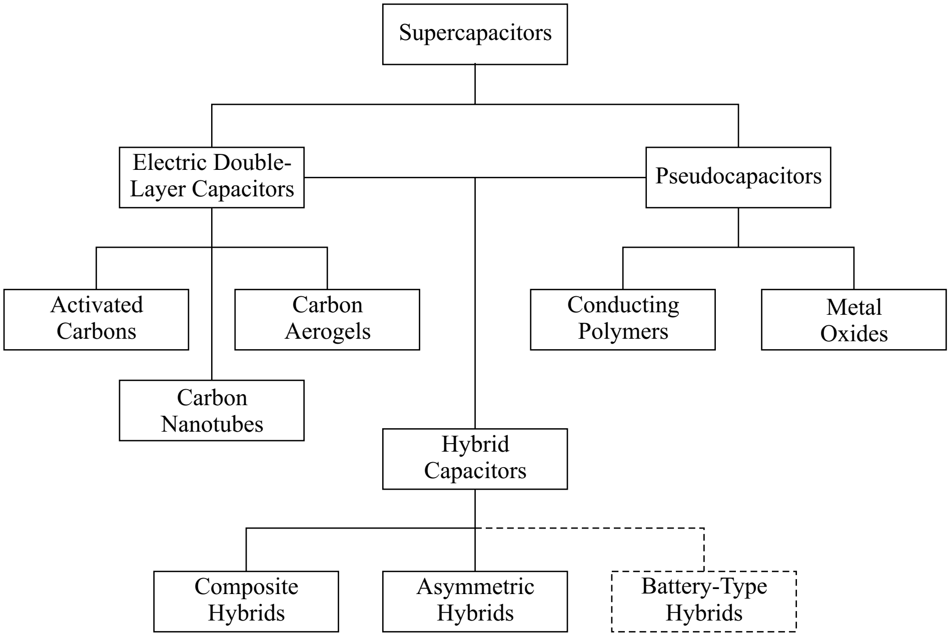


Figure 23.2 *Different types of supercapacitors.*

23.2 Types of Supercapacitors

23.2.1 Electrochemical Double-Layer Capacitors

A typical EDLC is shown in Fig. 23.3. Two carbon-based electrodes, an electrolyte and a separator are used for the construction of electrochemical double-layer capacitors (EDLCs). Like conventional capacitors, EDLCs store charge electrostatically, or nonfaradaically and there is no transfer of charge between the electrode and the electrolyte.

EDLCs store electrical energy as result of an electrochemical double-layer of charge. As voltage is applied, charge accumulates on the electrode surfaces. Due to electrostatic attraction of charges, ions in the electrolyte solution diffuse across the separator into the pores of the electrode of opposite charge. However, the electrodes are engineered to prevent the recombination of the ions. Thus, a doublelayer of charge is produced at each electrode as shown in Fig. 23.3. These double-layers, coupled with an increase in surface area and a decrease in the distance between electrodes, allow EDLCs to achieve higher energy densities than conventional capacitor.

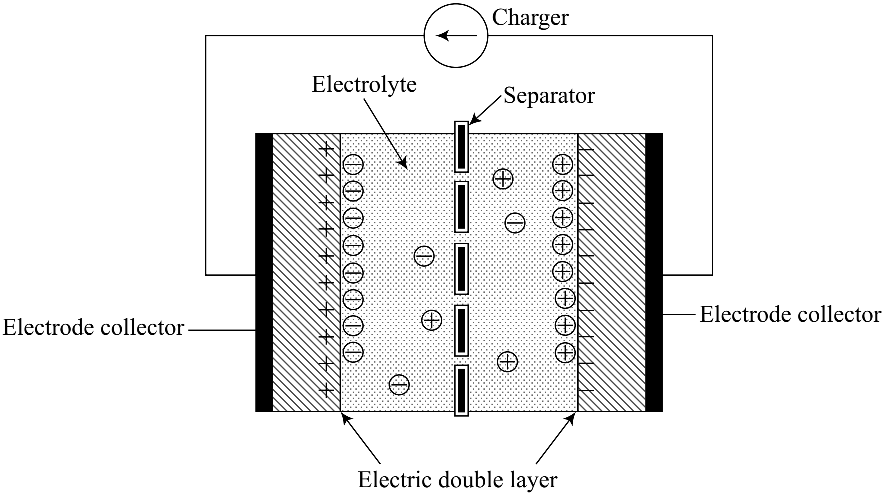


Figure 23.3 *A typical ELDC.*

Because there is no transfer of charge between the electrolyte and the electrode, there are no chemical or composition changes associated with nonfaradaic processes. For this reason, charge storage in EDLCs is highly reversible, which allows them to achieve very high cycling stabilities. EDLCs generally operate with stable performance characteristics for a great many charge–discharge cycles. Because of their cycling stability, EDLCs are well suited for applications that involve nonuser serviceable locations, such as deep sea or mountain environments.

It can be seen that the nature of electrolytes used also changes the characteristics of EDLC performance. An EDLC can utilize either an aqueous or organic electrolyte. When aqueous electrolytes are used, EDLCs generally have lower equivalent series resistance (ESR) and lower minimum pore size requirements compared to organic electrolytes. However, aqueous electrolytes also have lower breakdown voltages. Therefore, the selection of an electrolyte in EDLCs for the purpose depends on the intended application of the supercapacitor which is based on capacitance, ESR and voltage. Therefore, nature of the electrolyte is of great importance in supercapacitor design; and the types of EDLCs are distinguished primarily by the form of carbon they use as an electrode material. Carbon electrode materials generally have higher surface area, lower cost and more established fabrication techniques than other materials, such as conducting polymers and metal oxides. Energy storage in EDLCs is also achievable using different forms of carbon materials such as activated carbon, carbon aerogels and carbon nanotubes.

(i) Activated carbons

Activated carbon is the most commonly used electrode material in EDLCs. The activated carbon possesses a higher surface area than other carbon-based materials and is less expensive. It can be observed that the activated carbon can be obtained in different complex porous structures of micropores(<20 Å wide) to macropores (>500 Å) to achieve their high surface areas. It is evident that not all the high surface area of the activated carbon contributed to the capacitance; for reasons that the cause is due to the large ionic size of the electrolyte which does not diffuse through smaller micropores affecting the charge storage.

Research also suggests an empirical relationship between the distribution of pore sizes, the energy density and the power density of the device. Larger pore sizes correlate with higher power densities and smaller pore sizes correlate with higher energy densities. As a result, the pore size distribution of activated carbon electrodes is a major area of research in EDLC design. In particular, researchers have focused on determining the optimal pore size for a given ion size and upon improving the methods used to control the pore size distribution during fabrication.

(ii) Carbon aerogels

Carbon aerogels are also being used as an electrode material for EDLCs. Carbon aerogels are formed from a continuous network of conductive carbon nanoparticles with interspersed mesopores (mesopores (20–500 Å). Due to this continuous structure and their ability to bond chemically to the current collector, carbon aerogels do not require the application of an additional adhesive binding agent. As a binderless electrode, carbon aerogels have been shown to have a lower ESR (equivalent series resistance) than activated carbons. This reduced ESR, which yields higher power, is the primary area of interest in supercapacitor research involving carbon aerogels.

(iii) Carbon nanotubes

Carbon nanotubes as an ELDC electrode material with an open and accessible network of mesopores are utilized in supercapacitor. Unlike other carbon-based electrodes, the mesopores in carbon nanotube electrodes are interconnected, allowing a continuous charge distribution that uses almost all of the available surface area. Thus, the surface area is utilized more efficiently to achieve capacitances comparable to those in activated-carbon-based supercapacitors, even though carbon nanotube electrodes have a modest surface area compared to activated carbon electrodes. Because the electrolyte ions can more easily diffuse into the mesoporous network, carbon nanotube electrodes also have a lower ESR than activated carbon. In addition, several fabrication techniques have been developed to reduce the ESR even further. In particular, carbon nanotubes can be grown directly onto the current collectors, subjected to heat-treatment or cast into colloidal suspension thin films. The efficiency of the entangled mat structure allows energy densities comparable to other carbon-based materials and the reduced ESR allows higher power densities.

23.2.2 Pseudocapacitors

Energy storage in psuedocapacitors is through the transfer of charge between an electrode and an electrolyte faradaic process. The process of energy storage is accomplished by electrosorption, reduction-oxidation reactions and intercalation processes. These processes are responsible for achieving greater capacitances and energy densities in psuedocapacitors than those in EDLCs. Conducting polymers and metal oxides are, generally, the two electrode materials that are being used to store charge in pseudocapacitors.

1. Conducting polymers

Conducting polymers are also used as an electrode material for storage of electrical energy because they have a relatively high capacitance and conductivity. In addition, these materials have a relatively low ESR and cost compared to carbon-based electrode materials. A psuedocapacitor can be constructed utilizing both *n*- and *p*-type conducting polymeric electrode materials which have the greatest potential energy and power densities; however, a lack of efficient, *n*-doped conducting polymer materials has prevented these pseudocapacitors from reaching their potential. However, the mechanical stress on conducting polymers during reduction-oxidation reactions limits the stability of these pseudocapacitors and has limitation for charge–discharge cycles. This reduced cycling stability has hindered the development of conducting polymer pseudocapacitors.

2. Metal oxides

Metal oxides with high electrical conductivity are preferred as an electrode material for psuedocapacitors. Ruthenium oxide is used as an electrode material due to its comparable high capacitance, while other metal oxides have yet to obtain comparable capacitances. The reason for comparable high capacitance is due to the insertion and removal or intercalation of protons into its amorphous structure. In its hydrous form, the capacitance exceeds that of carbon-based and conducting polymer materials. Furthermore, the ESR of hydrous ruthenium oxide is lower than that of other electrode materials. As a result, ruthenium oxide pseudocapacitors may be able to achieve higher energy and power densities than similar EDLCs and conducting polymer pseudocapacitors. However, despite this potential, the success of ruthenium oxide has been limited due to its prohibitive cost. Thus, a major area of research is the development of fabrication methods and composite materials to reduce the cost of ruthenium oxide, without reducing the performance.

23.2.3 Hybrid Supercapacitors

It is realized that hybrid capacitors with better performance characteristics have been obtained and are observed to have relative advantages and mitigate the relative disadvantages of both EDLCs and psuedocapacitors. Utilizing both faradaic and nonfaradaic processes to store charge, hybrid capacitors have achieved energy and power densities greater than EDLCs without the sacrifices in cycling stability and affordability that have limited the success of pseudocapacitors. There are three different types of hybrid capacitors, distinguished by their electrode configuration: composite, asymmetric and battery-type respectively.

(a) Composite

A polymer composite electrode material obtained from a blend of carbon nanotube and polypyrrole (or metal oxide materials) is used to construct a hybrid capacitor to procure better performance characteristics which incorporate both physical and chemical charge storage mechanisms together in a single electrode. The carbon-based materials facilitate a capacitive double-layer of charge and also provide a high-surface-area backbone that increases the contact between the deposited pseudocapacitive materials and the electrolyte. The pseudocapacitive materials are able to further increase the capacitance of the composite electrode through faradaic reactions. Several experiments have demonstrated that this electrode is able to achieve higher capacitances than either a pure carbon nanotube or a pure polypyrrole polymer-based electrode. This is attributed to the accessibility of the entangled mat carbon nanotube structure, which allows a uniform coating of polypyrrole and a three-dimensional distribution of charge. Moreover, the structural integrity of the entangled mat has been shown to limit the mechanical stress caused by the insertion and removal of ions in the deposited polypyrrole. Therefore, unlike conducting polymers, these composites have been able to achieve a cycling stability comparable to that of EDLCs.

(b) Asymmetric

Asymmetric hybrids combine faradaic and nonfaradaic processes by coupling an EDLC electrode with a psuedocapacitor electrode. In particular, the coupling of an activated carbon negative electrode with a conducting polymer positive electrode has received a great deal of attention. The lack of an efficient, negatively charged, conducting polymer material has limited the success of conducting polymer pseudocapacitors. The implementation of a negatively charged, activated carbon electrode attempts to circumvent this problem. While conducting polymer electrodes generally have higher capacitances and lower resistances than activated carbon electrodes, they also have lower maximum voltages and less cycling stability. Asymmetric hybrid capacitors that couple these two electrodes mitigate the extent of this trade-off to achieve higher energy and power densities than comparable EDLCs. Also, they have better cycling stability than comparable pseudocapacitors.

(c) Battery type

A unique supercapacitor known as battery type is obtained by coupling a supercapacitor electrode with a battery electrode, which is similar to asymmetric hybrids consisting two different electrodes. This typical configuration reflects the demand for higher energy supercapacitors and higher power batteries, combining the energy characteristics of batteries with the power, cycle life and recharging times of supercapacitors. The construction of battery-type supercapacitor is focused primarily on using nickel hydroxide, lead dioxide and LTO ($Li_4Ti_5O_{12}$) as one electrode and activated carbon as the other. Results of preliminary studies on this battery-type hybrid indicate that these hybrids may be able to bridge the gap between supercapacitors and batteries. In the future, however, detailed research has to be carried out to exploit the full potential of battery-type hybrids.

23.2.4 Advantages of Supercapacitors

- Energy within a supercapacitor is quickly available—this is one of its greatest advantages
- much higher power density than batteries
- Virtually unlimited cycle life,10 to 12 year life, which reduces maintenance costs.
- self-discharge rate is much higher than with batteries.
- Low impedance enhances load handling when put in paralleled with a battery.
- Rapid charging—super capacitors charge in seconds.
- Environmentally friendly
- no danger of overcharge.
- Safe
- Simple charge methods—no full-charge detection is needed
- High cycle efficiency (95% or more)
- Lightweight

23.2.5 Limitations

- Linear discharge voltage prevents use of the full energy spectrum.
- Low energy density—typically holds one-fifth to one-tenth the energy of an electrochemical battery.
- Cells have low voltages—serial connections are needed to obtain higher voltages. Voltage balancing is required if more than three capacitors are connected in series.
- High self-discharge—the rate is considerably higher than that of an electrochemical battery.

23.2.6 Applications

- Reserve power
- Electrical toy
- Start high power
- Store the solar energy and wind energy
- Igniter, solar street light, etc.
- Starting diesel trucks and railroad locomotives, and in electric/hybrid-electric vehicles for transient load levelling and capturing the energy used in braking.
- In power systems, they are most likely to be used as bridging power sources in uninterruptible power supplies, much like flywheels.
- Smart grids including renewable energy sources.

Review Questions

1. What is a supercapacitor?
2. How does it differ from a capacitor and a battery?
3. Mention the mechanism of energy storage in a capacitor and a battery.
4. How are supercapacitors classified on the basis of mechanism of energy storage?
5. Mention the types of supercapacitors.
6. Write a note on electric double-layer capacitor (EDLC).
7. Write note on activated carbon, carbon aerogels and carbon nanotube EDLC.
8. What is a psuedocapacitor? How does it differ from an EDLC.
9. Write a brief note on the following:
   (i) Polymer composite and (ii) metal oxide-type capacitors.
10. What is a hybrid capacitor? Mention the types of hybrid capacitors.
11. Write a brief note on the following:
    (i) polymer composite hybrid and (ii) Asymmetric and battery type hybrids.
12. Mention the advantage of supercapacitors.
13. What are the limitations of supercapacitors.
14. Mention the applications of supercapacitors

Index