MATERIAL SCIENCE

S. L. KAKANI
AMIT KAKANI

NEW AGE INTERNATIONAL PUBLISHERS

MATERIAL SCIENCE

This page intentionally left blank

MATERIAL SCIENCE

S. L. KAKANI
and
AMIT KAKANI

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
New Delhi • Bangalore • Chennai • Cochin • Guwahati • Hyderabad
Jalandhar • Kolkata • Lucknow • Mumbai • Ranchi
Visit us at **www.newagepublishers.com**

Copyright © 2004, New Age International (P) Ltd., Publishers
Published by New Age International (P) Ltd., Publishers

All rights reserved.

No part of this ebook may be reproduced in any form, by photostat, microfilm, xerography, or any other means, or incorporated into any information retrieval system, electronic or mechanical, without the written permission of the publisher. *All inquiries should be emailed to* ***rights@newagepublishers.com***

ISBN (13) : 978-81-224-2656-4

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
4835/24, Ansari Road, Daryaganj, New Delhi - 110002
Visit us at **www.newagepublishers.com**

Preface

Material Science has become a very important subject as an interdisciplinary course in almost all universities. Keeping this in view we have developed the subject matter for *B.E.* (Material Science and Metallurgy/Production Engineering/Mechanical Engineering), *A.M.I.E.*, Diploma in engineering, *M.Sc.* (Material Science/Physics/Chemistry) and *B.Sc.* (Hons.) courses as per their latest syllabii. The book is also useful for *UPSC*, *GATE*, *NET*, *SLET* and other entrance examinations.

A reasonably wide coverage in sufficient depth has been attempted, giving the importance to the basic principles, essential theory and experimental details necessary for understanding the nature, properties and applications of materials. All efforts have been made to provide topics which are of great use to the readers, e.g. semiconductors, superconductors, polymers, composites, nanostructured materials, etc. Latest developments, e.g. quantum dots, spinotrics, MOSFET, Microelectronic circuits, MEMS, nanotechnology, etc. are also covered.

To make the text more useful, good number of worked out problems, review questions, problems, short-question answers, typical objective questions, suggested readings are provided with each chapter.

We are thankful to M/s *New Age International (P) Limited Publishers*, for their untiring efforts in bringing out the book with excellent printing and nice get up within the shortest possible time period.

Suggestions for the improvement of the book are most welcome.

Bhilwara **S.L. Kakani**
February 2004 **Amit Kakani**

This page intentionally left blank

Contents

3. Crystal Geometry, Structure and Defects 64

This page intentionally left blank

1

Classification and Selection of Materials

1. INTRODUCTION

Materials science and engineering plays a vital role in this modern age of science and technology. Various kinds of materials are used in industry, housing, agriculture, transportation, etc. to meet the plant and individual requirements. The rapid developments in the field of quantum theory of solids have opened vast opportunities for better understanding and utilization of various materials. The spectacular success in the field of space is primarily due to the rapid advances in high-temperature and high-strength materials.

The selection of a specific material for a particular use is a very complex process. However, one can simplify the choice if the details about (i) operating parameters, (ii) manufacturing processes, (iii) functional requirements and (iv) cost considerations are known. Factors affecting the selection of materials are summarized in Table 1.1.

Table 1.1 Factors affecting selection of materials

(i) *Manufacturing processes*	*(ii)* *Functional requirements*	*(iii)* *Cost considerations*	*(iv)* *Operating parameters*
• Plasticity	• Strength	• Raw material	• Pressure
• Malleability	• Hardness	• Processing	• Temperature
• Ductility	• Rigidity	• Storage	• Flow
• Machinability	• Toughness	• Manpower	• Type of material
• Casting properties	• Thermal conductivity	• Special treatment	• Corrosion requirements
• Weldability	• Fatigue	• Inspection	• Environment
• Heat	• Electrical treatment	• Packaging properties	• Protection from fire
• Tooling	• Creep	• Inventory	• Weathering
• Surface finish	• Aesthetic look	• Taxes and custom duty	• Biological effects

There are thousands and thousands of materials available and it is very difficult for an engineer to possess a detailed knowledge of all the materials. However, a good grasp of the fundamental principles which control the properties of various materials help one to make the optimum selection of material. In this respect, materials science and engineering draw heavily from the engineering branches, e.g. metallurgy, ceramics and polymer science.

The subject of material science is very vast and unlimited. Broadly speaking, one can sub-divide the field of study into following four branches: (i) Science of metals, (ii) Mechanical behaviour of metals (iii) Engineering metallurgy and (iv) Engineering materials. We shall discuss them in subsequent chapters.

2. ENGINEERING REQUIREMENTS

While selecting materials for engineering purposes, properties such as impact strength, tensile strength, hardness indicate the suitability for selection but the design engineer will have to make sure that the radiography and other properties of the material are as per the specifications. One can dictate the method of production of the component, service life, cost etc. However, due to the varied demands made metallic materials, one may require special surface treatment, e.g. hardening, normalising to cope with the service requires. Besides, chemical properties of materials, e.g. structure, bonding energy, resistance to environmental degradation also effect the selection of materials for engineering purposes.

In recent years polymeric materials or plastics have gained considerable popularity as engineering materials. Though inferior to most metallic materials in strength and temperature resistance, these are being used not only in corrosive environment but also in the places where minimum wear is required, e.g. small gear wheels, originally produced from hardened steels, are now manufactured from nylon or teflon. These materials perform satisfactorily, are quiet and do not require lubrication.

Thus, before selecting a material or designing a component, it is essential for one to understand the requirements of the process thoroughly, operating limitations like hazardous or non-hazardous conditions, continuous or non-continuous operation, availability of raw materials as well as spares, availability of alternate materials vis-a-vis life span of the instrument/equipment, cost etc. Different materials possess different properties to meet the various requirement for engineering purposes. The properties of materials which dictate the selection are as follows:

(a) Mechanical Properties The important mechanical properties affecting the selection of a material are:

(i) *Tensile Strength*: This enables the material to resist the application of a tensile force. To withstand the tensile force, the internal structure of the material provides the internal resistance.

(ii) *Hardness*: It is the degree of resistance to indentation or scratching, abrasion and wear. Alloying techniques and heat treatment help to achieve the same.

(iii) *Ductility*: This is the property of a metal by virtue of which it can be drawn into wires or elongated before rupture takes place. It depends upon the grain size of the metal crystals.

(iv) *Impact Strength*: It is the energy required per unit cross-sectional area to fracture a specimen, i.e., it is a measure of the response of a material to shock loading.

(v) *Wear Resistance*: The ability of a material to resist friction wear under particular conditions, i.e. to maintain its physical dimensions when in sliding or rolling contact with a second member.

(vi) *Corrosion Resistance*: Those metals and alloys which can withstand the corrosive action of a medium, i.e. corrosion processes proceed in them at a relatively low rate are termed corrosion-resistant.

(vii) *Density*: This is an important factor of a material where weight and thus the mass is critical, i.e. aircraft components.

(b) Thermal Properties The characteristics of a material, which are functions of the temperature, are termed its thermal properties. One can predict the performance of machine components during normal operation, if he has the knowledge of thermal properties. Specific heat, latent heat, thermal conductivity, thermal expansion, thermal stresses, thermal fatigue, etc. are few important thermal properties of materials. These properties play a vital role in selection of material for engineering applications, e.g. when materials are considered for high temperature service. Now, we briefly discuss few of these properties:

(i) *Specific Heat (c)*: It is the heat capacity of a unit mass of a homogeneous substance. For a homogeneous body, $c = C/M$, where C is the heat capacity and M is the mass of the body. One can also define it as the quantity of heat required to raise the temperature of a unit mass of the substance through 1°C. Its units are cal/g/°C.

(ii) *Thermal Conductivity (K):* This represents the amount of heat conducted per unit time through a unit area perpendicular to the direction of heat conduction when the temperature gradient across the heat conducting element is one unit. Truly speaking the capability of the material to transmit heat through it is termed as the thermal conductivity. Higher the value of thermal conductivity, the greater is the rate at which heat will be transferred through a piece of given size. Copper and aluminium are good conductors of heat and therefore extensively used whenever transfer of heat is desired. Bakelite is a poor conductor of heat and hence used as heat insulator.

The heat flow through an area *A* which is perpendicular to the direction of flow is directly proportional to the area (*A*) and thermal gradient (*dt*/*dx*). Thermal conductivity (*K*) is given by

$$K = \frac{Qx}{A(\theta_1 - \theta_2)t} \quad \text{k Cal/m/°C/s or J/m/s/k} \quad \text{or} \quad \text{W/m/k} \tag{1}$$

where $Q \rightarrow$ flow of heat (k cal), $A \rightarrow$ face area (m^2), $t \rightarrow$ time (second), θ_1 and θ_2 are temperatures of hot and cold side of the material (°C) and x is the distance between the two faces (*m*).

The thermal conductivity of a metal can be expressed as

$$K = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 T \, \frac{ne^2 \lambda}{2 m v_0} \tag{2}$$

where $\lambda \rightarrow$ mean free path, $k \rightarrow$ Boltzmann Constant, $m \rightarrow$ electron mass, $e \rightarrow$ electronic charge, $v_0 \rightarrow$ initial velocity of the electron. We must note that similar expression is used for electrical conductivity. The ratio of heat and electrical conductivity (k and σ respectively) is given by

$$\frac{K}{\sigma} = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 T \tag{3}$$

Obviously, the thermal conductivity (*K*) and electrical conductivity (σ) vary in the same fashion from one material to another. The ratio $\frac{\pi^2}{3}\left(\frac{k}{e}\right)^2$ is known as *Wridemann—Franz ratio*. Thermal conductivity for some of the materials is given in Table 1.1

Table 1.2 Thermal conductivity for some materials

Type of the material	*Material*	*Thermal conductivity (K) (W/m/k)*
(i) Metals	Copper	380
	Aluminium	230
	Cast iron	52
	Mild steel	54
	Stainless steel	16
(ii) Ceramics	Alumina	2.0
	Titanium Carbide	3.0
	Glass	1.0
(iii) Polymers	Bakelite	0.23
		0.0019
(iv) Composites	Concrete	1.4
	Wood	0.14

(iii) Thermal Expansion: All solids expand on heating and contract on cooling. Thermal expansion may take place either as linear, circumferential or cubical. A solid which expands equally in three mutually orthogonal directions is termed as *thermally isotropic*. The increase in any linear dimension of a solid, e.g. length, width, height on heating is termed as *linear expansion*. The *coefficient of linear expansion* is the increase in length per unit length per degree rise in temperature. The increase in volume of a solid on heating is called *cubical expansion*. The thermal expansion of solids has its origin in the lattice vibration and lattice vibrations increases with the rise in temperature.

(iv) Thermal Resistance (R_T): It is the resistance offered by the conductor when heat flow due to temperature difference between two points of a conductor. It is given by

$$R_T = \frac{\theta_1 - \theta_2}{H} \text{ second – °C/k Cal}$$

where $H \rightarrow$ rate of heat flow and θ_1 and θ_2 are temperatures at two points (°C).

(v) Thermal Diffusivity (h): It is given by

$$h = \frac{\text{Thermal conductivity } (K)}{\text{Heat capacity } (C_p) \times \text{density } (\rho)} \text{cm}^3/\text{s}$$

$$= \frac{K}{C_p\,\rho} \text{ represent heat requirement per unit volume}$$

A material having high heat requirement per unit volume possesses a low thermal diffusivity because more heat must be added to or removed from the material for effecting a temperature change.

(vi) Thermal Fatigue: This is the mechanical effect of repeated thermal stresses caused by repeated heating and cooling.

The thermal stresses can be very large, involving considerable plastic flow. We can see that fatigue failures can occur after relatively few cycles. The effect of the high part of the temperature cycle on the strength of material plays an important factor in reducing its life under thermal fatigue.

(c) Electrical Properties Conductivity, resistivity, dielectric strength are few important electrical properties of a material. A material which offers little resistance to the passage of an electric current is said to be a good conductor of electricity.

The electrical resistance of a material depends on its dimensions and is given by

$$\text{Resistance} = \text{Resistivity} \times \frac{\text{Length}}{\text{Cross-section area}}$$

Usually resistivity of a material is quoted in the literature. Unit of resistivity is Ohm-metre.

On the basis of electrical resistivity materials are divided as: (i) Conductors (ii) Semiconductors and (iii) Insulators. In general metals are good conductors. Insulators have very high resistivity. Ceramic insulators are most common examples and are used on automobile spark plugs, Bakelite handles for electric iron, plastic coverings on cables in domestic wiring.

When a large number of metals and alloys are sufficiently cooled below transition temperature, T_c, enter the state of superconductivity in which the dc resistivity goes to zero. The estimates of the resistivity in the super-conducting phase place it at less than 4×10^{-25} Ω-m, which is essentially zero for all practical purposes. The highest value of T_c upto 133 K has been reached for mercury cuprate.

(d) Magnetic Properties Materials in which a state of magnetism can be induced are termed magnetic materials. There are five classes into which magnetic materials may be grouped: (i) diamagnetic (ii) paramagnetic (iii) ferromagnetic (iv) antiferromagnetic and (v) ferrimagnetic. Iron, Cobalt, Nickel and some of their alloys and compounds possess spontaneous magnetization. Magnetic oxides like ferrites and garnets could be used at high frequencies. Because of their excellent magnetic properties alongwith their high electrical resistivity these materials today find use in a variety of applications like magnetic recording tapes, inductors and transformers, memory elements, microwave devices, bubble domain devices, recording hard cores, etc. Hysteresis, permeability and coercive forces are some of the magnetic properties of magnetic substances which are to be considered for the manufacture of transformers and other electronic components.

(e) Chemical Properties These properties includes atomic weight, molecular weight, atomic number, valency, chemical composition, acidity, alkalinity, etc. These properties govern the selection of materials particularly in Chemical plant.

(f) Optical Properties The optical properties of materials, e.g. refractive index, reflectivity and absorption coefficient etc. affect the light reflection and transmission.

(g) Structure of Materials The properties of engineering materials mainly depends on the internal arrangement of the atoms on molecules. We must note that in the selection of materials, the awareness regarding differences and similarities between materials is extremely important.

Metals of a single type atom are named *pure metals*. Metals in actual commercial use are almost exclusively alloys, and not pure metals, since it is possible for the designer to realize an infinite variety of physical properties in the product by varying the metallic composition of the alloy. Alloys are prepared from mixed types of atoms. Alloys are classified as *binary alloys*, composed of two components, as ternary alloys, composed of three components or as multi component alloys. Most commercial alloys are multicomponent. The composition of an alloy is described by giving the percentage (either by weight or by atoms) of each element in it.

The basic atomic arrangement or pattern is not apparent in the final component, e.g. a shaft or a pulley but the properties of the individual crystals within the metallic component, which are controlled by the atomic arrangement, are mainly responsible for their application in industry.

One can determine the strength of a piece of metal by its ability to withstand external loading. The structure of metal or alloy responds internally to the applied load by trying to counteract the magnitude of the applied load and thus tries to keep the constituent atoms in their ordered positions if however the load is higher than the force which holds the atoms in place, the metallic bond becomes ineffective and atoms in the metal are then forced into new displaced positions. The movement of atoms from their original positions in the metal is termed as *slip*. The ease with which atoms move or slip in a metal is an indication of *hardness*. We must note that the relative movement of atoms or slip within a material has a direct bearing on the mechanical properties of the material.

3. CLASSIFICATION OF ENGINEERING MATERIALS

The factors which form the basis of various systems of classifications of materials in material science and engineering are: *(i)* the chemical composition of the material, *(ii)* the mode of the occurrence of the material in the nature, *(iii)* the refining and the manufacturing process to which the material is subjected prior it acquires the required properties, *(iv)* the atomic and crystalline structure of material and *(v)* the industrial and technical use of the material.

Common engineering materials that falls within the scope of material science and engineering may be classified into one of the following six groups:

(i) Metals (ferrous and non-ferrous) and alloys
(ii) Ceramics
(iii) Organic Polymers
(iv) Composites
(v) Semi-conductors
(vi) Biomaterials
(vii) Advanced Materials

(i) Metals: All the elements are broadly divided into metals and non-metals according to their properties. Metals are element substances which readily give up electrons to form metallic bonds and conduct electricity. Some of the important basic properties of metals are: *(a)* metals are usually good electrical and thermal conductors, *(b)* at ordinary temperature metals are usually solid, *(c)* to some extent metals are malleable and ductile, *(d)* the freshly cut surfaces of metals are lustrous, *(e)* when struck metal produce typical sound, and *(f)* most of the metals form alloys. When two or more pure metals are melted together to form a new metal whose properties are quite different from those of original metals, it is called an *alloy*.

Metallic materials possess specific properties like plasticity and strength. Few favourable characteristics of metallic materials are high lustre, hardness, resistance to corrosion, good thermal and electrical conductivity, malleability, stiffness, the property of magnetism, etc. Metals may be magnetic, non-magnetic in nature. These properties of metallic materials are due to: *(i)* the atoms of which these metallic materials are composed and *(ii)* the way in which these atoms are arranged in the space lattice.

Metallic materials are typically classified according to their use in engineering as under:

(i) Pure Metals: Generally it is very difficult to obtain pure metal. Usually, they are obtained by refining the ore. Mostly, pure metals are not of any use to the engineers. However, by specialised and very expensive techniques, one can obtain pure metals (purity ~ 99.99%), e.g. aluminium, copper etc.

(ii) Alloyed Metals: Alloys can be formed by blending two or more metals or atleast one being metal. The properties of an alloy can be totally different from its constituent substances, e.g. 18-8 stainless steel, which contains 18%, chromium and 8% nickle, in low carbon steel, carbon is less than 0.15% and this is extremely tough, exceedingly ductile and highly resistant to corrosion. We must note that these properties are quite different from the behaviour of original carbon steel.

(iii) Ferrous Metals: Iron is the principal constituent of these ferrous metals. Ferrous alloys contain significant amount of non-ferrous metals. Ferrous alloys are extremely important for engineering purposes. On the basis of the percentage of carbon and their alloying elements present, these can be classified into following groups:

(a) *Mild Steels:* The percentage of carbon in these materials range from 0.15% to 0.25%. These are moderately strong and have good weldability. The production cost of these materials is also low.

(b) *Medium Carbon Steels:* These contains carbon between 0.3% to 0.6%. The strength of these materials is high but their weldability is comparatively less.

(c) *High Carbon Steels:* These contains carbon varying from 0.65% to 1.5%. These materials get hard and tough by heat treatment and their weldability is poor.

The steel formed in which carbon content is upto 1.5%, silica upto 0.5%, and manganese upto 1.5% alongwith traces of other elements is called *plain carbon steel.*

(d) *Cast Irons:* The carbon content in these substances vary between 2% to 4%. The cost of production of these substances is quite low and these are used as ferrous casting alloys.

(iv) Non-Ferrous Metals: These substances are composed of metals other than iron. However, these may contain iron in small proportion. Out of several non-ferrous metals only seven are available in sufficient quantity reasonably at low cost and used as common engineering metals. These are aluminium, tin, copper, nickle, zinc and magnesium. Some other non-ferrous metals, about fourteen in number, are produced in relatively small quantities but these are of vital importance in modern industry. These includes, chromium, mercury, cobalt, tungsten, vanadium, molybdenum, antimony, cadmium, zirconium, beryllium, niobium, titanium, tantalum and manganese.

(v) Sintered Metals: These materials possess very different properties and structures as compared to the metals from which these substances have been cast. Powder metallurgy technique is used to produced sintered metals. The metals to be sintered are first obtained in powered form and then mixed in right calculated proportions. After mixing properly, they are put in the die of desired shape and then processed with certain pressure. Finally, one gets them sintered in the furnace. We must note that the mixture so produced is not the true alloy but it possesses some of the properties of typical alloys.

(vi) Clad Metals: A ‘sandwich’ of two materials is prepared in order to avail the advantage of the properties of both the materials. This technique is termed as *cladding*. Using this technique stainless steel is mostly embedded with a thick layer of mild steel, by rolling the two metals together while they are red hot. This technique will not allow corrosion of one surface. Another example of the use of this technique is cladding of duralium with thin sheets of pure aluminium. The surface layers, i.e. outside layers of aluminium resist corrosion, whereas inner layer of duralumin imparts high strength. This technique is relatively cheap to manufacture.

4. ORGANIC, INORGANIC AND BIOLOGICAL MATERIALS

Organic materials are carbon compounds and their derivatives. They are solids composed of long molecular chains. The study of organic compounds is very important because all biological systems are composed of carbon compounds. There are also some materials of biological origin which do not possess organic composition, e.g., limestone.

Organic Materials

These materials are carbon compounds in which carbon is chemically bonded with hydrogen, oxygen and other non-metallic substances. The structure of these compounds is complex. Common organic materials are plastics and synthetic rubbers which are termed as *organic polymers*. Other examples of organic materials are wood, many types of waxes and petroleum derivatives. Organic polymers are prepared by polymerisation reactions, in which simple molecules are chemically combined into long chain molecules or three-dimensional structures. Organic polymers are solids composed of long molecular chains. These materials have low specific gravity and good strength. The two important classes of organic polymers are:

(a) Thermoplastics: On heating, these materials become soft and hardened again upon cooling, e.g., nylon, polythene, etc.

(b) Thermosetting plastics: These materials cannot be resoftened after polymerisation, e.g., urea-formaldehyde, phenol formaldehyde, etc. Due to cross-linking, these materials are hard, tough, non-swelling and brittle. These materials are ideal for moulding and casting into components. They have good corrosion resistance.

The excellent resistance to corrosion, ease of fabrication into desired shape and size, fine lusture, light weight, strength, rigidity have established the polymeric materials and these materials are fast replacing many metallic components. PVC (Polyvinyl Chloride) and polycarbonate polymers are widely used for glazing, roofing and cladding of buildings. Plastics are also used for reducing weight of mobile objects, e.g., cars, aircrafts and rockets. Polypropylenes and polyethylene are used in pipes and manufacturing of tanks. Thermo-plastic films are widely used as lining to avoid seepage of water in canals and lagoons.

To protect metal structure from corrosion, plastics are used as surface coatings. Plastics are also used as main ingredients of adhesives. The lower hardness of plastic materials compared with other materials makes them subjective to attack by insects and rodents.

Because of the presence of carbon, plastics are combustible. The maximum service temperature is of the order of 100°C. These materials are used as thermal insulators because of lower thermal conductivity. Plastic materials have low modulus of rigidity, which can be improved by addition of filters, e.g., glass fibres.

Natural rubber, which is an organic material of biological origin, is an thermoplastic material. It is prepared from a fluid, provided by the rubber trees. Rubber materials are widely used for tyres of automobiles, insulation of metal components, toys and other rubber products.

Inorganic Materials

These materials include metals, clays, sand rocks, gravels, minerals and ceramics and have mineral origin. These materials are formed due to natural growth and development of living organisms and are not biological materials.

Rocks are the units which form the crust of the earth. The three major groups of rocks are:

(i) Igneous Rocks: These rocks are formed by the consolidation of semi-liquid of liquid material (magma) and are called *Plutonic* if their consolidation takes place deep within the earth and *volcanic* if lava or magma solidifies on the earth's surface. Basalt is igneous volcanic where as granite is igneous plutonic.

(ii) Sedimentary Rocks: When broken down remains of existing rocks are consolidated under pressure, then the rocks so formed are named as sedimentary rocks, e.g., shale and sandstone rocks. The required pressure for the formation of sedimentary rocks is supplied by the overlying rocky material.

(iii) Metamorphic Rocks: These rocks are basically sedimentary rocks which are changed into new rocks by intense heat and pressure, e.g., marble and slates. The structure of these rocks is in between igneous rocks and sedimentary rocks.

Rock materials are widely used for the construction of buildings, houses, bridges, monuments, arches, tombs, etc. The slate, which has got great hardness is still used as roofing material. Basalt, dolerite and rhyolite are crushed into stones and used as concrete aggregate and road construction material.

Another type of materials, i.e. *Pozzolanics*, are of particular interest to engineers because they are naturally occurring or synthetic silicious materials which hydrate to form cement. Volcanic ash, blast furnace slag, some shales and fly ash are examples of pozzolanic materials. When the cement contains 10-20% ground blast furnace slag, then it is called pozzolans-portland cement, which sets more slowly than ordinary portland cement and has greater resistance to sulphate solutions and sea water.

Rocks, stone, wood, copper, silver, gold etc. are the naturally occurring materials exist in nature in the form in which they are to be used. However, naturally occurring materials are not many in number. Nowadays, most of the materials are manufactured as per requirements. Obviously, the study of engineering materials is also related with the manufacturing process by which the materials are produced to acquire the properties as per requirement.

Copper, silver, gold, etc. metals, which occur in nature, in their free state are mostly chemically inert and highly malleable and ductile as well as extremely corrosion resistant. Alloys of these metals are harder than the basic metals. Carbonates, sulphates and sulphide ores are more reactive metals.

Biological Materials

Leather, limestone, bone, horn, wax, wood etc. are biological materials. Wood is fibrous composition of hydrocarbon, cellulose and lignin and is used for many purposes. Apart from these components a small amount of gum, starch, resins, wax and organic acids are also present in wood. One can classify wood as *soft wood* and *hard wood*. Fresh wood contains high percentage of water and to dry out it, seasoning is done. If proper seasoning is not done, defects such as cracks, twist, wrap etc. may occur.

Leather is obtained from the skin of animals after cleaning and tanning operations. Nowadays, it is used for making belts, boxes, shoes, purses etc. To preserve the leather, tanning is used. Following two tanning techniques are widely used:

(a) Vegetable Tanning: It consist of soaking the skin in tanning liquor for several days and then dried to optimum conditions of leather.

(b) Chrome Tanning: This technique involves pickling the skin in acid solution and then revolving in a drum which contains chromium salt solution. After that the leather is dried and rolled.

Limestone is an important material which is not organic but has biological origin. It mainly consist of calcium carbonate and limestone. It is widely used to manufacture cement. In Iron and Steel Industries, limestone in pure form is used as flux.

In early days bones of animals were used to make tools and weapons. Nowadays bones are used for the manufacture of glue, gelatin etc. Bones are laminate of organic substances and phosphates and carbonates of calcium. These are stronger in compression as compared to tension.

Table 1.3 lists typical examples from each of the four groups of materials.

Table 1.3 Important grouping of materials

Material group (1)	*Important characteristics* (2)	*Typical examples of engineering use* (3)
1. Metals and Alloys	Lusture, hardness, thermal and electrical conductivity, resistance to corrosion, malleability, stiffness and the property of magnetism	Iron and steels, aluminium, copper, silver, gold, zinc, magnesium, brasses, bronzes, manganin, invar, super alloy, boron, rare-earth alloys, conductors, etc.

(*Contd.*)

(*Contd.*)

Material group *(1)*	*Important characteristics* *(2)*	*Typical examples of engineering use* *(3)*
2. Ceramics and Glasses	Thermal resistance, hardness, brittleness, opaqueness to light, electrical insulation abrasiveness, high temperature strength and resistance to corrosion	Silica, soda-lime-glass, concrete, cement, refractories, Ferrites and garnets, ceramic superconductors, MgO, CdS, Al_2O_3, SiC, $BaTiO_3$, etc.
3. Organic Polymers	Soft, light in weight, poor conductors of electricity and heat, dimensionally unstable, ductile, combustible, low thermal resistance	*Plastics:* PVC, PTFE, polyethylene, polycarbonate *Fibres:* terylene, nylon, cotton, natural and synthetic rubbers, leather *Other uses:* refrigerants, explosives, insulators, lubricants, detergents, fuels, vitamins, medicines for surface treatment, adhesives, fibre-reinforced plastics, etc.
4. Composites (i) Metals and alloys and ceramics (ii) Metals and alloys and organic polymers (iii) Ceramics and organic polymers	They are better than any of the individual components as regards to their properties like strength, stiffness, heat resistance, etc.	• Steel-reinforced concrete, dispersion hardened alloys. • Vinyl coated steel, whisker-reinforced plastics. • Fibre-reinforced plastics, carbon-reinforced rubber.

Some important properties for different groups of materials are summarized in Table 1.4.

Table 1.4 Important properties for different groupings of materials

Property	*Metals*	*Ceramics*	*Polymers*	*Composites (wood)*
1. Tensile strength (N/mm^2)	200–2000	10–400	30–100	20–110
2. Density ($10N/mm^2$)	$2–8 \times 10^3$	$2–17 \times 10^3$	$1–2 \times 10^3$	0.5×10^3
3. Hardness	medium	high	low	low
4. Tensile modulus ($10^3 N/mm^2$)	100–200	150–450	0.7–3.5	4–20
5. Melting point (°C)	200–3500	2000–4000	70–200	—
6. Thermal expansion	medium	low	high	low
7. Thermal conductivity	high	medium	low	low
8. Electrical conductivity	good conductors	insulator	insulator	insulator

5. SEMICONDUCTORS

These are the materials which have electrical properties that are intermediate between the electrical conductors and insulators. The electrical characteristics of semiconductors are extremely sensitive to the presence of minute concentrations of impurity atoms; these concentrations may be controlled over very small spatial regions. Semiconductors form the backbone of electronic industry. The semiconductors have made possible the advent of *integrated circuitary* that has totally revolutionized the electronics and computer industries. They affect all walks of life whether it is communications, computers, biomedical, power, aviation, defence, entertainment, etc. The field of semiconductors is rapidly changing and expected to continue in the next decade. Organic semiconductors are expected to play prominent role during this decade. Diamond as semiconductor will also be important. Optoelectronic devices will provide three-dimensional integration of circuits, and optical computing.

6. BIOMATERIALS

These are employed in components implanted into the human body for replacement of diseased or damaged body parts. Biomaterials must not produce toxic substances and must be compatible with body tissues (i.e., these materials must not cause adverse biological reactions). All the above materials, i.e., metals, ceramics, polymers, composites, and semiconductors—may be used as biomaterials.

7(A) CURRENT TRENDS AND ADVANCES IN MATERIALS

Timber, steel and cement are the materials which are widely used for engineering applications in huge quantities. The consumption of steel in any country is considered as an indicator of its economic well being. For high temperature applications, e.g. steam and gas turbines the design engineers keep creating the demand for various high steel alloy. However, alloys of chromium, nickel, molybdenum and tungsten alongwith iron are better suited for the said applications. Newer materials for combined resistance to high temperature and corrosion are increasing rapidly and material scientists and engineers are busy in developing such materials. Different kinds of ceramics, though difficult to shape and machine, are finding demand for their use at high temperatures.

Recently prepared new metallic materials in conjunction with new processing techniques as isostatic pressing and isothermal forging are capable of imparting better fatigue properties to aircraft components. Powder metallurgy technique while producing finished surfaces and cutting down metal cutting cost is much capable of imparting improved mechanical properties under different loading conditions. Surprisingly, rapid cooling technology achieving cooling rates in the vicinity of one million degree celcius per second and this is being used to produce metal powders which can be used in such product producing techniques as powder metallurgy and hot isostatic pressing to obtain temperature resistant parts. Nowadays, metallurgists have produced several molybdenum and aluminium alloys as well as alloys of titanium and nickel to meet anticorrosion properties at elevated temperatures.

Polymeric materials are growing at annual rate of 9% and have grown in volume more than any other material. In several applications plastics have replaced metals, wood, glass and paper. A new trend in plastic technology is the production of synergistic plastic alloys which have better properties than individual members producing the alloy. Recent discovery of plastic conductors may have wider impact in near future.

Ceramics are mainly used as high temperature low load carrying materials. The major drawback of ceramics is the brittleness and difficulty in cutting and shaping. When mixed with metal powder like molybdenum, ceramic produce *cerments*, which are expected to be useful cutting materials. Tool bits of cerments are expected to find various applications in attaining high cutting speeds and producing better surface finish. Alumina, a well known ceramic is expected to be successfully reinforced with fibres of molybdenum. Due to microcracking of molybdenum fibres, the attempts to achieve better strength in such composite ceramics have not been successful yet. However, such composites have been found to exhibit better impact and thermal shock resistance.

The advent of solar cells, electronic digital circuits and computers in factory automation and use of robots in several industrial applications is adding to the enormous demand of silicon chips and of such material as silicon. Today, semiconductors form the backbone of electronics and they affect all types of instruments/industries related to e.g. communications, computers, biomedical, power, aviation, defence, entertainment etc.

7(B) ADVANCED MATERIALS

The materials that are utilized in high-technology (or high-tech) applications are sometimes called advanced materials. By high technology we mean a device or product that operates or functions using relatively intricate and sophisticated principles; for example: electronic equipment (VCRs, CD players, etc.), computers, fiber optic systems, spacecraft, aircraft and military rocketry. These advanced materials are typically either traditional materials whose properties have been enhanced or newly developed high performance materials. Furthermore, advanced materials may be of all material types (e.g., metals, ceramics, polymers) and are normally relatively expensive. In subsequent chapters are discussed the properties and applications of a good number of advanced materials—for example, materials that are used for lasers, ICs, magnetic information storage, liquid crystal displays (LCDs), fiber optics, and the thermal projection system for the space shuttle orbiter.

7(C) SMART MATERIALS (MATERIALS OF THE FUTURE)

Smart or intelligent materials form a group of new and state-of-the art materials now being developed that will have a significant influence on many of our technologies. The adjective "smart" implies that these materials are able to sense changes in their environments and then respond to these changes in predetermined manners—traits that are also found in living organisms. In addition, the concept of smart materials is being extended to rather sophisticated systems that consist of both smart and traditional materials.

The field of smart materials attempts to combine the sensor (that detects an input signal), actuator (that performs a responsive and adaptive function) and the control circuit on as one integrated unit. Actuators may be called upon to change shape, position, natural frequency, or mechanical characteristics in response to changes in temperature, electric fields, and/or magnetic fields.

Usually, four types of materials are commonly used for actuators: shape memory alloys, piezoelectric ceramics, magnetostrictive materials, and electrorheological/magnetorheological fluids. Shape memory alloys are metals that, after having been deformed, revert back to their original shapes when temperature is changed. Piezoelectric ceramics expand and contract in response to an applied electric field (or voltage); conversely, these materials also generate an electric field when their dimensions are altered. The behaviour of magnetostrictive materials is analogous to that of the piezoelectric ceramic materials, except that they are responsive to magnetic fields. Also, electrorheological and magnetorheological fluids are liquids that experience dramatic changes in viscosity upon the application of electric and magnetic fields, respectively.

The combined system of sensor, actuator and control circuit on as one IC unit, emulates a biological system (Fig. 1.1).

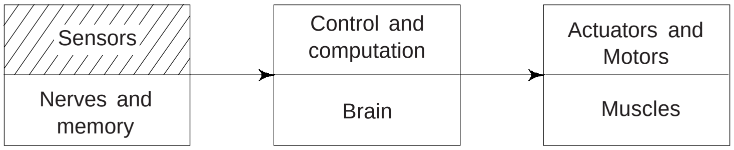


Fig. 1.1 Integrated sensor-actuator systems with controller are analogous to biological systems

These are known as smart sensors, microsystem technology (MST) or microelectromechanical systems (MEMS). Materials/devices employed as sensors include optical fibers, piezoelectric materials (including some polymers), and MEMS.

For example, one type of smart system is used in helicopters to reduce aero-dynamic cockpit noise that is created by the rotating rotor blades. Piezoelectric sensors inserted into the blades, monitor blade stresses and deformations; feedback signals from these sensors are fed into a computer controlled adaptive device, which generates noise cancelling antidose.

MEMS devices are small in size, light weight, low cost, reliable with large batch fabrication technology. They generally consist of sensors that gather environmental information such as pressure, temperature, acceleration etc., integrated electronics to process the data collected and actuators to influence and control the environment in the desired manner.

The MEMS technology involves a large number of materials. Silicon forms the backbone of these systems also due to its excellent mechanical properties as well as mature micro-fabrication technology including lithography, etching, and bonding. Other materials having piezoelectric, piezoresistive, ferroelectric and other properties are widely used for sensing and actuating functions in conjunction with silicon. The field of MEMS is expected to touch all aspects of our lives during this decade with revolution in aviation, pollution control, and industrial processes.

7(D) NANOSTRUCTURED MATERIALS AND NANOTECHNOLOGY

Nanostructured materials are those materials whose structural elements—clusters, crystallites or molecules—have dimensions in the range of 1-100 nm. These small groups of atoms, in general, go by different names such as nanoparticles, nanocrystals, quantum dots and quantum boxes. Substantial work is being carried out in the domain of nanostructured materials and nanotubes during the past decade since they were found to have potential for high technology engineering applications. One finds a remarkable variations in fundamental electrical, optical and magnetic properties that occur as one progresses from an 'infinitely extended' solid to a particle of material consisting of a countable number of atoms. The various types of nanostructured materials which has been considered for applications in opto-electronic devices and quantum-optic devices are nano-sized powders of silicon, silicon-nitride (SiN), silicon-carbide (SiC) and their thin films. Some of these are also used as advanced ceramics with controlled micro structures because their strength and toughness increase when the grain size diminishes. Carbon-based nanomaterials and nanostructures including fullerenes and nanotube plays an increasingly pervasive role in nanoscale science and technology. Today, nanotechnology is being heralded as the next enabling technology that will redesign the future of several technologies, products and markets. A brief account of nanostructured materials, particularly of carbon based nanomaterials and nanostructures is presented in a separate chapter.

7(E) QUANTUM DOTS (QDS)

Rapid progress in the fabrication of semiconductor structures has resulted into the reduction of three-dimensional bulk systems to two-dimensional, one-dimensional, and ultimately to zero dimensional systems. These reduced dimensional systems are used in future applications like semiconductor lasers and microelectronics. Quantum dots represent the ultimate reduction in the dimensionality of semiconductor devices. These are three dimensional semiconductor structures only nanometer in size confining electrons and holes. QDs operate at the level of single electron which is certainly the ultimate limit for an electronic device and are used as the gain material in lasers. QDs are used in quantum dot lasers, QD memory devices, QD photodetectors and even quantum cryptography. The emission wavelength of a quantum dot is a function of its size. So by making dots of different sizes, one can create light of different colours.

7(F) SPINTRONICS

A revolutionary new class of semiconductor electronics based on the spin degree of freedom could be created. The study of electron spin in materials is called spintronics. Spintronics is based on the direction of spin- and spin-coupling. Every appliance ranging from electric bulb to laptop computer works on the principle of transport of electric charge carriers-electrons, which cause electric current to flow through the wires. The electrons have both charge and spin. The spin of the electrons could greatly enhance the particles' usefulness. Presently, the semiconductor technology is based on the number of charges and their energy. The electronic devices, e.g. transistors work due to flow of charge. The electron can be assumed as tiny rotating bar magnet with two possible orientations: spin-up or spin-down. An applied magnetic field can flip electrons from one state to another. Obviously, spin can be measured and manipulated to represent the 0's and 1's of digital programming analogous to the "current on and current off" states in a conventional silicon chip.

The performance of conventional devices is limited in speed and dissipation whereas spintronics devices are capable of much higher speed at very low power. Spintronics transistors may work at a faster speed, are also smaller in size and will consume less power.

The electron spin may exist not only in the up or down state but also in infinitely many intermediate states because of its quantum nature depending on the energy of the system. This property may lead to highly parallel computation which could make a quantum computer work much faster for certain types of calculations. In quantum mechanics, an electron can be in both spin-up and spin-down states, at the same time. The mixed state could form the base of a computer, built around not binary bits but the quantum bits or qubit. It is any combination of a 1 or a 0. The simplest device using spin-dependent effect is a sandwich with two magnetic layers surrounding a non-magnetic metal or insulator. If the two magnetic layers are different, then the magnetization direction of one can be rotated with respect to the other. This leads to the utilization of these structures as sensor elements and for memory elements. Scientists are now trying to use the property of the electron-like spin rather than charge to develop new generation of microelectronic devices which may be more versatile and robust than silicon chips and circuit elements. Spins appear to be remarkably robust and move effectively easily between semiconductors. In case of electron transport from one material to another, the spins do not lose its orientation or scatter from impurities or structural effects.

7(G) FERMIONIC CONDENSATE MATTER

Very recently scientists had created a new form of matter called a *fermionic condensate matter* and predicted it could lead to the next generation of superconductors for use in electricity generation more efficient trains and countless other applications.

The new matter form is the sixth known form of the matter—after solids, liquids, gases, plasma and Bose-Einstein condensate, created only in 1995.

Fermionic condensate matter is a scientific breakthrough in providing a new type of quantum mechanical behaviour. It is related to Bose-Einstein condensate. However, new state is not a superconductor but it is really something in between these two states that may help us in science link these two interesting behaviours. New state of matter uses *fermions*—the building blocks of matter.

8. LEVEL OF MATERIAL STRUCTURE EXAMINATION AND OBSERVATION

One can examine and observe the internal structure of a material at various levels of observation. The degree of magnification required to study a material by various methods is called a measure of the level of observation. We may note that the details that are revealed or disclosed at a certain level of magnification are generally different from the details that are disclosed at some other level of magnification.

While considering the structures at various levels we may find it is useful to have some concept of relative sizes, and the so-called "*logarithmic scale of universe*" provides one method of size comparison (Table 1.5).

Table 1.5 Relative sizes within the universe: Logarithmic scale of the universe

Object	*Average size*	*Logarithmic scale of the universe*
1. Nearest galaxy distance	1 million light year	10^{21}
2. Nearest star distance	1 light year	10^{16}
3. Diameter of the sun	1 million kilometer	10^{9}
4. Diameter of the earth	10^3 kilometer	10^{7}
5. Man (average dimension)	1 metre	10^{0}
6. Sand grain	1 millimetre	10^{-3}
7. Dust particle	1 micrometre	10^{-6}
8. Molecule	1 nanometre	10^{-9}
9. Atom	10^{-1} nanometre	10^{-10}

9. MATERIAL STRUCTURE

In recent years, the number and variety of materials, which are of particular interest to an engineer have increased tremendously. Each type of material has a specific composition possessing specific properties for a specific use. It is not possible for one to explain the properties of all types of these materials. A knowledge of the structure of the material helps students and engineers to study the properties of the material. Material structure can be classified as: macrostructure, microstructure, substructure, crystal structure, electronic structure and nuclear structure.

(a) Macrostructure Macrostructure of a material is examined by low-power magnification or naked eye. It deals with the shape, size and atomic arrangement in a crystalline material. In case of some crystals, e.g., quartz, external form of the crystal may reflect the internal symmetry of atoms. Macrostructure may be observed directly on a fracture surface or on a forging specimen. The individual crystals of a crystalline material can be visible, e.g. in a brass doorknob by the constant polishing and etching action of a human hand and sweat. Macrostructure can reveal flaws, segregations, cracks etc. by using proper techniques and one can save much expenses by rejecting defective materials at an early stage.

(b) Micro Structure This generally refers to the structure of the material observed under optical microscope. Optical microscopes can magnify a structure about 1500 to 3000 times linear, without loss of resolution of details of the material structure. We may note that optical microscopes can resolve two lines separately when their difference of separation is 10^{-7} m (= 0.1 μm). Cracks, porosity, non-metallic insclutions within materials can be revealed by examining them under powerful optical microscope.

(c) Sub Structure When crystal imperfections such as dislocation in a structure are to be examined, a special microscope having higher magnification and resolution than the optical microscope is used. *Electron microscope* with magnifications 10^5 are used for this purpose. Another important modern microscope is *field ion microscope*, which can produce images of individual atoms as well as defects in atomic arrangements.

(d) Crystal Structure This reveals the atomic arrangement within a crystal. X-ray diffraction techniques and electron diffraction method are commonly used for studying crystal structure. It is usually sufficient to study the arrangement of atoms within a *unit cell*. The crystal is formed by a very large number of unit cells forming regularly repeating patterns in space.

(e) Electronic Structure This refers to the electrons in the outermost shells of individual atoms that form the solid. Spectroscopic techniques are commonly used for determining the electronic structure.

(f) Nuclear Structure This is studied by nuclear spectroscopic techniques, e.g., nuclear magnetic resonance (NMR) and Mössbauer spectroscopy.

10. ENGINEERING METALLURGY

This includes the study of metallurgy, which is of special interest to an engineer. At the time of taking final decision regarding the selection of suitable material for a particular job, it is essential for an engineer to have a thorough knowledge of engineering metallurgy. This helps him in deciding the treatment processes and their sequence, which are to be carried out on the finished components and structures. This includes the following processes:

(i) Iron-carbon alloy system: This deals with the structure of iron and steel as well as iron-carbon equilibrium diagrams. This also deals with the transformation of alloys and steels under various sets of conditions. A detailed and thorough study of this process helps an engineer to decide the suitability of process and the selection of iron alloy for various types of his jobs.

(ii) Heat Treatment: A thorough knowledge of this helps an engineer in deciding the type of process to be undertaken for the smooth and efficient working of the components and structures.

(iii) Corrosion of Metals: This deals with the colossal problem of corrosion of metals and its prevention. It also deals with the processes which help engineers in improving the life and outward appearance of the metal components and structures.

11. SELECTION OF MATERIALS

One of the most challenging task of an engineer is the proper selection of the material for a particular job, e.g., a particular component of a machine or structure. An engineer must be in a position to choose the optimum combination of properties in a material at the lowest possible cost without compromising the quality. The properties and behaviour of a material depends upon the several factors, e.g., composition, crystal structure, conditions during service and the interaction among them. The performance of materials may be found satisfactory within certain limitations or conditions. However, beyond these conditions, the performance of materials may not be found satisfactory.

One can list the major factors affecting the selection of materials as

(i) Component shape
(ii) Dimensional tolerance
(iii) Mechanical properties
(iv) Fabrication requirements
(v) Service requirements
(vi) Cost of the material
(vii) Cost of processing, and
(viii) Availability of the material.

All these major factors have a complex effect on the selection of materials. The shape and size of a component has great effect on the choice of the processing unit which ultimately effects the choice of the material. To make it more clear, we consider an example, let the best possible production method is selected, under given conditions, it is die casting, obviously, now the choice of the material becomes limited, i.e. one can only choose materials with lower melting points, e.g. aluminium, zinc, magnesium and thermoplastics.

There are some materials which can be finished to close tolerance while others cannot. Obviously, the required dimensional tolerance for finished components will, influence the choice of materials.

To select a suitable material for specific conditions, all mechanical properties, e.g., hardness, strength, etc. guide us. Method of processing of the material also affects the properties of a component, e.g., forged components can be stronger than the casted components. Different types of working processes may also give different types of fibre structure. However, investment casting can provide precise dimensions at low cost in comparison to machine operations.

Service requirements are dimensional stability, strength, toughness, heat resistance, corrosion resistance, fatigue and creep resistance, electrical and thermal conductivity etc. where as fabrication requirements are castability, i.e., ease in casting a material, weldability-ease in welding the material, machinability-ease to machine a material, formability-ease to form a material, hardenability etc.

In most of the cases, the cost of raw material accounts about 50 per-cent of the finished cost. Obviously, the cost of the material is a major factor which influences the choice of the material or process. We must note that the use of cheaper material will not always reduce the final cost of the component or product. Use of cheaper material may be associated with higher processing cost due to large number of operations to be performed and also more scrap. We can easily see that this sometimes makes the overall cost more than that of expensive raw material in combination with low processing cost due to lesser number of operations and lesser scrap.

The type of material affects the detailed aspect of design and hence the choice of material as well as the process is selected at the early design state, e.g., whether the material is to be joined by spot welding, screws or rivetes, must be decided at the design state.

In most of the industries, the processing cost (labour cost) and other costs such as overhead costs account for about 50% of the production cost. Overhead cost in automatic industries is much more than the other costs. If one can somehow reduce all such costs, the total production cost will automatically reduce. In comparison to conventional processes, sometimes injection moulding process is preferred because the conventional process involves many intermediate stages and several machining processes.

One finds that the cost of production of a component by rolling or forging operations is twice as compared to the production by powered metallurgy process because the rolling or forging are followed by several machining operations for the same finish and tolerances.

We may find that sometimes the availability of the material becomes a governing factor. When the desired material supply is limited, then a costly material which is available in ample quantity may be chosen.

In the light of above factors, sometimes it may be the case that two or more than two materials are found suitable for a particular component or job. Sometimes, it may also happen that the above factors may oppose each other. This shows that there may no exact or true solution and one has to compromise in the final selection. One can compromise by taking into consideration the relative merits and demerits, cost of finished component and its life.

Summarizing, we can say that the selection of material is a dynamic process and change in design may be progressive. Keeping in view, the availability and awareness of latest technological developments, one can always change the material from time to time. While making the selection of material amongst the available large range of different types of materials the wisest choice should be made keeping in view the above factors to achieve the efficient utilisation of materials.

Modern Materials' Need

In spite of the tremendous progress that has been made in the field of materials science within the past few years, there still remain technological challenges, including the development of more sophisticated and specialized materials, as well as the impact of materials production on environment.

Nuclear (fission as well as fusion) energy holds some promise, but the solutions to the many key problems that remain will necessarily involve materials, from fuels to containment structures to facilities for the disposal of radioactive waste. Fusion is the process that powers the sun and stars, has the potential to provide large-scale, safe energy production, without adding to the global warming and without the long-term radioactive waste associated with conventional fission power stations. Progress in fusion research has been incredibly rapid in recent years. There has been major progress in fusion materials and technology, with prototypes of the key components of a fusion power plant built and successfully tested.

Significant quantities of energy are involved in transportation vehicles (aircrafts, trains, automobiles, etc.), as well as increasing engine operating temperatures, will enhance fuel efficiency. Obviously, new high strength, low density structural materials remain to be developed, as well as materials that have higher temperature capabilities, for use in engine components.

Furthermore, there is a urgent need to find new, economical sources of energy and to make use of present energy resources more economically. Hydrogen, seems to be the fuel of the future. Hydrogen offers the greatest potential environment and energy supply benefits. Like electricity, hydrogen is a versatile energy carrier that can be made from a variety of widely available primary (i.e., naturally occurring) energy sources including natural gas, coal, biomass (agricultural or forestry residues or energy crops), wastes, sunlight, wind, and nuclear power. Available hydrogen technologies can dramatically reduce pollution and greenhouse emission. Although hydrogen production techniques do exist, further optimization is desirable for use in energy systems with zero carbon emissions. Side issues associated with various primary energy sources will be important. Materials will undoubtedly play a significant role in these developments, e.g., the direct conversion of solar energy into electrical energy has been demonstrated. Solar cells employ some rater complex and expensive materials. To ensure a viable technology, materials that are highly efficient in this conversion process yet less costly have to be developed.

We know that environment quality depends on our ability to control air and water pollution. Pollution control techniques employ various materials. There is a need to improve material processing and refinement methods so that they produce less environmental degradation i.e., less pollution and less despoilage of landscape from the mining of raw materials. Toxic substances are produced during manufacturing processes of some materials and therefore we have to consider the ecological impact of their disposal.

There are many materials which we use are derived from resources that are non renewable, i.e. not capable of being regenerated. These include polymers for which the prime raw material is oil, and some metals. These non renewable resources are gradually becoming depleted, which necessitates: (i) the search of additional reserves (ii) the development of new materials having comparable properties with less adverse environmental impact, and/or (iii) increased recycling efforts and the development of new recycling technologies. Obviously, as a consequence of the economics of not only production but also environmental impact and ecological factors, it is becoming more important to consider ‘cradle to grave’ life cycle of materials relative to overall manufacturing process.

Example 1 *What will be the selection criteria for 15 A electrical plugs?* [AMIE, Diploma]

Solution The selection criteria for 15 A electrical plugs are

(i) The said electrical plugs are designed for the required amperage.
(ii) The plugs are electrically safe to handle.
(iii) The electrical plugs are cheap to produce.
(iv) The plugs can be produced on mass scale.
(v) While in use, the said plugs should not cause injury.
(vi) The plugs should be rigid enough so that they do not break easily while in use, i.e., handling.
(vii) The size of the pins should be such that these can be easily inserted and removed from the terminal.

Example 2 *What will be the selection criteria for a flexible PVC hose pipe?* [AMIE, Diploma]

Solution The selection criteria for a flexible PVC hose pipe will be

(i) It should be light in weight.
(ii) It should be flexible and not rigid.
(iii) It should not wear out easily.
(iv) Its surface should be smooth and does not affects the hands while handling.
(v) It should not crack when rolled after use.
(vi) It can be produced on mass scale.
(vii) It should be cheap in comparison to the other products.

SUGGESTED READINGS

1. A. Street and W. Alexander, ‘*Metals in the Service of Man*’, Penguin Books (1976).
2. *Materials*, ‘*A Scientific American Book*’, W. H. Freeman and Co., San Francisco (1968).
3. D. Fishlock, ‘*The New Materials*’, John Murray, London (1967).
4. J. P. Schaffer, et al., ‘*The Science and Design of Engineering Materials*’, 2nd Ed., McGraw-Hill (2000).
5. J. F. Shackelford, ‘*Introduction to Materials Science for Engineers*’, 5th Ed., Prentice Hall (2000).
6. R. E. Hummel, ‘*Understanding Material Science*’, Springer-Verlag, New York (1998).

REVIEW QUESTIONS

1. Mention the main properties of metals, polymers and ceramics.
2. Write general properties and characteristics of metals, polymers, ceramics, semiconductors and composite materials. Give few examples belonging to each group.
3. Justify the statement that ‘selection of material is a compromise of many factors’?
4. How metals are classified according to their use?
5. How materials are classified according to their chemical composition?
6. What should be the criteria of selection of material for the construction for chemical process industries?
7. Explain in brief, why metals in general are ductile, whereas ceramics are brittle?
8. Why it is essential for a materials engineer to have the systematic classification of materials?
9. Write the properties required for a material to withstand high temperatures.
10. Explain the following terms as they relate to metals: alloys, sintered metal, coated metal, clad metal, and non-ferrous metals. Give atleast two examples of each.
11. Write the important features of inorganic materials.

12. What do you understand by pozzolanic material?

13. Compare organic and inorganic materials in terms of their structures, properties and characteristics.

14. Write short notes on

(i) Biological materials
(ii) Pozzolanic Material
(iii) Effects of tanning on animal hides
(iv) Metals
(v) Rocks
(vi) Composite Materials

PROBLEMS

1. Explain with specific reason the property you will consider while selecting the material for the following:

(a) Tyres for an aircraft wheels
(b) A screw driver
(c) Lining for oil fired furnace having temperature ~1200°C.
(d) A 100 mm diameter domestic water pipeline above the ground.

2. Explain, what are the service requirements for the following?

(a) A vacuum cleaner
(b) An air conditioner
(c) Electric iron
(d) A conveyor belt for handling crushed coal
(e) A car

SHORT QUESTION-ANSWERS

Fill in the blanks

1. Materials Science and engineering draw heavily from engineering sciences such as ________, ________, and ________. [metallurgy, ceramics and Polymer Science]

2. What are the three broad groups in which engineering materials can be classified according to their nature? [Metals and alloys, ceramics and glasses and organic polymers]

3. How the materials can be classified according to major areas in which they are used? [structures, machines, devices]

4. What are devices?
[These are the most recent addition to engineering materials and refer to such innovations as a photoelectric cell, transistor, laser, computer, ceramic magnets, piezoelectric pressure gauges, etc.]

5. Microstructure generally refers to the structure as observed under the ________. [optical microscope]

6. The optical microscope can resolve details upto a limit of about ________. [10^{-7} m = 0.1 μm]

7. The modern microscope which produces images of individual atoms and imperfections in atomic arrangements is called ________. [field ion microscope]

8. In an electron microscope, a magnification of ________ times linear is possible. [10^5]

9. What are nuclear spectroscopic techniques for studying nuclear structure? [NMR and Mössbauer spectroscope]

10. What do you understand by electronic structure of a solid?

11. What are the two categories in which polymers can be classified?

12. What are composites?

13. What are the criteria which affect the selection of a material?

14. What is the difference between macro and micro structures?

15. What is a unit cell in a crystal?

16. What do you understand by nonrenewable resources?

Ans. These are the resources which are not capable of regenerated. These includes polymers, for which the prime raw material is oil, and some metals. These nonrenewable are gradually being depleted, which necessitates: (i) the discovery of additional reserves, (ii) the development of new materials having comparable properties with less adverse environmental impact, and/or (iii) increased recycling efforts and the development of new recycling technologies.

Objective (Multiple-choice) Questions

1. Composite materials are
- *(a)* made mainly to improve temperature resistance
- *(b)* used for improved optical properties
- *(c)* made with strong fibres embedded in weaker and softer matrix to obtain strength better than strength of matrix.
- *(d)* made with strong fibres embedded in weaker and softer matrix to obtain strength better than strength of both matrix and filler.

2. Ceramic materials are
- *(a)* good conductors of electricity
- *(b)* basically crystalline oxides or metals
- *(c)* inorganic compounds of metallic and non-metallic elements
- *(d)* none of the above

3. Which one are inorganic materials
- *(a)* biological materials *(b)* minerals and ceramics
- *(c)* plastics *(d)* wood

4. Composite with molybdenum fibres and alumina matrix do not exhibit good strength because
- *(a)* molybdenum is weaker than alumina
- *(b)* molybdenum fibres develop micro cracks
- *(c)* molybdenum does not form any bond with alumina
- *(d)* molybdenum dissolves in alumina

5. Using cutting material which can sustain high temperature
- *(a)* cerment *(b)* high carbon steel alloy
- *(c)* composite of two metals *(d)* none of the above

6. Which one does not require paint on its surface
- *(a)* body of a railway wagon *(b)* body of a car
- *(c)* body of a machine tool *(d)* body of an aeroplane

7. The choice of material as well as process is made at the early design stage because type of material affects the detailed aspects of
- *(a)* design *(b)* cost
- *(c)* design as well as cost *(d)* none of the above

ANSWERS

1. *(c)* **2.** *(c)* **3.** *(b)* **4.** *(b)* **5.** *(a)* **6.** *(d)* **7.** *(c)*

2

Atomic Structure and Electronic Configuration

1. INTRODUCTION

All the materials available either in solid, liquid or gaseous form are made up of atoms, the smallest indivisible particles. Atoms of the same element are identical to each other in weight, size and all properties, whereas atoms of different elements differ in weight, size and other characteristics. The size of the atoms is of the order of 1 Å ($=10^{-10}$ m). A material which consists just one type of atoms is called element. Nitrogen, carbon, hydrogen, aluminium, copper, gold, iron etc. are few examples of elements. Group of atoms which tend to exist together in a stable form are called *molecules*, e.g. H_2, O_2, N_2 etc. Large number of molecules in nature exist as combination of atoms of different elements, e.g. water (H_2O), etc. Molecules containing one atom (known as monoatomic), two atoms (known as diatomic), three atoms (known as triatomic) or more atoms (known as polyatomic).

In the beginning of 20th century, scientists like Sir J.J. Thomson, Rutherford, Niels Bohr and many other have found that atom, though indivisible, yet consists of smaller particles called *electrons*, *protons* and *neutrons*. The electrons and protons are electrical in nature, where as neutrons are neutral.

Rutherford and coworkers have shown that the mass of the atom is concentrated at the centre of the atom, called *nucleus.* The atom essentially has an electrical structure and is made of smaller particles, the principal ones being electrons, protons and neutrons.

The atom is considered to be made up of a heavy nucleus, consisting of protons and neutrons, surrounded by highly structured configurations of electrons, revolving around the nucleus in shells or orbits at a relatively greater distance from nucleus. The size of the atomic nucleus is of the order of 10^{-14} m, whereas its density is found to be about 2×10^{17} kg/m^3. The densities of certain stars called *white dwarfs* approach that of a pure nuclear matter. As the mass of the electron is negligible as compared to that of protons and neutrons, the mass of the atom depends mostly on the protons and neutrons inside the nucleus.

2. THE ELECTRON

M. Faraday in 1983, in his experiments on the laws of electrolysis, provided the first experimental evidence that electrical charge was not infinitely divisible, but existed in discrete units.

In 1897, J.J. Thomson, while studying the passage of electricity through gases at low pressure, observed that the rays of light appear to travel in straight lines from the surface of the cathode and move away from it in the discharge tube. These rays are called *cathode rays* since they start from the cathode of the discharge tube. W. Crookes studied the properties of these cathode rays and showed that the rays,

(i) travel in straight line and cast shadows

(ii) carried negative charge and sufficient momentum

(iii) possess high kinetic energy and can induce some chemical reactions, excite fluorescence on certain substances.

These properties of the cathode rays were best explained by J.J. Thomson by his hypothesis that the cathode rays consist of a stream of particles, each of mass m and charge e (= 1.602×10^{-19} C), originating at the cathode of the discharge tube. These particles are called *electrons*. Thomson determined the specific charge (e/m), the ratio between the electronic charge and the mass of the electron and found its value as -1.76×10^{11} C/kg. Now electron is found to be a universal constituent of all matter. This is supported by the following facts:

(i) The electrons obtained from different sources throughout the world are found the same and have identical effects when used for any applications, e.g., radio and television (electromagnetic waves), X-ray applications, producing fluorescence effects, producing photoelectric effect, etc.

(ii) The electrons emitted in the cathode rays have been found to exhibit identical features in different modes of cathode ray discharge tube.

(iii) Electrons obtained from any source possess the same charge, e (= -1.602×10^{-19} C), same rest mass, m_0 (= 9.109×10^{-31} kg, equivalent energy = 0.51 MeV), same effective radius, r_e (= 4.6×10^{-6} nanometer). The mass of the electron, m, is usually taken as 9.1×10^{-31} kg. When electrons move with very high velocity, approaching to the velocity of light, c, the mass of the electron varies relativistically, in accordance with the relation

$$m = m_0 \Big/ \sqrt{1 - v^2 c^2} \qquad (1)$$

when $v = 0$, $m = m_0$. Hence m_0 is called the *rest* mass. The mass of the electron is negligible compared to that of protons and neutrons. The mass of the electron is 1/1836 th of the mass of the proton.

We have already stated that all the matter consists of atoms as the universal constituents where electrons form the structure of atom. Obviously, electrons are universal constituent of all matter.

The electrons revolve round the nucleus in various shells and orbits. As stated above, it carries negative electric charge equal in magnitude to the positive charge of a proton (= 1.602×10^{-19} C). Electron behaves as minute particle as well as an energy wave (de Broglie wavelength, $\lambda = h/mv$, experimental determination yield $\lambda = 1.66 \times 10^{-10}$ m = 0.166 nm for electron).

The electron surrounding the nucleus in an atom occupy different *orbitals*. According to *Heisenberg uncertainty principle*, the momentum and the position of an electron cannot be specified precisely. The product of the uncertainty in the momentum Δp and the uncertainty in the position Δx cannot be less than a certain value given by the relation

$$\Delta p \; \Delta x \geq h/2\pi \qquad (2)$$

where h (= 6.626×10^{-34} J-s) is Planck's constant. An outcome of the uncertainty principle is that one cannot visualize an electron orbital to be discrete path around the nucleus. Instead, one will have to think of the orbital as an *electron probability density* cloud surrounding the nucleus.

In Bohr's model of atom, the motion of an electron around the nucleus is limited to discrete orbits or quantum states, each of which represents a discrete level of energy—an integral multiple of quanta of fundamental energy units.

The electrons which are in the outermost shell or orbit account for most of the engineering properties of materials, e.g., bonding with other atoms, electrical and optical properties, chemical reactivity, etc.

3. PROTONS

The nucleus of hydrogen atom is called the *proton*. A proton has a unit positive charge of same magnitude as that of electron (= 1.602×10^{-19} C). The mass of a proton is 1.672×10^{-27} kg. The proton and the neutron are considered to be two different charge states of the same particle which is called a nucleon. The number of protons in a nucleus is called the charge Z of the given *nucleus*, or the charge number.

4. NEUTRONS

These are electrically neutral particles and 1.008 times heavier than protons. The mass of each neutron is 1.675×10^{-27} kg. Each neutron is composed of one proton and one electron, i.e.

Neutron = Proton + Electron

The number of neutrons in a nucleus is denoted by N. For all nuclei $N \geq Z$ (with the exception of ${}_1H^1$, ${}_2He^3$ and other neutron deficient nuclei). For light nuclei the ratio $\frac{N}{Z} \cong 1$; for the nuclei of elements at the end of periodic system $\frac{N}{Z} \cong 1.6$.

The total number of nucleons in a nucleus $A = N + Z$ is called the *mass number of the nucleus.*

In addition to the above mentioned three particles, there are other particles, e.g. positrons or positive electrons, neutrino and antineutrino, mesons, deutrons, alpha particle etc. Particles having mass intermediate between the electron and the proton are called *mesons.* Mesons are of two types: π (pie) and μ(mu) mesons, the former being somewhat heavier than the latter. Both the types of mesons may be either positively or negatively charged. There are also particles of small mass and zero charge called *neutrinos* (ν) and *antineutrinos* ($\bar{\nu}$). The existence of these particles has been postulated to account for energy changes during the radioactive emission of electrons and positrons respectively. Neutrino is supposedly associated with and shares the energy of electron, whereas antineutrino occupies the same position with respect to the positron. Table 2.1 gives the mass, charge and important properties of few elementary particles.

Table 2.1 Properties of few elementary particles

Particle	*Mass (amu)*	*Charge*	*Properties*
Electron (e^-)	0.0005486	–e	Stable, carrier of electric current, fundamental constituent of all elements
Proton (p)	1.00757	+e	Stable, a positively charged nuclear particle
Neutrion (n)	1.00893	0	A neutral nuclear particle
Positron (e^+)	0.0005486	+e	The positive counterpart of electron. Electron and positron mutually annihilate each other with the production of γ (Gamma) ray.
Meson (π)	0.0000019	+e	
Meson (μ)	0.0000025	+e	
Deutron (D)	2.0416	+e	Heavy isotope of hydrogen with mass about double that of ordinary hydrogen
Alpha particle (α)	4.00279	+2e	Doubly charged helium nucleus

However, these particles are not of much importance from the material science point of view.

5. ATOMIC NUMBER (*Z*)

This is a fundamental property of the atom and denoted by Z. The atomic number of an element is numerically equal to the number of protons present in the nucleus, i.e., the value of positive charge on the nucleus. We must note that all the atoms of the same element possess the same atomic number which identifies the position of the element in the periodic table of elements. A normal atom is electrically neutral and hence the number of protons and electrons are equal. For example, an iron atom contains 26 protons ($Z = 26$) and hence the balancing electrons are also 26. Obviously, 26 specifies the position of the iron element in the periodic table of elements.

6. ATOMIC WEIGHT AND MASS NUMBER

The atomic weight of an element is the average relative weight of its atom as compared to the weight of one atom of oxygen which is taken to be 16, i.e. it is the ratio between the weight of one atom of the element and 1/16 th of the weight of an atom of oxygen.

We should not confused this with the *mass number*. The mass number is equal to the sum of the number of protons and neutrons. The mass number is usually denoted by A.

$$A = \text{Number of protons} + \text{Number of neutrons}$$

For example chlorine atom has 17 protons and 18 neutrons in its nucleus. Obviously, mass number of chlorine is 35. We can see that the mass number of an element is always a whole number and its value is very close to the atomic weight. The mass of the nucleus is directly proportional to the mass number.

We can see that most elements have fractional atomic weights due to the existence of different *isotopes* of the same element.

The mass of an atom is usually expressed in terms of *atomic mass unit* (amu).

$$1 \text{ amu} = 1.6603 \times 10^{-24} \text{ g}$$

Hydrogen has one mass unit and carbon 12 amu respectively.

Once the atomic number (Z) and mass number (A) of an element are known, one can easily determine the number of neutron in its nucleus (see Table 2.2)

Table 2.2 Detailed atomic arrangement of some elements

Atomic characteristics/element	*Oxygen*	*Argon*	*Calcium*
Atomic number (Z)	8	18	20
Mass number (A)	16	40	40
No. of protons (p)	8	18	20
No. of electrons (e)	8	18	20
No. of neutrons (n) = (A – Z)	8	22	20

7. ISOTOPES

All atoms having different atomic weights but belonging to the same element are termed as *isotopes*, i.e. atomic number of isotopes of an element remains the same. Obviously, isotopes contain same number of protons and electrons. Thus the isotopes are atoms of different weight belonging to the same element and having the same atomic number. The difference in the masses of the isotopes of the same element is due to the different number of neutrons contained in the nuclei. For example, hydrogen exists in three isotopic forms. Atomic number of hydrogen is 1. Three isotopes of hydrogen are:

(i) Ordinary hydrogen ($_1H^1$) with atomic mass equal to 1.
(ii) Deuterium ($_1D^2$) with atomic mass equal to 2.
(iii) Tritium ($_1T^3$) with atomic mass equal to 3.

Similarly, chlorine has two isotopes, $_{17}Cl^{35}$ and $_{17}Cl^{37}$. These isotopes are available in the ratio of 3 : 1. Their average atomic weight is $\frac{37 \times 3 + 35 \times 1}{4}$ = 35.48

8. ISOBARS

Atoms with the same mass but belonging to different chemical elements are called isobars. Obviously, isobars possess different number of protons and electrons in their atoms. Total number of protons and neutrons in each of their nuclei is also same. The example of first pair of isobars is argon and calcium. Argon (atomic number 18) has 18p, 18e and 22 n in its atom. Calcium (atomic number 20) has 20p, 20e, and 20 n in its atom.

9. ISOTONES

These are the nuclides having the same numbers of neutrons (N) but a different Z and A. Example of isotones are $^{13}_{6}C$ and $^{14}_{7}N$. Isotones having a given value of N, obviously do not all correspond to the same chemical element.

The analysis of the properties of isotopes, isotones and isobars helps us to disclose several features of atomic nuclei. Such analysis helps us to predict that what will happen to the stability of a nucleus when an extra n or p is added to the nucleus.

10. AVOGADRO'S NUMBER (N)

The number of atoms per gm-mole of a substance is called as Avogadro's number and denoted by N. Its value is 6.023×10^{23} mol. It is a universal constant.

11. ATOMIC NUCLEUS

Rutherford in 1911, made use of α-particles emitted by a radioactive source to explore the structure of the atom. He directed a stream of α-particles through a thin gold foil and by measuring the angles through which the beam of α-particles scattered. Rutherford was able to conclude that most of the mass of the gold-atom resided in a volume called the nucleus, which carried a positive charge. The radius of the gold nucleus is found to be less than 2.8×10^{-14} m, well under 1/10,000 of the radius of the atom as a whole. For the metals studied, the radius of the nucleus is about 10^{-14} m. The *nuclear radius* or '*nuclear force radius*', represents the distance from the center of the nucleus at which an external uncharged nucleon first feels its influence. In the recent years, with more accurate techniques, it is found that the radius of an atomic nuclei lies within the range 10^{-14} m to 10^{-15} m. We must remember that a nucleus has no clear defined boundary. Obviously, the nuclear radius has an arbitrary meaning. The average density of nuclear matter is about 1.5×10^{18} kg/m^3. This is about 10^{15} times greater than the density of matter in bulk, and gives us an idea of the degree of compactness of the nucleons in a nucleus. It also shows that matter in bulk is essentially empty, since most of the mass is concentrated in the nuclei. We must also note that the supposition, that the nucleus is spherical, is not always true.

12. ATOMIC MODELS

In the foregoing sections, we have read about the existence of some elementary particles that are more fundamental than chemical elements. Electrons, for example, are common to all elements and form a common building block of all matter. In order to have a clear picture of the extra-nuclear electronic structure, mainly with the help of positive rays and mass spectroscopy, several atomic models about atomic structure have been advanced over the years after obtaining quantitative measurements on electrons and positive rays. We shall now discuss some of the important atomic models proposed by Sir J.J. Thomson, Rutherford, Niels Bohr, Sommerfeld and modern atomic model.

12.1 Thomson's Atomic Model

J.J. Thomson in 1911 proposed an atomic model known as plum-pudding model. According to this model,

(i) The electron is a constituent of all matter

(ii) The electron has a negative charge of 1.602×10^{-19} C and a rest mass of 9.1085×10^{-31} kg

(iii) Atoms are uniform spheres in which electrons were distributed alongwith positive charges, like plums in a pudding (Fig. 2.1). The atom is electrically neutral.

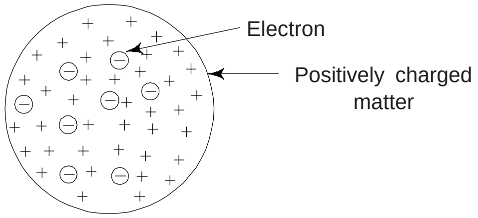


Fig 2.1 Thomson's plum-pudding Atomic model

(iv) The electrons possess vibratory motion about their equilibrium position and cause emission of light.

Thomson, by using various methods of spectroscopy also computed the total number of electrons in an atom.

Thomson’s proof for the existence of electrons in an atom was a major achievement and an essential prerequisite for the subsequent theories of atomic structure. By using different materials for the filament wire, Thomson established that the same value is always obtained for the charge to mass (e/m) ratio of the electron. This clearly support the statement that the electron is a constituent of all matter.

Thomson’s model could account for the periodicity in elements, spectral lines existence of ions and kinetic theory of gases. Thomson’s model failed to explain

(i) The scattering of α-particle incident on thin gold foil (Rutherford’s experiment)

(ii) The emission of spectral series by the atoms.

Thomson’s model of atom had finally to be given up with the development of Rutherford’s alpha particle scattering experiments.

12.2 Rutherford’s Nuclear Atomic Model

Rutherford and coworkers in 1911 performed differential scattering experiments and proposed a new atomic model. Rutherford directed α-particles emitted from a radioactive source on a thin gold foil 4×10^{-6} m thick. He observed that some α-particles passed through the film, while others were scattered all around. Those α-particles passing through the thin gold film also scattered over a wide area, and produced luminscence on a fluorescent screen of zinc sulphide (Fig. 2.2). Very small number of α-particles were deflected through large angles, i.e. about 1 in 10,000 particles suffered a deflection of more than 90°. Only very rarely a α-particle reversed back along its own path, i.e. suffered a deflection through 180°. (Fig. 2.3)

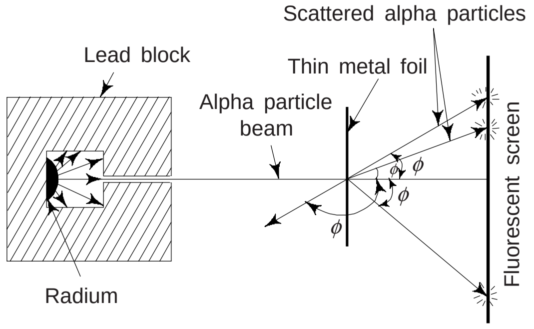


Fig. 2.2 Rutherford α-scattering experiment

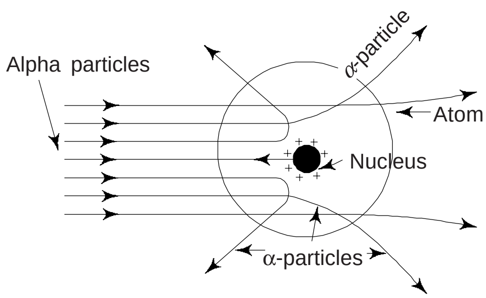


Fig. 2.3 Deflection of α-particles in the vicinity of a nucleus

Rutherford α-particle scattering experiments established the incorrectness of the uniform distribution of positive charge in a sphere of atomic dimensions in the Thomson’s atomic model. Rutherford’s results can be explained only if it is assumed that the positive charge in an sphere or atom is concentrated in the centre and the negatively charged particles surround it loosely, leaving enough space for α-particles to pass through, i.e. atoms must have a large empty space.

The large angle α-scattering observed can be explained on the basis that an electron is too light a particle, to cause any appreciable deflection of α-particle. It could only be possible if an α-particle has something equally or even more massive than itself, i.e. if the entire positive charge of an atom and the whole of its mass are concentrated in a small core, which Rutherford called the nucleus of the atom having very small dimensions. One can summarize the Rutherford’s observations as,

(i) The entire positive charge and mass of an atom are concentrated at the centre of the atom in a nucleus of very small dimensions.

(ii) The diameter of the nucleus is about 10^{-15} m whereas the diameter of an atom is about 10^{-10} m. Obviously, the nucleus of an atom occupies a volume which is a million-millionth part of an atom. Thus there is a lot of empty space in the atom.

(iii) An atom consists of a positive nucleus with the electrons moving very rapidly around different orbits (Fig. 2.4). The positive charge of the nucleus equals to total negative charge of electrons, making atom electrically neutral.

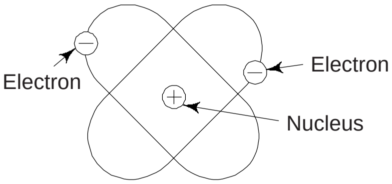


Fig. 2.4 Rutherford's model of the atom

In Rutherford's model, it was assumed that the force of attraction between the electrons and the nucleus was balanced by the centrifugal force attained by the electrons due to their motion in orbits round the nucleus. This model explains α-particle scattering in the following way.

The α-particles are positively charged and heavy. The forces between electrons and α-particles may be neglected. α-particles move with a large velocity and thus a strong electrical force of repulsion is required to deflect them at large angles. The α-particle which passes through the atom at relatively large distance from the centre, i.e., nucleus, experiences very little electrostatic repulsion and passes almost undeflected. This shows that there is much empty space in the atom. On the otherhand, α-particle which is passing close to the nucleus, experiences a very strong repulsive force that is exerted by the massive positive core (nucleus) and thus gets scattered through a large angle in a single encounter (Fig. 2.3). For the α-particle which is aimed at the nucleus, the repulsive force will slow down the speed of the α-particle and then finally stop it and then it will be repelled back along its own path.

The significant interaction, then is between the doubly charged α-particle (+2e) and the positive core nucleus situated at the centre of the atom having an integral number Z and charge $+Ze$. The kinetic energy of α-particle having a mass m and moving with velocity v, approaching directly towards the nucleus is

$$\text{K.E. of } \alpha\text{-particle, } E_k = \frac{1}{2} mv^2 \tag{3}$$

Rutherford assumed that α-particle and nucleus acted on each other with Coulomb force which, in this case repulsive. The equation for this force at distance D from the nucleus is given by

$$F = \frac{1}{4\pi\varepsilon_0} \frac{Ze \cdot 2e}{D^2} = \frac{2Ze^2}{4\pi\varepsilon_0 D^2} \tag{4}$$

where ε_0 (= 8.854×10^{-12} F/m) is the permittivity of the free space. $1/4\pi\varepsilon_0 = 9 \times 10^9$ MKS units. Z is the atomic number of the gold foil and e the charge on the electron. This force is inversely proportional to the square of distance between α-particle and the nucleus. The potential energy of the α-particle at distance D, is given by

$$E_p = \frac{1}{4\pi\varepsilon_0} \frac{Ze \cdot 2e}{D} = \frac{2Ze^2}{4\pi\varepsilon_0 D} \tag{5}$$

At the point of closest approach, $E_k = E_p$,

or $$E_k = \frac{2Ze^2}{4\pi\varepsilon_0 D}$$

or $$D = \frac{2Ze^2}{4\pi\varepsilon_0 E_k} \tag{6}$$

The closest distance of approach can be taken approximately as the radius (r) of the nucleus but the actual radius of the nucleus is slightly less than D. One can estimate the nuclear radius with the help of the following relation

$$r = r_0 \text{ A}^{1/3} \tag{7}$$

where A is the mass number of the atomic nucleus and r_0 (= 1.4×10^{-15} m) is a constant, same for all nuclei.

If the electron moves in a circular orbit of radius r with a constant linear velocity v round the nucleus, then it will be subjected to two forces: (i) one acting inwards will be the force of electrostatic attraction given by Coulomb's law

$$F = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2} = \frac{Ze \cdot 2e}{4\pi\varepsilon_0 r^2} \qquad (8)$$

where q_1 (= Ze) is the charge on the nucleus and q_2 (= 2e) is the charge on the α-particle and (ii) centripetal force, mv^2/r, which is supplied by the electrostatic attractive force of the nucleus on the electrons (Fig. 2.5).

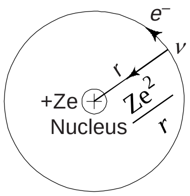


Fig. 2.5 **Rutherford's model**

12.3 Drawbacks of Rutherford's Atomic Model

Rutherford's model of the atom has a serious flaw in it. The following are its drawbacks:

(i) In this model of the atom, the electrons move in a Coulomb field of the nucleus in orbits, like a planetary system. We know that a particle moving on a curved trajectory is accelerating and an accelerating charged particle radiates electromagnetic radiations and continuously loses its energy. In accordance with Newton's law of motion and Maxwell's electromagnetic field equations, all the energy of the atom would be radiated away in a time of the order 10^{-10} seconds and ultimately the electron would collapse into the nucleus (Fig. 2.6). This would result in annihilation of atom. This means that there could not be any atom with the positively and negatively charged parts existing separately in it, which is contrary to the laws of classical mechanics and the observed facts.

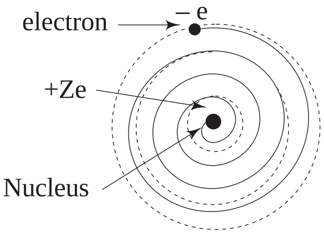


Fig. 2.6 **Sprial path of an atom**

(ii) According to electromagnetic theory of radiations, the revolving electron must emit continuously radiations of all increasing frequencies. This is contrary to the observed fact that elements emit spectral lines of fixed frequencies and not of all frequencies.

(iii) There is no rule for distribution of electrons in various orbits round the nucleus.

In order to remove the above drawbacks of the Rutherford atomic model, the famous Danish physicist Niels Bohr in 1913, put forward some new proposals based on the quantum idea which had been proposed by Max Planck a few years earlier.

12.4 Bohr's Atomic Model

Bohr considered the simplest of all atoms, viz., the hydrogen atom. He accepted the Rutherford model of the atom. According to the Rutherford model, there is only one electron revolving in the orbit of the hydrogen atom round the nucleus which is a single proton. Bohr assumed that the electron revolves in a circular orbit. It may be noted that this simple picture of the hydrogen atom is also applicable to some ionized atoms. In the helium atom, there are two electrons. So the removal of one electron from it leaves behind the singly charged helium ion, viz., He^+ which has only one electron revolving in the orbit round the nucleus which carries two electronic units of positive electricity. Thus the He^+ ion resembles the hydrogen atom. Besides He^+, other examples are Li^{++} (Z = 3), Be^{+++} (Z = 4), B^{+++} (Z = 5) etc.

In Bohr's model of the atom, the electrons are assumed to be in a definite planetary system (in circular orbits) of fixed energy. These stationary states are called as *energy levels.* Since there is a definite value of potential energy associated with each orbit. In this model of the atom, Bohr considered that more than one energy level to be possible for any electron of the atom.

One can determine the allowed energy levels by the condition that the angular momentum of the electron moving in a circular orbit can take one of values

$$L = nh/2\pi = n\hbar \qquad (9)$$

where n is a positive integer known as the *quantum* number, and its possible values are $n = 1, 2, 3, \ldots$ etc. Equation (9) is known as *Bohr's quantum condition*. The electron in this case is in a stationary state.

According to Bohr's model, an electron can either emit or absorb energy when making a transition from one possible orbit to another.

12.5 BOHR'S POSTULATES

(i) In the case of hydrogen atom, there is single electron which can revolve round the nucleus in certain definite orbits, known as stationary orbits. The electrons are permitted to have only those orbits for which the angular momenta of the planetary electron are integral multiple of $h/2\pi$ or $\hbar$, where h is the Planck's constant. Thus angular momentum of an electron is given by relation (9). This is Bohr's first postulate.

(ii) When the electron revolves in a stationary orbit, it does not emit electromagnetic radiation as predicted by the electromagnetic theory of light. Radiation occurs only when an electron falls from a higher energy state to a lower energy state. If the transition is from an orbit of higher energy E_2 to an orbit of energy E_1, then the energy $h\nu$ of the emitted radiation, according to Planck's law, will be

$$h\nu = \hbar\omega = E_2 - E_1 \tag{10}$$

where ν is the frequency of the emitted radiation; $\omega = 2\pi\nu$ is the angular frequency. $h\nu$, the energy difference ejected from the atom in the form of light radiations of energy called a *photon*. This reveals the origin of light waves from atom. Obviously, light is not emitted by an electron when moving in one of its stationary orbits, but it ejects light only when it jumps from one orbit to another. This is Bohr's second postulate.

We must note that the above postulates of Bohr are in direct contradiction to the laws of classical mechanics and to Maxwell's electromagnetic theory. According to Kepler's laws in classical mechanics, all orbits are permissible for the electron revolving round the nucleus, just as in the case of the planets revolving in orbits round the sun. Thus Bohr's first postulate regarding the stationary orbits contradicts the laws of classical mechanics. Again according to the electromagnetic theory of light, a revolving electron must emit electromagnetic radiation because its motion is under centripetal acceleration. Obviously, Bohr's second postulate that no emission or absorption of radiation takes place when the electron revolves in a stationary orbit contradicts the electromagnetic theory of light. We must note that the postulates of Bohr were introduced as *ad hoc* hypotheses. Bohr, accepted the classical laws of mechanics as the laws governing the motion of electrons. He imposed certain restrictions upon the permissible orbits through the quantum condition. Bohr made no attempts to propound any mechanics to describe the motion of the electron.

(iii) To keep the electron in its orbit around the small heavy central part of the atom called nucleus and prevent it from spiralling toward the nucleus or away from it to escape, Bohr next assumed that the centripetal force ($= mv^2/r$) required for rotation is provided by inward electrostatic force of attraction between the positively charged nucleus and the negatively charged electrons $\left(= \frac{1}{4\pi\varepsilon_0} \frac{Ze \cdot e}{r^2} \text{ in case of Hydrogen atom}\right)$. Bohr assumed that the electron was a negatively charged particle which orbited the nucleus along a circular path at a distance r (Fig. 2.7).

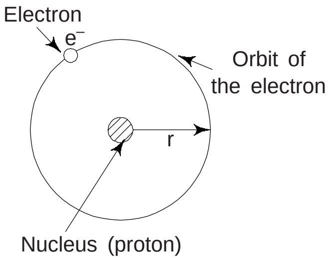


Fig. 2.7 Bohr's atomic model for hydrogen atom with circular orbit

12.6 Bohr's Theory of Hydrogen Atom

On the basis of the above three postulates, Bohr first developed the theory of hydrogen atom (which is also applicable to hydrogen like atoms). A hydrogen atom (Z = 1) is the simplest of all atoms consists of a nucleus with one positive charge e (proton) and a single electron of charge –e revolving around it in a circular orbit of radius r. Proton (i.e., nucleus of hydrogen atom) is 1836 times heavier than electron, the nucleus could be assumed at rest.

The centripetal force acting on the electron of mass m and moving with velocity v in a orbit of radius r around the nucleus = mv^2/r.

The electrostatic force of attraction between the positively charged nucleus (proton) and the electron is

$$F = \frac{1}{4\pi\varepsilon_0}\frac{Ze^2}{r^2} \quad (Z = 1 \text{ for hydrogen})$$

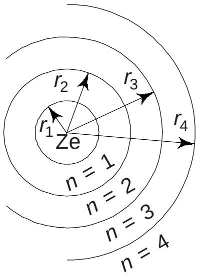


Fig. 2.8

In accordance with Newton's third law of motion

$$\frac{mv^2}{r} = \frac{Ze^2}{4\pi\varepsilon_0 r^2} \qquad (11)$$

From Bohr's quantum condition

$$mvr = L = mr^2\omega = \frac{nh}{2\pi} \qquad (12)$$

Where $n = 1, 2, 3, \ldots \infty$ is an integer, each value of n is associated with a different orbit (Fig. 2.8).

From Eq. (12),

$$v = \frac{nh}{2\pi mr} = \frac{n\hbar}{mr} \qquad (13)$$

From Eq. (11), we have

$$v^2 = \frac{1}{4\pi\varepsilon_0}\frac{Ze^2}{mr} \qquad (14)$$

Eliminating v from Eqs. (13) and (14), one obtains

$$\frac{1}{4\pi\varepsilon_0}\frac{Ze^2}{mr} = \frac{n^2\hbar^2}{m^2r^2} \qquad (\hbar = h/2\pi)$$

which yield

$$r = \frac{4\pi\varepsilon_0 n^2\hbar^2}{mZe^2} \qquad (15)$$

We can see that the *radii of the stationary orbits* are proportional to the square of the principal quantum number n, i.e., they are in the ratio of 1 : 4 : 9 : 16 : . . . The orbit with $n = 1$ has the smallest radius. For hydrogen (Z = 1), this radius is known as *Bohr radius* and is given by

$$a_0 = \frac{4\pi\varepsilon_0\hbar^2}{me^2} = 0.529 \times 10^{-10} \text{ m} = 0.529 \text{ Å}$$

Velocity of the revolving electron, i.e., orbital velocity

Substituting the value of radius, r from relation (15) in Eq. (13), one obtains

$$v = \frac{Ze^2}{2\varepsilon_0 nh} = \frac{Ze^2}{4\pi\varepsilon_0 n\hbar} \qquad (16)$$

We see that the electron velocity also depend on the quantum numbers n, i.e., velocity of the electron is inversely proportional n and is the highest in the smallest orbit having $n = 1$. The velocities become progressively less in the orbits of increasing radii. For hydrogen, the velocity of the electron in the Bohr orbit is

$$v_1 = \frac{e^2}{4\pi\varepsilon_0 \hbar} = 2.18 \times 10^6 \text{ m/s}$$

This velocity is about 1/137 times the velocity of light c (= 3×10^{10} m/s).

Orbital Frequency

We have $\omega = 2\pi\nu = v/r$

$$\therefore \qquad \nu = \frac{v}{2\pi r} \tag{17}$$

Substituting the values of v and r from Eqs. (16) and (15) in the above relation, one obtains

$$\nu = \frac{nZ^2e^4}{64\pi^2\varepsilon_0^2\hbar^3} = \frac{nZ^2e^4}{4\varepsilon_0^2 n^3 h^3} \tag{18}$$

Obviously, frequency ν is inversely proportional to n^3.

Energy of the Electron

The energy of an electron revolving around the nucleus in a orbit is of two parts:

(i) Kinetic Energy (E_k) The K.E. is

$$E_k = \frac{1}{2}mv^2 = \frac{Ze^2}{2\pi\varepsilon_0 r} \tag{19}$$

Substituting the value of r from Eq. (15) in Eq. (19), one obtains

$$E_k = \frac{mZ^2e^4}{8\varepsilon_0^2 n^2 h^2} \tag{20}$$

Potential Energy (E_p)

The potential energy is due to the position of nuclear electron in the orbit, i.e. the electron lies in the electric field of the positive nucleus, i.e. E_p = the product of electrical potential, ν and nuclear charge, e

$$\therefore \qquad E_p = -eV = -\frac{Ze^2}{4\pi\varepsilon_0 r} \tag{21}$$

$$\therefore \qquad V = -\frac{Ze}{4\pi\varepsilon_0 r} \tag{22}$$

Total Energy

Total orbital energy E is the sum of,

$$E = E_k + E_p = \frac{1}{2}mv^2 - \frac{Ze^2}{4\pi\varepsilon_0 r}$$

$$= \frac{Ze^2}{8\pi\varepsilon_0 r} - \frac{Ze^2}{4\pi\varepsilon_0 r} = -\frac{Ze^2}{8\pi\varepsilon_0 r} \tag{23}$$

Substituting the value of r from Eq. (15), one obtains,

$$E_n = -\frac{mZ^2e^4}{8\varepsilon_0 n^2 h^2} = -\frac{mZ^2e^4}{32\pi^2\varepsilon_0^2 n^2 \hbar^2} \tag{24}$$

Substituting the values of constants, i.e., $m = 9.1 \times 10^{-31}$ kg, e = 1.6×10^{-19}, $\varepsilon_0 = 8.854 \times 10^{-12}$ and $h = 6.62 \times 10^{-34}$ in relation to 24, we obtain

$$E_n = -\frac{9.1\times10^{-31}\times Z^2\,(1.6\times10^{-19})^4}{8(8.854\times10^{-12})^2\,n^2\,(6.62\times10^{-34})^2}\ \text{Joules}$$

$$= -\frac{21.7\times10^{-19}}{n^2}\,Z^2\ \text{Joules}$$

$$= -\frac{21.7\times10^{-19}}{1.6\times10^{-19}}\,\frac{Z^2}{n^2}\ \text{Joules}$$

$$= -13.6\,\frac{Z^2}{n^2}\ \text{eV} \qquad (\because\ 1\ \text{eV} = 1.6\times10^{-19}\ \text{J})$$

Thus the total energy of an electron in the first orbit

$$E_1 = -13.6\,\frac{Z^2}{(1)^2} = -13.6\,Z^2\ \text{eV}$$

Similarly, the total energy of an electron in the second orbit

$$E_2 = -13.6\,\frac{Z^2}{(2)^2} = -3.4\,Z^2\ \text{eV}$$

and in the third orbit

$$E_3 = -13.6\,\frac{Z^2}{(3)^2} = -1.51\,Z^2\ \text{eV}$$

Similarly $E_n = 0$ as $n \to \infty$

The energy of the electron given by Eq. (24) is negative. So as n increases, i.e., as the electron goes to orbits of larger radii, E_n increases and converges to the limiting value $E_n = 0$ as $n \to \infty$. E_n is the lowest for the smallest orbit for which $n \approx 1$. Further the negative sign of the total energy of an electron shows that the electron is bound to the nucleus by the attractive forces. To separate it from the bound of the nucleus, some work is required to be done. The total energy of an electron is known as *binding energy* or *Work function* and it is considered to be a positive quantity.

If the electron, initially in an orbit of quantum number n_i, makes a transition to a final orbit of quantum number n_f, then its energy changes from E_i to E_f where we have from Eq. (24)

$$E_i = -\frac{mZ^2e^4}{32\pi^2\varepsilon_0^2 n_i^2 \hbar^2}$$

and

$$E_f = -\frac{mZ^2e^4}{32\pi^2\varepsilon_0^2 n_f^2 \hbar^2}$$

If $n_i > n_f$, then $E_i > E_f$. According to Bohr's postulate, we have in this case

$$\hbar\omega = E_i - E_f = \frac{mZ^2 e^4}{32\pi^2 \varepsilon_0^2 \hbar^2} \left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \tag{25}$$

So that the frequency of the emitted radiation is

$$\nu = \frac{\omega}{2\pi} = \frac{mZ^2 e^4}{64\pi^3 \varepsilon_0^2 \hbar^3} \left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \tag{26}$$

If λ is the wavelength of the emitted radiation, then $c = \nu\lambda$. Since ν full waves are contained in a path length c, the number of full waves per unit length which is known as the *wave number* is given by

$$\frac{1}{\lambda} = \bar{\nu} = \frac{\nu}{\mathrm{c}} = \frac{mZ^2 e^4}{64\pi^2 \varepsilon_0^2 c\hbar^3} \left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

$$= R\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \tag{27}$$

where $R = \dfrac{me^4}{64\pi^3 \varepsilon_0^2 c\hbar^3} = \dfrac{me^4}{8\varepsilon_0^2 h^3 c}$ = 1.0937 × 10^7 m^{-1} is the *Rydberg constant* of the Bohr formula. The unit of wave number is m^{-1}. In terms of R, the energy of the electron in the nth orbit is given by

$$E_n = -\frac{2\pi c\hbar R Z^2}{n^2} \tag{28}$$

Example 1 *Calculate the distance of the closest approach of α-particles from the copper nucleus, when α-particles of 5* MêV *energy are scattered back by a thin sheet of copper. Given Z for Cu = 29, e = 1.6 × 10^{-19}C, ε_0 = 8.854 × 10^{-12} F/m.*

Solution

$$\text{Kinetic energy of } \alpha\text{-particles} = \frac{1}{2} mv^2 = 5 \text{ MeV}$$

$$= 5 \times 10^6 \times 1.6 \times 10^{-19} = 8 \times 10^{-13} \text{ J}$$

or $$mv^2 = 16 \times 10^{-13} \text{ J}$$

$$\phi = 180°$$

$$Z = 29$$

∴ $$d = \frac{Ze^2}{2\pi\varepsilon_0 mv^2}\left(1 + \operatorname{cosec}\frac{\phi}{2}\right)$$

$$= \frac{29 \times (1.6 \times 10^{-19})^2 \times 7}{2 \times 22 \times (8.85 \times 10^{-12}) \times 16 \times 10^{-13}} (1 + \operatorname{cosec} 90)$$

$$= 167.04 \times 10^{-16} \text{ m.}$$

Example 2 *Calculate the radius and frequency of an electron in the Bohr's first orbit in hydrogen atom. Given $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m, m = 9.1×10^{-31} kg, e = 1.6×10^{-19} C, h = 6.625×10^{-34} J-s*

Solution

$$r_n = \frac{\varepsilon_0 n^2 h^2}{\pi m Z e^2}$$

For the first Bohr orbit of hydrogen atom, $n = 1$ and $Z = 1$. Thus

$$r_1 = \frac{\varepsilon_0 h^2}{\pi \mathrm{me}^2} = \frac{8.854 \times 10^{-12} \times (6.625 \times 10^{-34})^2}{\frac{22}{7} \times 9.1 \times 10^{-31} \times (1.6 \times 10^{-19})^2}$$

$$= 5.3 \times 10^{-11} \text{ m} = 0.53 \text{ Å}$$

and orbital frequency

$$\nu = \frac{mZ^2 e^4}{4\varepsilon_0^2 n^3 h^3}$$

$$= \frac{9.1 \times 10^{-31} \times 1 \times (1.6 \times 10^{-19})^4}{4 \times (8.854 \times 10^{-12})^2 \times (1)^3 \times (6.625 \times 10^{-34})^3}$$

$$= 6.54 \times 10^{15} \text{ Hz.}$$

Example 3 *The radius of first orbit of electron in a hydrogen atom is 0.529 Å. Show that the radius of the second Bohr orbit in a singly ionized helium atom is 1.058 Å.* *[AMIE]*

Solution For hydrogen: Z = 1, $n = 1$ and $r_1 = 0.529$ Å
For helium: Z = 2, $n = 2$ and $r_2 = ?$
The radius of nth Bohr orbit of an atom,

$$r_n = \frac{\varepsilon_0 n^2 h^2}{\pi \mathrm{mZe}^2} = k\frac{n^2}{Z} \qquad \text{(i)}$$

Where

$$k = \frac{\varepsilon_0 h^2}{\pi m e^2} \text{ is a constant}$$

For the first Bohr orbit of electron in hydrogen atom, we have

$$(r_1)_H = 0.529 \text{ Å} = k\ \frac{(1)^2}{1}$$

$$\therefore \qquad k = 0.529 \text{ Å}$$

For helium (Z = 2), the radius of second Bohr orbit ($n = 2$), we have

$$(r_2)_{\mathrm{He}} = k\ \frac{n^2}{Z} = 0.529 \times \frac{2^2}{2} = 1.058 \text{ Å}$$

Example 4 *Show that the energy released by an electron jumping from orbit 3 to orbit 1 and the energy emitted by an electron jumping from orbit 2 to orbit 1 are in the ratio 32 : 27.* *[AMIE]*

Solution The energy of an electron in the nth Bohr orbit of an atom is given by

$$E_n = -\frac{13.6 Z^2}{n^2} \text{ eV}$$

Now, for the first orbit (n = 1),

$$E_1 = -13.6 \ \frac{Z^2}{(1)^2} \ \text{eV}$$

For the second orbit (n = 2),

$$E_2 = -13.6 \ \frac{Z^2}{(2)^2} \ \text{eV}$$

Similarly for the third orbit (n = 3),

$$E_3 = -\ 13.6 \ \frac{Z^2}{(3)^2} \ \text{eV}$$

Energy emitted by an electron jumping from orbit 3 to orbit 1,

$$E_3 - E_1 = -13.6 \ \frac{Z^2}{(3)^2} - \left(-\frac{13.6\,Z^2}{(2)^2}\right)$$

$$= -13.6 \ Z^2 \left(\frac{1}{9} - \frac{1}{1}\right) = -13.6 \ Z^2 \left(-\frac{8}{9}\right)$$

Similarly, energy emitted by an electron jumping from orbit 2 to orbit 1,

$$E_2 - E_1 = -\frac{13.6\,Z^2}{(2)^2} - \left(-\frac{13.6\,Z^2}{(1)^2}\right)$$

$$= -13.6 \ Z^2 \left(\frac{1}{4} - \frac{1}{1}\right) = -13.6 \ \left(-\frac{3}{4}\right)$$

$$\therefore \qquad \frac{E_3 - E_1}{E_2 - E_1} = \frac{8/9}{3/4} = 32 : 27$$

Example 5 *How many revolutions does an electron in the first Bohr orbit of hydrogen make per second.* *[B.E.]*

Solution The number of revolutions per second = $v/2\pi r$

Now, substituting for r and v, one obtains,

Number of revolutions per second

$$= \frac{Ze^2}{2\varepsilon_0 nh} \times \frac{\pi \text{mZe}^2}{2\pi\varepsilon_0 n^2 h^2} = \frac{mZ^2 e^2}{4\varepsilon_0^2 n^3 h^3}$$

Using $m = 9.1 \times 10^{-31}$ kg, $Z = 1$, $e = 1.6 \times 10^{-19}$ C,

$\varepsilon_0 = 8.85 \times 10^{-12}$ F/m, $n = 1$, and $h = 6.63 \times 10^{-34}$,

We obtain

$$\text{Number of rev./sec} = \frac{9.1 \times 10^{-31} \times (1.6 \times 10^{-19})^4}{4 \times (8.25 \times 10^{-12})^2 \times (6.63 \times 10^{-34})^2}$$

$$= 7.23 \times 10^{17}$$

Example 6 *Determine the orbital frequency of an electron in the first Bohr's orbit in a hydrogen atom.* *[AMIE]*

Solution Given $n = 1$, $Z = 1$ for hydrogen atom

The orbital frequency of the electron

$$\nu_n = \frac{mZ^2e^4}{4\varepsilon_0^2 h^3 n^3}$$

Substituting the values of various constants and simplifying, we obtain

$$\nu_n = 6.56 \times 10^{15} \frac{Z^2}{n^3}$$

Now, for first Bohr orbit of hydrogen atom

$$\nu_1 = 6.56 \times 10^{15} \frac{(1)^2}{(1)^3} = 6.56 \times 10^{15} \text{ Hz.}$$

Example 7 *Calculate the values of kinetic energy, potential energy and total energy of an electron in a hydrogen atom in its ground state. Given* $\varepsilon_0 = 8.854 \times 10^{-12}$ *F/m,*
$h = 6.625 \times 10^{-34}$ *J-s,* $m = 9.11 \times 10^{-31}$ *kg and*
$e = 1.6 \times 10^{-19}$ *c.*

Solution For an electron in the ground state orbit of the hydrogen; $n = 1$ $Z = 1$
Using relations derived earlier and values of physical constants given, we obtain,

Total energy
$$E = -\frac{9.11 \times 10^{-31} \times 6.554 \times 10^{-76}}{8 \times 78.39 \times 10^{-24} \times 43.89 \times 10^{-60}}$$
$$= -2 \times 10^{-18} \text{ J} = -13.6 \text{ eV}$$
$$\text{K.E.} = 13.6 \text{ eV}$$
$$\text{P.E.} = -27.2 \text{ eV}$$

Example 8 *Calculate the velocity of an electron in hydrogen atom in Bohr's first orbit. Given*
$h = 6.626 \times 10^{-34}$ *J-s,* $\varepsilon_0 = 8.825 \times 10^{-12}$ *F/m and*
$e = 1.6 \times 10^{-19}$ *c.*

Solution For hydrogen $n = 1$, $Z = 1$

We have
$$v_n = \frac{Ze^2}{2\varepsilon_0 nh}$$

∴
$$v_1 = \frac{e^2}{2\varepsilon_0 h} = \frac{(1.6 \times 10^{-19})^2}{2 \times 8.825 \times 10^{-12} \times 1.626 \times 10^{-34}} \text{ m/s}$$
$$= 2.189 \times 10^6 \text{ m/s.}$$

Example 9 *Electrons of energies 10.2 eV and 12.09 eV can cause radiation to be emitted from hydrogen atoms. Calculate in each case: (a) the principal quantum number of the orbit to which electron in the hydrogen atom is raised and (b) the wavelengths of radiation emitted if the electron drops back to the ground level.* *[AMIE]*

Solution The electron is excited from $n = 1$ (ground state) to the excited state $n = ?$ (higher energy level). During this process of excitation, electron acquires energy of 10.2 eV or 12.09 eV.

We have
$$E_{n_2} - E_{n_1} = \frac{me^4}{8\varepsilon_0^2 h^2}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

(a) $10.20 = 13.6 \left(\frac{1}{1^2} - \frac{1}{n^2}\right)$ eV $= 13.6 - 13.6/n^2$

$\therefore \quad n^2 = 13.6/3.41 = 4 \;\therefore\; n = 2$

Thus, when an electron of energy 10.2 eV excites the hydrogen electron from ground level it goes to second orbit.

When this electron falls back to ground level it will lose same amount of energy, i.e. 10.2 eV.

$$\because \quad E = h\nu$$

$$\therefore \quad \lambda = \frac{hc}{E} = \frac{6.625 \times 10^{-34} \times 3 \times 10^{-8}}{10.2 \times 1.6 \times 10^{-19}} = 12.16 \times 10^{-10} \text{ m}$$

$$= 1216 \text{ Å}$$

(b) $12.09 = 13.6 \left(\frac{1}{1^2} - \frac{1}{n^2}\right)$

or $\frac{13.6}{n^2} = 1.52$

or $n^2 = 9 \;\therefore\; \text{n} = 3$

Obviously, when 12.09 eV energy is absorbed by the electron of hydrogen atom in its ground level, it is excited to $n = 3$ orbit. While returning back to the ground level (from $n = 3$ to $n = 1$), electron will emit energy = 12.09 eV.

$$\text{Now,} \quad \lambda = \frac{\lambda c}{E} = \frac{6.625 \times 10^{-34} \times 3 \times 10^{8}}{12.09 \times 1.6 \times 10^{-19}}$$

$$= 10.26 \times 10^{-10} \text{ m} = 1026 \text{ Å}$$

12.7 The Hydrogen Spectrum

Towards the end of the 19th century it was discovered that the light which is discntinuous in frequency distribution forms a discrete set of light images that are termed as *spectral lines*. One can analyze these lines by the distribution of their frequency or colour. It was also discovered that the wavelengths of the lines of a spectrum fall into definite sets which are called *series*. An element may display several series. Each series can be represented by an empirical formula which has a similarity for all the series of a given element. While making a study of the visible spectrum of simplest element and probably the simplest structure, hydrogen, J.J. Balmer in 1885 discovered the first such spectral series. This series is now known as *Balmer series* of hydrogen (Fig. 2.9).

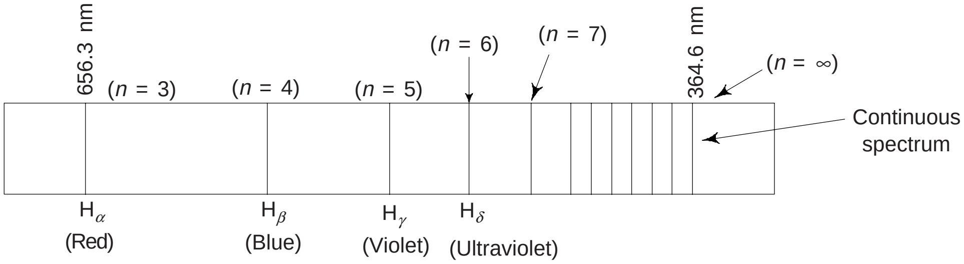


Fig. 2.9 Balmer series of hydrogen atom

Starting from the line of longest wavelength 656.3 nm the various lines in this series are designated as H_α, H_β, H_γ, etc. We see that towards the shorter wavelengths the lines get closer together until we reach what is known as the series limit at 364.6 nm. Beyond this series limit there are no discrete lines but only a faint continuous spectrum. Balmer enunciated an empirical relation to represent the wavelength of this series as

$$\bar{\nu} = \frac{1}{\lambda} = R\left[\frac{1}{2^2} - \frac{1}{n^2}\right] \qquad (n = 3, 4, 5, \ldots) \tag{29}$$

This series was the first to the observed simply because it is in the visible region.

The success of Balmer led Rydberg to attempt a general formulation which may be applied to heavier elements. Rydberg proposed a relation of the form.

$$\bar{\nu} = \frac{1}{\lambda} = A - \frac{R}{(n+\alpha)^2} \tag{30}$$

where n is running integer, A and α are constants which depend on the element and part of the spectrum or spectral series to which the relation (30) is applied. One can use this relation in many spectral series. The experimental determined value of constant R is 1.09737×10^7 m^{-1}. The relation (30) was further modified by Ritz to

$$\bar{\nu} = \frac{1}{\lambda} = \frac{R}{(m+\beta)^2} - \frac{R}{(n+\alpha)^2} \tag{31}$$

Where α and β are adjustment constants and depend mainly on the element. m takes on different integral values for different spectral series of a given element. One can compute the different lines within a series by changing the running integer n = 3, 4, 5, 6. We can easily see that $\alpha = \beta = 0$ and $m = 2$ in Eq. (31) reduces to Balmer formula (29) for hydrogen.

An examination of the ultraviolet and infrared regions revealed the existence of other series in hydrogen spectrum. These other series are

Lyman Series: $$\bar{\nu} = \frac{1}{\lambda} = R\left[\frac{1}{1^2} - \frac{1}{n^2}\right] \qquad (n = 2, 3, 4)$$

Paschen Series: $$\bar{\nu} = \frac{1}{\lambda} = R\left[\frac{1}{3^2} - \frac{1}{n^2}\right] \qquad (n = 4, 5, 6)$$

Brackett Series: $$\bar{\nu} = \frac{1}{\lambda} = R\left[\frac{1}{4^2} - \frac{1}{n^2}\right] \qquad (n = 5, 6, 7)$$

Pfund Series: $$\bar{\nu} = \frac{1}{\lambda} = R\left[\frac{1}{5^2} - \frac{1}{n^2}\right] \qquad (n = 6, 7, 8) \tag{32}$$

The value of R is the same for all the series. The origin of different spectral series of hydrogen due to transitions between the different orbits is shown in Fig. 2.10. We must note that the radii of the different orbits have not been drawn according to scale in this figure. The transitions are indicated by drawing arrows from the initial to the final orbits.

Bohr's model of the atom explain quite satisfactorily the observed hydrogen spectrum. Though the Bohr's postulates are quite arbitrary but the explanation of the hydrogen spectrum was one of its major triumphs. These postulates have now been replaced by more fundamental hypothesis in the modern quantum mechanics. However, Bohr's model provides a suitable transition from the unsatisfactory classical theory to more abstract quantum theory. Moreover the model provides a physical picture of the atom which even today is a useful concept.

12.8 Normal and Excited States of Atom

When the single electron of a hydrogen atom is in the ground state, i.e. the innermost orbit, n = 1, the atom is said to be in the normal or ground state. Now, if we allow to pass an electrical discharge through a vessel

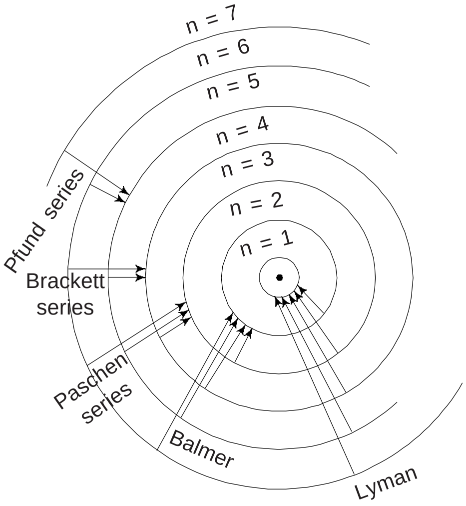


Fig. 2.10 Different spectral series of hydrogen atom

containing a hydrogen gas, cathode rays (electrons) moving at a high speed make frequent collisions with electrons, knocking some of the electrons out of the atom completely and some of the electrons into one of the outer allowed orbits, i.e., $n = 2, 3, 4, 5, 6, \ldots$

When the electron is completely removed from the atom, i.e. $n = \infty$, the atom is said to the *ionized*. The corresponding potential (V_i) is known as the *ionization potential*. Ionization energy or ionization potential is expressed in units of electron volt (eV : 1 eV = 1.6×10^{-19} J). The inert or nobel gases have the highest ionization energy as these elements have stable electronic configuration. The electrons fill completely a shell or sub-shell. We can see that alkali atoms, e.g. Lithium and Sodium have low ionization potential as they have one electron in outermost & sub-shell, beyond the stable configuration and hence can easily be removed out of the atom. The another term associated with ionization energy is *electron affinity*. This is the amount of energy released, when a neutral atom adds an electron. The energy required to transfer an electron from atom 1 to atom 2 is the difference between the ionization energy I_1 and the electronic affinity E_{12} of the respective atoms, i.e., $I_1 - E_{12}$. We see that Halogen atoms have the highest electron affinity.

However, when the electron is forced into an outer orbit (say $n = 2, 3, 4, 5, \ldots$) the atom is said to be excited. In the unexcited normal state, i.e. ground state, with n = 1, the electron is in its lowest energy state at the bottom. The electron moves in this orbit continuously without emitting or absorbing energy and is said to be in stable state. When electron is excited (say $n = 2, 3, 4, 5 \ldots$ orbits), it absorbs energy. When electron returns from the excited state to any of the lower states, it emits energy.

12.9 Energy Levels

According to Bohr's theory of hydrogen atom, there are only certain discrete energy levels in the hydrogen atom. When an electron makes a spontaneous jump from an outer orbit to an inner orbit, energy is radiated which is equal to the difference in energy between the orbits or levels. According to theory, the electron has definite energy in a stationary orbit, given by Eq. (28),

$$E_n = -\frac{chRZ^2}{n^2}$$

For hydrogen $Z = 1$, so that

$$E_n = -\frac{chR}{n^2} = -\frac{13.6}{n^2} \tag{33}$$

We can represent the energies of the electron in the different orbits of hydrogen (or other hydrogen like atoms) given by Eq. (33) by drawing a set of horizontal lines for different values of n. This has been done for hydrogen in Fig. 2.11. These are known as the *energy levels* of hydrogen. Electron energy levels are most conveniently expressed in electrons volts (eV).

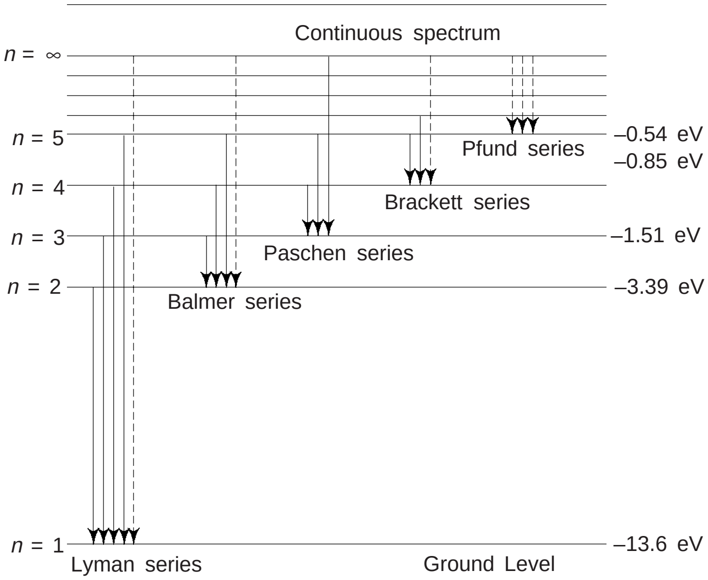


Fig. 2.11 Energy level diagram for hydrogen atom

$$\begin{aligned} 1\text{eV} &= \text{electronic charge (e)} \times 1 \text{ volt} \\ &= 1.6 \times 10^{-19} \text{ C} \times 1 \text{ Volt} \\ &= 1.6 \times 10^{-19} \text{ Joule} \end{aligned}$$

From relation (33), we can see that electron volt energies of shells corresponding to $n = 1, 2, 3, 4, 5, \ldots, \infty$ shells or orbits are

For ground state $n = 1$ (K-shell) : $E_1 = -13.6$ eV

$$n = 2 \text{ (L-shell)} : E_2 = -\frac{E_1}{2^2} = -3.4 \text{ eV}$$

$$n = 3 \text{ (M-shell)} : E_3 = -\frac{E_1}{3^2} = -1.51 \text{ eV}$$

$$n = 4 \text{ (N-shell)} : E_4 = -\frac{E_1}{4^2} = -0.86 \text{ eV}$$

$$n = 5 \text{ (O-shell)} : E_5 = \frac{E_1}{5^2} = -0.54 \text{ eV}$$

. . .

. . .

$n = \infty$ $E_\infty = 0$ eV.

We can see that the minimum energy required to set free an electron originally bound in the lowest energy ($n = 1$) of the hydrogen atom is 13.6 eV, which is ionization energy for hydrogen.

In the commonly known notations, an electron in a $n = 1$ is said to be in the K-shell. Correspondingly, if $n = 2$, $n = 3$ or $n = 4$, the electron is said to be in the L, M or N shells respectively. Obviously, K, L, M, N, O, P, and Q – shells correspond to quantum numbers n = 1, 2, 3, 4, 5, 6, and 7. The sub-shells in each main shell are designated by *s*, *p*, *d* and *f*, – with l = 0, 1, 2 and 3 respectively. We must remember that the number of sub shells in a given main shell is same as its principal quantum number (n). However, in none of the known elements does any of the last three shells O, P, Q contains more than three sub shells.

12.9(a) Energy Level Diagram of Multielectron Atoms

An atom containing more than one electron is called a multielectron atom and a standard energy level diagram for such types of atoms is shown in Fig. 2.11(a).

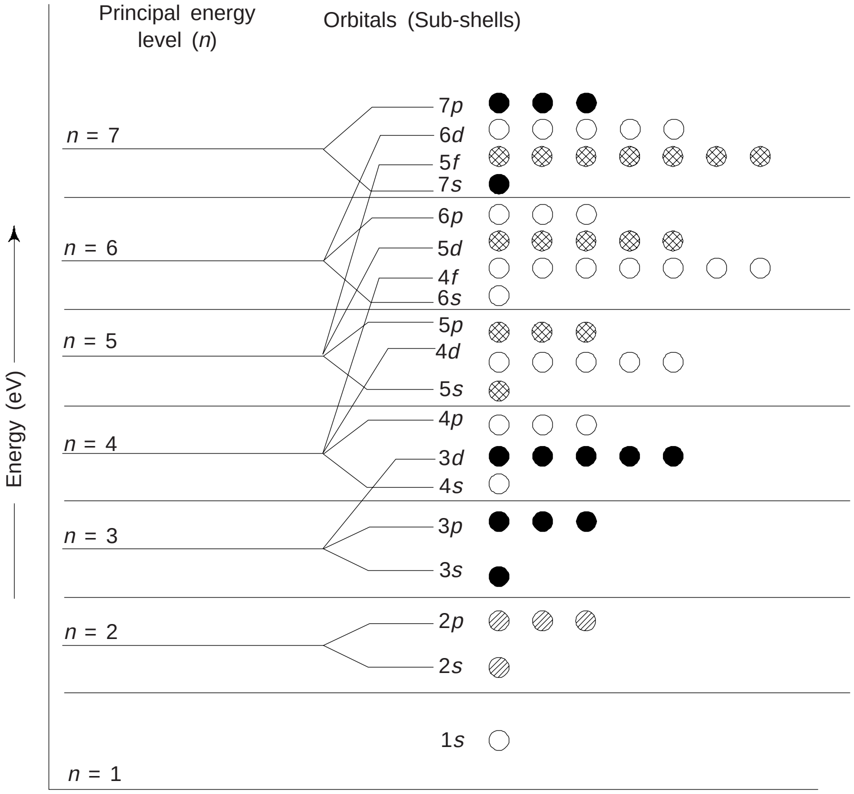


Fig. 2.11(a) A standard energy level diagram of a multielectron atom

There may be an infinite number of principal energy levels in such atoms. However, with the present state of knowledge, we have represented the distribution of electrons in seven principal levels only, i.e. n = 1 to 7, i.e. from bottom to top, up to n = 7 energy levels.

The following set of rules helps one to fill up electrons in various sub-shells or orbitals:

(i) As we move from one element to the next in the order of increasing atomic number, the electrons are added one by one to the orbitals.

(ii) The orbitals are filled successively in the order of increasing energy. The energy increases in the order as shown in Fig. 2.11(a), i.e., 1*s*, 2*s*, 2*p*, 3*p*, 3*s*, 4*s*, 3*d*, 4*p*, 5*s*, 4*d*, 5*p*, 6*s*, 4*f*, 5*d*, 6*p* and so on.

(iii) The pairing of electrons in any *s*, *p*, *d* or *f* orbitals is not possible, until all the available orbitals of the atom contain one electron each. This is called *Hund's rule* of maximum multiplicity.

(iv) In accordance with Pauli's exclusion principle, an orbital of a sub-shell in a given principle quantum number can accommodate a maximum of two electrons of opposite spin.

(v) Orbitals in the same sub-level tend to acquire full capacity or exactly half-filled electrons, as it represents a stable arrangement of electrons.

(vi) When two or more orbitals of a given set, i.e. *p*, *d* or *f* orbitals contain one electron each, they tend to have the same spin. However, no two electrons in an orbital can have same spin.

Now, we examine the sequence of filling up the electrons in various orbitals in the different energy levels of multielectron atoms:

(i) First Principal Energy Level (n = 1): This has only one orbital i.e., 1 *s*-orbital and accommodate only two electrons.

(ii) Second Principal Energy Level (n = 2): This has two orbitals 2*s* and 2*p*. 2*s*-orbtial has only one sub-sub-shell and can accommodate only two electrons. 2*p* orbital has three sub-sub-shells and can accommodate at the most 6 electrons and each sub-sub-shell cannot accommodate more than two electrons.

The maximum number of electrons that can be accommodated in $n = 2$ energy level is 8 and the total number of maximum electrons that can be accommodated upto the second principal energy level is 10 (2 in $n = 1$ level and 8 in $n = 2$ level). However, the second principal energy level ($n = 2$) will only start accommodating electrons only after the first principal energy ($n = 1$) is completely filled up.

(iii) Third Principal Energy Level (n = 3): This has three orbitals: 3*s*, 3*p* and 3*d* orbitals. From Fig. 2.11(a), we note that the 3*d*-orbital does not fall within the range of $n = 3$ principal energy level, as the energy of the 3*d*-orbital is greater that of 4*s* orbital in the fourth principal energy level. This is why the 3*d*-orbital is filled only after filling up of 4*s*-orbital of the 4th principal energy level. The sequence of filling up 3*s* and 3*p* orbitals is as follows:

3s orbital has only one sub-sub-shell and can accommodate only two electrons.

3p-orbital has three sub-sub-shells and can accommodate at the most 6 electrons.

Thus the maximum number of electrons filled up in the $n = 3$ level is 8 and the total number of electrons which can be accommodated up to $n = 3$ level is 18. We must remember that $n = 3$ level will start accommodating the electrons only after the second principal energy level is completely filled up.

(iv) Fourth Principal Energy level (n = 4): This has four orbitals: 4*s*, 4*p*, 4*d* and 4*f*. The energy of 4*d* and 4*f* orbitals are greater than 5*s* and 6*s* orbitals respectively (Fig. 2.11(a)) and they do not fall within the range of fourth principal energy level. Obviously, $n = 4$ energy level contains only 4*s*, 3*d* and 4*p* orbitals as shown. The sequence of filling these orbitals is as below:

4s-orbital: It has only one sub-shell and can accommodate only 2 electrons.

3d-orbtial: It has 5 sub-shells and can accommodate at the most 10 electrons but each sub-sub-shell of this orbital cannot accommodate more than 2 electrons.

4p-orbital: This has three sub-sub-shells and can accommodate at the most 6 electrons.

Thus $n = 4$ has the maximum capacity of 18 electrons and the total number of electrons which can be accommodated upto fourth principal energy level is 36. We must remember that $n = 4$ level will start accommodating the electrons only after the $n = 3$ level is completely filled up.

Similarly, 5th, 6th and 7th principal energy levels are filled with electrons. The maximum capacity of 5th, 6th and 7th levels are 18, 32 and 32 respectively. The sequence of filling up of the orbitals of $n = 5$ level is 5*s*, 4*d* and 5*p*; the sequence of filling up of the orbitals of $n = 6$ level is 6*s*, 4*f*, 5*d* and 6*p* and similarly that of $n = 7$ level is 7*s*, 5*f*, 6*d* and 7*p*. We must remember that each *f*-orbital has 7 sub-shells and each sub-shell cannot accommodate more than 2 electrons.

12.10 Limitations of Bohr's Theory of Hydrogen atom

Although the Bohr's model of hydrogen atom is rather successful and the idea of quantization of atomic energy is correct, the model is unsatisfactory in the following respects:

(i) The assumptions made that only circular orbits are allowed where as elliptical orbits are also possible. Obviously, these assumptions are inexplicable and arbitrary.

(ii) This model of the atom cannot be generalized to deal with complex atoms, i.e., systems with two or more planetary electrons.

(iii) The model gives no information regarding the arrangement of distribution of electrons in the atom.
(iv) The model unable to handle unbound systems.
(v) The model deals with the planetary model which introduces only one quantum number, i.e. principal quantum number, *n*
(vi) Model provides no explanation as to why certain spectral lines are more intense than others and also there is no method to calculate the rate of transition between different energy levels.
(vii) Model fails to explain wave nature of electrons.
(viii) Model fails to explain the fine structure lines for hydrogen.
(ix) The model fails to explain as to how the interaction of individual atoms take place, among themselves, resulting in aggerate of matter.
(x) Model does not provide any explanation for chemical bonding.

12.11 Sommerfeld-Wilson's Model of the Atom

In Bohr's model of the atom, the electron is assumed to be circular. Actually however, the orbit of a particle acted upon by an attractive force directed towards a fixed point (central force) is an *ellipse* with the centre of force at one of the foci of the ellipse, if the force $F \propto 1/r^2$ and the energy E of the particle is negative. The orbit is circular only as a special case. Sommerfeld and Wilson in 1916 showed how to remove the restriction to circular orbits and also made relativistic correction to the Bohr atomic model.

In order to explain the observed spectra of complex atoms having two or more planetary electrons as well as to account for the fine structure, Sommerfeld and de Broglie have taken into consideration a number of modifications. Sommerfeld extended the Bohr's theory to allow for elliptical orbits which are also possible according to Newtonian mechanics. For this he introduced a second quantum number l which is analogous to the angular momentum number l introduced earlier and also retained the quantum number n which is now redefined in terms of the total momentum. Obviously, Sommerfeld introduced two new quantum numbers. We shall discuss these quantum numbers alongwith the principal quantum number in subsequent sections of this chapter.

According to Sommerfeld's theory, an electron in any allowed energy level of hydrogen atom may move in any one of a number of orbits. For each energy level $n = 1$, $n = 2$, $n = 3$, etc., there are n possible orbits. The allowed orbits for the first three energy levels are as shown in Fig. 2.12.

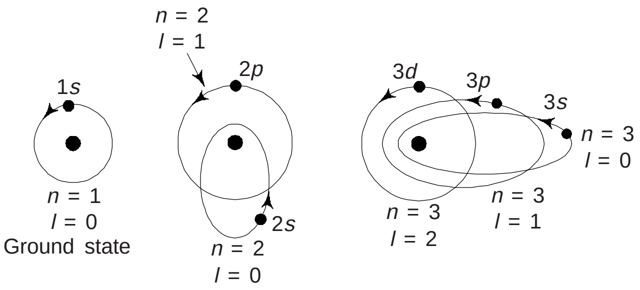


Fig. 2.12 Electron orbitals for an hydrogen atom according to Sommerfeld's atomic model

We note that for $n = 3$, there are three orbits with $l = 2$, $l = l$ and $l = 0$. We also note that for a particular value of 'n', according to Sommerfeld's model there are the same number of possible sub-shells out of which one is circular and remaining, i.e. $(n - 1)$ are elliptical in shape. The diameter of the circular orbit is given by Bohr's theory and is just equal to the major axis of the two elliptical orbits. The minor axes are 2/3 and 1/3 of the major axis. Let us examine this in detail.

In Sommerfeld's theory all the sub-shells are designated by '*l*'. For a particular value of '*n*', the different values of '*l*' are 0, 1, 2, . . . ($n - 1$). '*n*' is called the *principal quantum number* and '*l*' as *orbital* or *azimuthal quantum number*.

It was shown by Sommerfeld that the energy level '*n*' and the related sub-shells '*l*' are related through the relation.

$$\frac{b}{a} = \frac{l+1}{n} \tag{34}$$

where a and b are semi-major and semi-minor axis of an elliptical orbit.

When $n = l$, then l = 0, i.e. $a = b$.

In this case the two axes are equal and the orbit corresponding to $n = 1$ is circular. This sub-shell is designated as *s*-sub shell. Since this sub-shell belongs to $n = 1$ and hence it is designated as *s*-sub shell. For $n = 2$, l has two values 0 and 1. Using these values of l, one obtains from Eq. (34),

For $l = 0$, $\frac{b}{a} = \frac{0+1}{1} = \frac{1}{2}$ or $b = \frac{a}{2} = 0.5a$: obviously, this sub-shell is elliptical in shape and this is designated as 2*s*-sub-shell.

For $l = 1$, $\frac{b}{a} = \frac{1+1}{2} = 1$ or $b = a$: obviously, the sub-shell corresponding to $l = 1$ is circular in shape and is designated as 2*p*. Similarly, we can easily see that for $n = 3$, l has three values 0, 1, 2. Using these values of l in Eq. (34), one obtains,

For $l = 0$, $\frac{b}{a} = \frac{0+1}{3} = \frac{1}{3}$ or $b = \frac{a}{3}$, corresponding sub shell is elliptical in shape and designated as 3s.

For $l = 1$, $\frac{b}{a} = \frac{1+1}{3} = \frac{2}{3}$ or $b = \frac{2a}{3}$, corresponding sub-shell is elliptical in shape and designated as 3*p*.

For $l = 2$, $\frac{b}{a} = \frac{2+1}{3} = \frac{3}{3}$ or $b = a$ obviously, the sub-shell corresponding to this is circular and designated as 3*d*.

In accordance with Sommerfeld's model, one can designate the letters to the *l*-values, as follows:

$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$ . . .
s	*p*	*d*	*f*	*g* . . .

We must note that according to this model of atom, all orbits with the same value of *n* have the same total energy, the energy given by Bohr's theory of circular orbits. Each of the allowed orbits of the Sommerfeld's theory of the hydrogen atom becomes a sub-shell into which electrons are added to build up the elements of periodic table. We must note that the maximum number of allowed electrons in any one sub-shell is governed by the relation $2(2l + 1)$.

13. VECTOR ATOM MODEL

This is the widely accepted present day atomic model and is also known as *quantum model* of the atom. In this model of the atom all the principal quantized terms are represented by vectors. This model takes into account electron spin while retaining the feature of planetary movement of electrons in different orbits (circular or orbital) and movement of electron in different planes of Sommerfeld model. Based on this model several investigators have calculated the fine structure separations of various energy levels and studied the effects of electric and magnetic fields on spectral lines, which is not possible from Sommerfeld's model.

In Bohr's model of hydrogen atom, we have seen that the angular momentum of the rotating electron was an integer of $h/2\pi\,(=\hbar)$. When an atom is placed in an external magnetic or electric field there will be an interaction between the atom and the external field. As a result each energy level in the atom will split into two or more sub-levels. Transitions may then take place between the resulting levels according to prescribed selection rules. The resulting fine structure pattern is called the *Zeeman effect* in the case of a magnetic field and the *Stark effect* in the case of an electric field. These effects clearly reveal that there was *spining* of electrons due to orbital angular momentum and magnetic momentum under the influence of external magnetic and electric fields. The fine structures created by the effect of external magnetic and electric field show that electrons move in different planes (horizontal or vertical) in order to give spatial motion. We must note that these energy levels are different in different planes. With the introduction of this model of the atom, the two new quantum numbers, the magnetic spin quantum number (m_s) and the magnetic orbital quantum number (m_l), in addition to the principal quantum number (n) and the orbital or azimuthal quantum number (l) can be used to completely represent the energy state. At this stage it would be proper to summarize our present day understanding about the atomic model.

(i) An atom is made of protons, neutrons and electrons possessing positive, neutral and negative charge respectively. In a normal atom, the number of protons is equal to the number of electrons, i.e., atom is electrically neutral.

(ii) The protons and neutrons are located in the tiny nucleus of diameter about 10^{-14} m at the centre of the atom. The size of the atom is about 10^{-10} m. Due to the presence of protons, the nucleus is positively charged.

(iii) The electrons revolve round the nucleus in fixed circular as well as elliptical paths known as principal energy levels or main shells denoted by letters K, L, M, N, O, P, Q and so on. These levels are designated from the nucleus to outwards.

(iv) Each energy level or shell is further divided into *sub-shells* which are designated by the letters *s*, *p*, *d*, *f*, ... The table 2.3 shows these designations.

Table 2.3

Orbital quantum number	0	1	2	3	4	5	6
Electron designation	*s*	*p*	*d*	*f*	*g*	*h*	*i*
Shells in the atom	K	L	M	N	O	P	Q

(v) There is no change in the energy of an electron, so long as it keeps on revolving in the same shell or orbit and the atom is said to be stable. When an electron jumps from a higher energy level to lower energy level or excited from lower energy level to higher energy level, change in energy of an electron takes place in accordance with the quantum condition

$$E_{n_2} - E_{n_1} = h\nu$$

The energy level which is nearest to the nucleus, has minimum energy, where as the energy level which is farthest from the nucleus has maximum energy.

(iv) The set of four quantum numbers, n, l, m_l and m_s can be used to specify the state of electrons in an atom.

14. QUANTUM NUMBERS

(i) Principal Quantum Number (*n*)

This quantum number of any electron in an atom determines the main energy level or shell to which an electron belongs. This quantum number has integral positive value 1, 2, 3, 4, . . . ∞ but it is never zero. This has been used in Bohr model as well as Sommerfeld model of the atom. All the electrons that have the same value of n are at nearly the same distance from the nucleus and have the same energy states. These electrons occupy the same energy level/shell. In accordance with the value of n, the shells assigned a letter are given as under:

n	1	2	3	4	5	6
Shell designation letter	K	L	M	N	O	Q

With increase in distance of the shell from the nucleus, the energy in the energy level or shell increases. Obviously, different shells possess different energies.

(ii) Orbital or Azimuthal Quantum Number (l)

It specifies the number of units of angular momentum associated with an electron in a given orbit and determines the shape of the orbit and the energy of the sublevel. This quantum number is represented by a vector which is parallel to the axis of rotation of the electron and its direction is given by the direction of advance of the right hand screw (Fig. 2.13). This quantum number, l can have any integral value from 0 to $n - l$, e.g.

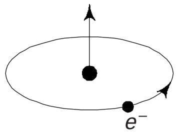


Fig. 2.13 Orbital quantum number

For $n = 1$ $l = 0$
$n = 2$ $l = 0, 1$
$n = 3$ $l = 0, 1, 2$
- - - - - - -
- - - - - - -

Clearly, the electron in the smallest orbit $n = 1$ will have no angular momentum, i.e., $l = 0$. We see that for $n = 3$ there are three possible orbital shapes or eccentries: *(i)* a straight line through the nucleus corresponding to $l = 0$, *(ii)* an ellipse with an angular momentum for $l = 1$ and *(iii)* a rounder ellipse with more angular momentum for $l = 2$.

We have seen that n is the principal quantum number and gives principal shell. l gives the possible orbital sub-shells. The sub-shells in the main shell are designated by *s*, *p*, *d*, *f*, *g* and *h* with quantum number $l = 0, 1, 2, 3, 4$ and 5 respectively. We can represent it as follows:

For $n = 1$, $l = 0$, the electron is said to be in 1*s* sub-shell
$n = 2$, $l = 1$, the electron is said to be in 2*p* sub-shell
$n = 2$, $l = 0$, the electron is said to be in 2*s* sub-shell
$n = 3$, $l = 1$, the electron is said to be in 3*p* sub-shell
$n = 3$, $l = 2$, the electron is said to be in 3*d* sub-shell

In designating an electron the letter designating it is written after the number that indicates the total quantum number n. For example a 3*p* electron is one for which $n = 3$ and $l = 1$. If an atom has more than one electron with particular values of n and l, this number is written as a superscript of the letter, e.g. if an atom has 6 electrons for which $n = 3$ and $l = 2$, the atom is said to have $3p^6$ electrons. We have stated above that for a given value of n the highest possible l value is $n - 1$. Hence we can have 4*d*, 5*f*, 2*p* and 2*s* electrons whereas 1*p*, 2*d*, 3*f* electrons do not exist.

(iii) Magnetic Quantum Number (m or m_l)

This quantum number determines the possible quantized space orientation of the electron's elliptical orbit without any effect on energy levels.

When an atom is subjected to an external strong magnetic field, B, each electronic orbit will be subjected to torque due to which vector $\vec{l}$ starts rotating around $\vec{B}$ (Fig. 2.14). The angle of rotation θ is called the angle of precession. Due to restrictions of orbits of electrons, θ has only *discrete* values. The permitted values of θ are governed by $l \cos \theta$, i.e., the projection of l along the direction of the magnetic field $\vec{B}$ (Fig. 2.15).

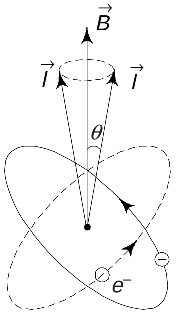


Fig. 2.14

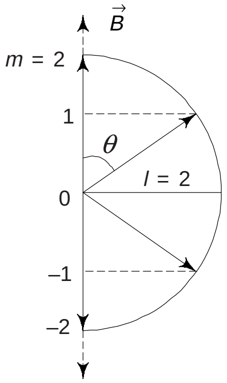


Fig. 2.15 Projection of orbital quantum number, l

We note that $l \cos \theta = m$ is also discrete (integer). m varies from $+l$ to $-l$, including zero. (Fig. 2.15). Obviously, there are $(2l + 1)$ values of m for a given l. Thus we have

For	$l = 0$	$m = 0$
	$l = 1$	$m = -1, 0, 1$
	$l = 2$	$m = -2, -1, 0 + 1, + 2$
	- - - -	- - - -

Concluding, we can say that in addition to n and l, the magnetic quantum number m or m_l must be introduced when the atom is under the influence of the external magnetic field.

(iv) The Magnetic Spin Quantum Number (m_s) When spectrum of hydroden or hydrogen like atoms, e.g. Na, K, etc. are examined with powerful instruments having high magnification, each spectral line is found to be a double line. These closely packed double lines are due to the fact that electron has a spin motion around its own axis. The electron can spin only in the clockwise and anticlockwise direction and obviously only two values of spin $+\frac{1}{2}$ and $-\frac{1}{2}$, depending upon the direction of spin, are possible. The spin of an electron is represented by s. The spin of the electron about its own axis as it revolves about the nucleus is analogous to the rotation of the earth along its orbit around the sun.

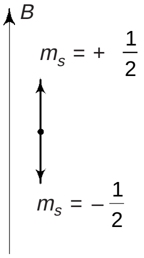


Fig. 2.16 Orientation of the spin angular momentum of the electron with respect to the magnetic field direction

The spin angular momentum has a numerical value $\frac{1}{2}(h/2\pi)$, i.e.

$$p_s = s(h/2\pi)$$

The projection of angular momentum along the direction of the magnetic field gives a new quantum number called the magnetic spin or electron spin quantum number, m_s. We must note that in the absence of a magnetic field there is no unique assignment to the direction of m_l or m_s. The spin angular momentum $\vec{p_s}$ of the electron can align either parallel or antiparallel to an applied magnetic field so that its components in the field direction are $mh/2\pi$, where $m_s = +\frac{1}{2}$ (parallel) and $m_s = -\frac{1}{2}$ (antiparallel) as shown in Fig. 2.16.

In the absence of any external field, the motion of an electron in an atom can be fully described in terms of the three quantum numbers n, l and s. In addition, the magnetic quantum number must be introduced when the atom is under the influence of the magnetic field. Obviously, one can clearly understand the state of an electron with the help of quantum numbers in terms of its position in the shell or the principal energy level (n), the subshell or the sub-level (l) or sub-subshell (m_l) and the direction of spin governed by m_s.

15. PAULI EXCLUSION PRINCIPLE AND ELECTRONIC CONFIGURATION OF ATOMS

Pauli in 1886 showed that no two electrons in an atom can exist in the same quantum state. This principle is known as Pauli's exclusion principle. The quantum state of an electron in an atom is defined by its quantum numbers. It two electrons in an atom cannot be in the same quantum state their quantum numbers must be different. Of the four quantum numbers at least one must be different for the two electrons. For example n, l, m_l may be the same for the two electrons in an atom but the fourth quantum number m_s must be different for the two electrons. If m_s is $+\frac{1}{2}$ for one electron then it must be $-\frac{1}{2}$ for the other.

Electrons which have the same value of the principal quantum number n are said to be in the same shell. Each shell is divided into subshells. The value of l is the same for all electrons in the same sub-shell but different for electrons in different sub-shells. For a given value n there are n permitted values of l. Hence each shell has n sub-shells. In each sub-shell no two electrons can have the same values of m_l and m_s. For each value of l there are $2l + 1$ permitted values of m_l. Obviously, in each sub-shell there can be $(2l + 1)$ electrons with m_s value $+\frac{1}{2}$ and $(2l + 1)$ electrons with m_s value of $-\frac{1}{2}$. Hence each sub-shell can accommodate $2(2l + 1)$ electrons. Hence the total number of electrons that can be accomodated in a shell with principal quantum number n is given by

$$\begin{aligned} 2\sum_{l=0}^{l=n-1} 2(2l+1) &= 2 \times 1 + 2 \times 3 + 2 \times 5 + 2 \times 7 + \ldots + 2 \times \{2(n-1)+1\} \\ &= 2 \times [\, 1 + 3 + 5 + \ldots \{2\,(n-1)+1\} \\ &= 2n^2 \qquad (35) \end{aligned}$$

Obviously, the maximum number of electrons that can be placed in a given atomic shell with a definite n is $2n^2$. These are listed in Table

Table 2.4 Electron capacity of main and subshells

Shell	*K*	*L*	*M*	*N*
n	1	2	3	4
Sub-level (or shell)	s	*s p*	*s p d*	*s p d f*
l	0	0 1	0 1 2	0 1 2 3
Maximum number of electrons in the sub-level	2	2 6	2 6 10	2 6 10 14
Maximum number of electrons in the shell	2	8	18	32

Thus Pauli exclusion principle together with the restrictions on the values of l, m_l and m_s leads to the following conclusions:

(i) The maximum number of electrons that can be contained in states characterized by n (principal quantum numbers i.e., 1, 2, 3, 4, . . ., n are equal to 2, 8, 18, 32, . . ., $2n^2$ respectively.

(ii) For a given value of n, the maximum number of electrons that can be contained in sub-shells s, p, d, f. . . . are 2, 6, 10, 14, . . . respectively, i.e., $2(2l + 1)$.

We can see that Pauli exclusion principle is very important. By placing a limit on the number of electrons in a given state, it leads to the concept of complete groups and sub-groups of electrons and thereby explains a periodic repetition.

We must note that the Pauli's exclusion principle also applicable to molecules and other assemblies of atoms with electrons occupy quantized energy states. However, in all such cases no states can contain more than one electron as defined by all the quantum numbers including the spin quantum number.

We have seen that each of the allowed orbits in the Bohr-Sommerfeld model of the hydrogen atom is designated a subshell. As the elements of the *periodic table* are built up the elements that are added fall into these subshells. The manner in which these elements are built up one after the other is called *Bohr-Stoner* scheme. These subshells are tabulated as shown below.

		Subshells				
	n/l	*0*	*1*	*2*	*3*	*4*
	1	1*s*				
shells	2	2*s*	2*p*			
	3	3*s*	3*p*	3*d*		
	4	4*s*	4*p*	4*d*	4*f*	
	5	5*s*	5*p*	5*d*	5*f*	5g

Now, we have a clear picture about various shells K, L, M, N, . . . according to their principal quantum numbers (*n*) 1, 2, 3, 4, . . ., their possible subshells *s*, *p*, *d*, *f*, . . . and their maximum possible number of electrons in various subshells. We must note that the capacity described in Table 2.4 apply only to the inner completed main shell. The outermost main shell never contains more than 8 electrons. Once this number is reached the next electron begins to fill the next main shell, occupying its subshell. Only after both vacancies in *s*-subshell are filled does the inner incomplete main shell go on filling up. This process of filling of shells produces, the commonly known as transitional elements. The order in which the electrons fill the subshell is illustrated in Fig. 2.17. Scientists have evolved a principle to denote systematically the distribution of electrons in shells and sub-shells of an atom,. This principle is known as electronic configuration. One can best understood this principle from the following few examples:

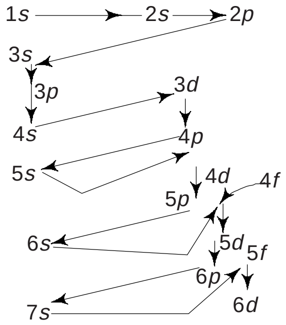


Fig. 2.17 Order of filling sub-shells by electrons

(i) Hydrogen (H) *with atomic number,* Z = 1. The hydrogen atom has only one electron and it is accommodated in the first main shell (1*s*). The electronic configuration of this is represented by $1s^1$. The prefix '1' represents the first main shell, '*s*' represents the *s* sub-shell and the script '1' represents the number of electrons in the sub-shell.

(ii) Helium (He: Z = 2): Helium atom has two electrons and these are accommodated in the first main shell. Obviously, the electronic configuration of the Helium atom is designated as $1s^2$.

(iii) Beryllium (Be: Z = 4): The Beryllium atom has 4 electrons, out of which two are accommodated in the first main shell and remaining two are accommodated in the '*s*' subshell of the second main shell. We can write the electronic configuration of Beryllium atom as $1s^2\ 2s^2$.

(iv) Nitrogen (N: Z = 7): Nitrogen has 7 electrons, out of which 2 electrons are accommodated in the first main shell, the other 2 are accommodated in the '*s*' subshell of second main shell and remaining 3 electrons are accommodated in the '*p*' subshell of the second main shell. We can write the electronic configuration of the Nitrogen as $1s^2\ 2s^2\ 2p^3$.

(v) Sodium (Na: Z = 11): The electronic configuration of Na can be written as $1s^2\ 2s^2\ 2p^6\ 3s^1$.

(vi) Copper (Cu: Z = 29): The Cu atom has 29 electrons. The first 10 electrons are accommodated in the first and second main shells. Out of remaining 19, 18 electrons are accommodated in the 3rd main shell and remaining one electron is accommodated in the '*s*' subshell of the fourth main shell. Thus we can write the electronic configuration of Cu (Z = 29) atom as $1s^2\ 2s^2\ 2p^6\ 3p^6\ 3d^{10}\ 4s^1$.

(vii) Silver (Z = 47): $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4d^{10}\ 5s^1$.

(viii) Uranium (Z = 92): $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6\ 4d^{10}\ 4f^{14}\ 5s^2\ 5p^6\ 5d^{10}\ 5f^3\ 6s^2\ 6p^2\ 6d^1\ 7s^2$.

The electronic configuration of all the known elements is summarized in Table 2.5.

Table 2.5 Electronic configuration of atoms

Atomic number	*Element*	*Symbol*	*K* *1*	*L* *2*		*M* *3*			*N* *4*				*O* *5*				*P* *6*				*Q* *7*			
			s	*s*	*p*	*s*	*p*	*d*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*
1	Hydrogen	H	1																					
2	Helium	He	2																					
3	Lithium	Li	2	1																				
4	Beryllium	Be	2	2																				
5	Boron	B	2	2	1																			
6	Carbon	C	2	2	2																			
7	Nitrogen	N	2	2	3																			
8	Oxygen	O	2	2	4																			
9	Fluorine	F	2	2	5																			
10	Neon	Ne	2	2	6																			
11	Sodium	Na	2	2	6	1																		
12	Magnesium	Mg	2	2	6	2																		
13	Aluminium	Al	2	2	6	2	1																	
14	Silicon	Si	2	2	6	2	2																	
15	Phosphorus	P	2	2	6	2	3																	
16	Sulphur	S	2	2	6	2	4																	
17	Chlorine	Cl	2	2	6	2	5																	
18	Argon	Ar	2	2	6	2	6																	
19	Potassium	K	2	2	6	2	6		1															
20	Calcium	Ca	2	2	6	2	6		2															
21	Scandium	Sc	2	2	6	2	6	1	2															
22	Titanium	Ti	2	2	6	2	6	2	2															
23	Vanadium	V	2	2	6	2	6	3	2															
24	*Chromium	Cr	2	2	6	2	6	5	1															
25	Manganese	Mn	2	2	6	2	6	5	2															
26	Iron	Fe	2	2	6	2	6	6	2															
27	Cobalt	Co	2	2	6	2	6	7	2															
28	Nickel	Ni	2	2	6	2	6	8	2															
29	*Copper	Cu	2	2	6	2	6	10	1															
30	Zinc	Zn	2	2	6	2	6	10	2															
31	Gallium	Ga	2	2	6	2	6	10	2	1														

(*Contd.*)

Table 2.5 (*Contd.*)

Atomic number	*Element*	*Symbol*	*K* *1*	*L* *2*		*M* *3*			*N* *4*				*O* *5*				*P* *6*				*Q* *7*			
			s	*s*	*p*	*s*	*p*	*d*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*
32	Germanium	Ge	2	2	6	2	6	10	2	2														
33	Arsenic	As	2	2	6	2	6	10	2	3														
34	Selanium	Se	2	2	6	2	6	10	2	4														
35	Bromine	Br	2	2	6	2	6	10	2	5														
36	Krypton	Kr	2	2	6	2	6	10	2	6														
37	Rubidium	Rb	2	2	6	2	6	10	2	6			1											
38	Strontium	Sr	2	2	6	2	6	10	2	6			2											
39	Yttrium	Y	2	2	6	2	6	10	2	6	1		2											
40	Zirconium	Zr	2	2	6	2	6	10	2	6	2		2											
41	*Niobium	Nb	2	2	6	2	6	10	2	6	4		1											
42	*Molybdenum	Mo	2	2	6	2	6	10	2	6	5		1											
43	Technetium	Tc	2	2	6	2	6	10	2	6	5		2											
44	*Ruthenium	Ru	2	2	6	2	6	10	2	6	7		1											
45	*Rhodium	Rh	2	2	6	2	6	10	2	6	8		1											
46	*Palladium	Pd	2	2	6	2	6	10	2	6	10													
47	*Silver	Ag	2	2	6	2	6	10	2	6	10		1											
48	Cadmium	Cd	2	2	6	2	6	10	2	6	10		2											
49	Indium	In	2	2	6	2	6	10	2	6	10		2	1										
50	Tin	Sn	2	2	6	2	6	10	2	6	10		2	2										
51	Antimony	Sb	2	2	6	2	6	10	2	6	10		2	3										
52	Tellurium	Te	2	2	6	2	6	10	2	6	10		2	4										
53	Iodine	I	2	2	6	2	6	10	2	6	10		2	5										
54	Xenon	Xe	2	2	6	2	6	10	2	6	10		2	6										
55	Caesium	Cs	2	2	6	2	6	10	2	6	10		2	6			1							
56	Barium	Ba	2	2	6	2	6	10	2	6	10		2	6			2							
57	Lanthanum	La	2	2	6	2	6	10	2	6	10		2	6	1		2							
58	Cerium	Ce	2	2	6	2	6	10	2	6	10	1	2	6	1		2							
59	Praseodymium	Pr	2	2	6	2	6	10	2	6	10	3	2	6			2							
60	Neodymium	Nd	2	2	6	2	6	10	2	6	10	4	2	6			2							
61	Promethium	Pm	2	2	6	2	6	10	2	6	10	5	2	6			2							
62	Samarium	Sm	2	2	6	2	6	10	2	6	10	6	2	6			2							
63	Europium	Eu	2	2	6	2	6	10	2	6	10	7	2	6			2							

(*Contd.*)

Table 2.5 (*Contd.*)

Atomic number	Element	Symbol	*K* 1	*L* 2		*M* 3			*N* 4				*O* 5				*P* 6				*Q* 7			
			s	*s*	*p*	*s*	*p*	*d*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*
64	Gadolinium	Gd	2	2	6	2	6	10	2	6	10	7	2	6	1		2							
65	Terbium	Tb	2	2	6	2	6	10	2	6	10	9	2	6			2							
66	Dysprosium	Dy	2	2	6	2	6	10	2	6	10	10	2	6			2							
67	Holmium	Ho	2	2	6	2	6	10	2	6	10	11	2	6			2							
68	Erbium	Er	2	2	6	2	6	10	2	6	10	12	2	6			2							
69	Thullium	Tm	2	2	6	2	6	10	2	6	10	13	2	6			2							
70	Ytterbion	Yb	2	2	6	2	6	10	2	6	10	14	2	6			2							
71	Lutetium	Lu	2	2	6	2	6	10	2	6	10	14	2	6	1		2							
72	Hafnium	Hf	2	2	6	2	6	10	2	6	10	14	2	6	2		2							
73	Tantalum	Ta	2	2	6	2	6	10	2	6	10	14	2	6	3		2							
74	Tungsten	W	2	2	6	2	6	10	2	6	10	14	2	6	4		2							
75	Rhenium	Re	2	2	6	2	6	10	2	6	10	14	2	6	5		2							
76	Osmium	Os	2	2	6	2	6	10	2	6	10	14	2	6	6		2							
77	Iridium	Ir	2	2	6	2	6	10	2	6	10	14	2	6	7		2							
78	Platinum	Pt	2	2	6	2	6	10	2	6	10	14	2	6	9		1							
79	*Gold	Au	2	2	6	2	6	10	2	6	10	14	2	6	10		1							
80	Mercury	Hg	2	2	6	2	6	10	2	6	10	14	2	6	10		2							
81	Thallium	Ti	2	2	6	2	6	10	2	6	10	14	2	6	10		2	1						
82	Lead	Pb	2	2	6	2	6	10	2	6	10	14	2	6	10		2	2						
83	Bismuth	Bi	2	2	6	2	6	10	2	6	10	14	2	6	10		2	3						
84	Polonium	Po	2	2	6	2	6	10	2	6	10	14	2	6	10		2	4						
85	Astatine	At	2	2	6	2	6	10	2	6	10	14	2	6	10		2	5						
86	Radon	Rn	2	2	6	2	6	10	2	6	10	14	2	6	10		2	6						
87	Francium	Fr	2	2	6	2	6	10	2	6	10	14	2	6	10		2	6			1			
88	Radium	Ra	2	2	6	2	6	10	2	6	10	14	2	6	10		2	6			2			
89	Actinum	Ac	2	2	6	2	6	10	2	6	10	14	2	6	10		2	6	1		2			
90	Thorium	Th	2	2	6	2	6	10	2	6	10	14	2	6	10		2	6	2		2			
91	Protactinum	Pa	2	2	6	2	6	10	2	6	10	14	2	6	10	2	2	6	1		2			
92	Uranium	U	2	2	6	2	6	10	2	6	10	14	2	6	10	3	2	6	1		2			
93	Neptunium	Np	2	2	6	2	6	10	2	6	10	14	2	6	10	4	2	6	1		2			
94	Plutonium	Pu	2	2	6	2	6	10	2	6	10	14	2	6	10	6	2	6			2			
95	Amerilium	Am	2	2	6	2	6	10	2	6	10	14	2	6	10	7	2	6			2			

(*Contd.*)

Table 2.5 *(Contd.)*

Atomic number	*Element*	*Symbol*	*K* 1	*L* 2		*M* 3			*N* 4				*O* 5				*P* 6				*Q* 7			
			s	*s*	*p*	*s*	*p*	*d*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*	*s*	*p*	*d*	*f*
96	Curium	Cm	2	2	6	2	6	10	2	6	10	14	2	6	10	7	2	6	1		2			
97	Berkelium	Bk	2	2	6	2	6	10	2	6	10	14	2	6	10	9	2	6			2			
98	Californium	Cf	2	2	6	2	6	10	2	6	10	14	2	6	10	10	2	6			2			
99	Einsteinium	Es	2	2	6	2	6	10	2	6	10	14	2	6	10	11	2	6			2			
100	Fermium	Fm	2	2	6	2	6	10	2	6	10	14	2	6	10	12	2	6			2			
101	Mendelevium	Mv	2	2	6	2	6	10	2	6	10	14	2	6	10	13	2	6			2			
102	Nobelium	No	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6			2			
103	Lawrencium	Lw	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6	1		2			
104	Unniquadium	Unq	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6	2		2			
105	Unnilpentium	Unp	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6	3		2			
106	Unnihexium	Unh	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6	4		2			
107	Unniseptium	Uns	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6	5		2			
108	Unniloctium	Uno	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6	6		2			
109	Unnilenium	Une	2	2	6	2	6	10	2	6	10	14	2	6	10	14	2	6	7		2			

*We must note that the electronic configurations of the elements such as Chromium, Copper, Molybdenum, Silver, Gold, etc., differs from the normal sequence of filling up the electrons in the energy levels of their atoms, it is due to the inter electronic repulsions. As a result of this internal electronic repulsions, the electrons differ slightly in their energies have a preference to enter into those sub-shells which either get completely filled or just half filled.

Example 10 *Write the four quantum numbers for each of the electrons in the outermost shell of a boron atom.*

Solution For boron, Z = 5. Obviously, it has 5 electrons in it. Out of these 2 are in K shell and remaining 3 in the L-shell. Of the 3 electrons in the L shell 2 are S electrons and 1 is a *p* electron. Hence the quantum numbers of the electrons in the *L*-shell are as follows:

n	l	m_l	m_s
2	0	0	$+\frac{1}{2}$
2	0	0	$-\frac{1}{2}$
2	1	0	$\pm\frac{1}{2}$

Example 11 *How are electrons distributed in the various sub-shells for n = 3?*

Solution For $n = 3$, l can have +2, +1, 0. When $l = 2$, m_l can be –2, –1, 0, +1, +2. For each of l, m_s can be $\pm\frac{1}{2}$. When $l = 1$, m_l can be –1, 0 or +1. Hence the electron distribution of $n = 3$ can be expressed as

n	l	m_l	m_s
3	2	2	$\frac{1}{2}$
3	2	2	$-\frac{1}{2}$
3	2	1	$\frac{1}{2}$
3	2	1	$-\frac{1}{2}$
3	2	0	$+\frac{1}{2}$
3	2	0	$-\frac{1}{2}$
3	2	–1	$\frac{1}{2}$
3	2	–1	$-\frac{1}{2}$
3	2	–2	$\frac{1}{2}$
3	2	–2	$-\frac{1}{2}$
3	1	1	$\frac{1}{2}$
3	1	1	$-\frac{1}{2}$
3	1	0	$\frac{1}{2}$
3	1	0	$-\frac{1}{2}$
3	1	–1	$\frac{1}{2}$

3	1	–1	$-\frac{1}{2}$
3	0	0	$\frac{1}{2}$
3	0	0	$-\frac{1}{2}$

Example 12 *Write the electronic configuration of Sn (Z = 50).*

Solution Sn (Z = 50): $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6\ 4d^{10}\ 5s^2\ 5p^2$

Example *13* Which elements have the following levels filled in the ground state?

(a) K and L shells, the 3*s* subshell and one half of 3*p* subshell.
(b) K, L and M shells, the 4*s*, 4*p*, 4*d* and 5*s* subshells. [BE]

Solution

(a) From Table 2.1, we can have the following table

Shell	*Principal Quantum No.*	*Subshell*	*No. of electrons*	
K	1	*s*	2	Total 15
L	2	*s*	2	
		p	6	
M	3	*s*	2	
		p	3	

For filling of subshell we must follow the rule cited earlier. In this case, however, the subshells 1*s*, 2*s*, 2*p*, 3*s* and 3*p* fill in sequence.

(b) The subshells will fill in the following order 1*s*, 2*s*, 2*p*, 3*s*, 3*p*, 4*s*, 4*d*, 5*s*, 4*d*
We can prepare the table as follows:

Shell	*Principal Quantum No.*	*Subshell*	*No. of electrons*	
K	1	*s*	2	Total 48
L	2	*s*	2	
		p	6	
M	3	*s*	2	
		p	6	
		d	10	
N	4	*s*	2	
		p	6	
		d	10	
O	5	*s*	2	

16. WAVE MECHANICAL PICTURE OF THE ATOM

In 1924 de Broglie suggested that particles in motion should exhibit properties characteristic of waves. He further suggested that certain basic formulae should apply both to waves and particles. The wavelength of such particles, e.g., electron, proton, neutron, etc. is given by the relation

$$\lambda = \frac{h}{mv} \tag{36}$$

where h is Planck's constant, m is mass of the particle and v is the velocity of the particle. de Broglie called these waves as matter waves. Relation (36) provides the mathematical relationship between the momentum ($p = mv$) of a particle which is a dynamical variable characteristic of a corpuscle and the wavelength which is characteristic of the associated wave.

de Broglie's hypothesis was verified experimentally by Davisson and Germer in 1926 and G.P. Thomson in 1928.

Since the mechanics of the particle is described in terms of wave and hence this gave birth to a new mechanics, now known as *wave mechanics*. In atomic phenomena, this new mechanics replaces the orbital theory in order to explain many details of atomic and molecular spectra, electron diffraction and reflection, nuclear properties, etc.

Heisenberg remarked that the wave representation of the particle implies some uncertainty Δx of the position x of the particle and a corresponding uncertainty Δp in specifying its momentum p simultaneously. He proposed the following relationship between uncertainties Δx and Δp:

$$\Delta x\ \Delta p \geq h/2\pi \tag{37}$$

This is known as Heisenberg's uncertainty relation. This principle asserts that it is impossible to determine precisely simultaneously both the position and momentum of a body in motion. This relation can be used to decide about the applicability of classical and quantum concepts. The quantum considerations are necessary when the uncertainty in the momentum of a particle is comparable to or greater than the total momentum. In accordance with the de Broglie theory, we find that only those orbits are permitted which contain an integral number of full de Broglie waves. We have from Bohr theory

$$mvr = nh/2\pi$$

From de Broglie relation

$$2\pi r = nh/mv = n\lambda$$

This provides an 'explanation of the quantum condition on the basis of de Broglie's theory. (Fig. 2.18).

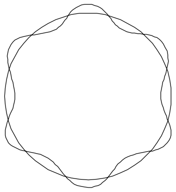


Fig. 2.18 **de Broglie wave in atomic orbit**

The origin of quantized motion cannot be explained by classical mechanics. The old quantum theory is also inadequate to give a logical explanation for the quantization of certain dynamical quantities in the atomic or subatomic domain. Schrödinger in 1926 showed that the following time-dependent wave equation

$$\frac{h^2}{8\pi^2 m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)-V\psi=\frac{h}{2\pi i}\frac{\partial\psi}{\partial t} \tag{37}$$

can be used to solve problems on atomic scale through the wave aspects of particles. Here $\psi(x, y, z, t)$ is the wave function or state function and represents the probability amplitude with its coordinates in space and time: x, y, z and t and V is the potential energy of the particle. When $V = 0$, particle is said to be a *free particle*. Equation (37) is also written as

$$\frac{h^2}{8\pi^2 m}\nabla^2\psi - \mathrm{V}\psi = \frac{h}{2\pi i}\frac{\partial\psi}{\partial t}$$

where $$\nabla^2 = \frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}$$

When the potential energy of a particle does not explicitly depend on time, the steady state form or time independent form of Schrödinger equation for a particle is

$$\nabla^2\psi + \frac{8\pi^2 m}{h^2}(E - V)\ \psi = 0 \tag{38}$$

where E is the total energy of the particle. The Schrödinger wave equation is quite useful in studying the motion of electron when subjected to forces which hold it in a certain region of space, as in atom or a metal. We shall discuss in detail in the chapter on *Free electron theory*.

17. PERIODIC TABLE

Scientists have discovered 109 elements. Mandeleev in 1869, devised a way of tabulating all the known elements on the basis of the periodicity of their chemical properties. This table is known as Mandeleev's periodic table.

Mandeleev's Periodic Table

Mandeleev discovered that if the elements are arranged in the order of their increasing atomic weights, the elements with their similar properties occur at regular intervals. Based on this observation, Mandeleev gave a law which states that the properties of the elements are periodic functions of their atomic weights.

Mandeleev arranged all the known elements at that time in the order of their increasing atomic weights in horizontal rows in such a way that the elements having similar properties fall directly under one another in the same vertical columns or groups. In Mandeleev's periodic table, there were seven horizontal rows or periods and eight vertical columns or groups. To make it sure that the elements with similar properties fall in the same vertical column or group, he left some gaps in his periodic table. These gaps were left probably for the elements not discovered at that time.

Though the Mandeleev's periodic table of elements was of great significance in the study of elements, yet it failed to explain the position of isotopes. No separate place for isotopes were provided in the periodic table. It was also observed that in certain cases, elements with higher atomic mass appear before the elements having lower atomic mass. On the basis of these drawbacks, later workers concluded that atomic weight cannot be taken as the basis for the classification of elements.

Modern Periodic Table

Moseley in 1913 found that the properties of elements are periodic functions of their atomic numbers and not atomic mass. We know that atomic number of an element is fixed and no two elements have same atomic number. Obviously, the atomic number is more a fundamental property than the atomic weight for the classification of elements. Moseley proposed a periodic table in which the elements were arranged in the increasing order of atomic number. This table has seven horizontal rows known as *periods* and 18 vertical columns known as *groups*. These groups are further divided into groups IA to VIIA, IB to VIIB, group VIII divided into three parts and group O. Later this periodic table was modified by Bohr, which is now known as '*Modern Periodic Table'*. In this periodic table, the elements are also arranged in order of their increasing atomic numbers (Fig. 2.19). This 'perodic table is divided into following four blocks:

(i) s-block: The elements whose valence electrons lie in the *s*-sub shell, are known as *s*-block elements. This block of elements comprises of groups IA and IIA elements and is shown in the left side of the periodic table. This block comprises of two groups IA and IIA elements. Group IA elements have one valence electron and atomic numbers 1, 3, 11, 19, 37, 55 and 87. These elements have similar electronic configuration. The elements in this group are known as *Alkali* metals. Similarly group IIA have two valence electrons and consists of elements of atomic numbers 4, 12, 20, 38, 56 and 88. These elements have similar electronic configuration.

(ii) p-block: The elements whose valence electrons lie in the *p*-subs-shell are known as elements of *p*-block. The *p*-block elements are shown in the right hand side of the periodic table and this block consists of 6 groups namely IIIA, IVA, VA, VIIA and zero. The group number indicates the number of valence electrons. The outermost shell of zero group is completely filled and hence this has no valence electron in its outermost shell. The elements in zero group are known as *inert gases* or *noble gases*.

(iii) d-block: The elements, whose valence electrons lie in the *d*-sub-shell are known as *d*-block elements and these are shown in the centre of the periodic table. Group IB to VII B and group VIII, i.e. these 8 groups come under this block. These elements are known as *transition elements*.

Periods	s-Block		d-Block										p-Block					
	Group IA	Group IIA	Group IIIB	Group IVB	Group VB	Group VIB	Group VIIB	Group VIIIB			Group IB	Group IIB	Group IIIA	Group IVA	Group VA	Group VIA	Group VIIA	Group 0
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57 La	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Ti	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	89 Ac	104 Unq	105 Unp	106 Unh	107 Uns	108 Uno	109 Une									

f-Block														
Lanthanides	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
Actinides	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

Fig. 2.19 Modern Periodic Table (After Bohr)

(iv) f-block: The elements, whose valence electrons lie in the *f*-sub-shell are known as *f*-block elements. These are shown at the bottom of the periodic table.

The modern periodic table is also divided into seven horizontal rows or periods. These are:

(i) First Period: This contains two elements having atomic numbers 1 and 2. The element with Z = 1 lies in group IA of *s*-block and that with Z = 2 in group zero of *p*-block. This period is also known as *short period*.

(ii) Second Period: This contains eight elements with atomic number 3 to 10. The elements with Z = 3, 4 lie in group IA and IIA of s-block and those with atomic number 5 to 10 in groups IIIA to zero in *p*-block. This is also known as *short period*.

(iii) Third Period: This contains eight elements with Z = 11 to 18. The elements with Z = 11, 12 lie in groups IA and IIA of *s*-block and those with Z = 13 to 18 in groups IIIA to zero in *p*-block. It is also called as a *short period*.

(iv) Fourth Period: This contains 18 elements with Z = 19 to 36. The elements with Z = 19 and 20 lie in *s*-block, elements with Z = 21 to 30 in *d*-block and Z = 31 to 36 in *p*-block. This is known as a *long period*.

(v) Fifth Period: This contains 18 elements with Z = 37 to 54. The elements with Z = 37 and 38 lie in *s*-block, with Z = 39 to 48 in *d*-block and Z = 49 to 54 in *p*-block. This is also known as a *long-period*.

(vi) Sixth Period: This contains 32 elements with Z = 55 to 86. The elements with Z = 55, 56 lie in the *s*-block, Z = 57 and 72 to 80 in *d*-block, Z = 58 to 71 in *f*-block and Z = 81 to 86 in *p*-block. This is known as *very long period.*

(vii) Seventh Period: This contains the remaining elements with Z = 87 to 105. The elements with Z = 87, 88 lie in *s*-block, Z = 89, 104 and 105 lie in *d*-block and Z = 90 to 103 in *f*-block. This period of elements is still *incomplete*.

The elements with Z = 58 to 71 in *f*-block are called *lanthanides,* all the elements in *d*-block are called *transition elements* and those with Z = 90 to 103 are actinides. The elements belonging to lanthanides and actinides have similar properties and hence they have been placed at the bottom of the periodic table. The actinides do not occur in nature and are prepared artificially.

Elements on the left hand side of the diagonal dividing band are known as '*metals*' whereas those on the right are called '*non-metals*' and those within the band are called '*semi-metals*' or '*metalloides*'.

The periodic table is very useful in solving the problem of scarcity, e.g. during World War II, scientists and engineers solved the problem of scarcity 'tungsten' by engineering application of the Periodic Table. 18% tungsten tool steel was substituted by 8% molybdenum and 2% tungsten, which is even used today. This was possible, because molybdenum is placed directly above the tungsten in the periodic table of elements.

Example 13 *Assuming that the weight of an electron is negligible compared to the weight of proton and neutron, calculate the weight of copper atom. Assuming that weight of one proton is equal to that of one neutron, find the weight of one proton.* *[AMIE]*

Given, *Atomic weight of copper = 63.54,*
Avogadro's numbers (N) = 6.023 $\times 10^{23}$ *atoms/gm-mole.*

Solution N= 6.023×10^{23} atoms/gm-mole

$$\therefore \quad \text{Weight of 1 atom} = \frac{63.54}{6.023 \times 10^{23}} = 1.054 \times 10^{-22} \text{ gm}$$

Now, Atomic Weight = No. of protons + No. of neutrons
We have (no. of protons + no. of neutrons)/atom = 63

$$\therefore \quad \text{Weight of 1 proton} = \frac{1.054 \times 10^{-22}}{63} = 1.673 \times 10^{-23} \text{ gm}$$

Example 14 *Using conventional s, p, d, f, . . . notations, write the electronic configuration of Fe atom (Z = 26) and* Fe^{2+} *and* Fe^{3+} *ions.* *[BE, AMIE]*

Solution We have the capacity of subshells as

$s - 2, \quad p - 6, \quad d - 10 \quad \text{and} \quad f - 14$

Electronic configuration of Fe atom is

$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^6\ 4s^2$

Fe^{2+} *ions* means the Fe atom has lost 2 electrons of the outermost orbit which is 4*s*. Thus the electronic configuration of Fe^{2+} is

$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^6$

Fe^{3+} *ions* means the Fe atom has lost 3 electrons, i.e. one more than Fe^{2+} ion. Obviously the loss of one electron will occur from 3*d* electrons. We can write the electronic configuration as

$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^5$

Example 15 *Atomic weight of Cu and Si are 63.54 and 28.09 respectively. Find the percentage of Si in* Cu_5Si. *[AMIE, BE]*

Solution The percentage of Si in Cu_5 Si is = $\dfrac{28.09}{63.54 + 28.09} \times 100 = 8.12\%$

SUGGESTED READINGS

1. L. Solymar and D. Walsh, '*Lectures on the Electrical Properties of Materials*', Oxford University Press, Oxford (1984)
2. H.H. Sisler, *Electronic structure, Properties and the Periodic Law*, Reinhold Pub. Cooperation, New York (1963)
3. W.L. Masterton and C.N., Hurley, *Chemistry, Principles and Reactions*, 2nd Ed. Harcourt, Fort Worth, TX, (2000)
4. S.L. Kakani, *Modern Physics*, Viva Books Publishers, New Delhi (In Press)

REVIEW QUESTIONS

1. What is an atom? Describe briefly the important constituents of an atom.
2. Explain the simplified concept of an atom and the significance of the same in relation to properties of bulk materials. [AMIE, B.E.]
3. State and explain Bohr's model of an atom. [AMIE]
4. In Bohr's theory of hydrogen atom the principal quantum number *n* cannot take the value zero. Explain.
5. State the basic postulates of Bohr's atom model.
6. Deduce an expression for the radius of the electron orbit in the hydrogen atom.
7. Deduce an expression for the binding energy of an electron in a hydrogen like atom according to Bohr's theory.
8. How does Bohr's model account for the different spectral series of the hydrogen atom?
9. Obtain an expression for the energy of the electron in the *n*th orbit in hydrogen atom.
10. According to Bohr's theory, the potential energy of electron in a hydrogen atom is negative and larger in magnitude than the kinetic energy. Explain the significance of this. [BE]
11. Give a brief account of Sommerfeld's modification of Bohr's theory. What was the need for this modification.

12. Explain atomic mass and mass number. What is an isotope? [AMIE]

13. What is Pauli's exclusion principle and how it is applied to atom's electronic structure? [AMIE]

14. What is modern concept about the structure of an atom? Describe briefly.

15. What do you understand by the electronic configuration? [AMIE]

16. Write electronic configuration of iron. [AMIE]

17. Explain the difference between atomic weight and atomic number and their importance in the periodic table. [AMIE]

18. Explain how atomic shells and sub-shells are formed. [AMIE]

19. Justify the Mosley's law on the basis of Bohr's theory. [AMIE]

20. Explain, how the experiment on α-particle scattering led to the concept of the nuclear model of the atom. [AMIE, BE]

21. Explain how the modern periodic table is different from Mandeleev's periodic table. [AMIE]

22. What is the basis on which periodic table is divided into four blocks?

23. Differentiate between periods and groups in a periodic table.

24. What is deuterium. Write its electronic configuration.

25. Why elements with zero valence electrons are called '*inert gases*'? [AMIE]

26. Show that potential energy = –2 kinetic energy for an electron in a circular orbit in Coulomb field. [AMIE]

27. Define the following

(i) Atomic structure
(ii) Atomic number
(iii) Molecule
(iv) Nucleus
(v) Proton
(vi) Neutron
(vii) Electron [AMIE]

28. State the fundamental postulates of Bohr's theory and explain hydrogen spectrum. [AMIE]

29. Write short notes on:

(i) Ruther-ford's nuclear atom model
(ii) The spectrum of hydrogen [AMIE]
(iii) Bohr's postulates [AMIE]
(iv) Sommerfeld-Wilson atomic model
(v) Pauli's exclusion principle
(vi) Vector Atom model
(vii) Non-acceptance of Thomson's model of an atom. [AMIE]

30. Explain the following

(a) Isotopes
(b) Isobars
(c) nucleon
(d) atomic mass
(e) atomic number
(f) alpha particle
(g) meson
(h) nucleus.

PROBLEMS

1. Calculate the radius of the first orbit of the electron in the hydrogen atom. Given, $h = 6.626 \times 10^{-34}$ J.s, $m = 9.11 \times 10^{-31}$ kg and $e = 1.6 \times 10^{-19}$ C. [*Ans.* 5.315×10^{-11}m]

2. Calculate the velocity of the electron in the innermost orbit of the hydrogen atom. [*Ans.* 2.2×10^{3} m/s]

3. The wavelength of the first line of the Balmer series of hydrogen is 656.3 nm. Calculate the Rydberg constant. [*Ans.* 1.097×10^{7} m^{-1}]

4. The radius of the first orbit in a hydrogen atom is 0.053 nm. Calculate the velocity of the electron in this orbit. [*Ans.* 2.186×10^{6} m/s]

5. The Rydberg constant for hydrogen is 10967700 m^{-1}. Find the short and long wavelength limits of the Lyman series. [*Ans.* 91.16 nm; 121.5 nm]

SHORT QUESTIONS

1. How does the Thomson atom model differ from a random distribution of protons and neutrons in a spherical region?
2. Write objections to the Thomson model of the atom.
3. The scattering of α-particles at very small angles disagrees with the Rutherford's formula for such angles. Explain.
4. Did Bohr postulate the quantization of energy? What did he postulate?
5. Does the Rydberg constant vary with the nucleus? Explain.
6. Can a hydrogen atom absorb a photon whose energy is more than the binding energy?
7. Explain the need for introducing the concept of electron spin.

OBJECTIVE QUESTIONS

1. The charge carriers in the discharge tube at very low pressures are ________ and ________ (electrons, positive ions)
2. Rydberg's constant varies with the ________ of a given element. (mass number)
3. Photons do not have a finite ________ (rest mass)
4. The different lines in the Lyman series have their wavelength lying between ________ Å and ________ Å. (911, 1215)
5. If elements with principal quantum number $n > 4$ were not allowed in nature, then the number of possible elements would be ________ (60)
[**Hint:** The maximum number of electrons allowed in an orbit being $2n^2$. The required number is $2 \times (1^2 + 2^2 + 3^2 + 4^2) = 60$].
6. If the angular momentum of the earth due to its motion around the sun were quantized according to Bohr's relation $L = nh/2\pi$, then the quantum number corresponding to this quantization would be ________ (2.5×10^{74})
7. State whether the following statements are *True* or *False*:
(a) The mass of an electron is greater than the mass of a proton. (False)
(b) Atomic number of an element is equal to the number of protons present in the nucleus. (True)
(c) The mass number of an element is the sum of the number of protons and neutrons present in the nucleus of the element. (True)
(d) All atoms having different atomic weights but belonging to the same element are called isobars. (False)
(e) The nucleus is normally composed of protons and neutrons. (True)
(f) The orbital quantum number (l) is also known as the azimuthal quantum number. (True)
(g) Transition elements have higher melting points and higher densities as compared to light metals. (True)
(h) The elements which occupy the seventh row of the periodic table but do not occur in nature and are prepared artificially are called actinides. (True)
8. Fill in the blanks
(a) 1 eV = __________J. (1.6×10^{-19})
(b) Rydberg constant (R) of hydrogen formula = __________ m^{-1} (1.097×10^{7})
(c) The mass of an electron is __________ kg. (9.1×10^{-31})

(d) The charge on the proton is + __________ C. (1.6×10^{-19})
(e) Chlorine atom has __________ protons and __________ neutrons. (17, 18)
(f) Potassium has __________ protons and __________ electrons. (19, 19)

9. A photon is a quantum of energy equal to
(1) $h\nu$ (2) $h\lambda$
(3) $1/h\nu$ (4) ch (1)

10. Which of the following transitions in a hydrogen atom emits the photon of lowest frequency.
(1) $n = 2$ to $n = 1$ (2) $n = 4$ to $n = 2$
(3) $n = 4$ to $n = 3$ (4) $n = 3$ to $n = 1$ (3)

11. The different lines in the Lyman series have their wavelengths lying between
(1) 0 to ∞ (1) 911 Å to 1215 Å
(3) 1000 Å to 1500 Å (4) 500 Å to 1000 Å (2)

12. The shortest wavelength in the Lyman series is 911.6 Å. Then the longest wavelength in the Lyman series is
(1) 2450 Å (2) 1215 Å
(3) 600 Å (4) ∞ (2)

13. If elements with principal quantum number $n = 3$ were not allowed, the number of possible element would be
(1) 28 (2) 90
(3) 32 (4) 64 (1)

14. The atomic number of an atom is equal to
(1) atomic weight (2) atomic mass
(3) number of protons (4) mass number (3)

15. The force which holds neutrons and protons together in a nucleus is
(1) electrostatic force (2) gravitational force
(3) magnetic force (4) nuclear force (4)

16. The energy of an electron in an orbit of radius r of an atom and atomic number z is proportional to
(1) z^2 (2) z
(3) zr (4) zr^2 (1)

17. The angular momentum of an electron, revolving round the nucleus, in the nth orbit is
(1) $m\omega$ (2) $m\omega r$
(3) $m\omega r^2$ (4) $nh\ 2\pi$ (4)

18. The maximum number of electrons in a shell of principal quantum number n is
(1) n^2 (2) $2n^2$
(3) $8n^2$ (4) $16n^2$ (2)

19. For quantum number $l = 2$, the quantum number m_l can have the values
(1) 2 (2) 5
(3) 7 (4) 9 (2)

20. If there are six electrons in the d-orbital of a transition metal, the number of unpaired electrons are
(1) 2 (2) 4
(3) 6 (4) none (2)

21. The order of increasing energy of orbitals in the sixth row of periodic table is
(1) $6s\ 6p\ 4f\ 5d$ (2) $6s\ 6p\ 6d\ 6f$
(3) $6s\ 5f\ 5d\ 6p$ (4) $6s\ 4f\ 5d\ 6p$ (4)

22. The outer electron configuration of Nickel which is to the left of Cu ($3d^{10}\ 4s^1$) is
(1) $3d^{10}\ 4s^2$ (2) $3d^{11}\ 4s^1$
(3) $3d^9\ 4s^1$ (4) $3d^8\ 4s^2$ (4)

23. If an electron transition occurs across the energy gap in Ga As of 1.43 eV, the wavelength of the radiation absorbed or emitted is ($h = 6.626 \times 10^{-34}$ J-s, $c = 3 \times 10^8$ m/s)

(1) 11200 Å (2) 8670 Å
(3) 12400 Å (4) 9892 Å (2)

24. Which one of the following combination of all the four quantum numbers n, l, m_l, m_s is not possible

(1) $3, 1, 1, -\frac{1}{2}$ (2) $3, 1, 1 +\frac{1}{2}$
(3) $1, 0, 0, -\frac{1}{2}$ (4) $1, 1, 1, -\frac{1}{2}$ (4)

25. The number of electrons in a d sub-shell of M shell is

(1) 10 (2) 8
(3) 12 (4) 16 (1)

26. Four quantum numbers n, l, m_l, m_s of four energy states of p-sub-shell of M-shell are as follows. Which one is wrong combination:

(1) $3\ 1\ 1\ \frac{1}{2}$
$3\ 1\ 0\ \frac{1}{2}$
$3\ 1\ 1\ \frac{1}{2}$
$3\ 1\ 0 -\frac{1}{2}$

(2) $3\ 1\ 0\ \frac{1}{2}$
$3\ 1\ 1\ \frac{1}{2}$
$3\ 1\ {-1}\ \frac{1}{2}$
$3\ 1\ {-1} -\frac{1}{2}$

(3) $3\ 1\ 1\ \frac{1}{2}$
$3\ 1\ {-1} -\frac{1}{2}$
$3\ 1\ 0\ \frac{1}{2}$
$3\ 1\ 0 -\frac{1}{2}$

(4) $3\ 1\ 1\ \frac{1}{2}$
$3\ 1\ {-1}\ \frac{1}{2}$
$3\ 1\ 0\ \frac{1}{2}$
$3\ 1\ 1 -\frac{1}{2}$ (1)

SHORT QUESTION-ANSWERS

1. How you view Bohr model and wave mechanical model of an atom?

Ans. Bohr model assumes electrons to be particles orbiting the nucleus in discrete paths, whereas in wave mechanical model of atom we consider electrons to be wave-like and treat electron position in terms of a probability distribution.

2. How electron energy states are specified?

Ans. These are specified in terms of quantum numbers that give rise to electron shells and subshells. The electron configuration of an atom corresponds to the manner in which these shells and subshells are filled with electrons in compliance with the Pauli exclusion principle.

3. How the periodic table of the elements is generated?

Ans. By arrangement of the various elements according to valence electron configuration.

4. What is ionization potential?

Ans. It is the energy required to remove an electron from the outer orbit of an atom.

5. What is electron affinity?

Ans. It is the energy released, when a free electron is added to the outer orbital.

6. What is the electronegativity of an atom?

Ans. It is a measure of the tendency of an atom to attract bonding electrons to itself.

3

Crystal Geometry, Structure and Defects

1. INTRODUCTION

Solids exist in nature in two principal forms: *crystalline* and *non-crystalline (amorphous)*, which differ substantially in their properties. Most crystalline solids are made up of millions of tiny single crystals called grains which constitute what is called microstructure and are said to be polycrystalline. These grains are oriented randomly with respect to each other. Any single crystal, however, no matter how large, is a single grain. Single crystals of metals many cubic centimeters in volume are relatively easy to prepare in the laboratory. Single crystals are regular polyhedrons whose shape depends upon their chemical composition.

Crystalline bodies remain solid, i.e., retain their shape, upto a definite temperature (melting point) at which they change from the solid to liquid state (Fig. 3.1). During cooling, the inverse process of solidification takes place, again at the definite solidifying temperature, or point. In both cases, the temperature remains constant until the material is completely melted or respectively solidified.

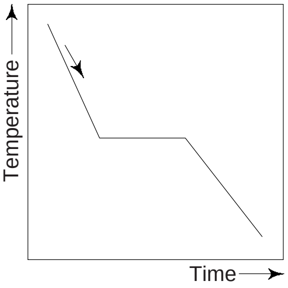


Fig. 3.1 Cooling curve of a Crystalline Substance

Amorphous substances have no crystalline structure in the condensed state (Fig. 3.2(a)) ordinary glass, sulphur, selenium, glycerine and most of the high polymers can exist in the amorphous state. Amorphous bodies, when heated, are gradually softened in a wide temperature range and become viscous and only then change to the liquid state. In cooling, the process takes place in the opposite direction. On repeated heating, long holding at 20-25°C or, in some cases, deformation of an amorphous body, the instability of the amorphous state may result in a partial or complete change to the crystalline state. Examples of such changes from amorphous to crystalline state are the turbidity effect appearing in inorganic glasses on heating or in optical glasses after a long use, partial crystallization of molten amber on heating, or additional crystallization and strengthening of nylon fibres on tension.

Crystalline bodies are characterized by an ordered arrangement of their ions, atoms or molecules (Fig. 3.2(b)). The properties of crystals depend on the electronic structure of atoms and the nature of their interactions in the crystal, on the spatial arrangement of their ions, atoms or molecules, and on the composition, size and shape of crystals.

A crystalline material may be either in the form of *single crystal* or an aggregate of many crystals usually known as polycrystalline separated by well-defined boundaries called as grain boundaries. Polycrystalline material is stronger than ordinary one because crystals in polycrystalline material have different orientations with respect to each other and grain boundaries obstruct the movement of dislocations. Such crystals are called *isotropic* because they exhibit same properties in every plane and direction, whereas single crystal is called *anisotropic*.

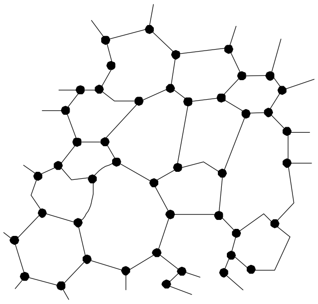


Fig. 3.2(a) **Amorphous, or non-crystalline structure**

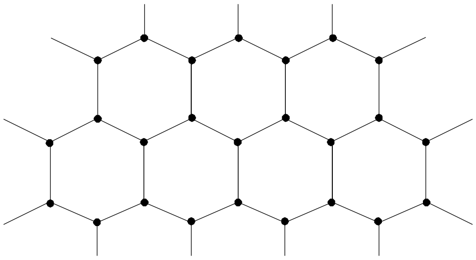


Fig. 3.2(b) **Crystalline (or highly ordered) structure**

2. CRYSTALS

These are solids which have a regular periodic arrangement in their component parties, bounded by flat faces, orderly arranged in reference to one another, which converge at the edges and vertices. A crystal is symmetrical about its certain elements like points, lines or planes and if it rotated about these elements, it is not possible to distinguish its new position from the original position. This symmetry is an important characteristic based on internal structure of crystal. Symmetry helps one to classify crystals and describing their behavior. At temperatures below that of crystallization, the crystalline state is stable for all solids.

3. SINGLE CRYSTAL

Most of the materials exist in polycrystalline form, but there are some materials, which exist in the form of single crystals, e.g. sugar, sodium chloride (common salt), diamond, etc.

Single crystals represent a material in its ideal condition and are produced artificially from their vapour or liquid state. These crystals help us in studying behaviour and defects of the material in ideal conditions.

4. WHISKERS

These are very thin filaments, hair-like single crystals of about 13 mm length and perhaps 10^{-4} cm diameter. These are produced as dislocations of free crystals and are without any structural defect. This is why whiskers are far stronger than polycrystals of same material. These are used as reinforcements in materials to increase strength by embedding fibres of one material in a matrix of another. The properties of these fibre or whisker-reinforced composites can often be tailored for a specific application. The increase in diameter of the whiskers decreases its strength and increases its ductility. The cost of whiskers and the expensive fabrication is the major disadvantage of the method.

Whiskers are the most defect-free crystalline solids available today. The best-known composite are probably, fibre glass, which consists of glass-reinforcing fibres in a matrix of either an epoxy polymer or polyester whiskers can bear considerably high stresses both at low and relatively elevated temperatures. Single crystals of SiC, Al_2O_3, S-Glass, graphite, boron, iron, silver, copper and tin can be produced by means of special techniques.

Whiskers of a wide variety of substances, e.g., mercury, graphite, sodium and potassium chlorides, copper, iron, and aluminium oxide, have been grown from super saturated media. Whiskers grown in this way are usually a few micrometers in diameter and up to a few inches long. Some are exceptionally strong, both in bend tests and in tension tests.

In addition to exceptional strength, whiskers often have unique electrical, magnetic, or surface properties. This behaviour can be interpreted to mean that the crystal structure of whiskers is virtually perfect, particularly with respect to line defects. Actually it appears that some whiskers contain line defects whereas others do not. However, no general correlation between whisker properties and whisker structure have been established.

5. LATTICE POINTS AND SPACE LATTICE

The atomic arrangement in crystal is called the crystal structure. In perfect crystal, there is a regular arrangement of atoms. In a model of a crystal, ions, atoms or molecules that constitute its structure can be imagined to be spheres which touch one another and are arranged regularly in different directions. In a simple model of crystal structure, spheres are replaced by points representing the centres of ions, atoms or molecules. The periodicity in the arrangement of ions, atoms or molecules generally varies in different directions. It is very convenient to imagine points in space about which these atoms, ions or molecules are located. Such points in space are called *lattice points*. The totality of lattice points forms a *crystal lattice* or *space lattice*. If all the atoms, molecules or ions at the lattice points are identical, the lattice is called a *Bravias lattice*. The space lattice of a crystal is described by means of a three-dimensional co-ordinate system in which the coordinate axis coincide with any three edges of the crystal that intersect at one point and do not lie in a single plane. Obviously, the three-dimensional space-lattice may be defined as a finite array of points in three dimensions in which every point has identical environment as any other point in the array. The space lattice is very useful as a reference in correlating the symmetry of actual crystals. A space lattice provides the framework with reference to which a crystal structure can be described. It is essential to distinguish a lattice from a crystal structure; a crystal structure is formed by associating with every lattice point an assembly of atoms identical in composition, arrangement, and orientation. The space-lattice concept was introduced by R.J. Hauy as an explanation for the special geometric properties of crystal polyhedrons. It was postulated that an elementary unit, having all the properties of the crystal, should exist, or conversely that a crystal was built up by the juxaposition of such elementary units. If the mathematical points forming the vertices of a parallelopiped *OABC* (defined by three vectors ($\overrightarrow{OA}, \overrightarrow{OB}$ and $\overrightarrow{OC}$) are considered (Fig. 3.3), a space lattice is obtained by translations parallel to and equal to $\overrightarrow{OA}, \overrightarrow{OB}$ and $\overrightarrow{OC}$. The parallelopiped is called the *unit cell*.

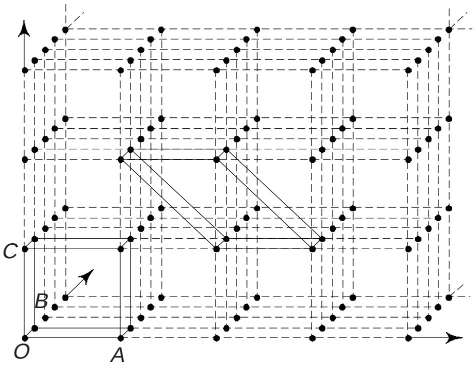


Fig. 3.3 A space lattice, two possible unit cells, and the environment of a point

In metals, we frequently encountered with the complex lattices-comprise of several primitive translation lattices displaced in relation to each other. Most of metallic crystals have highly symmetrical structures with closed packed atoms. The most common types of space lattices are: Body centred cubic (BCC) lattices, Face centred cubic (FCC) lattices and Hexagonal closed packed (HCP) lattices.

5(A). BASIS

The space lattice has been defined as an array of imaginary points which are so arranged in space that each point has identical surroundings. We must note that the crystal structure is always described in terms of atoms rather than points. Thus in order to obtain a crystal structure, an atom or a group of atoms must be placed on each lattice point in a regular fashion. Such an atom or a group of atoms is called the *basis* and this acts as a building unit or a structural unit for the complete crystal structure. Obviously, a lattice combined with a basis generates the *crystal structure*. Mathematically, one can express it as

$$\text{Space lattice} + \text{Basis} \rightarrow \text{Crystal Structure}$$

Thus, whereas a lattice is a mathematical concept, the crystal structure is a physical concept.

The generation of a crystal structure from a two-dimensional lattice is illustrated in Fig 3.3(a). The basis consists of two atoms, represented by ○ and •, having orientation as shown in Fig. 3.3(a). The crystal structure is obtained by placing the basis on each lattice point such that the centre of the basis coincides with the lattice point. We must note that the number of atoms in a basis may vary from one to several

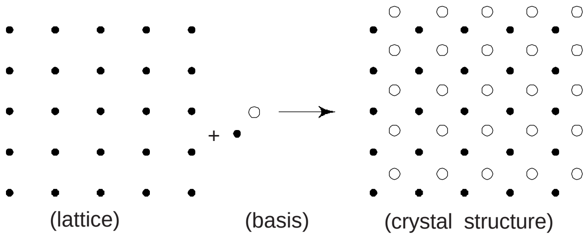


Fig. 3.3(a) Generation of crystal structure from lattice and basis

thousands, whereas the number of space lattices possible is only fourteen as described in a later section. Obviously, one can obtain a large number of crystal structures from just fourteen space lattices simply because of the different types of basis available. If the basis consists of a single atom only, a monoatomic crystal structure is obtained. Copper is an example of monoatomic face-centred cubic structures. Examples of complex bases are found in biological materials.

6. UNIT CELL

We have seen that the atomic order in crystalline solids indicates that the smallest groups of atoms form a repetitive pattern. Thus in describing crystal structures, it is often convenient to subdivide the structure into repetitive small repeat entities called *unit cells*, i.e. in every crystal some fundamental grouping of particles is repeated. Obviously, a unit cell is the smallest component of the space lattice. The unit cell is the basic structural unit or *building block* of the crystal structure by virtue of its geometry and atomic positions within. We must remember that space lattices of various substances differ in the size and shape of their unit cells. Figure 3.4 shows a unit cell of a three-dimensional crystal lattice. The distance from one atom to another atom measured along one of the axis is called the *space* constant. The unit cell is formed by primitives or intercepts a, b and c along *X*, *Y* and *Z* axes respectively.

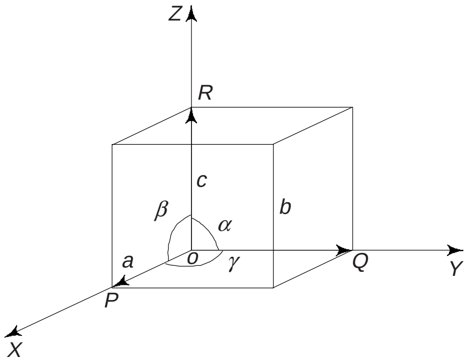


Fig. 3.4 Lattice parameters of a unit cell

A unit cell can be completely described by the three vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ ($\overrightarrow{OP}$, $\overrightarrow{OQ}$ and $\overrightarrow{OR}$) when the length of the vectors and the angles between them (α, β, γ) are specified. The three angles α, β and γ are called *interfacial angles*. Taking any lattice point as the origin, all other points on the lattice, can be obtained by a repeated of the lattice vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$. These lattice vectors and interfacial angles constitute the lattice parameter of a unit cell. Obviously, if the values of these intercepts and interfacial angles are known, one can easily determine the form and actual size of the unit cell.

7. PRIMITIVE CELL

This may be defined as a geometrical shape which, when repeated indefinitely in 3-dimensions, will fill all space and is equivalent of one lattice point, i.e. the unit cell that contains one lattice point only at the corners. We must note that in some cases the unit cell may coincide with the primitive cell, but in general the former differs from the latter in that it is not restricted to being the equivalent of one lattice point. The units cells, which contain more than one lattice point are called non-primitive cells. The unit cells may be primitive cells, but all the primitive cells need not to be unit cells.

8. CRYSTAL CLASSES

The atoms or molecules or ions in crystalline state are arranged in a regular, repetitive and symmetrical pattern, but the crystal will have the external symmetrical shape only, if no restraint is imposed during crystal growth.

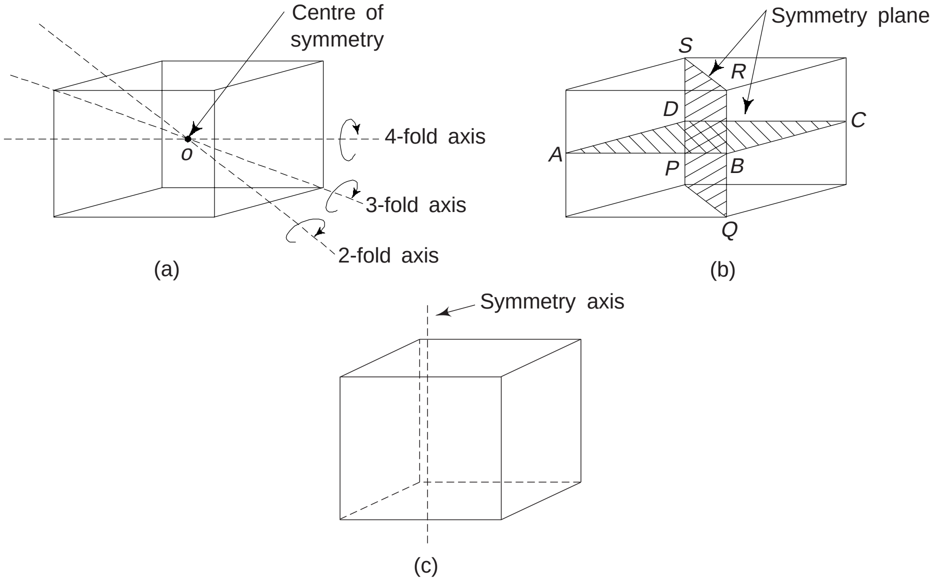
Crystals possess *symmetry* in that any given direction in the crystal corresponds to one or more directions which are exactly the same with respect to properties being considered. The symmetry of crystals is investigated by means of *symmetry operations*, as a result of which the crystal coincides with itself in various positions. The simplest of such operations (*rotation*, *reflection* and *translation*-parallel displacement) are associated with the *elements of symmetry*. The simplest elements of symmetry are the *axis and planes of symmetry*. The shape of the crystal is said to be symmetrical if it possesses one or more elements of symmetry. A group of symmetry operations, consisting commonly of a combination of rotations, refections and rotations with reflection, is called a *symmetry class*. The elements of symmetry are:

(i) The Symmetry Plane: The shape of the crystal is said to be symmetrical about a plane if it divides the shape into two identical halves or into two halves which are mirror images of one another. We must note that only in an ideal crystal the faces are of exactly same size.

(ii) The Symmetry Axis: If we can rotate the shape about an axis so that the shape occupies the same relative position in space more than once in a complete revolution, such an axis is called to be axis of symmetry. Such axes may be either 2, 3, 4 or 6 fold. The axis of symmetry causes the crystal to occupy more than one congruent position during rotation about that axis during rotation by 360°.

(iii) The Centre of Symmetry: Within a crystal, there is some point about which crystallographically similar faces are arranged in parallel and corresponding positions, e.g., the centre of the cube is a centre of symmetry. We must note that a tetrahedron has no such centre. A cube has highly symmetrical shape and contains many planes and axis of symmetry (Fig. 3.5).

The axis of the symmetry is the imaginary line, passing through the centre of the crystal, about which the crystal may be rotated so that it presents an identical environment, i.e. appearance more than once in the course of its rotation. When rotation around its axis presents the same appearance once, we call it as *one-fold symmetry*. One can determine the folds of symmetry by rotating the crystal through 360°, 180°, 90°, 60°, and so on to obtain 1, 2, 3, 4, . . . fold symmetry. We must note that the limit for crystalline solid is six-fold symmetry.



(a) (b)

(c)

Fig. 3.5 The elements of symmetry (a) axis and (b) planes

The principal axes of a cube are four fold, i.e. during each complete rotation about the axis, the crystal passes four through identical positions. The body diagonal axes are three fold and there are six two-fold axes. The vertical axis of hexagonal prism is a six-fold axes.

A symmetry operation is one that leaves the crystal and its environment invariant. Symmetry operations performed about a point or a line are called *point group symmetry* operations and symmetry operation performed by translations as well are called *space group symmetry operations.* We must note that crystals exhibit both types of symmetries independently and in compatible combinations. The following are the different point group symmetry elements that are exhibited by crystals: *(i)* centre of symmetry or inversion centre, *(ii)* reflection symmetry and *(iii)* rotation symmetry.

9. CRYSTAL SYSTEMS

If all the atoms at the lattice points are identical, the lattice is said to be *Bravais lattice.* There are four systems and five possible Bravais lattices in two dimensions (Fig. 3.6(a)). The four crystal systems of two dimensional space are oblique, rectangular, square and hexagonal. The rectangular crystal system has two Bravais lattices, namely, rectangular primitive and rectangular centered. In all, there are five Bravais lattices which are listed in Table 3.1 alongwith the corresponding point groups.

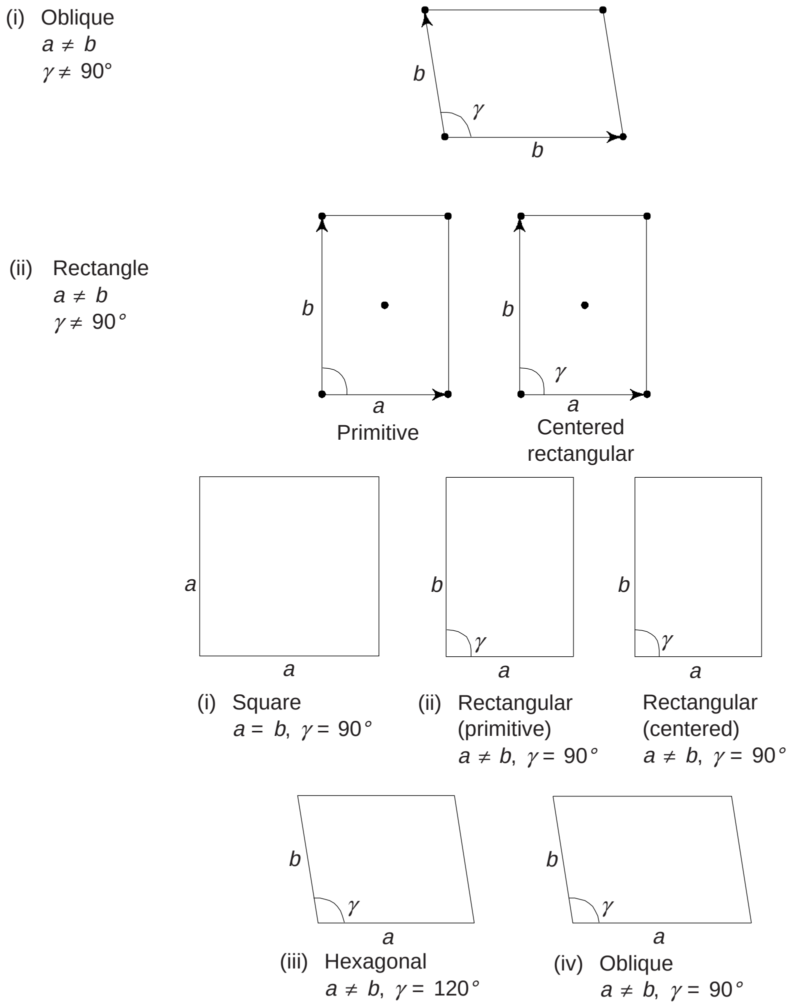


Fig. 3.6(a) Bravais lattices in two dimensions

Table 3.1 Crystal systems and Bravais lattices in two dimensions

S. No.	*Crystal system*	*Characteristic point group symmetry*	*Bravais lattice*	*Conventional unit cell*	*Unitl cell characteristics*
1.	Oblique	1, 2	oblique	Parallelogram	$a \neq b$, $\gamma \neq 90°$
2.	Rectangular	1m, 2 mm	1 Rectangular primitive 2 Rectangular centered	Rectangle	$a \neq b$, $\gamma = 90°$
3.	Square	4, 4 mm	Square	Square	$a = b$, $\gamma = 90°$
4.	Hexagonal	3, 3 m, 6, 6mm	Hexagonal	60° Rhombus	$a = b$, $\gamma = 120°$

Based on pure symmetry considerations, there are only fourteen independent ways of arranging points in three-dimensional space, such that each arrangement is in accordance or in confirmation with the definition of a space lattice. These 14 space lattices with 32 point groups and 230 space groups are called *Bravais lattices* (Fig. 3.6(b)). If considered as solids, the combination of symmetry elements they exhibit can be determined. Each space lattice can be defined by reference to a unit cell which, when repeated in space an infinite number of times, will generate the entire space lattice. To describe basic crystal structures, the 14 types of unit cells are grouped in seven different classes of crystal lattices, i.e. to describe basic crystal structures, seven different co-ordinate systems of reference axes are required (Fig. 3.7).

A description of the characteristics of 14 Bravais lattices of three dimensions along with the axial relationship for the class of crystal lattices, i.e. seven systems to which each belongs are summarized in Table 3.2.

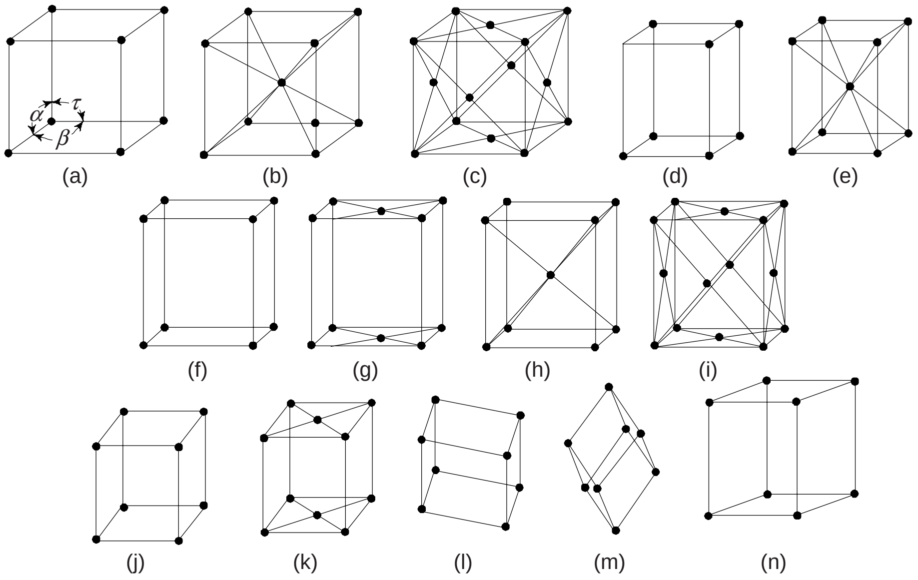


Fig. 3.6(b) The 14- Bravais space lattices:

(a) Simple cubic
(b) Body centered cubic
(c) Face-centered cubic
(d) Tetragonal
(e) Body-centered tetragonal
(f) Orthorhombic
(g) Base-centered orthorhombic
(h) Body-centered orthorhombic
(i) Face-centered orthorhombic
(j) Monoclinic
(k) Base-centered monoclinic
(l) Triclinic
(m) Trigonal
(n) Hexagonal

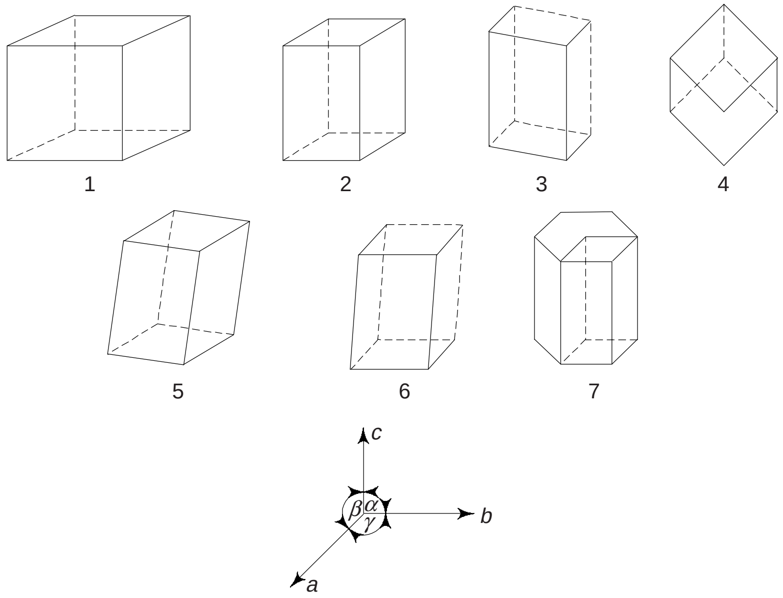


Fig. 3.7 Seven basic crystal systems with reference axis

Table 3.2 The Bravais lattices in three dimensions (the seven basic crystal systems)

Crystal system	*Lattice type*	*No. of lattices*	*Relation between primitives*	*Interface angles*	*Examples*
Cubic	P, F, C	3	$a = b = c$	$\alpha = \beta = \gamma = 90°$	Au, NaCl, CaF$_2$, CrCl, CaO (I)}
Monoclinic	P, B	2	$a \neq b \neq c$	$\alpha = \beta = 90° \neq \gamma$	2H$_2$O, NaSO$_4$, CaSO$_4$, FeSO$_4$}
Triclinic	P	1	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90°$	K$_2$Cr$_2$O$_7$, CuSO$_4$, K$_2$S$_2$O$_8$
Tetragonal	P, C	2	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$	NiSO$_4$, Sn, TiO$_3$, and SnO$_2$
Orthogonal	P, B, F, C	4	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$	MgSO$_4$, KNO$_3$, and BaSO$_4$
Rhombohedral (Trigonal) (orthorhombic)	P or R	1	$a = b = c$	$\alpha = \beta = \gamma \neq 90°$	SiO$_2$, CaSO$_4$, and CaCO$_3$
Hexagonal	P	1	$a = b \neq c$	$\alpha = \beta = 90°\ \gamma = 120°$	AgCl, SiO$_2$, Zn and Graphite

Representation of symbols: P → primitive, B → base centred
C → body centered and
F → face centered

The number of lattice points in unit cell can be calculated by appreciating the following:

Contribution of lattice point at the corner = $\frac{1}{8}$th of the point

Contribution of the lattice point at the face = $\frac{1}{2}$ of the point

Contribution of the lattice point at the centre = 1 of the point

Every type of unit cell is characterized by the number of lattice points (not the atoms) in it. For example, the number of lattice points per unit cell for simple cubic (SC), body centered cubic (BCC) and face centered cubic (FCC) lattices are 1, 2 and 4, respectively. We must note that our knowledge about unit cell may not be complete without having a quantitative estimate of its volume. This can be calculated with the help of the relation

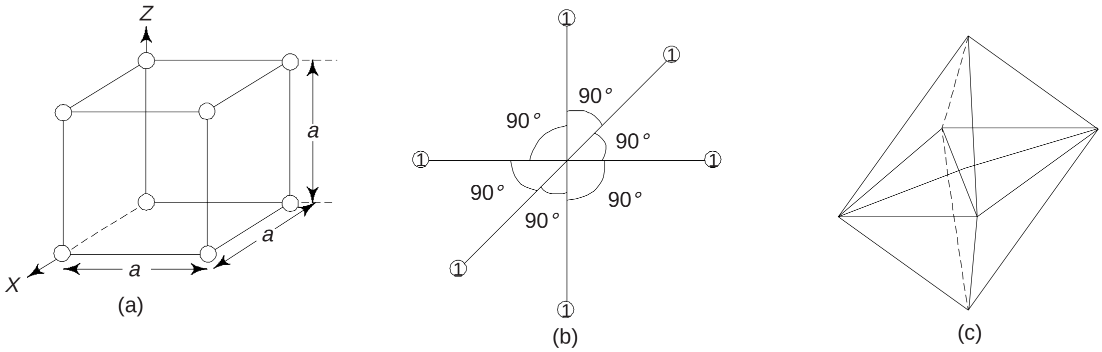
$$V_c = \vec{a}\cdot[\vec{b}\times\vec{c}] \tag{1}$$

where V_c stands for the volume of the cell and $\vec{a}$, $\vec{b}$ and $\vec{c}$ defined so far as the measure of the unit cell edges, are commonly known as lattice parameters.

Now, we shall discuss about the seven type of basic systems mentioned in Table 3.1.

(i) Cubic Crystal System: ($a = b = c$, $\alpha = \beta = \gamma = 90°$): All those crystals which have three equal axes and are at right angles to each other and in which all the atoms are arranged in a regular cube are said to be cubic crystals (Fig. 3.8). The most common examples of this system are cube and octahedron as shown in Fig. 3.8(a) and (c). In a cubic crystal system, we have

$$a = b = c,\ \alpha = \beta = \gamma = 90°$$



(a) (b) (c)

Fig. 3.8 Cubic Crystal System

Atomic Packing Factor (APF): This is defined as the ratio of total volume of atoms in a unit cell to the total volume of the unit cell. This is also called relative density of packing (RDP).
Thus

$$\text{APF} = \frac{\text{No. of atoms} \times \text{volume of one atom}}{\text{volume of unit cell}} = \frac{v}{V} \tag{2}$$

In a simple cubic cell, no. of atoms in all corners

$$= \frac{1}{8} \times 8 = 1$$

Radius of an atom = r and volume of cubic cell = $a^3 = (2r)^3$ $\qquad (\because a = 2r)$

$$\therefore \qquad \text{APF} = \frac{1 \times \frac{4}{3}\pi r^3}{(2r)^3} = \frac{\pi}{6}\ (= 0.52) = 52\%$$

(ii) Tetragonal Crystal System: ($a = b \neq c$, $\alpha = \beta = \gamma = 90°$). This includes all those crystals, which have three axes at right angles to each other and two of these axes (say horizontal) are equal, while the third (say vertical) is different (i.e., either longer or shorter than the other two). Figure 3.9(b) shows a tetragonal crystal system. The most common examples of this system of crystals are regular tetragonal and pyramids (Fig. 3.9(a) and (c))

(iii) Hexagonal Crystal System: ($a = b \neq c$, $\alpha = \beta = \gamma = 90°$, $\gamma = 120°$): All those crystals which have four axes falls under this system. Three of these axes (say horizontal) are equal and meet each other at an angle of 60° and the fourth axis (say vertical) is different, i.e. either longer or shorter than the other three axes. (Fig. 3.10(a)).

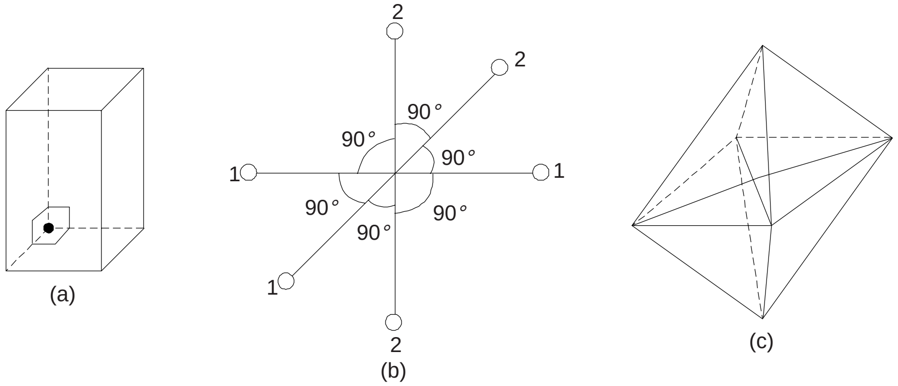


Fig. 3.9 Tetragonal crystal system

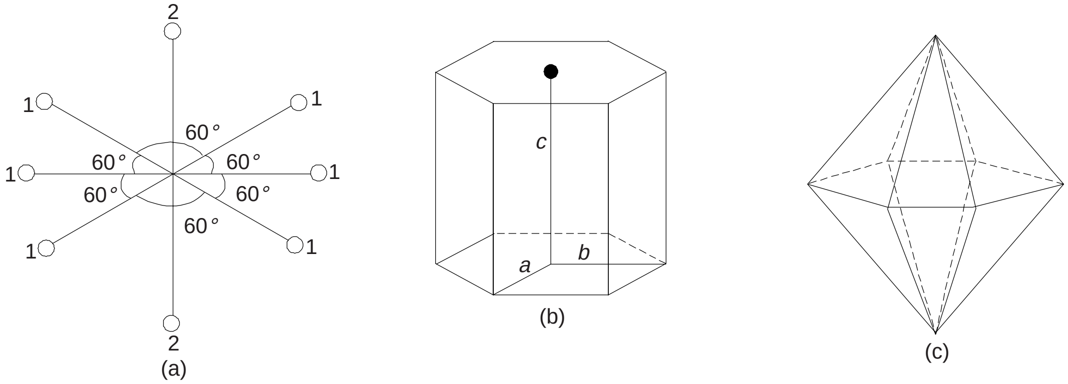


Fig. 3.10 Hexagonal crystal system

A simple hexagonal unit cell contains one atom at each corner of hexagon and one each at the centre of hexagonal faces. This has hcp structure. There are three atoms at the interstices between two hexagonal faces. The total number of atoms inside the hcp structure is 6 (Fig. 3.11a).

Each corner atom is has $\frac{1}{6}$ atom, i.e. shared by 6 other unit lattices.

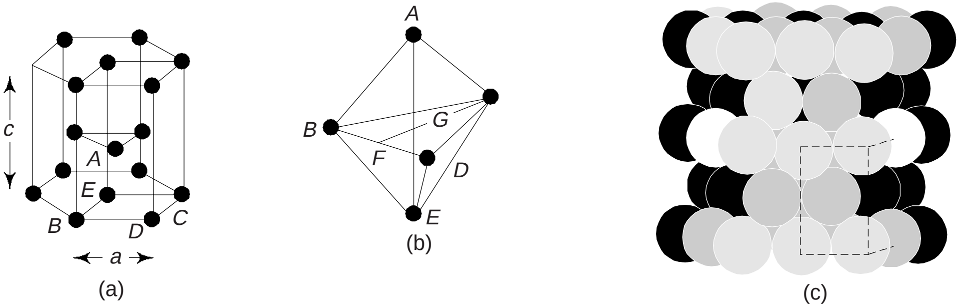


Fig. 3.11 Hexagonal closed packed structure
(a) Three unit cells showing how the hexagonal axis results. One of the cell is fully outlined.
(b) Calculation of the ratio c/a. Distance *AE* is the height of the cell.
(c) Hexagonal close packing

Number of atoms in upper hexagonal plane = $\frac{1}{6} \times 6 = 1$ and number of atoms in lower hexagonal plane $= \frac{1}{6} \times 6 = 1$.

We can see that each central atom is shared by two unit cells which means upper and lower planes contain $\frac{1}{2}$ atom each.

Thus the total number of central atoms in both, upper and lower planes

$$= \frac{1}{2} \times 2 = 1$$

and we note that there are 3 intersititial atoms.

∴ Total number of atoms in hcp crystal

$$= 1 + 1 + 1 + 3 = 6$$

APF may be calculated and found to be $\pi\sqrt{2}/6$ (= 0.74). This is identical to the packing factor of FCC metal because each has a co-ordination number of 12. The most common examples of such systems are regular hexagonal prisms and hexagonal pyramids as shown in Fig. 3.10(b) and (c). HCP is found in such metals as Ca, Mg, Be, Zn, Cd, Ti and others.

(iv) Orthorhombic Crystal System ($a \neq b \neq c$, $\alpha = \beta = \gamma = 90°$): The space lattice is simple. The crystal axes are perpendicular to one another but all the three axes are essentially of unequal lengths (Fig. 3.12). Orthorhombic lattices may be simple, base centered, body centered or face-centered. Orthorhombic prisms and pyramids (Fig. 3.12(b) and (c)) are most common examples of orthorhombic crystal system.

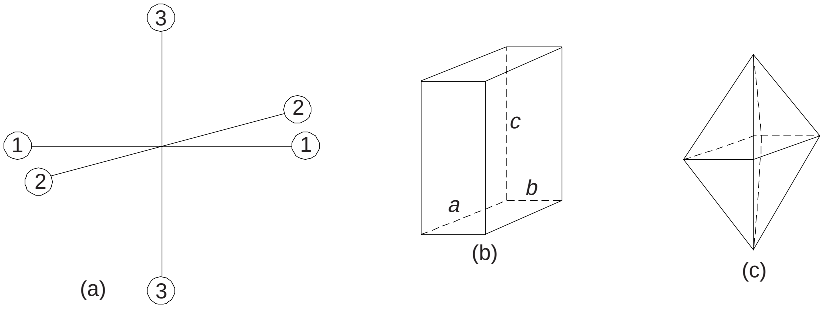


Fig. 3.12 Orthorhombic crystal system

(v) Rhom-bohedral or Trigonal Crystal System ($a = b = c$, $\alpha = \beta = \gamma \neq 90°$): Three axes are equal and are equally inclined to each other at an angle. Other 90° rhombohedral prisms and pyramids (Fig. 3.13(b) and (c)) are the most common examples of this crystal system.

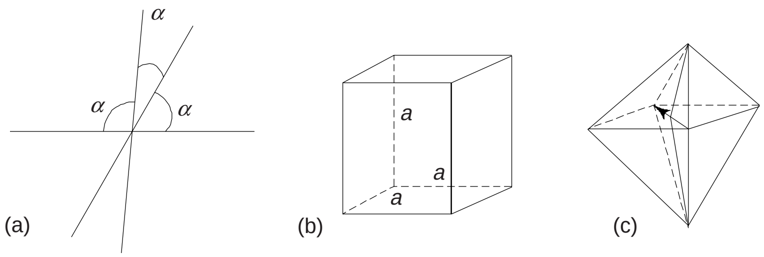


Fig. 3.13

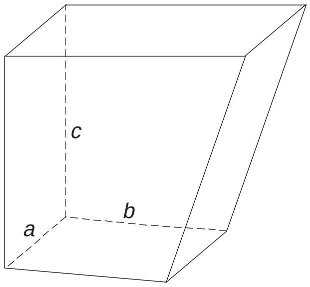


Fig. 3.14 Monoclinic crystal system

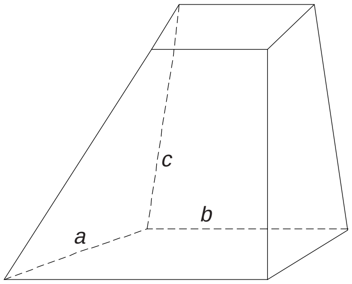


Fig. 3.15 Triclinic crystal system

(vi) Monoclinic Crystal System: ($a \neq b \neq c$, $\alpha = \beta = 90° \neq \gamma$): Two of the crystal axes are perpendicular to each other, but the third is obliquely inclined. The repetitive intervals are different along all the three axes. Monoclinic lattices may be simple or base-centered.

(vii) Triclinic Crystal System: ($a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90°$): All those crystals, which have three unequal axes and none of them is at right angles to the other two axes (Fig. 3.15) are included in this crystal system. The repetitive intervals are different along all the three axes. All irregular crystals belong to this class.

10. CRYSTAL STRUCTURE FOR METALLIC ELEMENTS

The most common types of space lattice or unit cells with which most common metals crystallise, are

(i) Body-centered cubic structures (BCC)
(ii) Face-centered cubic structures (FCC)
(iii) Hexagonal closed-packed structures (HCP)

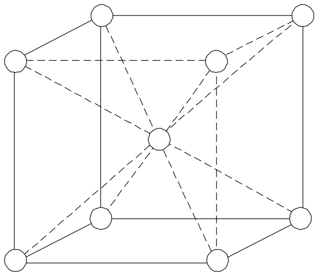


Fig. 3.16 BCC structure

(i) Body-centered Cubic Structure (BCC): The unit cell of BCC system has an atom at each corner, which are shared by the adjoining eight cubes and one at the body centre (Fig. 3.16). Obviously, each unit cell share 8 atoms one on each of its corners in addition to one atom at the body centre. Hence the share of each cube = $\frac{1}{8}$th of each corner atom.

$\therefore$ Total no. of atoms = $\frac{1}{8} \times 8 = 1$ atom

BCC crystal has one atom at the centre = 1 atom

$\therefore$ Total atoms in BCC unit cell = 1 + 1 = 2 atoms

$$(\text{Body diagonal})^2 = \text{Sum of the squares of all the three sides}$$

$$\therefore \quad (4r)^2 = a^2 + a^2 + a^2$$

$$\text{or} \quad a = \sqrt{\frac{16r^2}{3}} = \frac{4r}{\sqrt{3}}$$

$$\text{Volume} = 2 \times \frac{4}{3}\pi r^3$$

$$= 2 \times \frac{4}{3}\pi\left(\frac{a\sqrt{3}}{4}\right)^3$$

$$\therefore \quad r = \frac{a\sqrt{3}}{4}$$

$$\text{Volume of unit cell} = a^3 = \left(\frac{4r}{\sqrt{3}}\right)^3$$

$$\therefore \quad \text{APF} = \frac{(\text{Number of atoms}) \times (\text{Volume of one atom})}{\text{Volume of unit cell}}$$

$$= \frac{2 \times \frac{4}{3}\pi r^3}{\left(4r/\sqrt{3}\right)^3} = \frac{\sqrt{3}}{8}\pi = 0.68$$

The lattice constant of BCC structure is 9. Some of the metals possessing this structure are α-Fe (below 910°C), Mo, V, Mn, Nb, W, α – Cr, δ-iron (1400°C to 1539°C), etc.

(ii) Face-centered Cubic Structure (FCC): In this type of lattice structure, also known as cubic packed structure, the atoms are located at the eight corners of the cube and at the centre of each face (Fig. 3.17). This type of structure is more common among metals than the BCC structure. It is also formed in ceramic crystals. Cu, Al, Ag, Au, α-Fe, Ca, β-Co, γ-Iron (910°C to 1400°C) are common examples of metals possessing this type of structure.

A metal with FCC structure has four times as many atoms as it has unit cells. This means the FCC structure is more densely packed than BCC structure.

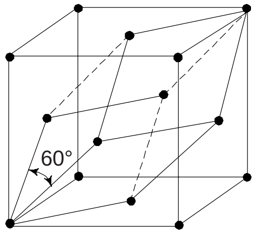


Fig. 3.17 Face centered cubic structure

Each face atom in FCC structure is shared by two unit cells, or we can say that one face of the FCC unit contains 1/2 atom.

$$\text{No. of atoms in all six faces} = \frac{1}{2} \times 6 = 3$$

$$\text{No. of atoms in all corners} = \frac{1}{8} \times 8 = 1$$

$\therefore$ Total atoms in FCC unit cell = 1 + 3 = 4 atoms

From the geometry of FCC structure

$$a^2 + a^2 = 4r^2$$

$$\therefore \quad a = \sqrt{8}\, r = 2\sqrt{2}\, r$$

$$\text{Volume of 4 atoms} = 4 \times \frac{4}{3}\pi r^3$$

$$= 4 \times \frac{4}{3}\frac{\left(a\sqrt{2}\right)^3 \pi}{(4)^3}$$

$$= \frac{16}{3}\pi a^3 \times \frac{2\sqrt{2}}{64} = \frac{\pi}{6}\sqrt{2}\, a^3$$

$$\therefore \quad \text{APF} = \frac{4 \times \frac{4}{3}\pi r^3}{\left(2\sqrt{2}\, r\right)^3} = \frac{\pi}{3\sqrt{2}} = 0.74$$

We must note that no other structure possesses such a large number of closed-packed planes and directions. This is why, metals with an FCC structure can be deformed critically. The coordination number in an FCC metal is 12 which accounts for the higher atomic packing fraction factor.

As an example structure of few solids are described hereunder:

(i) Sodium Chloride Structure: This is an arrangement in which each positive ion (Na$^+$) is surrounded by six negative ions (Cl$^-$) and vice versa (Fig. 3.18). The ionic radius of Cl$^-$ and Na$^+$ ions is 1.82 Å and 0.98 Å respectively. Obviously, the ratio of Na$^+$ ions is approximately half of Cl$^-$ ions. This is why, for such a high ratio, only 6 Cl$^-$ ions be packed around one Na$^+$ ion. Since positive and negative ions must be combined in such a manner so as to keep the net charge zero, i.e. they must be equal in number. This is why they form a cubic structure. The central cube is a FCC unit cell of sodium, while the central plane is the face plane of an overlapping FCC unit cell of chlorine. Obviously, the NaCl structure can, therefore, be viewed as two interpenetrating FCC sublattices, one belonging to Na$^+$ ions with its origin at the point (0, 0, 0) and the other belonging to Cl$^-$ ions with its origin at the point (a/2, 0, 0).

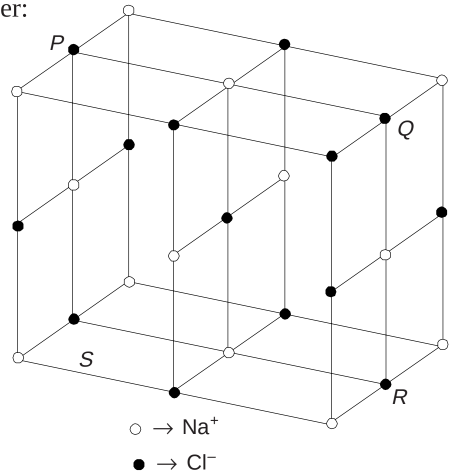


Fig. 3.18 Sodium chloride structure

(ii) Silica Structure: The ratio of silicon to oxygen atoms in Silica is 1 : 2. One silicon atom is surrounded by four oxygen atoms placed at the corners of the tetrahedra. When all the corners one silicon oxygen tetrahedra are shared by the other tetraohedra, the structure is said to be complete. Obviously, every oxygen atom is common to two tetrahedra. The melting point of silica is 1710 °C due to strong primary bonds between silicon and oxygen. The structure of silica is electrically neutral, open and having lower density.

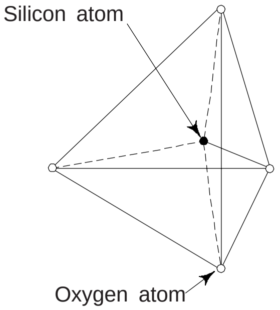
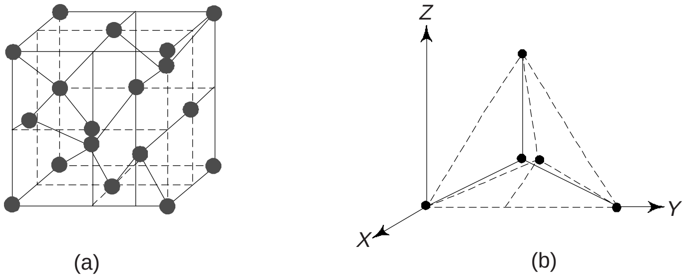


Fig. 3.19 Silica structure

(iii) Diamond Structure: Diamond exhibits both cubic and hexagonal type structures. The diamond cubic structure is more common. In diamond structure, each carbon atom is surrounded by other four carbon atoms (co-ordination-4). The limitation on account of number of bonds prevents the formation of closed packed structures. From Fig. 3.20, it is clear that the four nearest carbon atoms (neighbours) are at equal distance from the central atom and they form a tetrahedra. The coordination number of each carbon atom is 4 and the nearest neighbour distance is equal to $\sqrt{3}a/4$, where a is the lattice parameter. The angles made by the lines joining corner atoms with central atom are 109. 5°. The complete crystal is formed by joining a number of these tetrahedra.



(a) (b)

Fig. 3.20 Diamond structure

(iv) Hexagonal Closed Packed (HCP) Structures: In HCP structures, the unit cell contains one atom at each corner of the hexagonal prism, one atom each at the centre of the hexagonal faces and three more atoms within the body of the cell (Fig. 3.21). The HCP structure do not constitute a space lattice because the surroundings of the interior atoms and the corner atoms are different. This type of structure is more dense than the simple hexagonal structure. The total number of atoms inside the NCP structure is 6, as shown below.

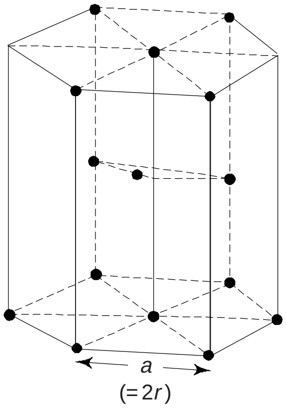


Fig. 3.21 HCP structure

Each corner atom is shared by six other unit lattices or each corner has 1/6 atom.

Number of atoms in upper hexagonal plane

$$= \frac{1}{6} \times 6 = 1$$

Number of atoms in lower hexagonal plane

$$= \frac{1}{6} \times 6 = 1$$

We note that each central atom is shared by two unit cells which means upper and lower planes contain $\frac{1}{2}$ atom each.

∴ Total number of central atoms in both, upper and lower planes = $\frac{1}{2} \times 2 = 1$

and there are three intersititial atoms.

∴ Total number of atoms in HCP crystal = 1 + 1 + 1 + 3 = 6

$$\text{APF} = \frac{2(4/3)\pi r^3}{a(a \sin 60°)\, c}$$

where r is the atomic radius. Using $c = 1.633a$ and $a = 2r$, one obtains,

$$\text{APF} = \frac{\pi\sqrt{2}}{6} = 0.74$$

Obviously, in an ideal HCP structure, 74% of the total volume is occupied by atoms. We further note that the APF value for HCP structure is identical to the APF of an FCC metal because each has a co-ordinate number of 12. The common examples of HCP structure are Be, Mg, Zn, Cd, Ti, Co, Hf, Se, Te, etc.

11. ATOMIC RADIUS

This is defined as half the distance between the nearest neighbours in the crystal structure of a pure element. Atomic radius is denoted by r and expressed in terms of the cube edge element a. One can calculate the atomic radius by assuming that atoms are spheres in contact in a crystal. Calculation of atomic radius in various crystal structures is illustrated as below:

(i) Simple Cubic (SC) Structure: In SC structure, atoms touch each other along the lattice (Fig. 3.22(a)). We have

$$a = 2r$$

∴ $r = a/2$

Area $a^2 = 4r^2$

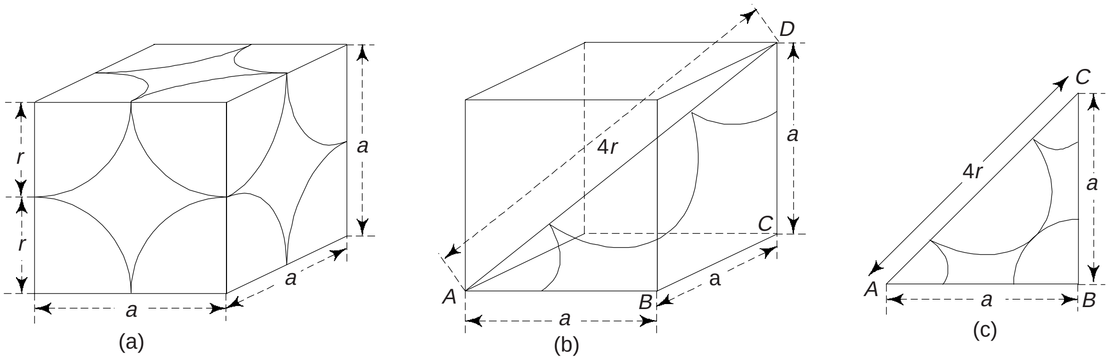


Fig. 3.22 Atomic radii for three unit cells (a) Simple Cubic (SC) (b) Body-centred cubic (BCC) and (c) Face-centred cubic (FCC)

(ii) BCC Structure: The atoms touch each other along the diagonal of the cube as shown in Fig. 3.22(b). Obviously, the diagonal in this case is 4*r*. Also,

$$AC^2 = AB^2 + BC^2 = a^2 + a^2 = 2a^2$$

$$AD^2 = AC^2 + CD^2 = 2a^2 + a^2 = 3a^2$$

$$\therefore \quad (4r)^2 = 3a^2$$

or $$a = \frac{4r}{\sqrt{3}} \text{ and } 2r = \frac{a\sqrt{3}}{2}$$

$$\therefore \quad r = \frac{a\sqrt{3}}{4}$$

area $$a^2 = \frac{16r^2}{3}$$

(iii) FCC Structure: Atoms within this structure touch along the diagonal of any face of the cube (Fig. 3.22 (c)). The diagonal has a length of 4*r*.

$$AC^2 = AB^2 + BC^2$$

or $$(4r)^2 = a^2 + a^2$$

$$\therefore \quad r^2 = \frac{2a^2}{16}$$

or $$r = \frac{\sqrt{2}a}{4} = \frac{a}{2\sqrt{2}}$$

and area $$= a^2 = 8r^2$$

12. DENSITY OF CRYSTAL

This is defined as:

$$\text{Density of the crystal } (\rho) = \frac{\text{Mass of unit cell}}{\text{Volume of unit cell}}$$

$$\text{Mass of unit cell} = \frac{\text{Atomic mass}}{\text{Avogadro number}} \times \text{Number of atoms per unit cell}$$

$$= \frac{M}{N_A} \times n$$

(n → number of atoms per unit cell and N_A → Avogadro's number, M → the atomic weight, and a → side of a cubic unit cell).

$$= \frac{nM}{a^3 N_A}$$

(i) Linear density: This is defined as the number of atoms per unit length along a specific crystal direction.

(ii) Planar density: This is defined as the number of atoms per unit area on a crystal plane. This affect significantly the rate of plastic deformation.

13. DIRECTIONS, LATTICE PLANES AND MILLER INDICES

In a crystal there exists directions and lattice planes which contain a large concentration of atoms. Various properties of crystals, particularly mechanical are connected with the structure of the crystal though the help of crystal directions. A complete description of the crystal structure can be obtained from the study of atomic positions in a unit cell. For crystal analysis it is necessary to locate directions and lattice planes.

(i) Crystal Directions: To specify the direction of a straight line joining lattice points in a crystal lattice, we choose any lattice point on the line as the origin and express the vector joining this to any other lattice point on the line as follows:

$$\vec{r} = n_1'\vec{a} + n_2'\vec{b} + n_3'\vec{c} \qquad (3)$$

The direction of the line is represented by the set of integers n_1', n_2' and n_3'. If these integer numbers have common factors, they are removed and the direction is denoted by [n_1, n_2, n_3]. Moreover, this line also denotes all lines parallel to this line. In Fig. 3.22(d), three different directions are shown in the orthorhombic lattice. The direction [111] is the line passing through the origin O and point P. It may be noted that the point P is at a unit cell distance from each axis. The direction [100] is the line passing through origin O and point Q. Obviously, the point Q is at a distance 1, 0, 0 from x, y and z axes respectively. The direction [101] is the line passing through the origin O and the point R. Again, the point R is at a unit cell distance of 1, 0, 1 from x, y and z axes respectively.

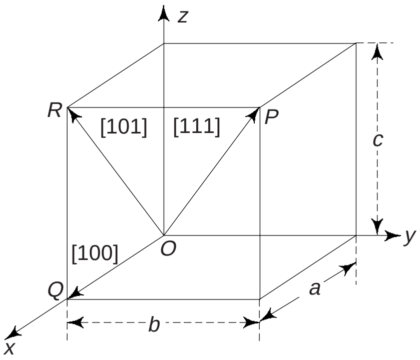


Fig. 3.22(d) Crystal directions in a orthorhombic lattice

In specifying crystal directions, we have taken crystal axes as base directions. We must note that directions [333] or [222] are identical to the direction [111]. As stated earlier, in such cases lowest combination of integers, i.e. [111] is used to specify the direction.

We must note that there are other directions, not parallel to the one under consideration which are equivalent to the given direction by virtue of rotation symmetry. Thus, the equivalent directions of [100] are [010], [001], [$\bar{1}00$], [$0\bar{1}0$] and [$00\bar{1}$], where the bars denote the negative values. By all possible positive and negative combinations of indices, we obtain a family of directions. In the present example, all these six equivalent directions are grouped together in the symbol <100>, where the bracket < > represents the whole family.

(ii) Crystallographic Planes: The crystal lattice may be regarded as made-up of an aggregate of a set of parallel equidistant planes, passing through the lattice points, which are known as lattice planes, or atomic planes. For a given lattice, one can choose the lattice planes in different ways as shown in Fig. 3.23. These crystal planes and crystal directions play an important role in hardening reaction, plastic deformation and other properties as well as behaviour of metal. Crystal planes in cubic structures are shown in Fig. 3.24.

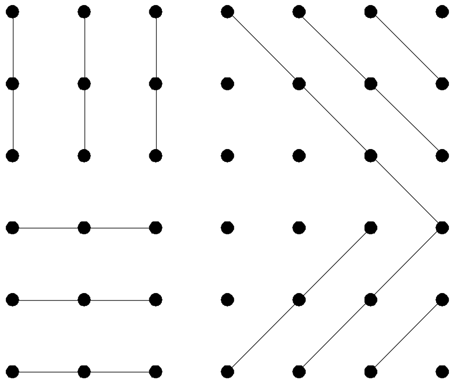


Fig. 3.23 Different crystal planes

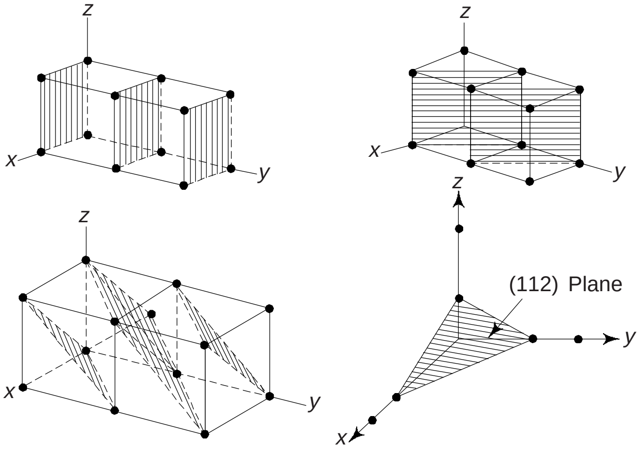


Fig. 3.24 Crystal planes in cubic structures (a) (010) plane (b) ($\bar{1}10$) plane and (c) (110) plane and (d) (112) plane

The problem is that how to designate these planes in a crystal. Miller evolved a method to designate a set of parallel planes in a crystal by three numbers *h*, *k* and *l*, usually written within brackets thus (*h*, *k*, *l*) known as *Miller indices* of the plane.

(iii) Determination of Miller Indices:

(a) Determine the coordinates of the intercepts made by the plane along the three crystallographic axes, i.e. *x*, *y*, *z* axes (Fig. 3.25)

x	y	z
$2a$	$3b$	c
pa	qb	rc

$(p = 2,\ q = 3,\ r = 1)$

(b) Now, express the above intercepts as multiples of the unit cell dimensions, or lattice parameters along the axes, i.e.

axes	x	y	z
	$\frac{2a}{a}$	$\frac{3b}{b}$	$\frac{c}{c}$
Intercept	2	3	1

(c) Take the reciprocal of intercepts

$$\frac{1}{2} \qquad \frac{1}{3} \qquad \frac{1}{1}$$

(d) Now, reduce these reciprocals to the smallest set of integral numbers and enclose within a bracket

$$6 \times \frac{1}{2} \qquad 6 \times \frac{1}{3} \qquad 6 \times \frac{1}{1}$$

$$(3 \qquad 2 \qquad 6)$$

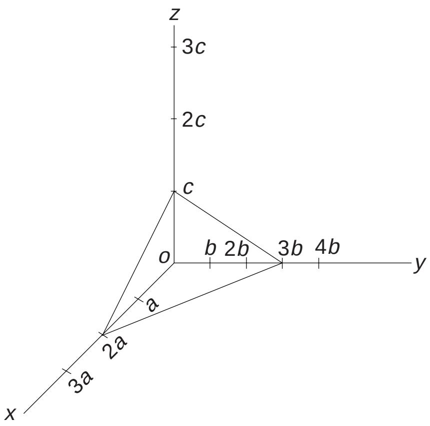


Fig. 3.25 Intercepts of plane on three crystallo-graphic axes and Miller indices

These integers within the bracket are Miller indices. In general it is denoted by (*h*, *k*, *l*). Thus, we note that

$$\frac{1}{p} : \frac{1}{q} : \frac{1}{r} = h : k : l$$

$$\frac{1}{2}:\frac{1}{3}:\frac{1}{1} = 3 : 2 : 6$$

Obviously, *Miller indices* are defined as the reciprocals of the intercepts made by the plane on the crystallographic axes when reduced to smallest numbers. We must remember that all planes have same indices. If negative sections are cut off by the plane, this is indicated by a bar above the corresponding index, e.g. $\bar{1}$ (Fig. 3.26).

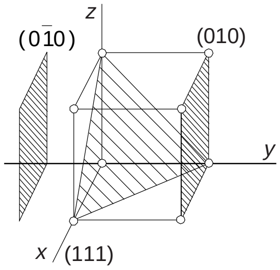


Fig. 3.26 Miller indices

Figure 3.27 shows the planes in a cubic structure. We can easily see that the Miller indices of the sides of a unit cell of a cubic lattice are (100), ($\bar{1}$00), (010), (0$\bar{1}$0), (001) and (00$\bar{1}$). These planes are planes of the same form, i.e. equivalent planes and are collectively represented by {100} and are called families of planes.

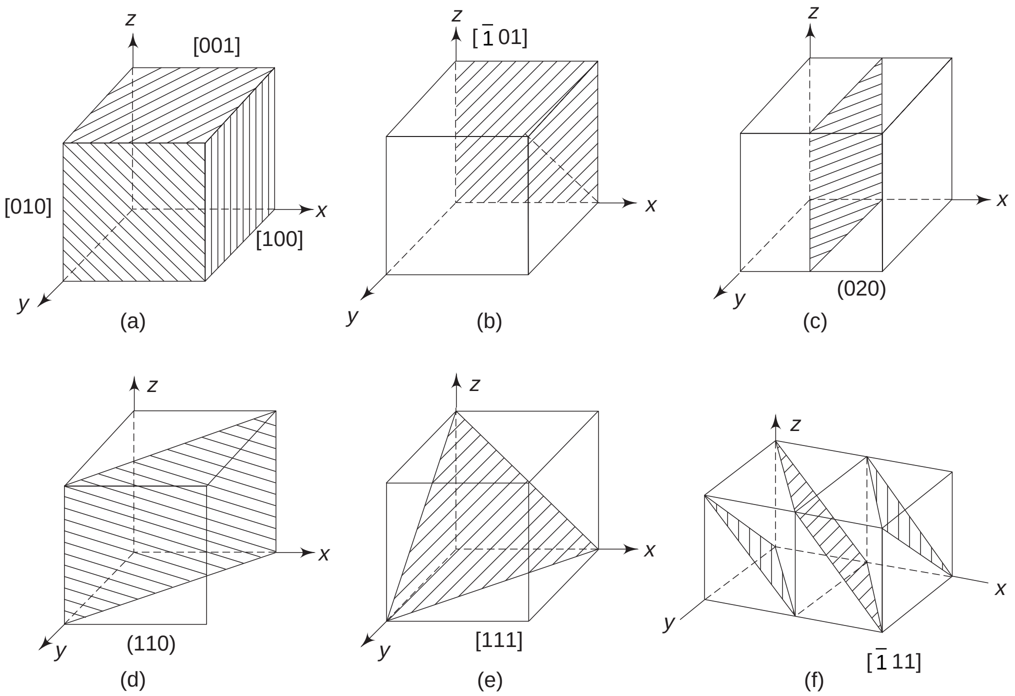
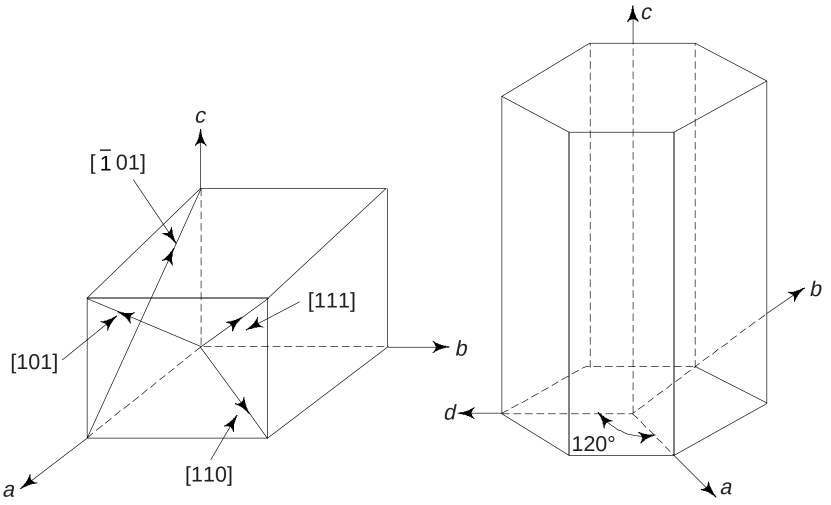


Fig. 3.27 Lattice planes in a cubic system. Negative intercepts are indicated on negative coordinates

The lattice or crystallographic direction can be defined as a line joining any two points of the lattice. Using a similar notation, we can describe the direction of a line in a lattice with respect to the unit vectors. We know that Miller indices of a direction are simply the vector components of the direction resolved along each of the co-ordinate axis, expressed as multiples of the unit cell parameters and reduced to their simplest form. They are denoted by [*hkl*] (to distinguish it from the (*hkl*) plane).

Just like the principal planes of importance, the directions with which one is mainly concerned are [110], [100] and [111]. We note that these are, respectively, a cube face diagonal, a cube edge and a body diagonal. We can label the families of directions by special brackets as are families of planes. Obviously, <100> denotes the family of directions which includes [100], [010], [001], [$\bar{1}$00], [0$\bar{1}$0] and [00$\bar{1}$]. Figure 3.27 (g) and (h) shows the Miller indices for directions: *(i)* cubic lattice system and *(ii)* Hexagonal lattice system.



(g) Cubic lattice system (h) Hexagonal system

Fig. 3.27 Miller indices for directions

(iv) Important Features of Miller Indices of Crystal Planes: A few important features of Miller indices of crystal planes are:

- *(a)* All the parallel equidistant planes have the same *Miller indices*. Obviously, Miller indices define a set of parallel planes.
- *(b)* A plane parallel to one of the coordinate axis has an intercept of infinity (∞) and therefore the Miller index is zero.
- *(c)* If the *Miller indices* of the two planes have the same ratio (i.e., say 844 and 422 or 211), then the planes are said to be parallel to each other. In other words, all equally spaced parallel planes with a particular orientation have the same index number [*hkl*].
- *(d)* Only the ratio of indices is important.
- *(e)* Miller indices of planes are denoted by (*hkl*), i.e., the plane cuts the axes into *h*, *k* and *l* equal segments. The directions in space are represented by square bracket [*xyz*].
- *(f)* The common inside brackets are used separately and not combined. Obviously, (111) is read as one-one-one and not 'one hundred eleven'.
- *(g)* Miller indices may also be negative and negative indices are represented by putting a bar over the digit, e.g. $(0\bar{1}0)$.

14. INTERPLANAR SPACINGS

These are the distances between planes and are represented by a number of parts of the body diagonal of a unit cell. In terms of Miller indices, the distance between planes can be calculated. Let us consider a plane with Miller indices (*hkl*). This has intercepts a/h, b/k and c/l on the three axes *x*, *y* and *z* respectively as shown in Fig. 3.28(a). If *d* is the length of the normal from the origin, *O* to the plane and α', β' and γ' are angles which the normal makes with the coordinates axes, considered orthogonal, then we have

$$d = \frac{a}{h}\cos\alpha' = \frac{b}{k}\cos\beta' = \frac{c}{l}\cos\gamma'$$

Since $\cos^2\alpha' + \cos^2\beta' + \cos^2\gamma' = 1$, we have

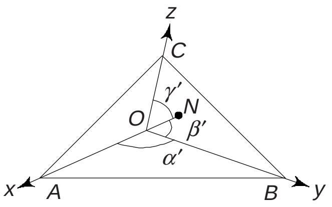


Fig 3.28(a) Calculation of interplanar spacings

$$d_{hkl} = \frac{1}{\sqrt{\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2}}} \qquad (4)$$

d_{hkl} gives the distance between two successive (hkl) places.
For a cubic system: $a = b = c$

$$\therefore \qquad d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \qquad (5)$$

For a cubic system lattice, the direction [hkl] is perpendicular to the plane (hkl).

There are three d_{111} interplanar spacings per long diagonal (body diagonal) of a unit cell in an FCC structure. We must note that relation (4) is valid for orthogonal lattices only. For non-orthogonal lattices, such an expression may not be obtained easily.

14(a). ANGLE BETWEEN TWO PLANES OR DIRECTIONS

Let us consider a cube having two planes *ABCD* and *EFCD* inclined at an angle θ with each other (Fig. 3.28(b)). Let h_1, k_1, l_1 are Miller indices of plane *ABCD* and h_2, k_2 and l_2 are Miller indices of plane *EFCD*.

The angle between these two planes is given by the relation

$$\cos\,\theta = \frac{h_1 h_2 + k_1 k_2 + l_1 l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2} \times \sqrt{h_2^2 + k_2^2 + l_2^2}} \qquad (6)$$

Similarly, the angle ϕ between the two directions having Miller indices (h_1, k_1, l_1) and (h_2, k_2, l_2) respectively is given by the relation

$$\cos\,\phi = \frac{h_1 h_2 + k_1 k_2 + l_1 l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2}\,\sqrt{h_2^2 + k_2^2 + l_2^2}} \qquad (7)$$

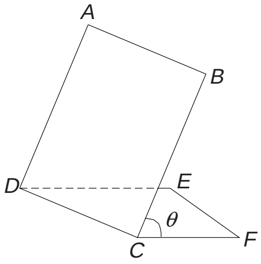
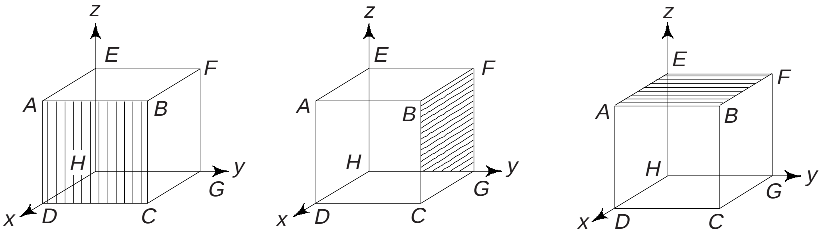


Fig. 3.28(b) Angle between two planes

15. REPRESENTATION OF CRYSTAL PLANES IN A CUBIC UNIT CELL

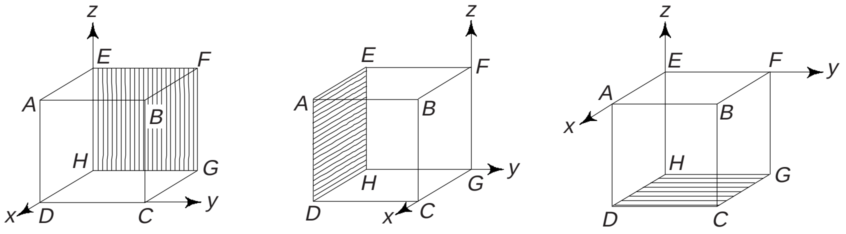
(100), (010) and (001) represent the Miller indices of the cubic planes *ABCD*, *BFGC* and *AEFB* respectively (Fig. 3.29).



(a) Plane (100) (b) Pane (010) (c) Plane (001)

Fig. 3.29 Crystal planes of a cubic unit cell

Obviously, the above mentioned three planes represent the three faces of the cubic unit cell. We can represent the other three faces of the cube by shifting the origin of the coordinate system to another corner of the unit cell, e.g. the plane *EFGH* may be represented by shifting the origin from point *H* to point *D*

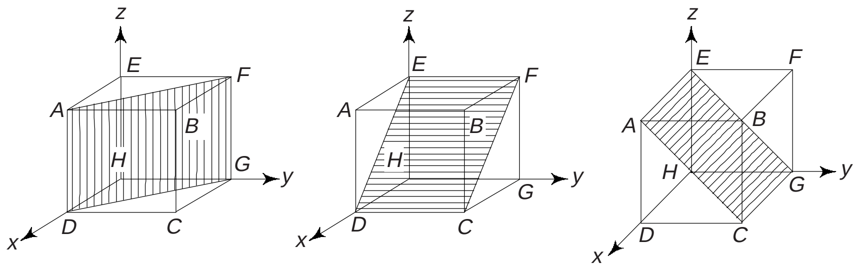


(a) Plane $(\bar{1}00)$ (b) Plane $(0\bar{1}0)$ (c) Plane $(00\bar{1})$

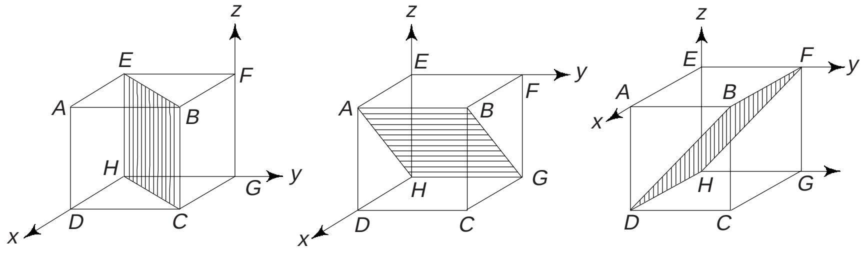
Fig. 3.30 Crystal planes obtained after shifting the origin

(Fig. 3.30(a)). We note that now the values of the intercepts are—1 along x-x axes, ∞ along y-y and z-z axes. Thus the Miller indices are $(\bar{1}00)$. Similarly, we obtain Miller indices for the planes *AEHD* and *DHGC* as $(0\bar{1}0)$ and $(00\bar{1})$ respectively by shifting the origin to the points *G* and *E* (Fig. 3.30(b) and (c)).

All the six faces of the cubic unit cell have same geometry, i.e. they are of the same form. Thus Miller indices of all the six planes are represented by {100}. This represents a set of six planes (100), (010), (011), $(\bar{1}00)$, $(0\bar{1}0)$ and $(00\bar{1})$. Similarly {110} represents a set of six planes (110), (101), (011), $(\bar{1}10)$, $(10\bar{1})$, and $(01\bar{1})$ (Fig. 3.31).



(a) Plane (110) (b) Plane (101) (c) Plane (011)

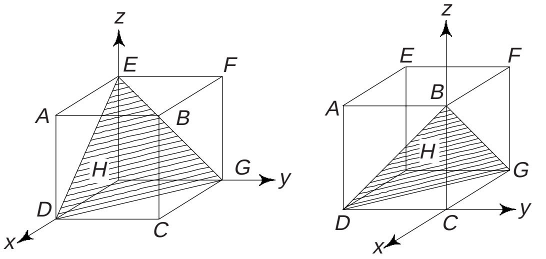


(d) Plane $(\bar{1}10)$ (e) Plane $(10\bar{1})$ (f) Plane $(01\bar{1})$

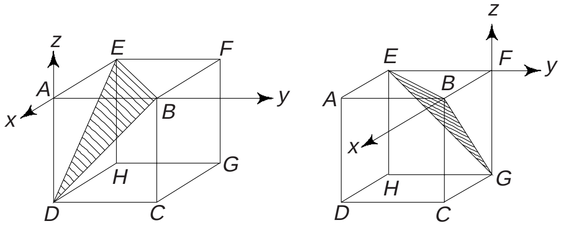
Fig. 3.31 Crystal planes corresponding to {110}

We must note that these planes represent diagonal planes, where each diagonal plane is formed by joining two opposite edges of the unit cell. There are 12 edges of a cubic unit cell and hence there are 6 diagonal faces represented by {110} (Fig. 3.31).

Similarly, {111} represents a set of four planes (111), $(\bar{1}\bar{1}1)$, $(\bar{1}1\bar{1})$ and $(1\bar{1}\bar{1})$, which are four triangular planes of the unit cubic cell (Fig. 3.32).



(a) Plane (111) (b) Plane $(\bar{1}\bar{1}1)$



(c) Plane $(\bar{1}1\bar{1})$ (d) Plane $(1\bar{1}\bar{1})$

Fig. 3.32 Crystal planes corresponding to {111}

16. SKETCHING THE PLANE FROM THE GIVEN MILLER INDICES

From a given Miller indices, one can sketch the planes. The procedure is as follows:

(i) First, take the reciprocal of the given Miller indices. These reciprocals represent the intercepts in terms of the axial units along x-x, y-y and z-z axes respectively. For example, if the given Miller indices are (211), then its reciprocals or intercepts will be 1/2, 1/1 and 1/1 or 0.5, 1, 1 respectively.

(ii) Now, we should sketch the plane with intercepts. Here with 0.5, 1, 1 along *x-x*, *y-y* and *z-z* axes respectively (Fig. 3.33).

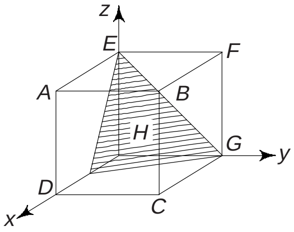
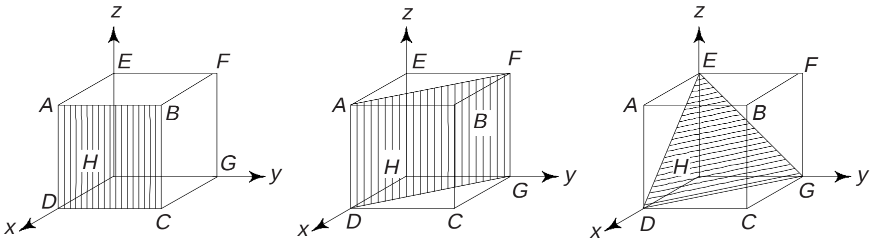


Fig. 3.33 Plane (211)

17. COMMON PLANES IN A SIMPLE CUBIC STRUCTURE

One can draw a number of planes in a simple cubic structure, but the most common planes are (100), (110) and (111). One can draw these common planes as follows:

(i) Plane of (100): In this case: $h = 1$, $k = 0$ and $l = 0$. The reciprocals of h, k and l are 1/1, 1/0 and 1/0 = 1, ∞, ∞. On sketching the plane with intercepts 1, ∞, ∞ along *x-x*, *y-y* and *z-z* axis respectively, we obtain the plane as shown in Fig. 3.34(a).



(a) Plane (100) (b) Plane (110) (c) Plane (111)

Fig. 3.34 Common planes in a simple cubic structure

(ii) Plane of (110): Here $h = 1$, $k = 1$ and $l = 0$. The reciprocals of h, k, l are 1/1, 1/1 and 1/0 = 1, 1, ∞. Now, sketching the plane with intercepts 1, 1 ∞, one obtains the plane as shown in Fig. 3.34(b).

(iii) Plane of (111): We have $h = 1$, $k = 1$ and $l = 1$. The reciprocals of h, k, l are 1/1, 1/1, 1/1 i.e. 1, 1, 1. Sketching the plane with intercepts 1, 1, 1, one obtains the plane as shown in Fig. 3.34(c).

18. CO-ORDINATION NUMBER

This is defined as the number of nearest atoms directly surrounding the given atom. The value of co-ordination number is 6 for simple cubic, 8 for BCC and 12 for FCC (Fig. 3.34d). In closely packed structures this number is 12. The coordination number of carbon is 4, i.e. number of nearest neighbours of carbon atom is 4. This low coordination, results in a relatively inefficient packing of the carbon atoms in the crystal.

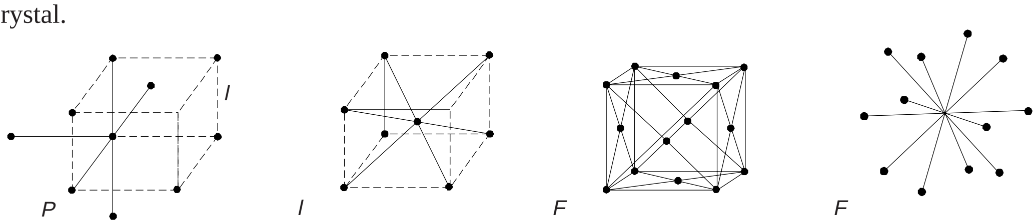


Fig. 3.34(d) **The coordination number in three cubic lattices**

Example 1 *Determine the relationship between the lattice parameter a and the atomic radius r for monoatomic, SC, BCC and FCC structures*

Solution In simple cubic (SC) structure, the atoms touch one another along the cube edges,

$$\therefore \qquad a = 2r$$

In BCC, the atoms touch along the body diagonals,

$$\therefore \qquad \sqrt{3}\, a = 4r$$

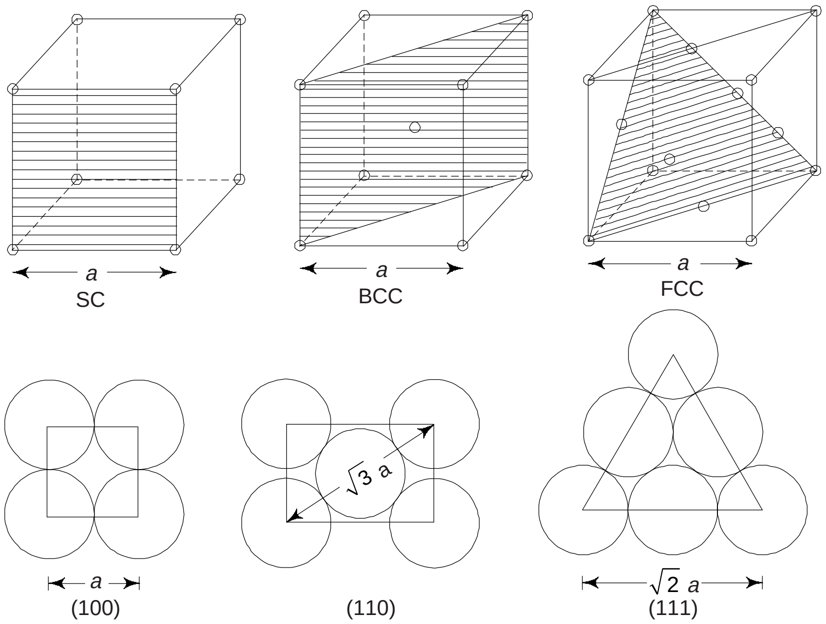


Fig. 3.35 **Monoatomic, SC, BCC and FCC structures along with their (100), (110) and (111) type planes respectively**

or
$$a = \frac{4r}{\sqrt{3}}$$

In FCC structure, the atoms touch along the face diagonals,

$$\therefore \qquad \sqrt{2}\, a = 4r$$

or
$$a = 2\sqrt{2}\, r.$$

Example 2 *An atomic plane in a crystal lattice makes intercept of 3a, 4b and 6c with the crystallographic axes where a, b and c are the dimensions of the unit cell. Show that the Miller indices of the atomic plane are (432).*

Solution In terms of the lattice constants, the intercepts are 3, 4 and 6. Their reciprocals are 1/3, 1/4, 1/6. On reducing to a common denominator, they become 4/12, 3/12 and 2/12. Obviously, the Miller indices of the plane are (432).

Example 3 *In a single cubic crystal find the ratio of the intercepts on the three axes by (123) plane.*

Solution The reciprocals of Miller indices are 1/1, 1/2 and 1/3. On reducing to a common denominator they become 6, 3 and 2. Intercepts on the three axes are $6a$, $3b$ and $2c$, where a, b and c are the lattice constants along the three axes.

Example 4 *Draw (101) and (111) planes in a unit cubic cell. Find the Miller indices of the directions which are common to both the planes.*

Solution Intercepts of the plane (101) with the three axes are 1/1, 1/0 and 1/1 i.e. 1, ∞ and 1. Similarly, the intercepts of the (111) with the three axes are 1/1, 1/1 and 1/1, i.e. 1, 1 and 1.

Now, taking the point O as origin and lines OA, OB and OC as the axes a, b and c respectively, the plane with the intercepts 1, ∞ and 1 is the plane $ADGC$ and that with intercepts 1, 1 and 1 is plane ABC (Fig. 3.36). From figure, it is obvious that the line common to both the planes is the line AC. This line corresponds to two directions, i.e., AC and CA.

Projections of the direction AC on the axes are –1, 0 and 1. Projections of direction CA on the axes are 1, 0 and –1.

Thus the required indices are $[\bar{1}01]$ and $[10\bar{1}]$.

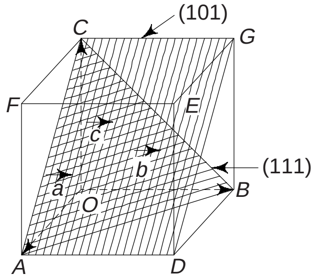
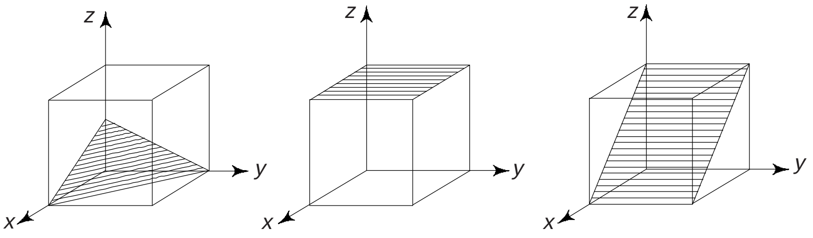


Fig. 3.36 Planes (101) and (111) in a simple cubic lattice

Example 5 *Draw the following planes and directions in the case of a FCC structure: (112), (001) and (101).*

Solution *(i) Plane and direction (112):* In this, we have $h = 1$, $k = 1$ and $l = 2$. The reciprocals of h, k and l are 1/1, 1/1, 1/2, i.e. 1, 1, 0.5.

Now, we can sketch the plane with intercepts 1, 1, 0.5 along x-x, y-y and z-z axes respectively (Fig. 3.37(a).



(a) Plane (112) (b) Plane (001) (c) Plane (101)

Fig. 3.37

(ii) Plane of (001): Here $h = 0$, $k = 0$, $l = 1$. The reciprocals are 1/0, 1/0, 1/1 i.e., ∞, ∞, 1. We obtain the sketch of the plane as shown in Fig. 3.37(b).

(iii) Plane of (101): Here $h = 1$, $k = 0$ and $l = 1$. The reciprocals are 1/1, 1/0 and 1/1, i.e. 1, ∞ and 1. We obtain the sketch as shown in Fig. 3.37(c).

Example 6 *A plane makes intercepts of 1, 2 and 0.5 Å on the crystallographic axes of an orthorhombic crystal with a : b : c = 3 : 2 : 1. Find the Miller indices of this plane.*

Solution Taking the lengths of the axes *OA*, *OB* and *OC* as 3, 2 and 1Å respectively, we obtain the plane with intercepts of 1, 2 and 0.5 Å on the axes is the plane *DBE* (Fig. 3.38). The intercepts of this plane relative to full lengths of the axes are: 1/3, 2/2 and 0.5/1 or 1/3, 1 and 1/2. Reciprocals are: 3, 1 and 2. Obviously, the Miller indices of the plane DBE are (312).

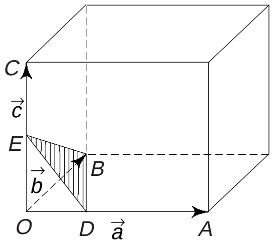


Fig. 3.38 **(312) plane in an orthorhombic lattice**

Example 7 *Sketch (111) plane of a unit cell of simple tetragonal crystal. Given c/a = 0.62.*

Solution Fig. 3.39 shows the plane (111). The plane cuts the three axes at unit distances. We note that the unit distance along *z*-axis is shorter than the unit distance of other two axes.

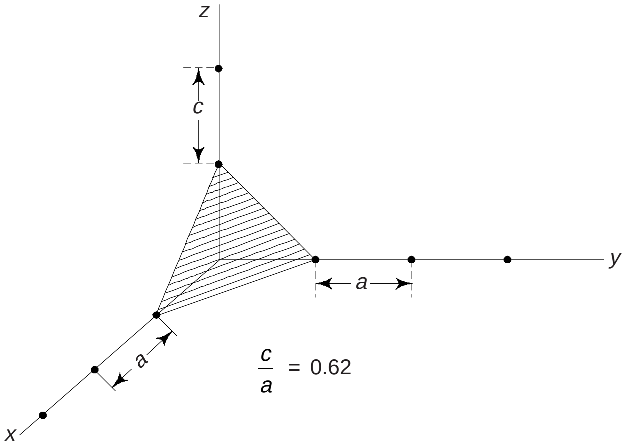


Fig. 3.39 **The plane (111) *z*-axis shorter**

Example 8 *Draw the planes (020), (120) and (220) in a FCC structure.*

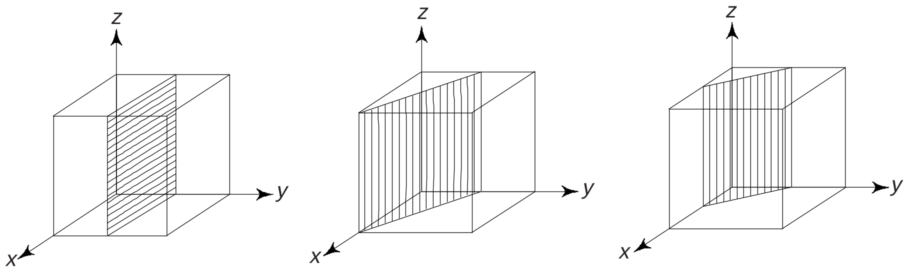
Solution *(i) Plane of (020):* Given $h = 0$, $k = 2$ and $l = 0$. The reciprocals are 1/0, 1/2, 1/0, i.e. ∞, 0.5, ∞.

Sketching the plane with intercepts ∞, 0.5, ∞ along *x-x*, *y-y* and *z-z* axes respectively, we obtain the plane (020) (Fig. 3.40(a)).

(ii) Plane of (120): Here $h = 1$, $k = 2$, and $l = 0$. The reciprocals of h, k, l are 1/1, 1/2, 1/0, i.e. 1, 0.5, ∞.

Sketching the plane, we obtain as shown in Fig. 3.40(b).

(iii) Plane of (220): Here $h = 2$, $k = 2$, and $l = 0$. The reciprocals of h, k, l are 1/2, 1/2, 1/0, i.e. 0.5, 0.5, ∞. The plane obtained is as shown in Fig. 3.40(c).



(a) Plane (020) (b) Plane (120) (c) Plane (220)

Fig. 3.40

Example 9 *Draw the planes and directions of FCC structures (321), (102), (201) and (111).*

[B.E. 2001]

Solution *(i) Plane and Direction of (321):* Here $h = 3$, $k = 2$, and $l = 1$. The reciprocals are 1/3, 1/2, 1/1, i.e. 0.3, 0.5, 1. The sketch of the plane with intercepts 0.3, 0.5 and 1 is as shown in Fig. 3.41(a). A line drawn normal to this sketched plane and passing through the origin gives the required direction.

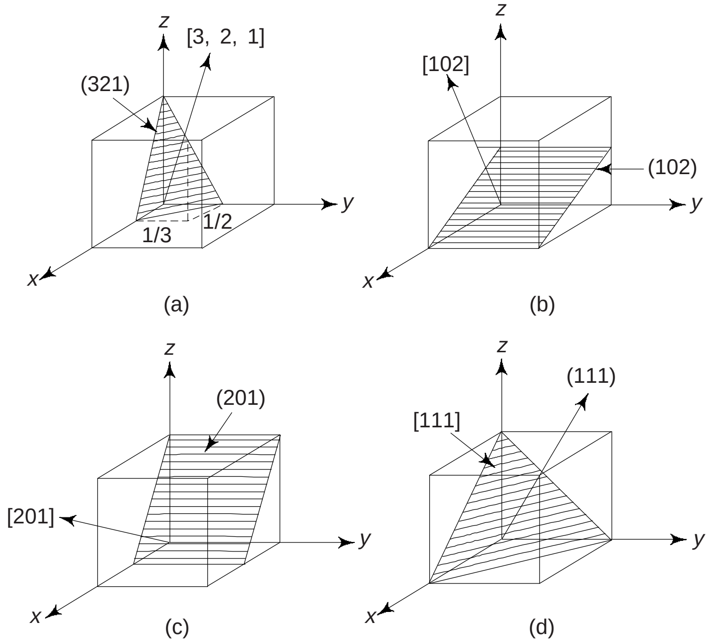


Fig. 3.41

(ii) Plane and Direction of (102): Here $h = 1$, $k = 0$, and $l = 2$. The reciprocals of these are: 1/1, 1/0 and 1/2, i.e. 1, ∞, 0.5. The sketch of the plane with these intercepts is as shown in Fig. 3.41(b). A line drawn normal to this sketched plane and passing through the origin gives the required direction.

(iii) Plane and Direction of (011): Here $h = 0$, $k = 1$, and $l = 1$. The reciprocals of h and k, l are: 1/0, 1/1, 1/1, i.e. ∞, 1.1. The sketch of the plane with these intercepts is shown in Fig. 3.41(c). A line drawn normal to this sketched plane and passing through the origin gives the required direction.

(iv) Plane and Direction of (001): Here $h = 0$, $k = 0$, and $l = 1$. The reciprocals of h, k and l are: 1/0, 1/0, 1/1, i.e. ∞, ∞, 1. The sketch of the plane with these intercepts is shown in Fig. 3.41(d). A line drawn normal to this sketched plane and passing through the origin gives the required direction.

Example 10 *In a cubic unit cell, find the angle between normals to the planes (111) and (121).*

Solution Since the given crystal is cubic, the normals to the planes (111) and (121) are the directions [111] and [121] respectively. If θ be the angle between the normals, then

$$\cos\theta = \frac{h_1 h_2 + k_1 k_2 + l_1 l_2}{\left(h_1^2 + k_1^2 + l_1^2\right)^{1/2} \left(h_2^2 + k_2^2 + l_2^2\right)^{1/2}}$$

$$= \frac{1\times 1 + 1\times 2 + 1\times 1}{(1^2+1^2+1^2)^{1/2}\,(1^2+2^2+1^2)^{1/2}}$$

$$= 0.9428$$

$$\therefore \qquad \theta = 19.47° \quad \text{or} \quad 19° \; 28'$$

Example11 *Determine the packing efficiency and density of sodium chloride from the following data: (i) radius of the sodium ion = 0.98 Å, (ii) radius of chlorine ion = 1.81 Å (iii) atomic mass of sodium = 22.99 amu and atomic mass of chlorine = 35.45 amu.*

Solution The unit cell structure of NaCl is shown in Fig. 3.18. We can see that the Na^+ and Cl^- ions touch along the cube edges.

$$\therefore \qquad \text{Lattice parameter, } a = 2\,(\text{radius of } Na^+ + \text{radius of } Cl^-)$$

$$= 2(0.98 + 1.81) = 5.58 \text{ Å}$$

$$\text{Atomic Packing Fraction} = \frac{\text{Volume of ions present in the unit cell}}{\text{Volume of the unit cell}}$$

$$= \frac{4(4/3)\pi r_{Na^+}^3 + 4\left(\frac{4}{3}\right)\pi r_{Cl^-}^3}{a^3}$$

$$= \frac{16\pi}{3}\left[\frac{(0.98)^3 + (1.81)^3}{(5.58)^3}\right]$$

$$= 0.663 \text{ or } 66.3\%$$

$$\text{Density} = \frac{\text{Mass of the unit cell}}{\text{Volume of the unit cell}}$$

$$= \frac{4(22.99 + 35.45)\times 1.66\times 10^{-27}}{(5.58\times 10^{-10})^3} \text{ kg/m}^3$$

$$= 2234 \text{ kg/m}^3 \text{ or } 2.23 \text{ gm/cm}^3$$

Example 12 *Aluminium has FCC structure. Its density is 2700 kg/m³. Find the unit cell dimensions and atomic diameter. Given at. weight of Al = 26.98.* [Roorkee]

Solution

$$\text{Density} = \frac{nm}{a^3}\,N_A$$

$$= 2700 \text{ kg/m}^3 = 2.7 \text{ gm/cm}^3$$

$$= \frac{4\times 26.98}{a^3\times 6.023\times 10^{23}}$$

$$\therefore \qquad a^3 = \frac{4\times 26.98}{2.7\times 6.023\times 10^{23}} = 6.6\times 10^{-23}/\text{cm}^3$$

$$\therefore \qquad a = 4.048\times 10^{-10} \text{ m} = 4.048 \text{ Å}$$

For FCP structure, $r = \dfrac{a}{2 \times 1.414} = \dfrac{4.048\,\text{Å}}{2 \times 1.414}$ = 1.43 Å

$\therefore$ Diameter = $2r$ = 2.86 Å

Example 13 *Find the interplanar distance of (200) plane and (111) plane of Nickel crystal. The radius of Nickel atom is 1.245 Å.* [Jodhpur]

Solution Nickel has FCC structure. Given radius of Nickel = r = 1.245 Å

$$\text{Lattice constant} = a = \frac{4r}{\sqrt{2}} = \frac{4 \times 1.245}{\sqrt{2}} = 3.52\ \text{Å}$$

$$\therefore \quad d_{200} = \frac{3.52}{\sqrt{2^2 + 0^2 + 0^2)}} = 1.76\ \text{Å}$$

$$d_{111} = \frac{3.52}{\sqrt{1^2 + 1^2 + 1^2}} = 2.03\ \text{Å}$$

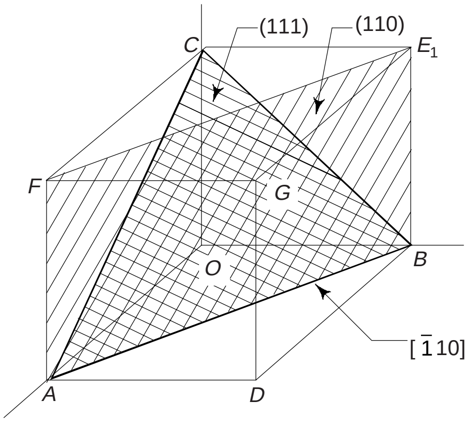


Fig. 3.42 Plane in a unit cell of FCC Nickel

Example 14 *The lattice constant of a unit cell of KCl crystal is 3.03 Å. Find the number of atoms/ mm^2 of planes (100), (110) and (111). KCl has simple cubic structure.* [B.E]

Solution a = 3.03 Å = 3.03×10^{-7} mm.

(100) plane The number of atoms in the (100) plane of a simple cubic structure

$$= \frac{1}{a^2} = \frac{1}{(3.03 \times 10^{-7})^2} = 10.9 \times 10^{12}$$

(110) plane The number of atoms in (110) plane of a simple cubic structure

$$= \frac{0.707}{a^2} = \frac{0.707}{(3.03 \times 10^{-7})^2} = 7.7 \times 10^{12}$$

(111) plane The number of atoms in (111) plane of a simple cubic structure

$$= \frac{0.58}{a^2} = \frac{0.58}{(3.03 \times 10^{-7})^2}$$

$$= 6.3 \times 10^{12}$$

Example 15 *Determine the planar density of Ni (FCC structure) in (100) plane. Given, the radius of Ni atom = 1.245 Å.*

Solution From Fig. 3.43, we have
Number of atoms in (100) plane

$$= 1 + \frac{1}{4} \times 4 = 2$$

Area of plane = a^2,

$$= a = \frac{4r}{\sqrt{2}} = 3.52\ \text{Å}$$

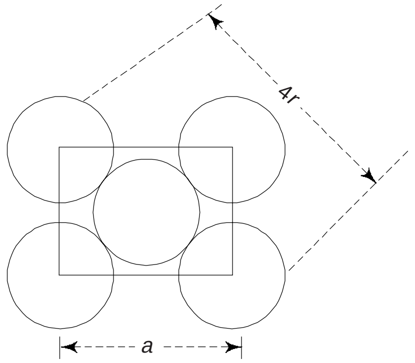


Fig. 3.43 Interplanar distances in Ni crystal

$\therefore$ Planar density $= \dfrac{2}{(3.52 \times 10^{-7}\ \text{mm}^2)} = 16.1 \times 10^{12}$ atoms/mm^2

Example 16 *Calculate the planar atomic densities of planes (100), (110) and (111) in FCC unit cell and apply your result for lead (FCC form).* [AMIE]

Solution (i) *Plane (100)*: Fig. 3.44(a) shows an FCC unit cell with planes (111) and (111). Figure 3.44(b) shows the plane (100) with atoms on it. Similarly Figs. 3.44(c) and (d) shows the planes (110) and (111) with atoms contained in them respectively.

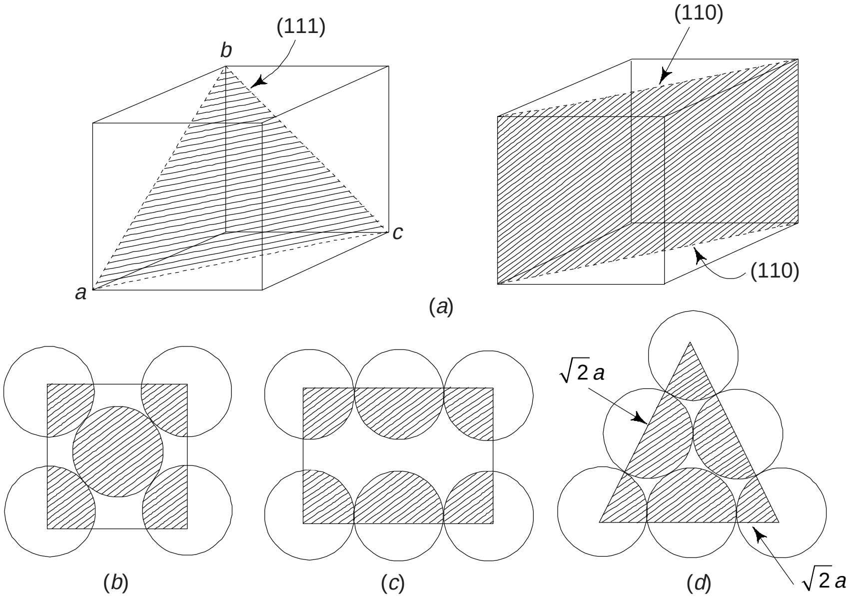


Fig. 3.44 Distribution of atoms in planes (100), (110) and (111) in FCC unit cell

Number of atoms contained in (100) plane is $4 \times \dfrac{1}{4} + 1 = 2$

Let a be the edge of the unit cell and r the radius of the atom, then

$$a = 2\sqrt{2}\ r$$

$\therefore$ Planar density of plane (100)

$$= \frac{2}{4 \times 2r^2} = \frac{0.25}{r^2}$$

The radius of lead atom is 1.75 Å. The planar density of (100) plane of lead

$$= \frac{0.25}{(1.75 \times 10^{-7})^2} = 8.2 \times 10^{12} \text{ atoms/mm}^2$$

$$= 8.2 \times 10^{18} \text{ atoms/m}^2$$

(ii) Plane (110): From Fig. 3.44(c), we have the number of atoms contained in plane (110)

$$= 4 \times \frac{1}{4} + 2 \times \frac{1}{2} = 2$$

The top edge of the plane (110) is $4r$, whereas the vertical edge = $a = 2\sqrt{2}\, r$. Thus the planar density of (110)

$$= \frac{2}{8\sqrt{2r^2}} = 0.177/r^2$$

In case of (110) plane in lead, we have planar density = $\dfrac{0.177}{(1.75 \times 10^{-2})^2}$

(iii) From Fig. 3.44(d), we have the number of atoms contained in the plane (111)

$$= 3 \times \frac{1}{6} + \frac{3}{2} = 2$$

Area of (111) plane = $\dfrac{1}{2}\sqrt{\dfrac{3}{2}}\, a\sqrt{2}\, a = 4\sqrt{3r^2}$

∴ Planar density of (111) = $\dfrac{2}{4\sqrt{3r^2}} = \dfrac{0.29}{r^2}$

For lead crystal, we obtain the value 9.5×10^{12} atoms/mm^2.

Example 17 *Determine the linear atomic density in the [110] and [111] directions of copper crystal lattice. Lattice constant of copper (FCC) is 3.61* $\times 10^{-10}$ *m.* [AMIE]

Solution From Fig. 3.45 and earlier discussions. We note that the face diagonal along [110] direction intersects two half diameters and one full diameters. Thus the number of diameters of atom along [110] direction

$$= \frac{1}{2} + 1 + \frac{1}{2} = 2$$

Let a be the lattice constant, then the length of the face diagonal = $\sqrt{2}\, a$

∴ Linear density of [110] within unit cell

$$= \frac{2}{\sqrt{2}a} = \frac{\sqrt{2}}{a}$$

Since $a = 3.61 \times 10^{-10}$ m $= 3.61 \times 10^{-7}$ mm

$$\therefore \quad \rho_{110} = \frac{\sqrt{2}}{3.61 \times 10^{-7}} = 3.92 \times 10^{6} \text{ atoms/mm}$$

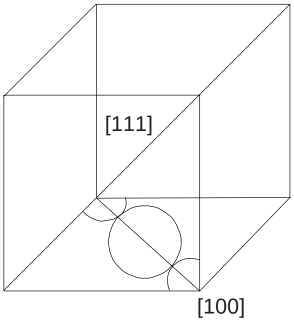


Fig. 3.45

The direction [111] is along the body diagonal. From Fig. 3.45, the length of the diagonal along [111]

$$= \sqrt{\left(\sqrt{2}a\right)^2 + a^2} = \sqrt{3}a$$

The number of atomic diameters intersected by diagonal along [111] is $\dfrac{1}{2} + \dfrac{1}{2}$ at two ends. Thus the linear density along [111] within the crystal unit cell

$$= \frac{1}{\sqrt{3}\,a} = \frac{1}{\sqrt{3} \times 3.61 \times 10^{-7}} = 1.6 \times 10^{6} \text{ atoms/mm}$$

Obviously, in FCC the linear density along [110] direction is greater than that along [111] direction.

Example 18 *The density of $\alpha - Fe$ is 7.87 $\times 10^3$ kg/m^3. Atomic weight of Fe is 55.8. If $\alpha - Fe$ crystallizes in BCC space lattice, find lattice constant. Given Avogadro's number (N) = 6.02 $\times 10^{26}$/kg/mole.*

[AMIE]

Solution Lattice constant *(a)* can be obtained from the relation,

$$a^3 = \frac{An}{N\rho} = \frac{55.8 \times 2}{6.02 \times 10^{26} \times 7.87 \times 10^3} = 2.355 \times 10^{-29}$$

$$\therefore \quad a = (2.355 \times 10^{-29})^{1/3} = 2.866 \times 10^{-10} \text{ m} = 2.866 \text{ Å}$$

Example 19 *Show that the number of atoms per unit cell of a metal having a lattice parameter of 2.9 Å and density of 7.87 gm/cc is 2. Given atomic weight of metal = 55.85 and N = 6.023 $\times 10^{23}$*

Solution

$$a^3 = \frac{An}{N\rho}$$

or

$$(2.9 \times 10^{-3})^3 = \frac{55.85\text{n}}{6.023 \times 10^{23} \times 7.87} = 1.18 \times 10^{-23} \text{ n}$$

$$\therefore \quad n = \frac{(2.9 \times 10^{-8})^3}{1.18 \times 10^{-23}} = 2$$

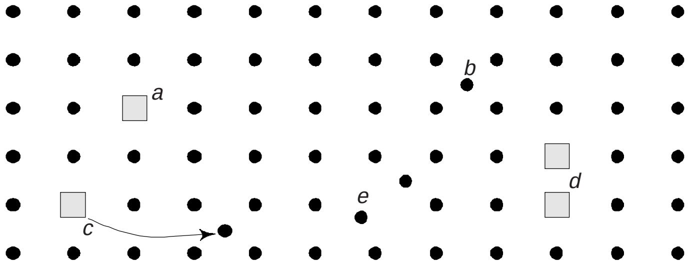
19. DEFECTS OR IMPERFECTIONS IN CRYSTALS

Up to now, we have described perfectly regular crystal structures, called *ideal* crystals and obtained by combining a basis with an infinite space lattice. In ideal crystals atoms were arranged in a regular way. However, the structure of real crystals differs from that of ideal ones. Real crystals always have certain defects or imperfections, and therefore, the arrangement of atoms in the volume of a crystal is far from being perfectly regular.

Natural crystals always contain defects, often in abundance, due to the uncontrolled conditions under which they were formed. The presence of defects which affect the colour can make these crystals valuable as gems, as in ruby (chromium replacing a small fraction of the aluminium in aluminium oxide : Al_2O_3). Crystal prepared in laboratory will also always contain defects, although considerable control may be exercised over their type, concentration, and distribution.

The importance of defects depends upon the material, type of defect, and properties which are being considered. Some properties, such as density and elastic constants, are proportional to the concentration of defects, and so a small defect concentration will have a very small effect on these. Other properties, e.g. the colour of an insulating crystal or the conductivity of a semiconductor crystal, may be much more sensitive to the presence of small number of defects. Indeed, while the term defect carries with it the connotation of undesirable qualities, defects are responsible for many of the important properties of materials and much of material science involves the study and engineering of defects so that solids will have desired properties. A defect free, i.e. ideal silicon crystal would be of little use in modern electronics; the use of silicon in electronic devices is dependent upon small concentrations of chemical impurities such as phosphorus and arsenic which give it desired properties. Some simple defects in a lattice are shown in Fig. 3.46.

There are some properties of materials such as stiffness, density and electrical conductivity which are termed structure—insensitive, are not affected by the presence of defects in crystals while there are many properties of greatest technical importance such as mechanical strength, ductility, crystal growth, magnetic



key:
a = vacancy (Schottky defect)
b = interstitial
c = vacancy—interstitial pair (Frenkel defect)
d = divacancy
e = split interstitial
☐ = vacant site

Fig. 3.46 Some simple defects in a lattice

hysteresis, dielectric strength, condition in semiconductors, which are termed structure sensitive are greatly affected by the relatively minor changes in crystal structure caused by defects or imperfections. Crystalline defects can be classified on the basis of their geometry as follows:

(i) Point imperfections
(ii) Line imperfections
(iii) Surface and grain boundary imperfections
(iv) Volume imperfections

The dimensions of a point defect are close to those of an interatomic space. With linear defects, their length is several orders of magnitude greater than the width. Surface defects have a small depth, while their width and length may be several orders larger. Volume defects (pores and cracks) may have substantial dimensions in all measurements, i.e. at least a few tens of Å. We will discuss only the first three crystalline imperfections.

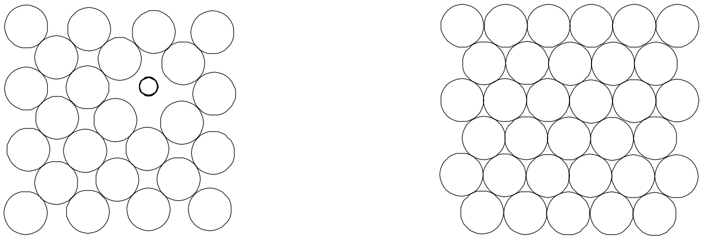
20. POINT IMPERFECTIONS

The point imperfections, which are lattice errors at isolated lattice points, take place due to imperfect packing of atoms during crystallisation. The point imperfections also take place due to vibrations of atoms at high temperatures. Point imperfections are completely local in effect, e.g. a vacant lattice site. Point defects are always present in crystals and their present results in a decrease in the free energy. One can compute the number of defects at equilibrium concentration at a certain temperature as,

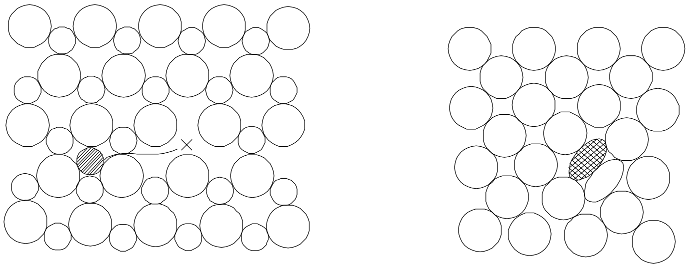
$$n = N \exp\left[\frac{-E_d}{kT}\right] \qquad (8)$$

Where $n \rightarrow$ number of imperfections, $N \rightarrow$ number of atomic sites per mole, $k \rightarrow$ Boltzmann constant, $E_d \rightarrow$ free energy required to form the defect and $T \rightarrow$ absolute temperature. E is typically of order l eV; since $k = 8.62 \times 10^{-5}$ eV/K, at $T = 1000$ K, $n/N = \exp[-1/(8.62 \times 10^{-5} \times 1000)] \simeq 10^{-5}$, or 10 parts per million. For many purposes, this fraction would be intolerably large, although this number may be reduced by slowly cooling the sample.

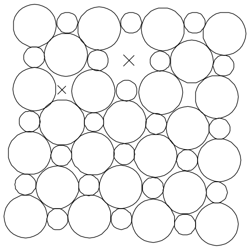
(i) Vacancies: The simplest point defect is a vacancy. This refers to an empty (unoccupied) site of a crystal lattice, i.e. a missing atom or vacant atomic site [Fig. 3.47(a)] such defects may arise either from imperfect packing during original crystallisation or from thermal vibrations of the atoms at higher temperatures. In the latter case, when the thermal energy due to vibration is increased, there is always an increased probability that individual atoms will jump out of their positions of lowest energy. Each temperature has a



(a) Vacancy defect (b) Interstitial defect



(c) Frenkel defect (d) Substitutional defect



(e) Schottky defect

Fig. 3.47 Point defects in a crystal lattice

corresponding equilibrium concentration of vacancies and interstitial atoms (an interstitial atom is an atom transferred from a site into an interstitial position). For instance, copper can contain 10^{-13} atomic percentage of vacancies at a temperature of 20–25°C and as many as 0.01% at near the melting point (one vacancy per 10^4 atoms). For most crystals the said thermal energy is of the order of l eV per vacancy. The thermal vibrations of atoms increases with the rise in temperature. The vacancies may be single or two or more of them may condense into a di-vacancy or trivacancy. We must note that the atoms surrounding a vacancy tend to be closer together, thereby distorting the lattice planes. At thermal equilibrium, vacancies exist in a certain proportion in a crystal and thereby leading to an increase in randomness of the structure. At higher temperatures, vacancies have a higher concentration and can move from one site to another more frequently. Vacancies are the most important kind of point defects; they accelerate all processes associated with displacements of atoms: diffusion, powder sintering, etc.

(ii) Interstitial Imperfections: In a closed packed structure of atoms in a crystal if the atomic packing factor is low, an extra atom may be lodged within the crystal structure. This is known as *interstitial position*, i.e. *voids*. An extra atom can enter the interstitial space or void between the regularly positioned atoms only when it is substantially smaller than the parent atoms [Fig. 3.47(b)], otherwise it will produce atomic distortion. The defect caused is known as interstitial defect. In close packed structures, e.g. FCC and HCP, the largest size of an atom that can fit in the interstitial void or space have a radius about 22.5% of the radii of parent atoms. Interstitialcies may also be single interstitial, di-interstitials, and tri-interstitials. We must note that vacancy and interstitialcy are inverse phenomena.

(iii) Frenkel Defect: Whenever a missing atom, which is responsible for vacancy occupies an interstitial site (responsible for interstitial defect) as shown in Fig. 3.47(c), the defect caused is known as Frenkel defect. Obviously, Frenkel defect is a combination of vacancy and interstitial defects. These defects are less in number because energy is required to force an ion into new position. This type of imperfection is more common in ionic crystals, because the positive ions, being smaller in size, get lodged easily in the interstitial positions.

(iv) Schottky Defect: These imperfections are similar to vacancies. This defect is caused, whenever a pair of positive and negative ions is missing from a crystal [Fig. 3.47(e)]. This type of imperfection maintains a charge neutrality. Closed-packed structures have fewer interstitialcies and Frenkel defects than vacancies and Schottky defects, as additional energy is required to force the atoms in their new positions.

(v) Substitutional Defect: Whenever a foreign atom replaces the parent atom of the lattice and thus occupies the position of parent atom (Fig. 3.47(d)], the defect caused is called substitutional defect. In this type of defect, the atom which replaces the parent atom may be of same size or slightly smaller or greater than that of parent atom.

(vi) Phonon: When the temperature is raised, thermal vibrations takes place. This results in the defect of a symmetry and deviation in shape of atoms. This defect has much effect on the magnetic and electric properties.

All kinds of point defects distort the crystal lattice and have a certain influence on the physical properties. In commercially pure metals, point defects, increase the electric resistance and have almost no effect on the mechanical properties. Only at high concentrations of defects in irradiated metals, the ductility and other properties are reduced noticeably.

In addition to point defects created by thermal fluctuations, point defects may also be created by other means. One method of producing an excess number of point defects at a given temperature is by quenching (quick cooling) from a higher temperature. Another method of creating excess defects is by severe deformation of the crystal lattice, e.g., by hammering or rolling. We must note that the lattice still retains its general crystalline nature, numerous defects are introduced. There is also a method of creating excess point defects is by external bombardment by atoms or high energy particles, e.g. from the beam of the cyclotron or the neutrons in a nuclear reactor. The first particle collides with the lattice atoms and displaces them, thereby causing a point defect. The number of point defects created in this manner depends only upon the nature of the crystal and on the bombarding particles and not on the temperature.

21. LINE DEFECTS OR DISLOCATIONS

Line imperfections are called dislocations. A linear disturbance, i.e. one dimensional imperfections in the geometrical sense of the atomic arrangement, which can very easily occur on the slip plane through the crystal, is known as dislocation. The most important kinds of linear defects are *edge* and *screw dislocation.* Both these types are formed in the process of their deformation. Both these defects are the most striking imperfections and are responsible for the useful property of ductility in metals, ceramics and crystalline polymers.

(i) Edge Dislocation: This type of dislocation is formed by adding an extra partial plane of atoms to the crystal [Fig. 3.48(a)]. An edge dislocation in its cross-section is essentially the edge of an 'extra' half-plane in the crystal lattice. The lattice around dislocation is elastically distorted.

Figure 3.49(a) shows a cross-section of a crystal where atoms (shown by dots) arranged in a perfect orderly manner. When an extra half plane is inserted from the top, the displacement of atoms is shown in Fig. 3.49(b). We note from Fig. 3.49(b) that top and bottom of the crystal above and below the line XY appears perfect. When the extra half plane is inserted from the top, the defects so produced is represented by $\perp$ (inverted tee) and if the extra half plane is inserted from the bottom, the defects so produced is represented by T (Tee).

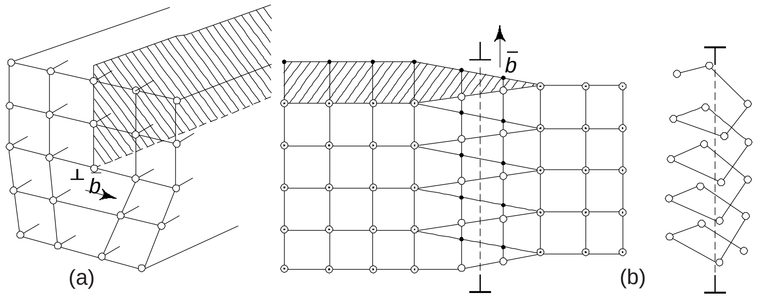


Fig. 3.48 Schemes of (a) edge dislocation (b) screw dislocation

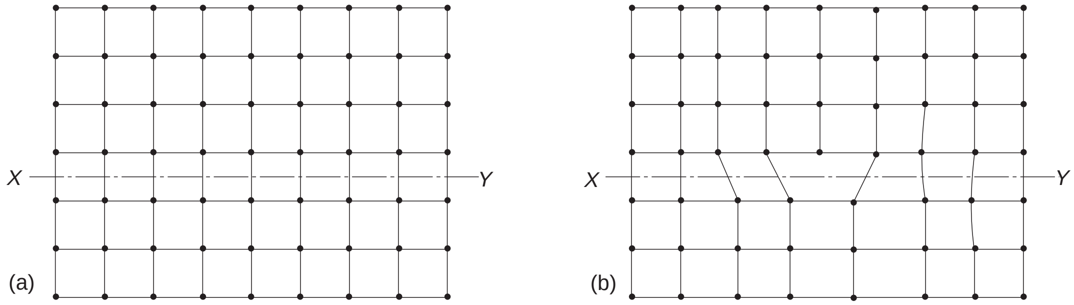


Fig. 3.49 Edge dislocation caused by an extra partial plane of atoms in the crystal

Near the dislocation, the distortion in the crystal is due to the presence of zones of compression and tension in the crystal lattice. The lattice above the line of dislocation is in a state of compression, whereas below this line, it is in tension. We must note that the dislocation line is a region of higher energy than the rest of the crystal.

The criterion of distortion is what is called the *Burgers vector*. It can be determined if a closed contour is drawn around a zone in an ideal crystal by passing from one site to another as shown in Fig. 3.50(a) and then the procedure is repeated a zone in a real crystal containing a dislocation. As may be seen from Fig. 3.50(b), the contour described in real crystal turns out to be unclosed. The vector required for the closing the contour is the Burgers vector. The Burgers vector of an edge dislocation is equal to the interatomic space and perpendicular to the dislocation line.

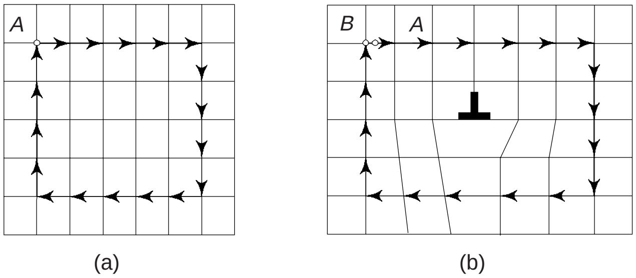
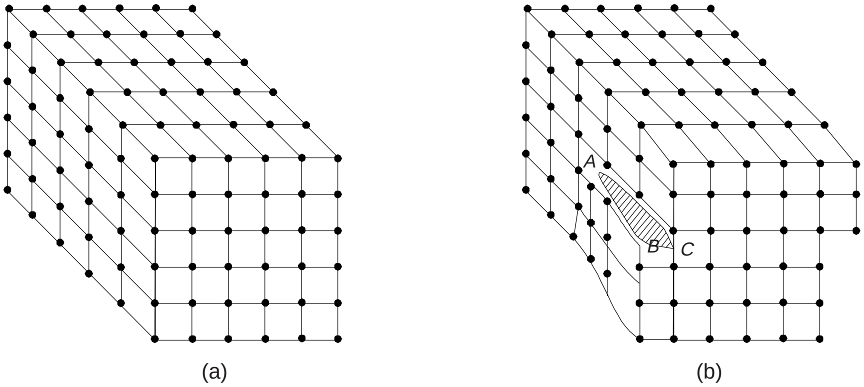


Fig. 3.50 Determination of Burgers vector (*BA*)

(ii) Screw Dislocation: The formation of screw dislocation is shown in Fig. 3.48(b). In this, the atoms are displaced in two separate planes perpendicular to each other. An isomeric view of a perfect crystal is shown in Fig. 3.51(a). The displacement of atoms in the region *ABC* is shown in Fig. 3.51(b). The arrangement of atoms in screw dislocations appear like that of a screw or a helical surface.



(a) (b)

Fig. 3.51 Scheme of screw dislocation

A screw dislocation does not exhibit climb motion.

The following effects of screw dislocation are of great importance.

(i) Probably, the force required to form and move a screw dislocation is somewhat greater than that required to initiate an edge dislocation.

(ii) Without breaking the continuity of the lattice, the plastic deformation is possible under low stress.

(iii) Screw dislocation causes distortion of the crystal lattice for a considerable distance from the centre of the line and takes the form of spiral distortion of the planes. We must remember that the dislocations of both types, i.e., combinations of edge and screw, are closely associated with the crystallization as well as deformation.

The total length of all dislocation lines in a unit of volume is called the *dislocation density*. It may be equal to $10^4 - 10^5$ cm^{-2} in semiconductor crystals and $10^6 - 10^8$ cm^{-2} in annealed metals. After cold plastic deformation, the dislocation density may rise up to 10^{11}–10^{12} cm^{-2}. Attempts to raise the dislocation density above 10^{12} cm^{-2} end quickly in cracking and failure of the metal. Dislocation appear on crystallization; their density may then be quite high and they influence substantially the properties of materials. Along with other defects, dislocations participate in phase transformations and recrystallization and may serve as nuclei for precipitation of a secondary phase from solid solution. The rate of diffusion along dislocation lines is several orders of magnitude greater than that through a crystal lattice without defects. Dislocations serve as places for concentration of impurity atoms, especially of interstitial impurities, since this decreases lattice distortions. Impurity atoms can concentrate around dislocations and form the *Cottrell atmospheres* which impede dislocation movement and strengthen the metal.

The effect of dislocations is especially pronounced on the strength of crystals. The experimentally measured yield strength of metals turns out to be the only one-thousandth of its theoretical value, the loss being mainly attributed to the effect of mobile dislocations. By increasing substantially the dislocation density and decreasing the dislocation mobility, the strength of a metal can be raised several times compared with its strength in the annealed state. Faultless pieces of metals (in particular, long and thin 'whiskers' obtained by crystallization from the gaseous phase) exhibit a strength approaching the theoretical value (Fig. 3.52).

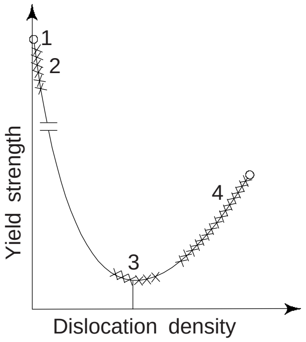


Fig. 3.52 Effect of dislocation density on yield strength: 1–a perfect faultless crystal; 2–faultless 'whisker' crystals; 3–annealed metals; 4–metals with an elevated stacking fault density after treatment.

Dislocations have certain geometrical characteristics, e.g. (i) The vectorial sum of the Burgers vectors of dislocations meeting at a point called the node must be zero, analogous to Kirchhoff's law for electrical currents meeting at a junction and (ii) A dislocation line cannot end abruptly within the crystal. It either ends at a node or at the surface. Alternatively, a dislocation line can close on itself as a loop.

Dislocations have distortional energy associated with them, as is evident from the compressive and the tensile strains around an edge dislocation or the shear strains around a screw dislocation. As a first approximation, one can take these strains to be elastic strains. The elastic strain energy E per unit length of a dislocation of Burgers vector b can be expressed approximately by

$$E \simeq \frac{\mu b^2}{2} \tag{9}$$

Here μ is the shear modulus of the crystals. For BCC iron (a = 2.87 Å), E is of the order 2.5×10^{-9} J-m^{-1}.

In real crystals, the dislocations can be classified as *full* and *partial* dislocations. For a *partial dislocation*, the Burgers vector is a fraction of a lattice translation, whereas for a *full dislocation*, the Burgers vector is an integral multiple of a lattice translation. From relation (9), it is obvious that the elastic strain energy of a dislocation is proportional to the square of the Burgers vector, dislocations tend to have as small a Burgers vector as possible.

22. SURFACE AND GRAIN BOUNDARY DEFECTS

Surface and grain boundary imperfections of a structural nature arise from a change in the stacking of atomic planes on or across a boundary and are two-dimensional. The change may be one of the orientations or of the stacking sequence of the planes. The most important kinds of surface defects are high-angle and low angle boundaries, stacking faults, and twin boundaries. Basically surface imperfections are of two types: *external* and *internal*.

The *external surface* of a material is an *imperfection* itself because the atomic bonds do not extend beyond it. The surface atoms have neighbours on one side only while atoms inside the crystal have neighbours on either side of them (Fig. 3.53). Since surface atoms are not entirely surrounded by others, they possess higher energy than that of internal atoms. For most metals, the energy of surface atom is of the order 1 J/m^2.

The *internal surface imperfections* are manifested by such defects as grain boundaries, tilt boundaries, twin boundaries, and stacking faults.

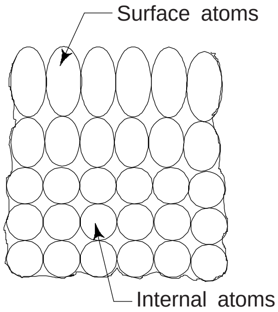
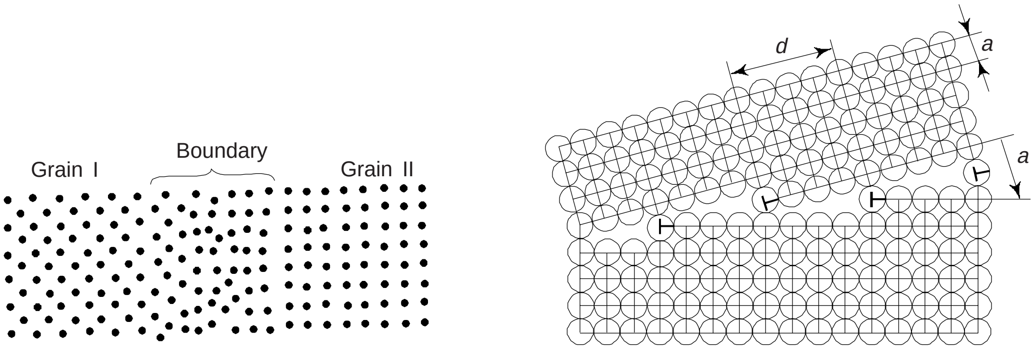


Fig. 3.53 External surface defects

Grain Boundaries

Engineering materials may be either polycrystalline or single crystal type. A polycrystalline alloy contains an enormous quantity of fine grains. Grain boundary imperfections are those surface imperfections which separate crystals or grains of different orientation in a polycrystalline aggregation during nucleation or crystallization. The shape of a grain is usually influenced by the presence of surrounding grains. The lattices of adjacent grains are oriented at random and differently (Fig. 3.54) and a boundary between any two grains is essentially a transition layer of thickness of 1–5 nm. This layer may have a disordered arrangement of atoms, dislocation of clusters, and an elevated concentration of impurities. In grain boundaries the atomic packing is imperfect and between two adjacent grains, there is a transition zone that is not aligned with either grain (Fig. 3.55). As a result, we note that boundary atoms in such cases cannot have a perfect complement of surrounding atoms. The mismatch with orientation of the adjacent grain produces a less efficient packing of atoms along the boundary. Thus, the atoms along the boundary have a higher energy than those within the grains. The higher energy of the boundary atoms is also important for the nucleation of polymorphic phase changes. The boundary between two crystals which have different crystalline arrangements or different compositions, is called an *interphase* boundary or commonly called an *interface.*



(a) (b)

Fig. 3.54 Schemes of (a) high angle and (b) low-angle boundaries

The lower atomic packing along the boundary favours atomic diffusion. When the orientation difference between neighbouring grains is more than 10°–15°, boundaries are called *high angle grain boundaries* (Fig. 3.54(a)). Each grain in turn consists of *subgrains or blocks.*

A subgrain is a portion of a crystal of a relatively regular structure. Subgrain boundaries are formed by walls of dislocations which divide a grain into a number of subgrains or blocks [Fig. 3.54(b)]. Angle of misorientation between adjacent subgrains are not large (not more than 5°), so that their boundaries are termed 'low angle'. Low angle boundaries can also serve as places of accumulation of impurities.

Tilt Boundaries

This is another type of surface defect called low-angle boundary as the orientation difference between two neighbouring crystals is less than 10°. This is why the disruption in the boundary is not so drastic as in the high angle boundary. This type of boundary is associated with relatively little energy and is composed of edge dislocations lying one above the other. In general, one can describe low-angle boundaries by suitable arrays of dislocation. The angle or tilt, $\theta = \dfrac{b}{D}$ [Fig. 3.55(a)], where b is the magnitude of Burgers vector and B is the average vertical distance between dislocations.

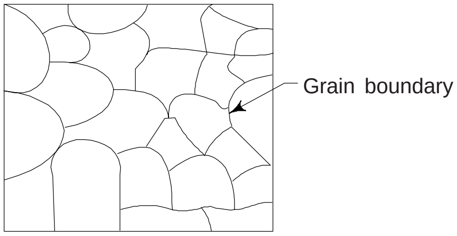


Fig. 3.55 Area of disorder at grain boundaries

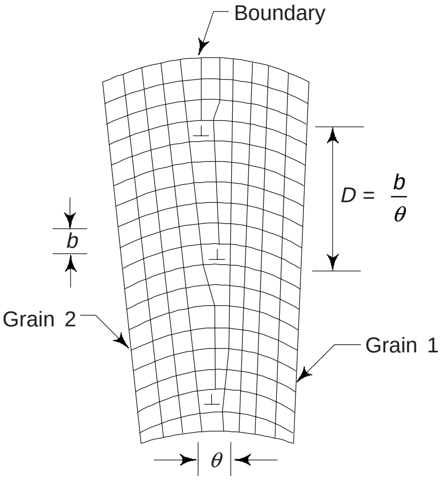


Fig. 3.55(a) Tilt boundary

Twin Boundaries

This is another planar surface imperfection. The atomic arrangement on one side of a twin boundary is a mirror reflection of the arrangement on the other side. Twinning may result during crystal growth or deformation of materials. Twin boundaries occur in pairs, such that the orientation change introduced by one boundary is restored by the other. The region between the pair of boundaries is termed as the *twinned region*. One can easily identify twin boundaries under an optical microscope. Twins which form during the process of recrystallization, i.e., in the process of mechanical working are known as mechanical twins, whereas twins formed as a result of annealing after plastic deformation are known as annealing twins (Fig. 3.56).

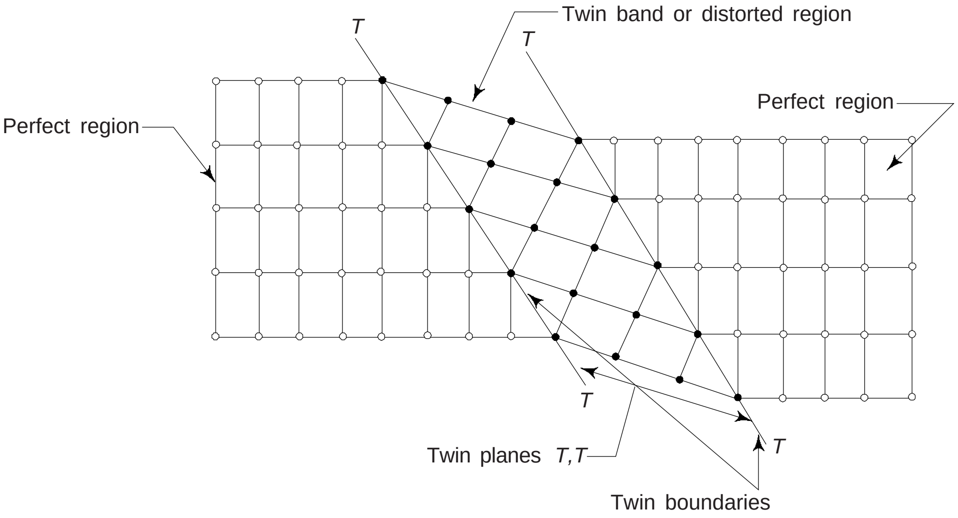


Fig. 3.56 Twin boundaries

Stacking Faults

This is a part of an atomic plane confined between dislocations and in which the regular order of alternation of atomic layers is disturbed. This type of fault arise from the stacking of one atomic plane out of sequence on another while the lattice on either side of the fault is perfect. For instance, in FCC lattice alloys, close-packed atomic layers are normally in an alternation *ABC ABC AB* …. On passing through a stacking fault, their alternation may change, for instance, to *ABC BCA BC* …. The BCBC alternation is typical of an HCP lattice, and therefore the stacking fault in the case considered represents, as it were, a thin plate of an HCP lattice in the structure of an FCC lattice. Figure 3.57 shows the faults in the structure of an FCC crystal. Here the stacking sequence of an ideal FCC crystal is described as *ABC ABC* . . . and the stacking fault may change the sequence *ABC ACAB*.

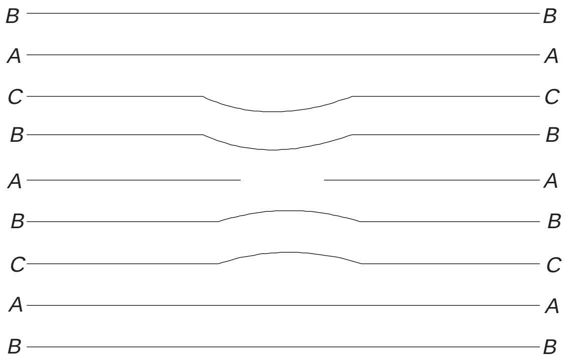


Fig. 3.57 Fault in stacking sequence in FCC crystal

Surface defects influence the mechanical and physical properties of materials of especially large importance are grain boundaries. The yield strength σ_y is associated with the grain size d by the relation.

$$\sigma_y = \sigma_0 + kd^{-1/2} \tag{10}$$

where σ_0 and k are constants for a given material. With finer grain, a material has higher yield strength and toughness and is less susceptible to brittle failure. The size of subgrains has a similar, though less strong, effect on mechanical properties.

Diffusion along boundaries of grains and subgrains occur many times quicker than in other directions in a crystal, especially on heating. Interactions of defects, their displacements in crystals, and variations of their concentrations all these factors can change the properties of metals and are of high practical significance.

23. VOLUME IMPERFECTIONS

Volume imperfections, e.g. cracks may arise when there is only small electrostatic dissimilarity between the stacking sequences of close packed planes in metals. Moreover, when clusters of atoms are missing, a large vacancy or void is got which is also a volume imperfection. Foreign particle inclusions, large voids or non-crystalline regions which have the dimensions of the order of 0.20 nm are also termed volume imperfections.

24. LIQUID CRYSTALS

These are liquids in which the molecular structure has a certain degree of order. Owing to the ordered arrangement of molecules, liquid crystals occupy an intermediate position between crystals and common liquids with disordered molecules. Liquid crystals are fluid as common liquids, but on the other hand, exhibit anisotropic properties as crystals.

A large number of liquid crystals are known today; important among them are certain organic substances with molecules of an elongated form. On melting, these elongated molecules are arranged in a certain order by virtue of some pecularities in molecular interaction. The intermediate state of a liquid crystal with ordered structure is retained in the temperature interval from the melting point to a transition point at which the liquid crystal changes to isotropic liquid. During this transition, thermal oscillations increase and the ordered molecular structure disappears completely, making the substance transparent. This is why, the upper temperature limit of the existence of a liquid crystal is called the *brightening point*.

From structure point of view, liquid crystals can be classified into three classes:
(i) nematic *(ii)* smectic and *(iii)* cholesteric. These are shown in Fig. 3.58.

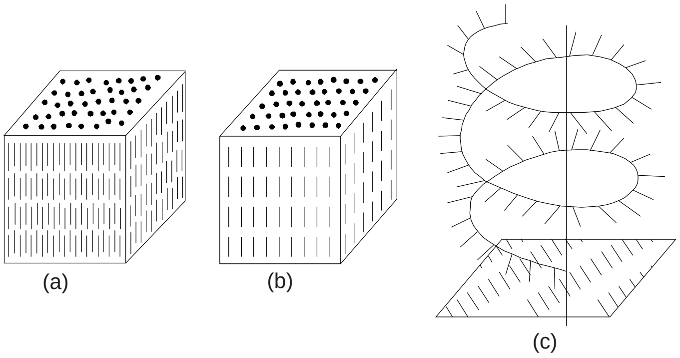


Fig. 3.58 Three types of structure of liquid crystals

In liquid crystals of class I (Fig. 3.58(a)), molecules are arranged into chains and the direction of preferable orientations of molecules is the optic axis of a liquid crystal. In type II (Fig. 3.58(b)) liquid crystals, molecules form parallel layers which can be easily displaced relative to one another. In type III (Fig. 3.58(c)), the structure is more complex, with molecules arranged on a spatial spiral. Long molecules form parallel layers, with each layer having the structure of a liquid crystal of type-I. The direction of preferable orientation varies smoothly from one layer to another, thus forming a spiral of a definite pitch.

The orientation order of molecular arrangement is responsible for anisotropy of properties of liquid crystals; refractive index, dielectric permittivity, specific electric resistance, viscosity, and several other properties turn out to depend on the direction in which they are measured, say parallel to or across molecular axes. In particular, in dielectric liquid crystals, the specific electric resistance measured across molecular chains attains $10^{12} – 10^{14}$ Ohm-m and is several orders of magnitude lower along molecular chains.

The structure of liquid crystals, is changed easily under the action of pressure, electric fields or heating. These effects make it possible to control the properties of liquid crystals by means of weak actions and make liquid crystals indispensable in the manufacture of especially sensitive indicators.

The ability of type-I and type-III liquid crystals to change their optical properties under the action of electric fields and at temperature variations is widely utilized in instrument engineering. A structural change in a liquid crystal caused by an external action is accompanied with displacements of molecules which take 1–10 ms; a still longer time, 20-200 ms, is required to restore their original state upon removal of the external action. This pecularity of liquid crystals restricts their applications to low frequencies (not more than 2–5 kHz). In industry, use is made of individual liquid-crystal organic compounds and their eutectic mixtures which offer a wider temperature range of existence of liquid crystals.

Type-I liquid crystals exhibit the electro-optical effect of dynamical light scattering. When an electric field of a sufficient intensity is applied, the original order of molecules is disturbed, molecules are stirred turbulently, and the substance becomes more transparent. Liquid crystals are used in *colour indicators* and other colour devices. Colour pictures are obtained by mixing liquid crystals with certain pigments whose molecules are also of an elongated form. At a low field intensity, liquid crystal molecules are arranged perpendicular to the cell electrodes and cause pigment molecules to arrange in the same order. In this position of molecules, pigmenting is not observed. When, however, molecules are rotated by the field of a higher intensity, pigment molecules produce a certain colour image.

On heating, the spiral pitch increases in liquid crystals of type-III. This changes the conditions of light interference on the crystals, resulting in a different colour of reflected light. The optical characteristics of liquid crystals are quite diverse, but the wavelength of reflected light changes on the average by 1–2 nm on heating per degree C.

This property of liquid crystals is utilized for recording and measurement of stationary and slowly varying temperature fields. Devices for the purpose may be based on the effect of increasing transparency as a liquid crystal changes to isotropic liquid or on that of variations of colour of reflected light flux.

Liquid crystals find use in clinical thermometers, temperature sensors for control of overheating of machine parts and units, and transducers of invisible infrared rays into visible light. In the last application, a liquid crystal absorbs infrared light and gets heated so that the colour of the reflected light is changed. Liquid crystals also find applications in modulators, information displays: calculators, watches, etc., instruments of automobiles, devices for deviation of light flow, etc.

25. ANISOTROPY

A body is said to be *isotropic* if its physical properties are not dependent upon the direction in the body along which they are measured. A body displaying isotropy has only one refractive index, one dielectric constant, and so on. Most but not all liquids and aggregates made up of small crystals randomly oriented in space are isotropic in all their properties. Depending on their symmetry, single crystals may or may not be isotropic with respect to a given property, e.g. single crystals with cubic structure are isotropic with respect to electrical resistivity but not with respect to elastic deformability.

The properties of crystals are different in various crystallographic directions, which is associated with an ordered arrangement of atoms (ions, molecules) in space. The phenomenon is called *anisotropy.*

The properties of crystals are determined by interactions of atoms. In crystals, the spacings between atoms are different in various crystallographic directions, because of which their properties are also different.

Virtually all properties of crystals are anisotropic. The phenomenon is however more pronounced in crystals with structures of a poor symmetry. The resistivity of certain single crystals measured with the electric field along a particular crystallographic direction may be higher than along directions perpendicular to it. Thus such crystals are anisotropic with respect to resistivity.

Magnetic properties are found to be anisotropic in cubic crystals as well. For instance, the magnetization of ferromagnetics with a cubic lattice is different in various crystallographic directions. For α-Fe (BCC lattice), the direction of easy magnetization is [100], for Ni (FCC lattice), it is [111], and for Co (HCP lattice), it is [110].

Examples of bodies which are anisotropic in some of their properties are liquid 'crystals', single crystals, and aggregates of polycrystals with a preferred orientation.

26. FRANK-READ SOURCE

It is observed that slip occurs intensely on a small number of crystal planes and, during the process, some hundreds of dislocations move. Obviously, there must be some effective creators or some mechanisms by which these numerous dislocations are produced on a given slip plane. These are called as Frank-Read sources. A single Frank-Read dislocation source can form hundreds of new dislocations.

It is observed in practical tests that under relatively high stresses, the incidence of plastic deformation is high in most metals through the combined movement of many hundred thousand of dislocations in individual crystals. Plastic deformation in real crystals is effected by successive movements of dislocations.

Frank-Read mechanism helps to explain the existence of so called dislocation mills or multiplication of dislocations (Fig. 3.59). Let us consider a dislocation line, shown by *AB* in a crystal. We see that the dislocation line or Frank-Read source consists of two nodes *A* and *B*. In the first situation, when the Burger vector is perpendicular to the line *AB*, a shear stress parallel to the plane of the figure will exert a force on the dislocation line *AB*. Due to the action of the shear stress the dislocation line bent outward and produces slip. A slip plane usually contains tens of dislocations. For a given stress, the dislocation line *AB* assume a certain radius of curvature. On further increasing the stress, the dislocation line becomes unstable and expands indefinitely. Figure 3.59 illustrates the successive stages.

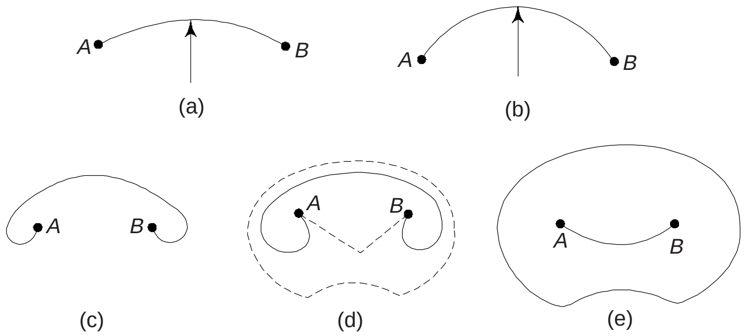


Fig. 3.59 Sequence of formation of a new dislocation by Frank-Read source. A dislocation is pinned by two pinning sites. Under the action of the applied shear stress a fixed (sessile) dislocation is bent outward until it becomes hemispherical. From that moment on, the dislocation will spontaneously pass through the configurations as shown in figure and a dislocation loop will be generated

Under the action of the applied stress, the fixed dislocation line *AB* is bent outward until it becomes hemispherical (Fig. 3.59(a) and (b)) of radius *AB*/2. We must note that if the stress is removed at any stage upto this point *AB* will regain its original shape. If the stress is further increased to the stage (Fig. 3 59(c)) at which the bulge becomes greater than a semicircle, a new system of balance exists and a lower strain energy will be attained by the loop becoming larger (i.e., radius of loop increasing again). From that moment on, the bent dislocation propagates spontaneously in the form of two spirals. As the spirals meet, they give rise to an expanding dislocation loop and a dislocation section. The dislocation section occupies an initial position and the dislocation source is ready to repeat the cycle. A single Frank-Read dislocation source can form hundreds of new dislocations.

The strengthening effect produced in metals by deformation (strain hardening) is based primarily on an increase of dislocation density.

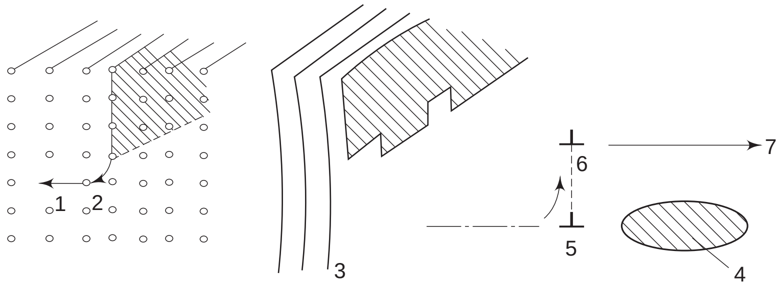
27. THEORY OF DISLOCATIONS: SALIENT FEATURES

The following terms are invariably used in the study of dislocations: *(i)* Burger vector, *(ii)* Dislocation climb, *(iii)* Cross slip, *(iv)* Jogs in dislocation and *(v)* Perfect and Partial dislocations. About Burger vector, we have already discussed earlier.

Dislocation Climb

Dislocation movements are impeded by various obstacles, such as grain boundaries, interface planes or dislocations intersecting the slip plane. Some obstacles can be overpassed by dislocations at an elevated stress. These are, for instance, the dislocations intersecting the slip plane. On the contrary, other obstacles, such as grain boundaries, are unsurmountable for dislocations and the latter accumulate before them. An accumulation (pile-up) of dislocations creates a stress field which repulses any approaching dislocation. The greater the number of dislocations in a cluster, the stronger the repulsion and the metal is more difficult to deform. When the dislocation density in a cluster attains a definite value, a crack nucleates in that point.

On heating to temperatures above 0.3 T_m, another mechanism of dislocation movement becomes operative. This is what is called dislocation climb. It is essentially the diffusional displacement of a dislocation into adjacent planes of the lattice by means of attachment of dislocation into adjacent planes of the lattice by means of attachment of vacancies (Fig. 3.60). Vacancies are attached successively to the edge of an excessive half-plane, which is equivalent to the movement of the edge by one atomic row upwards. Vacancies ‘attack’ the dislocation in various points and form steps on it. As vacancies are being attached, the dislocation shifts in an appreciable part of its length through tens of lattice planes. The dislocation climb weakens the retarding effect of the secondary phase particles. The climbed dislocations are then moved by slip under the action of the stress. On heating above 0.3 T_m, vacancies become quite mobile, and the required number of vacancies can be formed by plastic deformation. Dislocation climb is one of the important mechanisms that take place in the annealing process.



(a) (b) (c)

Fig. 3.60 Dislocation climb scheme 1-Displacement of an atom into vacancy site; 2-departure atom from dislocation; 3-beginning of dislocation climb; 4-secondary-phase particle blocking dislocation slip; 5-dislocation; 6-new position of dislocation after climb; 7-direction of dislocation slip

Cross-Slip

In comparison to edge dislocation, a screw dislocation can change its slip orientation more readily. Burger vector in a screw dislocation is parallel to the line of dislocation, therefore the possible slip planes of screw dislocation need not be parallel as is desired for edge dislocation. The dislocation simply changes slip planes, whereas the screw dislocations are capable of avoiding obstacles by a process known as cross-slip. When the planes of easy slip intersect each other, a screw dislocation can cross-slip from one easy slip-plane to another. However, the resolved shear stress is not the same on all easy slip-planes, i.e., it is of lower value on the plane of cross-slip than that on the slip-plane containing the obstacle. As a result, cross-slip occurs at higher stresses in comparison to normal slip.

We must note that the cross-slip occurs most readily in crystals with a large number of slip systems. There are four and six planes of easy slip in FCC and BCC structures respectively, and these planes intersect. Obviously, they fulfill the cross-slip condition. The work-hardening characteristics of various materials used for engineering purposes depend on the presence of cross-slip or occurrance of higher stress values. Cross slip is illustrated in Fig. 3.61.

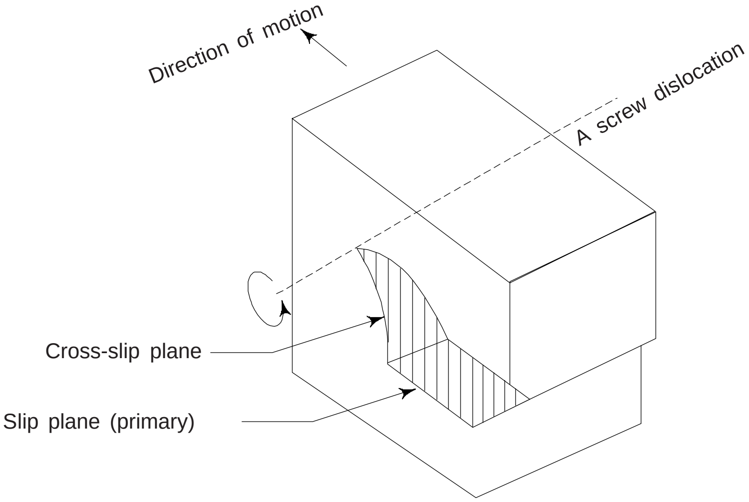


Fig. 3.61 Cross-slip

Jogs in Dislocation

One can regulate a jog in a dislocation as a short length of dislocation not lying in the same plane as the main plane but having the same Burger vector. A jog is defined as the vector dislocation jumping from one plane to another. This is due to the fact that dislocation may not be confined to a single plane but may move from one slip plane to another slip plane. Jogs in a crystal may be formed at the utilisation of two edge or screw dislocations. Jogs in dislocation is illustrated in Fig. 3.62.

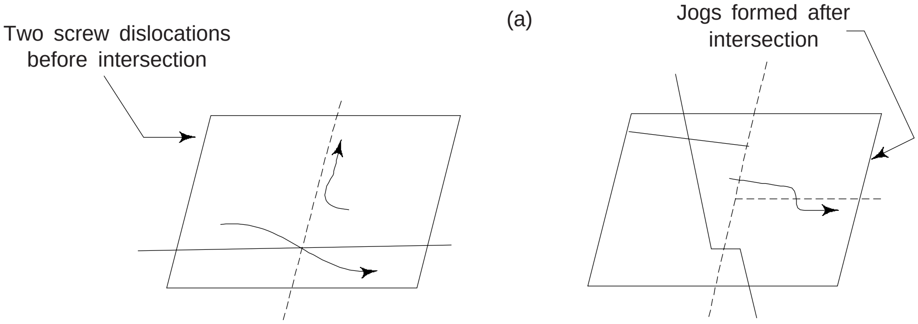


Fig. 3.62 Jogs in a crystal dislocation

Types of Dislocations

Dislocations in real crystals are of two types: *(i)* Partial or imperfect dislocations, and *(ii)* Perfect or full dislocation. Partial or imperfect dislocations have a Burger vector, which is only a fraction of the lattice translation vector. Partial dislocation is always associated with a surface imperfection or fault in the stacking arrangement of planes in the crystal. Perfect or full dislocations are surrounded by good regions of the crystal, i.e., on both the sides of the dislocation, the vertical planes match across the slip plane. This is possible because the Burger vector of perfect or full dislocation is an integral multiple of the lattice translation vector.

Example 20 *Calculate the line energy of dislocations in BCC iron. Given, the Burgers vector in iron is of the* $\frac{1}{2}$ *<111> type and shear modulus of iron is 80.2 GN/m*^{2} [BE]

Solution For BCC iron, the lattice parameter, a = 2.87 Å. Magnitude of Burgers vector,

$$b = 2.87\ \sqrt{3}/2\ = 2.49\ \text{Å}.$$

We have the relation for the line energy of the dislocation,

$$E \simeq \frac{1}{2}\,\mu b^2$$

$$= \frac{80.2\times 10^{9}\times (2.49)^{2}\times 10^{-20}}{2}$$

$$= 2.49\times 10^{-9}\ \text{J/m.}$$

Example 21 *A circular dislocation loop has edge character all round the loop. Show that the surface on which this dislocation can glide should be a cylindrical surface containing the loop.*

[M.Sc. (MS), AMIE]

Solution We know that the Burgers vector ($\vec{b}$) is perpendicular to an edge dislocation line and Burgers vector is invariant. We also know that the edge dislocation can glide only on a surface which contains both the Burgers vector ($\vec{b}$) and the direction vector ($\vec{t}$). These considerations are satisfied only when the dislocations moves, i.e. glide, on a cylindrical surface containing the loop.

Example 22 *There are 10*^{10} *m*$^{-2}$ *of edge dislocations in a simple cubic crystal. How much would each of these climb down on an average, when the crystal is heated from 0 to 1000 K. Given, the lattice parameter = 2 Å, volume of 1 mole of the crystal is 5.5* $\times 10^{-6}$ *m*^{3} *(= 5.5 cm*^{3}*) and the enthalpy of formation of vacancies is 100 kJ/mol.* [BE]

Solution We know that at equilibrium, there are no vacancies in the crystal at 0 K. Now, to maintain the equilibrium concentration, the number of vacancies that must be created on heating from 0 to 1000 K is given by

$$n = \text{N}\exp\left[\frac{-\Delta H_f}{RT}\right]$$

Substituting the proper values in the above relation, one obtains

$$n = 6.023\times 10^{23}\ \exp\left[-\frac{100\times 1000}{8.314\times 1000}\right]$$

$$= 3.6\times 10^{18}/\text{mol.} = 6.54\times 10^{23}\ \text{m}^{-3}$$

Example 23 *The surface of a copper crystal is of the {111} type. Show that the surface energy, i.e. enthalpy of copper is about 2.49 J-m*$^{-2}$*.*

Solution We have the bond energy per atom of copper

$$= \text{bond energy per bond} \times \text{no. of bonds per atom} \times \frac{1}{2}$$

$$= \frac{56.4 \times 1000 \times 12}{6.023 \times 10^{23} \times 2} = 5.62 \times 10^{-19} \text{ J}$$

Out of the twelve bonds, three bonds are broken at the surface. Energy of broken bonds per atom

$$= 5.62 \times 10^{-19} \times \frac{1}{4} \text{ J}$$

Number of atoms on {111} planes in copper

$$= 1.77 \times 10^{-19} \text{ m}^{-2}$$

Surface energy (enthalpy) of copper

$$= 5.62 \times 10^{-19} \times \frac{1}{4} \times 1.77 \times 10^{19}$$

$$= 2.49 \text{ J/m}^2 \ (= 2490 \text{ erg/cm}^2)$$

Example 24 *Find the equilibrium concentration of vacancies in aluminium at 0 K and 300 K.* [AMIE]

Solution For Aluminium $\Delta H_f = 68 \text{ kJ mol}^{-1}$

and $\frac{eV}{\text{vacancy}} = 0.70$

we have $\frac{n}{N} = \exp\left[-\frac{\Delta H_f}{RT}\right]$

(i) At 0 K,

$$\frac{n}{N} = \exp\left[-\frac{68 \times 10^3}{8.314 \times 0}\right] = \exp(-\infty) = 0$$

(ii) At 300 K

$$\frac{n}{N} = \exp\left[-\frac{68 \times 10^3}{8.314 \times 300}\right] = \exp(-27.26)$$

$$= 10 \ \frac{-27.26}{2.303} = 10^{-11.84}$$

$\therefore$ $\text{Log } \frac{n}{N} = -11.84$

$\therefore$ $\frac{n}{N} = 1.45 \times 10^{-12}$

28. DETERMINATION OF CRYSTAL STRUCTURE BY X-RAY DIFFRACTION

X-rays provide a powerful tool for the study of crystal structure. X-rays, being electromagnetic radiations, also undergo the phenomenon of diffraction as observed for visible light. The ordered arrangement of atoms in a crystal with interatomic spacing of the order of few angstroms behaves like a three-dimensional diffraction grating for X-rays. One can easily verify this. Let us consider sodium metal whose density ρ and molecular weight, M are 1.013×10^3 kg/m^3 and 23 respectively. The structure of sodium is BCC and number of atoms/unit volume = $N\rho/M = 2/a^3$, where N is Avogadro number (= 6.023×10^{26}) and a is cell constant. So, we have

$$a = \left(\frac{2M}{N\rho}\right)^{1/3} = \left[\frac{2 \times 23}{6.023 \times 10^{26} \times 1.013 \times 10^3}\right]^{1/3} = 4.225 \text{ Å}$$

Thus the wavelength of the radiation required for diffraction by sodium metal is of the order of a few angstroms which is clearly in the range of X-rays. We can easily show that the required voltage in the X-ray tube to produce this wavelength is 10 – 50 keV.

Bragg's Law of X-ray Diffraction

In 1912, Laue suggested that a crystal which consisted of a three-dimensional array of regularly spaced atoms could serve as a grating. The crystal differs from the ordinary grating in the sense that the diffracting centres in the crystal are not in one plane, i.e. crystal acts as a space grating rather than plane grating used for the study of light diffraction.

On the suggestion of Laue, his associates succeeded in diffracting X-rays by passing them through a thin crystal of Zinc blende. The diffraction pattern obtained by them consists of a central spot and a series of spots arranged in a definite pattern around the central spot. This symmetrical pattern of spots is known as Laue pattern. The explanation for this was given by Bragg. According to Bragg, the spots are produced due to the reflection of some of the incident X-rays from the various sets of parallel crystal planes (called as Bragg's planes), which contain a large number of atoms.

Consider a ray PA reflected at atom A in the direction AR from plane I and another ray QB reflected at another atom B in the direction BS (Fig. 3.63). Now, draw two perpendiculars AC and AD from the atom A on QB and BS respectively. The two reflected rays AR and BS will be in phase or out of phase depending on the path difference. When the path difference, $CB + BD$ is a whole wavelength λ, or multiple of whole wavelength ($n\lambda$), then the two rays will reinforce each other and produce an intense spot due to constructive interference. Obviously, the condition of reinforcement, i.e. constructive interference is:

$$CB + BD = n\lambda$$

$$CB = BD = d \sin \theta$$

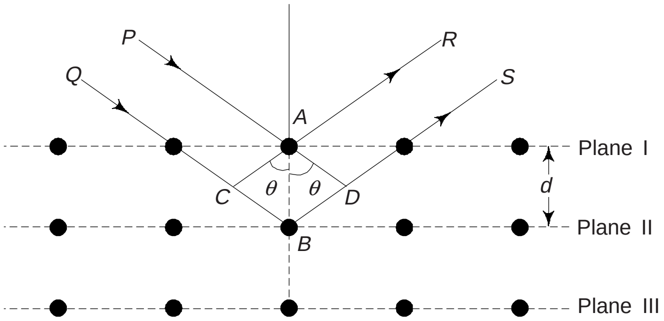


Fig. 3.63 Reflection of *X*-rays from lattice planes in a crystal and derivation of Bragg's law

where, θ is the angle between the incident ray and the planes of reflection and called as *glancing angle*. Therefore, we have

$$2d \sin \theta = n\lambda \tag{11}$$

where d is the interplanar spacing of planes, i.e. distance between the same set of planes and $n = 1, 2, 3, \ldots$ stand for first order, second order, third order maxima respectively. Equation (11) is known as Bragg's law. We must note that the Bragg's reflection can occur only for $\lambda \leq 2d$.

The directions along which the maxima are observed can be obtained from Bragg's law, by giving different values of θ; i.e.,

for first maximum, $\sin \theta_1 = \dfrac{\lambda}{2d}$

for second maximum, $\sin \theta_2 = \dfrac{2\lambda}{2d}$

for third maximum, $\sin\theta_3 = \dfrac{3\lambda}{2d}$ and so on.

. . . .

As the order of spectrum increases, the intensity goes on decreasing.

We must note that the Bragg equation is not limited to atomic planes parallel to the crystal surface. Sets of parallel atomic planes can be constructed through the crystal at different angles. The value of the grating space d is different for each set of parallel planes.

Reflection rules of X-ray diffraction for common metal structures are summarized in the following table 3.3

Table 3.3 Reflection rules of X-ray diffraction for the common metal structures

Crystal structure	*Diffraction does not occur for*	*Diffraction occurs for*
BCC	$h + k + l$ = odd number	$h + k + l$ = even number
FCC	h, k, l can have both even and odd integer values	h, k, l can be all even or all odd numbers
HCP	$h + 2k = 3n$, l odd ($n \to$ integer)	All other cases

29. BRAGG'S X-RAY SPECTROMETER

The X-rays from the target T of an X-ray tube pass through two slits S_1 and S_2 in thick lead sheets (Fig. 3.64). They then fall on the single crystal D mounted on a turn table. The crystal is rotated by means of the turn-table so as to increase the glancing angle at which X-rays are incident at the exposed face of the crystal. The position of the crystal can be read with the help of the vernier v and a circular scale. An ionisation chamber I is mounted on an arm which can rotate about the same axis as that of the crystal table. The ionisation chamber or photographic plate is used for measuring the intensities of reflected rays. The angles, for which reflection intensities are maximum, give the values of θ of equation $n\lambda = 2d \sin\theta$. One can carry out the process for each plane of the crystal. $n = 1$ creates the condition for the lowest angle at which the lowest reflection occurs. $n = 2$ creates the condition for the next higher angle at which the maximum reflection occurs and so on. Thus, we have

For $n = 1$ $\lambda = 2d\sin\theta_1$

$n = 2$ $2\lambda = 2d\sin\theta_2$,

$n = 3$ $3\lambda = 2d\sin\theta_3$, etc.

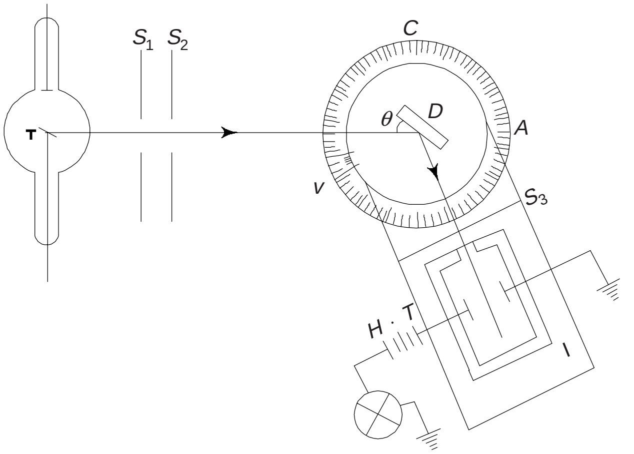


Fig. 3.64 Bragg's X-ray spectrometer

where θ_1, θ_2 and θ_3 are glancing angles for $n = 1, 2$ and 3 respectively. Thus

$$\sin\theta_1 : \sin\theta_2 : \sin\theta_3 = 1 : 2 : 3 \tag{12}$$

Obviously, by measuring glancing angles at which reflection occurs, one can determine the interplanar spacings knowing the wavelength of X-rays.

The variation of ionization current against glancing angle θ is shown in Fig. 3.65. This curve is obtained when X-rays from a tube containing a platinum target were reflected by a crystal of rock salt (NaCl).

The experimental results have shown that the first order reflection maxima occurred at 5.9°, 8.4° and 5.2° for (100), (110) and (111) planes respectively.

For first order reflection, $n = 1$ and hence

$$d = \frac{\lambda}{2\sin\theta}$$

or

$$d \propto \frac{1}{\sin\theta}$$

$$\therefore \quad d_{100} : d_{110} : d_{111} = \frac{1}{\sin\theta_1} : \frac{1}{\sin\theta_2} : \frac{1}{\sin\theta_3}$$

or

$$d_{100} : d_{110} : d_{111} = \frac{1}{\sin 5.9°} : \frac{1}{\sin 8.4°} : \frac{1}{\sin 5.2°}$$

$$= 9.73 : 6.84 : 11.04$$

$$\therefore \quad d_{100} : d_{110} : d_{111} = 1 : \frac{1}{\sqrt{2}} : \frac{2}{\sqrt{3}}$$

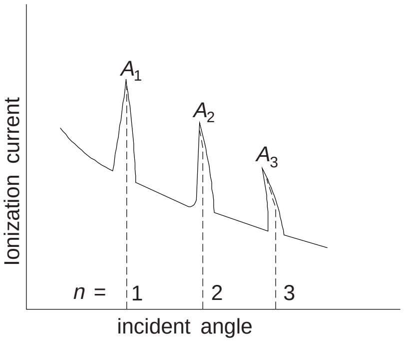


Fig. 3.65 **Variation of ionization current with glancing angle**

One can follow this procedure for other crystals. Following table summarizes the values of ratios between interplanar distance of planes (100), (110) and (111) for the cubic structures.

Table 3.4 Ratio between interplanar distances

S. No.	*Type of cubic structure*	$\frac{1}{d_{100}}$	:	$\frac{1}{d_{110}}$	:	$\frac{1}{d_{111}}$
1	Simple cubic	1	:	$\sqrt{2}$	:	$\sqrt{3}$
2	BCC	1	:	$\frac{1}{\sqrt{2}}$	:	$\sqrt{3}$
3	FCC	1	:	$\sqrt{2}$	:	$\frac{\sqrt{3}}{2}$

Example 25 *Determine the interplanar spacing when a beam of X-ray of wavelength 1.54 Å is directed towards the crystal at angle 20.3° to the atomic plane.* [AMIE]

Solution $2d\sin\theta = n\lambda$ | $\lambda = 1.54$ Å

$\therefore$ $2d\sin 20.3° = 1 \times 1.54$ | $\theta = 20.3°$

$$\therefore \quad d = \frac{1.54}{2\sin 20.3°} = \frac{1.54}{2 \times 0.3469} = 2.22 \text{ Å}$$

Example 26 *X-rays with wavelength of 0.58 Å are used for calculating d_{200} in nickel. The reflection angle is 9.5°. What is the size of unit cell?* [AMIE]

Solution

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

$\lambda = 0.58$ Å

$\theta = 9.5°$

$$\therefore \quad d_{200} = \frac{a}{\sqrt{2^2 + 0^2 + 0^2}} = \frac{a}{2} = 0.5a$$

Now, from Bragg's law

$$2d \sin \theta = n\lambda, \text{ we have}$$

$$2\ d_{200} \sin 9.5° = 1 \times 0.58$$

$$2 \times 0.5a \times 0.165 = 0.58$$

$$\therefore \quad a = \frac{0.58}{1.165} = 0.52 \text{ Å}$$

Example 27 *Calculate the Bragg angle if (111) planes of a cube (a = 3.57 Å) crystal are exposed to X-rays (wavelength = 1.54 Å)* [AMIE]

Solution We have, Miller indices of the (111) planes, $h = 1$, $k = 1$ and $l = 1$; $a = 3.57°$ and $\lambda = 1.54$ Å. Let θ be the Bragg's angle for the first order reflection.

We have

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

$$d_{111} = \frac{3.57}{\sqrt{(1)^2 + (1)^2 + (1)^2}} = 2.06 \text{ Å}$$

Now, from Bragg's law, we have

$$2d_{111} \sin \theta = n\lambda$$

$$\therefore \quad 2 \times 2.06 \times \sin \theta = 1 \times 0.54$$

$$\sin \theta = \frac{1 \times 0.54}{2 \times 2.06} = 0.131$$

$$\theta = 7° \ 32'$$

Example 28 *For a certain BCC crystal, the (110) plane has a separation of 1.181 Å. These planes are indicated with X-ray of wavelength 1.540 Å. Show that the maximum order of the Bragg's reflection that can be observed is n = 1.*

Solution

$$2d \sin \theta = n\lambda$$

$d = 1.181$ Å

$\lambda = 1.540$ Å

$$\therefore \quad n = \frac{2\,d \sin \theta}{\lambda}$$

$$= \frac{2 \times 1.181 \sin 90°}{1.540} = 1.53$$

Since the value of 'n' can be integer only, and hence the highest possible value of n in this case is 1.

Example 29 *The glancing angle of 10° is observed in first order Bragg's reflection. Show that the angle for 3rd order reflection is 31.39°.*

Solution For $n = 1, \quad \theta_1 = 10°$

For $n = 3, \quad \theta_3 = ?$

From Bragg's law, we have

$$2d \sin \theta_1 = 1\lambda \quad \text{(First order)}$$

$$2d \sin \theta_3 = 3\lambda \quad \text{(Third order)}$$

$$\therefore \quad \frac{\sin \theta_1}{\sin \theta_3} = \frac{1}{3}$$

or $$\sin \theta_3 = 3 \sin \theta_1 = 3 \times \sin 10° = 3 \times 0.1736 = 0.5209$$

$$\therefore \quad \theta_3 = 31.39°.$$

Example 30 *A diffraction pattern of a cubic crystal of lattice parameter a = 3.16 Å is obtained with a monochromatic X-ray beam of wavelength 1.54 Å. The first line on this pattern was observed to have θ = 20.3°. Obtain the interplanar spacing and Miller indices of the reflecting plane.*

[B.E.; B.Sc. (Hons.)]

Solution

$$2d \sin \theta = n\lambda$$

$$\therefore \quad d = \frac{n\lambda}{2 \sin \theta}$$

$$= \frac{1 \times 1.54}{2 \times 0.3469}$$

$$= 2.22 \text{ Å}$$

$\theta = 20.3°$
$\lambda = 1.54$ Å
$n = 1$
$a = 3.16$ Å

For cubic crystal, we have

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

or $$h^2 + k^2 + l^2 = \frac{a^2}{d^2} = \frac{(3.16)^2}{(2.22)^2} = 2$$

Since $(h^2 + k^2 + l^2) = 2$ and hence the Miller indices are (110) or (101) or (011).

Example 31 *Determine interatomic spacing when glancing angle of 30° is observed during first order reflection in a crystal having Miller indices as (111). The wavelength of X-rays is 2 Å.*

Solution

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

$$\therefore \quad d = \frac{a}{\sqrt{(1)^2 + (1)^2 + (1)^2}} = \frac{a}{\sqrt{3}}$$

$\lambda = 2$ Å
$\theta = 30°$
$n = 1$
$h = k = l = 1$

Now, $$2d \sin \theta = n\lambda$$

$$2 \frac{a}{\sqrt{3}} \sin 30° = 1 \times 2 \text{ Å}$$

$$a = 2\sqrt{3} \text{ Å}$$

30. RECIPROCAL LATTICE

We have seen that the diffraction of X-rays occurs from various sets of parallel planes having different orientations (slopes) and interplanar spacings. In certain situations involving the presence of a number of sets of parallel planes with different orientations, it becomes difficult to visualize all such planes because of their two-dimensional nature. To solve the problem, Ewald introduced the new type of lattice known as the *reciprocal lattice*. The idea underlying the development was that each set of parallel planes could be represented by a normal to these planes having length equal to the reciprocal of the interplanar spacing. Thus the direction of each normal represents the orientation of the corresponding set of parallel planes and its length is proportional to the reciprocal of the interplanar spacing.

The normals are drawn with reference to an arbitrary origin and points are marked at their ends. These points form a regular arrangement which is called a reciprocal lattice. Obviously, each point in a reciprocal lattice is a representative point of a particular parallel set of planes and it becomes easier to deal with such points than with sets of planes.

One can construct a reciprocal lattice to a direct lattice using the following procedure:

(i) Take origin at some arbitrary point and draw normals to every set of parallel planes of the direct lattice

(ii) Take length of each normal equal to the reciprocal of the interplanar spacing for the corresponding set of planes. The terminal points of these normals form the reciprocal lattice

Let us consider, as an example for the construction of reciprocal lattice, the unit cell of a monoclinic crystal ($a \neq b \neq c$; $\alpha = \beta = 90° \neq \gamma$) looking along its unique axis, i.e. c-axis. For simplicity, we may consider planes belonging to a single zone. This makes the normal to all parallel planes belonging to the zone axis, i.e., normal to the c-axis. In Fig. 3.66, we can see that the zone axis lies perpendicular to the plane of the diagram and hence all normals to the parallel planes of the family of the zone will lie in the plane of the diagram. Figure 3.66 also shows the edge views of four ($h\ k\ l$) planes, viz., (100), (110), (120) and (010) all belonging to [001] zone. We must note that the normals to the family of planes (100), (200), (300) etc. are parallel. Further $d_{100} = 2d_{200} = 3d_{300} = 4d_{400}$ etc.
Obviously, while

$$\sigma_{100} = k\ (1/d_{100}),$$
$$\sigma_{200} = 2\sigma_{100};$$
$$\sigma_{300} = 3\sigma_{100} \text{ and}$$

so. Clearly, the distance of the reciprocal lattice point representing the set of parallel planes (nh 00) from the origin will be n times greater than the distance of the reciprocal lattice point from the origin representing

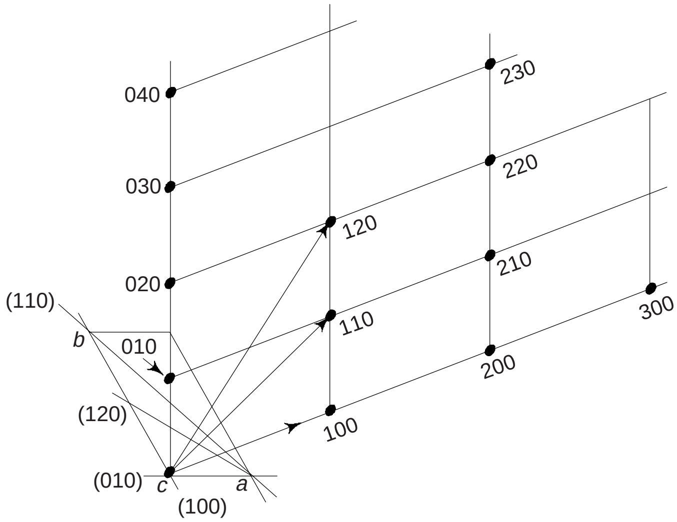


Fig. 3.66 Graphical construction of a two-dimensional reciprocal lattice

the set of parallel planes (300) will be at a distance from the origin equal to three times the distance of the reciprocal lattice point representing the set of parallel planes (100). In reciprocal lattice, it is the convention to represent the reciprocal lattice point representing the set of parallel planes with Miller indices (*h*, *k l*) by the same Miller indices without the parenthesis. We can see in Fig. 3.66 that the reciprocal lattice points representing the set of parallel planes (100), (200), (300) etc . . ., are indicated by 100, 200, 300 etc . . . . As stated earlier, all the set of parallel planes (100), (200), (300) etc., have common normal all the reciprocal lattice points 100, 200, 300 etc. all lie along the same straight line. We have taken the planes belonging to [001] zone and after plotting the reciprocal lattice points of all the (*h k o*) planes, we obtained a two dimensional lattice (Fig. 3.66). Ewald provided a vector-algebraic proof that a three-dimensional reciprocal lattice is similarly formed in real crystal.

A reciprocal lattice vector, σ_{hkl}, is defined as a vector having magnitude equal to the reciprocal of the interplanar spacing d_{hkl} and direction conciding with normal to (*h k l*) planes. Thus, we have

$$\sigma_{hkl} = \frac{1}{d_{hkl}}\hat{n} \tag{13}$$

where $\hat{n}$ is the unit vector normal to the (*h k l*) planes.

Denoting the fundamental reciprocal vector σ_{100}, σ_{010} and σ_{001} by a^*, b^* and c^* respectively, one obtains

$$\begin{aligned} \vec{a}^* &= \sigma_{100} = \frac{\vec{b}\times\vec{c}}{\vec{a}\cdot\vec{b}\times\vec{c}} \\ b^* &= \sigma_{010} = \frac{\vec{c}\times\vec{a}}{\vec{a}\cdot\vec{b}\times\vec{c}} \\ c^* &= \sigma_{001} = \frac{\vec{a}\times\vec{b}}{\vec{a}\cdot\vec{b}\times\vec{c}} \end{aligned} \tag{14}$$

where $\vec{a}\cdot\vec{b}\times\vec{c} = \vec{b}\cdot\vec{c}\times\vec{a} = \vec{c}\cdot\vec{a}\times\vec{b}$ is the volume of the direct cell. Obviously, the reciprocal translation vectors bear a simple relationship to the crystal translation vectors as

$$\left.\begin{aligned} &\vec{a}^* \text{ is normal to } \vec{b} \text{ and } \vec{c} \\ &\vec{b}^* \text{ is normal to } \vec{c} \text{ and } \vec{a} \\ &\vec{c}^* \text{ is normal to } \vec{a} \text{ and } \vec{b} \end{aligned}\right] \tag{15}$$

In literature, sometime the primitive translation vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ of a direct lattice are related to the primitive translation vectors $\vec{a}^*$, $\vec{b}^*$ and $\vec{c}^*$ of the reciprocal lattice as

$$\left.\begin{aligned} \vec{a}^* &= 2\pi\,\frac{\vec{b}\times\vec{c}}{\vec{a}\cdot\vec{b}\times\vec{c}} \\ \vec{b}^* &= 2\pi\,\frac{\vec{c}\times\vec{a}}{\vec{a}\cdot\vec{b}\times\vec{c}} \\ \vec{c}^* &= 2\pi\,\frac{\vec{a}\times\vec{b}}{\vec{a}\cdot\vec{b}\times\vec{c}} \end{aligned}\right] \tag{16}$$

31. METHODS OF DETERMINING CRYSTAL STRUCTURE

The following methods are widely used for determining crystal structure

(i) X-ray diffraction: (a) Bragg's spectrometer method (b) Laue's method (c) Powder method

(ii) Electron diffraction and *(iii)* Neutron diffraction

We have already discussed Bragg's method. Now, we will discuss other methods.

(b) Laue's Method

This is one of the principal method to study X-ray diffraction and to determine crystal structure. In this method, a single crystal is placed in the path of X-rays beam and the glancing angle θ is kept constant. A white radiation having wide range of wavelengths is allowed to fall on the crystal. Each set of crystal plane selects appropriate value of wavelength, satisfying Bragg's law. Obviously, each set of plane will produce a spot which fulfils Bragg's law.

A beam of polychromatic *X*-rays strike at 90°on the plane of the crystal (Fig. 3.67). The rays, while passing through the crystal meet Bragg's planes having different interplanar spacings *d*. These different sets of planes make different angles θ with the direction of X-rays beam. We can see that certain combinations of θ and *d* satisfy the Bragg's condition, with the result that the intensity of diffracted X-rays increase.

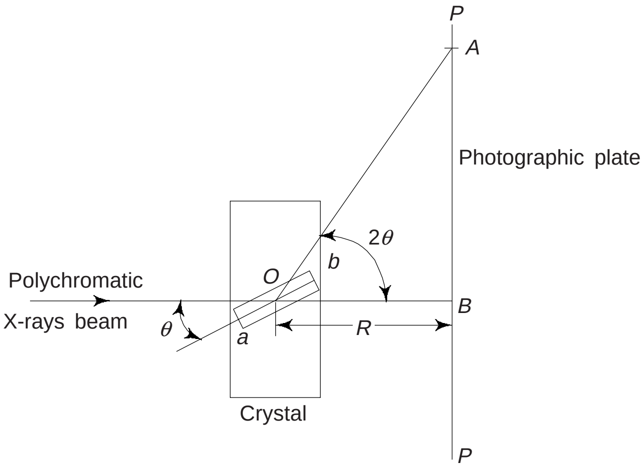


Fig. 3.67 Laue's Method

When primary rays are allowed to pass through the crystal, they produce a black spot at *B* (Fig. 3.67) on photographic plate PP. Different wavelengths are included in the primary X-rays and hence they will produce spots around this central spot, which are less pronounced.

Let *ab* represent the position of one of the possible Bragg's plane, θ the corresponding angle and *A* any spot (other than central spot) on *PP*. From ΔAOB, one finds

$$AB = R \tan 2\theta \qquad (17)$$

Knowing *AB* and *R*, one can find θ for the corresponding plane. Laue's method is quite useful for single large size crystal.

(c) Powder Method

This method is quite useful when single crystals of large size are not available. Method is used to determine the structure of crystalline solids. The sample used is in the form of a fine powder containing a large number of tiny crystallites with random orientations. It is prepared by crushing the commonly available polycrystalline material, thus eliminating the tedious process of growing the single crystals.

The powder is placed in a capillary tube or pasted on a thin wire or pressed and cemented into a thin spindle. This powder specimen is mounted at the centre around which a strip of circular photographic film is positioned (Fig. 3.68(a)).

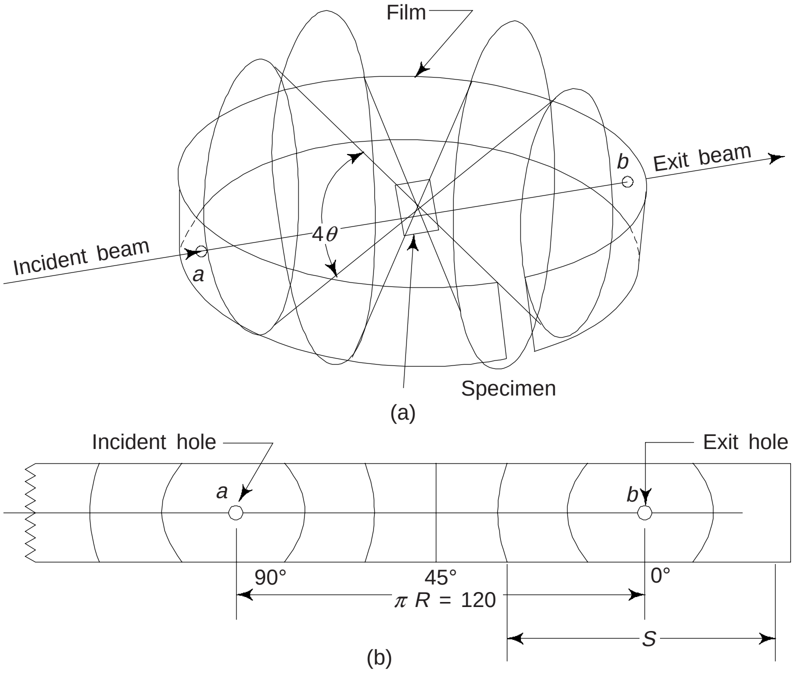


Fig. 3.68 Powder method
(a) Photographic film position
(b) Flattened photographic film after developing and indexing of different lines

A collimated beam of monochromatic X-rays is produced by passing the X-ray through a filter and a collimator. This is allowed to strike the specimen (in capillary tube or wire or spindle) through a small hole. The specimen (spindle etc.) is slowly revolved inside the specially constructed powder camera. The X-ray beam enters through hole '*a*', passes through the specimen and the unused part of beam exit through hole '*b*'.

Let us consider that the incident beam make an angle θ with a set of parallel crystal plane. If Bragg condition is satisfied, then there will be reflection. Since there are a large number of randomly oriented crystals in the powder sample and hence there are several possible orientations of this set of planes of same angle θ. Consequently, the reflected rays will not be in the form of parallel beam but they will lie on the surface of a cone with its apex at the sample and the semivertical angle 2θ (Fig. 3.69).

Different cones are observed for different sets of d and θ for a particular value of n, and also for different combinations of θ and n for a particular value of d. As the film is of narrow width only a part of each

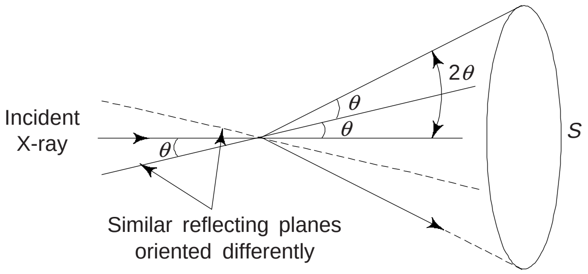


Fig. 3.69 A cone produced by reflection of X-rays from identical planes having different orientations

reflected cone is recorded on the film which is positioned around the camera. The recorded lines are in the form of pair of arcs (Fig. 3.69(b)).

The angle between the surface of cone, i.e. reflected ray and exit of beam is 2θ, i.e., the apex angle is 4θ. The angle θ corresponding to a particular pair of arcs is related to the distance S between the arcs is

$$4\theta \text{ (radiations)} = \frac{S}{R}$$

where R is the radius of the camera.

$$\therefore \qquad \underset{\text{(circumference)}}{2\pi R} \propto 360°$$

or

$$\frac{S}{2\pi R} = \frac{4\theta}{360°} = \frac{\theta}{90°}$$

or

$$S = \frac{2\pi R\theta}{90°}$$

If θ is measured in degrees, the above equation is modified as

$$4\theta \text{ (degrees)} = \frac{57.296\,S}{R}$$

The calculations can be made simpler by taking the radius of the camera in multiples of 57.296. For example, taking $R = 57.296$ mm, we get

$$\theta \text{ (degrees)} = S\text{(mm)}/4$$

Obviously, one fourth of the distance between the corresponding arcs of a particular pair in mm is a measure of the angle θ in degrees. Knowing all the possible θ's and considering only the first order reflections from all the possible planes, relation $2d \sin \theta = \lambda$ is used to calculate the interplanar spacings for various sets of parallel planes which contribute to these reflections. Thus, we have

$$d = \lambda/(2 \sin \theta)$$

These d values are used to determine the space lattice of the crystal structure.

The experimental arrangement of powder method (Debye-Scherrer Camera) is shown in Fig. 3.70. It consists of a cylindrical camera, called the Debye-Scherrer Camera , whose length is small as compared to the diameter.

Figure 3.71 shows the diffraction pattern of tungsten metal. The numbers on the peak are the indices of the planes which produce that peak. One can compute the interplanar spacings by measuring the value of θ and using Bragg's law.

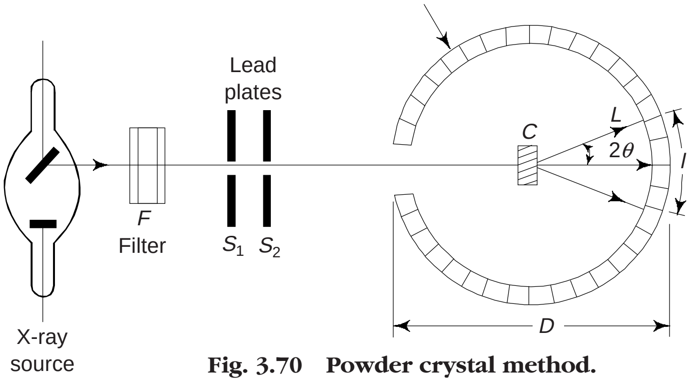


Fig. 3.70 Powder crystal method.

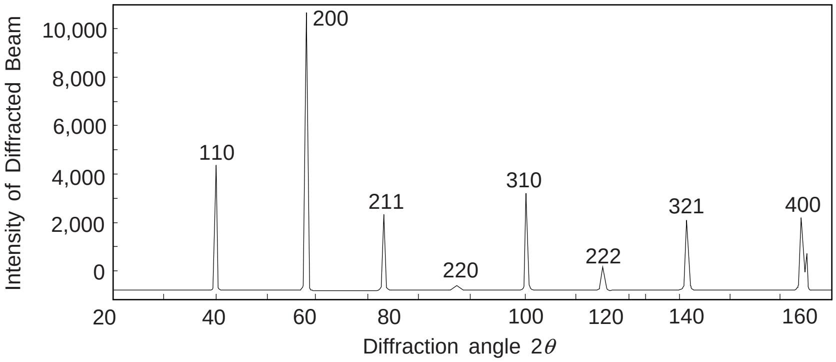


Fig. 3.71 Diffraction pattern intensity at various angles for tungsten

32. ELECTRON AND NEUTRON DIFFRACTION

Since waves are diffracted by a crystal, we can also use electrons and neutrons instead of X-rays, provided the wavelengths of matter waves associates with these particles are in the range of 1 to 10 Å.

Wavelength for neutron

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m_n E}} = \frac{6.626 \times 10^{-34}}{\sqrt{2 \times 1.675 \times 10^{-27} E}}$$

$$= \frac{1.145 \times 10^{-20}}{\sqrt{E}} = \frac{0.286}{\sqrt{E}}$$

where λ is in Angstroms and E in eV. Thus the energy of a neutron of wavelength of 1 Å is ≈ 0.08 eV. The thermal neutrons in a atomic reactor have the energy ≈ 0.08 eV and hence can be used for the study of crystal diffraction.

Wavelength for electron

$$\lambda = \frac{h}{\sqrt{2mE}} = \frac{6.626 \times 10^{-34}}{\sqrt{2 \times 9.109 \times 10^{-31} E}} = \frac{4.91 \times 10^{-19}}{\sqrt{E}}$$

$$= \frac{1.227}{\sqrt{E}}$$

where λ is in Angstroms and E is in kilo eV.

The thermal neutrons from the nuclear reactors have a large spread in energy and therefore we have to monochromatize them. These neutrons are allowed to fall on a crystal and a particular reflected beam is selected. This is allowed to fall on the test sample under examination. The intensity of the scattered beam is measured through a counter.

We must note that though the neutron diffraction studies are very similar to those of X-rays, there are many differences. X-rays are scattered by electrons, whereas neutrons are scattered by nuclei. Using this property of neutrons, one can study many low atomic weight elements by neutron diffraction where X-rays are incapable of giving precise results. We know that a neutron possesses a magnetic moment and hence has an additional scattering due to magnetic ordering. Obviously, neutron diffraction studies have become important in magnetic structure determinations, X-ray or electron diffraction studies do not reveal this structure. One can also study lattice vibrations using neutron diffraction technique.

Like X-rays or neutrons, electrons can also be used. However, the electrons are much more easily scattered and hence the penetration depth is small. This is why electron diffraction method is used in surface studies-bonds, defects, energy states, etc.

A comparison of X-rays and neutron diffraction techniques is made in Table 3.5. We can see that the essential difference between the two techniques lies in the magnetic order and isotope detection by neutron technique. On the other hand, X-rays are easier to produce as compared to neutrons which require a nuclear reactor.

Table 3.5 A comparative study of X-ray and neutron diffraction techniques

Property	*X-ray technique*	*Neutron diffraction technique*
1. Wavelength (λ)	~1.54Å (CuK_α)	~1.1Å (thermal neutrons in a reactor)
2. Energy for λ = 1Å (eV)	10^4	10^{-1}
3. Scattering amplitude	• Scattering amplitude increases regularly with atomic number	• Scattering amplitude varies irregularly with atomic number and depends on nuclear structure
4. Isotope detection	• Scattering among isotopes exhibit no difference	• Scattering amplitude is different for different isotopes
5. Magnetic Scattering	• No additional scattering is observed	• Additional scattering is observed

33. CRYSTAL GROWTH

Crystal growth generally occurs by means of following sequence of process: *(i)* diffusion of the molecules of the crystallizing substance through the surrounding environment (or solution) to the surface of crystal, *(ii)* diffusion of these molecules over the surface of the crystal to special sites on the surface, *(iii)* incorporation of molecules into the crystal of these sites, and *(iv)* diffusion of the heat of crystallization away from the crystal surface. The rate of crystal growth may be limited by any of these four steps. The initial formation of the centres from which crystal growth proceeds is known as *nucleation*.

Increasing the supersaturation of the crystallizing component or increasing the temperature independently increases the rate of crystal growth. However, in many physical situations the supersaturation is increased by decreasing the temperature. In these circumstances the rate of crystal growth increases with decreasing temperature at first, goes through a maximum, and then decreases. Often the growth is greatly retarded by traces of certain impurities.

After nucleation, the crystals in the medium grow isolated from one another for a time. However, if several differently oriented crystals are growing, they may finally impinge on one another, and intercrystalline (grain) boundaries will be formed. At relatively high temperatures, the average grain size in these polycrystalline aggregates increases with time by a process called grain growth, whereby the larger grains grow at the expense of the smaller.

Crystal Growth Methods

The advent of semiconductor-based technology generated a demand for large, high quality single crystals, not only of semiconductors but also of associated electronic materials. With increasing sophistication of semiconductor devices, an added degree of freedom in materials properties was obtained by varying the composition of major components of the semiconductor crystal over very short distances. Thin, multilayered single-crystal structures, and even structures that vary in composition both normal and lateral to the growth direction, are often required.

Description of various crystal growth techniques is beyond the scope of this book.

SUGGESTED READINGS

1. S.L. Kakani and C. Hemrajani, *Solid State Physics* (4th ed., 2004) Sultan Chand, New Delhi-2.
2. C. Kittel, *Introduction to Solid State Physics*, John Wiley, (7th ed.).
3. R.C. Evans, *An Introduction to Crystal Chemistry*, Cambridge Univ. Press (1964).
4. P.J. Brown and J.E. Forsyth, *The Crystal Structure of Solids* (1973).
5. F.C. Phillips, *Introduction to crystallography*, 4th ed. (1979).
6. J.H. Crawford. Jr. and L.M. Slifkin (eds.) *Point Defects in Solids*, vol. 1, 2 and 3 (1972, 75, 78).
7. B.D. Cullity and S.R. Stock, *Elements of X-ray Diffraction*, 3rd ed., Prentice Hall, Upper Saddle River, NJ (2001).
8. D. Brandon and W.D. Kaplan, '*Microstructural Characterization of Materials*', Wiley, New York, (1999).

REVIEW QUESTIONS

1. Explain the concepts of lattice, basis and crystal structure. How are they related?

2. Draw primitive cells corresponding to BCC and FCC unit cells.

3. Describe, how atoms are arranged in cubic crystal?

4. What is Bravais lattice? What is the maximum number of Bravais lattices possible? How will you account for the existence of thousands of structures from these lattices?

5. Write a short note on space lattice. [AMIE]

6. Draw net sketches of unit cells of simple cubic, BCC and FCC crystal structures. Calculate the number of atoms in each case. [AMIE]

7. The end-centred orthorhombic is one of the Bravais lattices but the end-centred tetragonal is not. Give reasons.

8. What is the number of atoms per unit cell in the FCC structure? [AMIE]

9. Give a brief account of crystal directions and crystallographic planes.

10. Draw the following:

(i) $[1\bar{1}1]$, $[1\bar{2}1]$ and $[0\bar{1}2]$ directions in cubic and tetragonal lattices.

(ii) $[\bar{1}11]$, $[1\bar{1}2]$ and $[\bar{2}10]$ planes in cubic and orthorhombic lattices.

11. Prove that in a cubic lattice the direction $[hkl]$ is perpendicular to the plane $(h\ k\ l)$.

12. Calculate the atomic packing factor of simple cubic, BCC and FCC crystals. [B.E.]

13. Draw a hexagonal closed packed unit cell. Calculate the number of atoms per unit cell. Calculate its atomic packing factor.

14. Draw a plan view of sodium chloride structure. In how many ways can this structure be interpreted?

15. Write the points of similarity and differences of the monoatomic SC, monoatomic BCC and CsCl structure.

16. What are Miller indices? How are they determined? [AMIE]

17. Draw atomic arrangement in planar surface imperfection with a boundary. [AMIE]

18. Discuss in brief the point defects. How these defect affect the properties of materials? [BE]

19. Explain Schottky and stacking faults.

20. Explain Miller Bravais system? How is this system used to find the index of different planes?

21. Using net sketches show (a) vacancy defect and interstitial defect in crystal. [AMIE]

22. State and explain the Bragg's law of X-ray diffraction. What are its uses? [AMIE]

23. Discuss Bragg's law of X-ray diffraction. Describe how the powder method is used for the determination of crystal structure. [AMIE]

24. Explain how with the help of X-ray diffraction techniques, the lattice dimensions are determined.

25. Write the different diffraction methods of determining crystal structure and describe any one in detail.

26. Write short notes on

(i) Bravais Lattices
(ii) Coordination number
(iii) Crystal defects
(iv) Crystal growth
(v) Crystal Whiskers
(vi) Grain boundaries
(vii) Crystallographic indices
(viii) Liquid crystals
(xi) Bragg's law
(x) X-ray spectrometer
(xi) Electron and neutron diffraction

PROBLEMS

1. Draw a $(1\bar{1}0)$ plane in a cubic unit cell. Show all the <111> directions that lie on this plane and give the Miller indices of each direction. ($[111]$, $[11\bar{1}]$, $[\bar{1}\bar{1}1]$ and $[\bar{1}\bar{1}\bar{1}]$)

2. A plane makes intercepts of 1, 2 and 3 Å on the crystallographic axes of an orthorhombic crystal with $a : b : c = 3 : 2 : 1$. Determine the Miller indices of this plane (9 3 1)

3. Aluminium has an FCC structure. Its density is 2700 kg/m^3 (sp. gravity = 2.7). Find the unit cell dimension and the atomic diameter. [4.05Å, 2.86 Å]

4. Show that the shortest centre to centre distance between any two atoms among the four atoms located inside the diamond-cubic unit cell is $a/\sqrt{2}$.

5. Determine the number of the nearest neighbours and the closest distance of approach in terms of lattice parameter for monoatomic SC, BCC and FCC structures. [6, a; 8, $a\sqrt{3}/2$; 12, $a/\sqrt{2}$]

6. Prove that c/a ratio for an ideal HCP structure is 1.633.

7. Zinc (HCP) has lattice parameters a and c as 2.66 Å and 4.95 Å respectively. Calculate the packing fraction and density of zinc, given the atomic radius and the atomic mass of Zn as 1.31 Å and 65.37 amu respectively. [62%, 7155 kg-m^{-3}]

8. Calculate the density of CsCl crystal. Given radii of Cs^+ = 1.65 Å and Cl^- = 1.81 Å [4380 kg-m^{-3}]

9. A crystal plane makes intercepts $4a$, $2b$ and $-2c$ on the crystallographic axes of a crystal whose unit cell has dimensions a, b and c. Find the Miller indices of the plane. [(122)]

10. The grating space for the (100) plane in a cubic crystal is found to be 2.84×10^{-10} m. Calculate the grating space for the (100) and (111) planes. [2.008×10^{-10} m, 1.603×10^{-10} m]

11. Show that the maximum occupied volumes by hard spheres arranged in the structures given below are:

Simple cubic	$\pi/6$
FCC	$\pi\, 2^{1/2}/6$
Diamond	$\pi\, 3^{1/2}/16$

12. Copper (FCC) has density of 8960 kg-m^{-3}. Calculate the unit cell dimension and radius of Cu atom, given the atomic mass of Cu = 63.54 amu. [3.61Å, 1.28 Å]

13. A crystal reflects monochromatic X-rays strongly when the Bragg's glancing angle for a first order reflection is 15°. Show that the glancing angles for the second and third order reflections of the same type are 31.2° and 50.9° respectively.

14. Copper (FCC) has a lattice parameter of 3.61Å. The first order Bragg reflection from (111) planes appears at an angle 21.7°. Show that the wavelength of X-rays used is 1.54 Å.

15. An aluminium crystal has dislocation density of 10^{10} m^{-2}. The shear modulus of aluminium is 25.94 GN-m^{-2}. Show that the elastic energy of line imperfections stored in the crystal is 10.6 J-m^{-3}.

16. Calculate the spacing between dislocations in a tilt boundary in FCC nickel. Given, the angle of tilt is 2°. [71.3 Å], [B.E.]

17. Calculate the dislocation energy per m^3 of copper (FCC) with a dislocation density of 10^{10} m^{-2}. Given, shear modulus of copper = 45 GN m^{-2} and lattice parameter = 3.61 Å [14.7 J-m^{-2}]

OBJECTIVE QUESTIONS

1. The number of lattice points in a primitive cell are
(1) 3/2 (2) 1/2 (3) 2 (4) 1 (4)

2. The number of lattice points in a rhombohedral unit cell is
(1) 1 (2) 2 (3) 4 (4) 16 (1)

3. The number of atoms present in the unit cell of HCP structure is
(1) 2 (2) 4 (3) 6 (4) 12 (3)

4. The minimum number of ions in the unit cell of an ionic crystal with FCC space lattice is
(1) 16 (2) 12 (3) 8 (4) 2 (3)

5. If r be the radius of the atom in a crystal crystallizing in the simple cubic structure, then the nearest neighbour distance is
(1) $8r$ (2) $4r$ (3) $2r$ (4) r (3)

6. The space lattices with two lattice parameters does not belong to the crystal system
(1) triclinic (2) rhombohedral (3) hexagonal (4) tetragonal (1)

7. The number of diad axes of symmetry elements that are present in a cubic crystal are
(1) 6 (2) 3 (3) 2 (4) 1 (1)

8. The nearest neighbour distance in BCC structure is
(1) $a/\sqrt{2}$ (2) $2a/\sqrt{3}$ (3) $a\sqrt{3}/2$ (4) $2a/\sqrt{3}$ (3)

9. The unit cell with three lattice parameter is
(1) monoclinic (2) tetragonal (3) triclinic (4) orthorhombic (3)

10. Which one of the following metals crystallizes in FCC structure
(1) zinc (2) sodium (3) aluminium (4) caesium chloride (3)

11. The atomic diameter of an FCC crystal (lattice parameters a) is
(1) $\frac{a}{2}$ (2) $a\sqrt{2}$ (3) $\frac{a\sqrt{2}}{2}$ (4) $\frac{a}{2\sqrt{2}}$ (3)

12. The number of Bravais lattices with two lattice points are
(1) 5 (2) 6 (3) 3 (4) 1 (1)

13. The packing factor of diamond cubic crystal structure is
(1) 90% (2) 45% (3) 34% (4) none of these (4)

14. The unit cell has a = 5 Å, b = 8 Å, c = 3 Å, $\alpha = 90°$, $\beta = 65°$ and $\gamma = 54°$. The space lattice for this unit cell is
(1) monoclinic (2) rhombohedral (3) triclinic (4) orthorhombic (3)

15. The faces in a tetragon are
(1) 24 (2) 12 (3) 6 (4) 2 (2)

16. Magnesium crystallizes in HCP structure. If the lattice constant is 0.32 nm, the nearest neighbour distance in magnesium is
(1) 0.64 nm (2) 0.32 nm (3) 0.16 nm (4) 0.8 nm (2)

17. Zinc crystallizes in HCP structure. If r is the radius of zinc atom, the height of the unit cell is
(1) $r\sqrt{3/8}$ (2) $2r\sqrt{8/3}$ (3) $r\sqrt{\frac{8}{3}}$ (4) $2r$ (2)

18. The lattice constant of a BCC unit cell with atomic radius of 1.24 Å is
(1) 1.432 Å (2) 2.864 Å (3) 1.754 Å (4) 1.432 Å (2)

19. The correct order of coordination number in BCC, FCC and HCP unit cells is
(1) 12, 8, 6 (2) 8, 12, 12 (3) 6, 8, 12 (4) 12, 8, 24 (2)

20. The number of sulphide atoms in the unit cell of zinc sulphide crystal is
(1) 6 (2) 4 (3) 2 (4) 1 (2)

21. The Miller indices of the line of intersection of a $(1\bar{1}1)$ and a $(1\bar{1}0)$ plane are
(1) [110] (2) $[1\bar{1}1]$ (3) $[\bar{1}\bar{1}0]$ (4) $[1\bar{1}0]$ (1, 3)

22. The Miller indices of the plane parallel to Y and Z axes are
(1) (0 0 1) (2) (1 1 1) (3) (0 1 0) (4) (1 0 0) (4)

23. The Miller indices of a plane in an orthorhombic unit cell ($a : b : c = 1 : 2 : 3$) making intercepts of 3, 2, 1Å on a, b, c axes respectively are
(1) (1 2 3) (2) (3 2 1) (3) (9 3 1) (4) (1 3 9) (4)

24. If (3 2 6) are the Miller indices of a plane, the intercepts made by the plane on the three crystallographic axes are
(1) (a, b, c) (2) $(2a, 3b, c)$ (3) $(a, 2b, 3c)$ (4) $(2a, b, 3c)$ (2)

25. In a simple cubic lattice $d_{100} : d_{110} : d_{111}$ is
(1) $\sqrt{6}:\sqrt{3}:\sqrt{2}$ (2) $\sqrt{6}:\sqrt{3}:2\sqrt{2}$
(3) 6 : 3 : 4 (4) $6:\sqrt{3}:\sqrt{2}$ (1)

26. THe acute angle between [101] and $[\bar{1}01]$ directions in a tetragonal crystal with $c/a = 1.5$ is
(1) 67.38° (2) 45° (3) 33.69° (4) 90° (1)

27. The maximum radius of the interstitial sphere that can fit into the void between the body centred atom of BCC structure is
(1) $r\left[\sqrt{3}-1\right]$ (2) $r\left[\sqrt{2}-1\right]$ (3) $r\left[(2\sqrt{3})-1\right]$ (4) $r/\sqrt{2}$ (3)

28. The interplanar spacing of the first reflecting plane (lowest θ) in an FCC crystal is
(1) $a/\sqrt{3}$ (2) $2a/\sqrt{3}$ (3) a (4) $2a/\sqrt{5}$ (2)

29. A plane in a unit cell is described by its Miller indices (632). The plane intersects x, y, z respectively at points whose distances from origin are
(1) 6, 3 and 2 units (2) 1/3, 2/3 and 1 units
(3) 1/3, 1/2 and 1/1 units (4) 1/6, 1/3 and 1/2 units (3)

30. If the first reflection from an FCC crystal has a Bragg angle $\theta = 21.5°$, the θ corresponding to second reflection is
(1) 13.5° (2) 18.5° (3) 25° (4) 36.8° (3)

31. The Miller indices of the fifth reflection in an FCC crystal is
(1) 111 (2) 222 (3) 311 (4) 331 (2)

32. The number of atoms per unit area of the plane (010) of a simple cubic crystal is
(1) $1/2r$ (2) $1/4r^2$ (3) $2r^2$ (4) $4r^2$ (2)

33. The highest linear density, i.e., the densest direction in FCC and BCC crystals are respectively
(1) <110> and <111> (2) <111> and <111>
(3) <100> and <110> (4) <111> and <110> (1)

34. In a unit cell whose lattice constant is a, the distance between two $\{hkl\}$ planes is
(1) $\sqrt{h^2+k^2+l^2}$ (2) $a/\sqrt{h^2+k^2+l^2}$
(3) $1/\sqrt{h^2+k^2+l^2}$ (4) $hkl/\sqrt{h^2+k^2+l^2}$ (2)

35. No reflection will be observed from a DC crystal ($a = 3$Å), if the wavelength of X-rays is greater than
(1) 2.5 Å (2) 3.0 Å (3) 3.46 Å (4) 3.15 Å (3)

36. The second order reflection from (200) planes coincides with the first order reflection from
(1) (400) (2) (300) (3) (200) (4) (100) (1)

SHORT QUESTIONS ANSWERS

1. What is difference between crystalline and non crystalline solids?

Ans. Atoms in crystalline solids are positioned in an orderly and repeated pattern whereas in non crystalline or amorphous materials there is random and disordered atomic distribution.

2. How various crystal structures are specified?

Ans. The various crystal structures are specified in terms of parallelopiped unit cells, which are characterized by geometry and atom positions within.

3. What is a space lattice?

Ans. A space lattice is an infinite array of points, all with identical surroundings.

4. How a crystal structure is obtained?

Ans. A crystal structure is obtained by combining a space lattice with a basis. The basis must give the number of atoms per lattice point, their types, mutual orientations and distances of separation.

5. How many types of space lattices are there?

Ans. Fourteen. However, crystal structure run into thousands.

6. What are the three relatively simple crystal structures in which most common metals exist?

Ans. FCC, BCC and HCP.

7. Mention two features of crystal structure.

Ans. *(i)* Coordination number (or number of nearest-neighbour atoms) and *(ii)* atomic packing (the fraction of solid sphere volume in the unit cell). Coordination number and atomic packing fraction are the same for both FCC and HCP crystal structures, each of which may be generated by the stacking of closed-packed planes of atoms.

8. How crystal directions and crystal planes are denoted?

Ans. These are denoted by Miller indices. A family of crystal directions or planes includes all possible combinations of the indices, both positive and negative. For a given crystal structure, planes having identical atomic packing yet different Miller indices belong to the same family.

9. What is a single crystal?

Ans. Single crystals are materials in which the atomic order extends uninterrupted over the entirety of the specimen; under some circumstances, they may have flat faces and regular geometric shapes. We may note that vast majority of crystalline solids, however, are polycrystalline, being composed of many small crystals or grains having different crystallographic orientations.

10. How the space lattice and its dimensions can be determined from the powder method for X-ray diffraction?

Ans. From the position of lines in a powder pattern and from the extinction rules for different cubic crystals, the space lattice and its dimensions can be determined.

11. What do understand by crystal imperfections?

Ans. All solid materials contain large number of imperfections or deviations from crystalline perfection, and several types of them may be classified according to their geometry, and size.

12. What are point defects?

Ans. Points defects are those associated with one or two atomic positions, including vacancies (or vacant lattice sites), self-interstitials (host atoms that occupy interstitial sites), and impurity atoms. They are Frenkel and Schottky imperfections in ionic crystals.

13. What are the edge and screw dislocations?

Ans. These are the limiting types of line imperfections. One can resolve any general dislocation into edge and screw components.

14. How a solid solution may be formed?

Ans. A solid solution may be formed when impurity atoms are added to a solid, in which case the original crystal structure is retained and no new phases are formed.

15. What do you understand by substitutional solid solutions?

Ans. For substitutional solid solutions, impurity atoms substitute for host atoms, and appreciable solubility is possible only when atomic diameters and electronegativities for both atom types are similar, when both elements have the same crystal structure, and when the impurity atoms have a valence that is the same as or less than the host material.

16. What are interstitial solid solutions?

Ans. These form for relatively small impurity atoms that occupy interstitial sites among the host atoms.

17. How the composition of an alloy may be specified?

Ans. This may be specified in weight percent or atom percent. The basis for weight percent computations is the weight (or mass) of each alloy constituent relative to the total alloy weight. Atom percents are calculated in terms of the number of moles for each constituent relative to the total moles of all the elements in the alloy.

18. What are dislocations?

Ans. These are one-dimensional crystalline defects of which there are two pure types: edge and screw. Edge and screw dislocations are the limiting types of line imperfections. An edge may be thought of in terms of the lattice distortion along the end of an extra half plane of atoms; a screw, as a helical planar ramp. Any general dislocation can be resolved into edge and screw components. For mixed dislocations components of both pure edge and screw are found.

19. What do you understand by Burgers vector?

Ans. The magnitude and direction of lattice distortion associated with a dislocation are specified by its Burgers vector i.e., the elastic strain energy.

20. What are the relative orientations of Burgers vector and dislocation line are?

Ans. *(i)* Perpendicular for edge, *(ii)* Parallel for screw, and *(iii)* neither perpendicular nor parallel for mixed.

21. What are the other imperfections?

Ans. These include interfacial defects [external surfaces, grain boundaries (both small and high angle), twin boundaries, etc.], volume defects (cracks, pores, etc.), and atomic vibrations. Each type of imperfection has some influence on the properties of a material.

22. How microscopic defects and structural elements of materials are examined?

Ans. These are observed with the help of microscopes. Both optical and electron microscopes are employed, usually in conjunction with photographic equipment. Transmissive and reflective modes are possible for each microscope type; preference is detected by the nature of the specimen as well as the structural element or defect to be examined.

23. What is the main advantage of *scanning probe microscopic techniques?*

Ans. These techniques generate topographical maps representing the surface features and characteristics of the specimen. Using these techniques examinations on the atomic and molecular levels are possible.

24. How grain size of polycrystalline materials determined?

Ans. Grain size of polycrystalline materials is frequently determined using *photomicrographic* techniques. Two methods are commonly employed: *(i)* intercept and *(ii)* standard comparison charts.

4

Bonds in Solids

1. INTRODUCTION

In nature one comes across several types of solids. Many solids are aggregates of atoms. The arrangement of atoms in any solid material is determined by the character, strength and directionality of the chemical binding forces, cohesive forces or chemical bonds. We call these binding forces as *atomic interaction* forces. The atoms, molecules or ions in a solid state are more closely packed than in the gaseous and liquid states and are held together by strong mutual forces of attraction and repulsion. One can describe the atomic arrangement in elements and compounds on the basis of this.

The type of bond that appears between atoms in crystal is determined by the electronic structure of interacting atoms. Atoms in a crystal approach one another to certain distances at which the crystal is in the state of the highest thermodynamic stability. These distances depend on the interaction forces that appear in crystals. The attractive forces between atoms are basically electrostatic in origin. Its magnitude is proportional to some power of the interatomic distance r. The different types of bonding depends on the electronic structure of the atoms concerned and hence directly related to the periodic table. The type of bonding within a material plays a major role in determining the electrical, chemical and physical properties of the material. The repulsive forces which come into existence when the distance between the atoms is decreased to such an extent that they are very close to one another and increase more intensively than attractive forces do. The repulsion between positively charged nuclei also contributes to the repulsive forces. The magnitude of the total repulsive force is also proportional to some power of r. A state of equilibrium is reached when these attractive and repulsive forces balance and this happens when two atoms or molecules are at an equilibrium distance r_0.

2. TYPES OF BONDS

Seitz in 1940 classified solids into five types according to the bonding of atoms, which has become a generally adapted classification (Table 4.1)

Table 4.1 Classification of solids according to the bonding of atoms

Solids	*Type of bond*	*Formation*	*Binding energy (eV/atom)*	*Typical examples*
1. Covalent	Covalent, atomic or homopolar bonds	Electron shared between two atoms	2–6	Carbon(diamond) Ge, Si, SiC, BN etc.
2. Ionic	Ionic or electrostatic bonds	Electron transfer and Coulomb interaction between cations and anions	0–2	Alkali halides
3. Metals	Metallic	Freely moving electrons in an array of positive ions	1–5	All metals and alloys

(*Contd.*)

Table 4.1 *(Contd.)*

	Solids	*Type of bond*	*Formation*	*Binding energy (eV/atom)*	*Typical examples*
4.	Molecular (Vander waals)	Molecules between pairs	Weak attractive forces due to dipole-dipole interaction	0.002–0.1	Nobel gases
5.	Hydrogen bond	Hydrogen atom attracted between two other atoms	Electrostatic bond of H-atom with an electronegative atom	0.5	ice, organic compounds, biological materials

The atoms of different types come closer and join together during chemical reaction and usually these bonds are referred to as chemical bonds. There are basically two groups which classify common bonds on the basis of strength, directionality of bonding forces, cohesive forces (chemical bonds) and the character of any solid material: These are

(i) Primary Bonds: These are inter-atomic bonds in which electrostatic force holds the atoms together. Relatively large atomic forces develop in these bonds, making them more stable and imparting high strength. They have bond strength varying from 1-1.5 eV. these bonds are also known as attractive bonds. The inter-atomic distance is usually 1-2*r*. Following three types of primary bonds are found in different materials

(i) Ionic or electrostatic bonds
(ii) Covalent, atomic or homopolar bonds
(iii) Metallic bonds

The principal types of bonding is shown in Fig. 4.1.

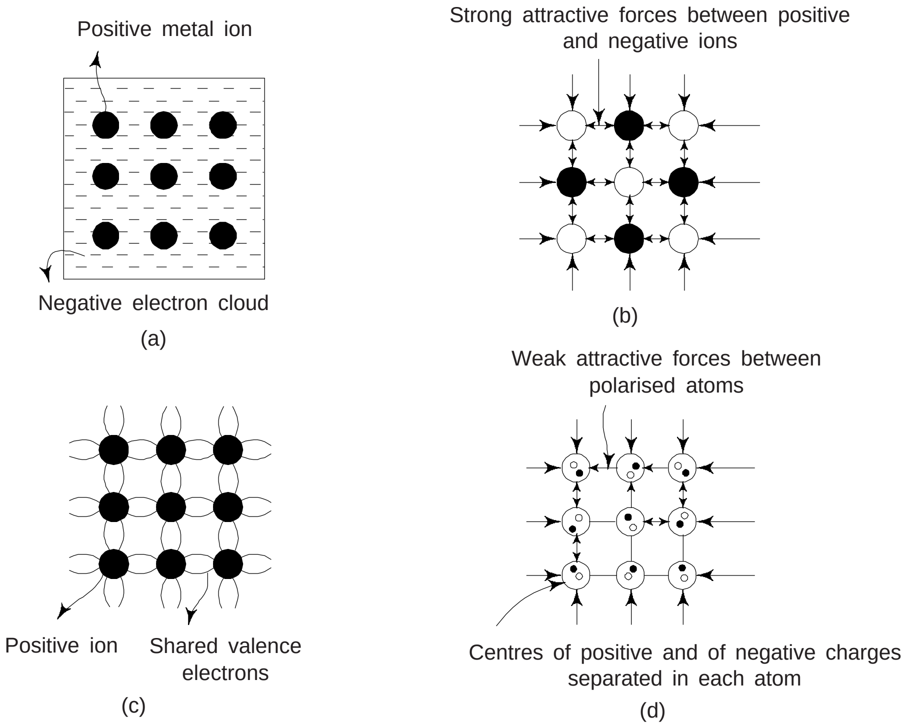


Fig. 4.1 The four principal types of chemical bonding in solids (a) Metallic bonding (b) Ionic bonding (c) Covalent bonding (d) Vander Waals bonding

(ii) Secondary Bonds: A *secondary bond* is an intermolecular bond and weaker and less stable than primary bond. In this type of bonds, the forces hold the molecules together. These secondary bonds result from intermolecular or dipole attractions. Asymmetrical electron distribution in molecules and atoms creates *dipoles*. Due to dipoles the molecules are attracted towards other molecules having opposite dipoles. Similar asymmetric electron distribution in atoms causes dipole to exist in them. We must note that the atomic dipolar bonds are weaker than molecular dipole bonds and moreover atomic dipolar bonds keep on fluctuating as electron distribution in atomic outer shells keeps changing. Common examples of secondary bonds are Vander Waals bonds and hydrogen bonds.

3. FORCES BETWEEN ATOMS: MECHANISM OF BOND FORMATION AND BOND ENERGY

We have already stated that the interatomic forces exist amongst the atoms of a crystal and are responsible for holding atoms together to form solid structure. We have also said that the forces between atoms can be of two kinds: *(i)* attractive forces which keep the atoms together, and *(ii)* repulsive forces which come into play when the solid is compressed. The force of attraction between atoms brings them closer together until the individual electron clouds begin to overlap and a strong repulsive force arises to comply with Pauli's exclusion principle. The attractive forces decrease the potential energy of the system and repulsive forces increase it. Obviously, the net energy of the system is equal to the algebraic sum of these two energies. When the attractive force and the repulsive force between any two atoms are equal, the two atoms should be in a stable situation with a minimum potential energy. The forces between two atoms or ions in a solid as a function of their distance of separation r are shown in Fig. 4.2.

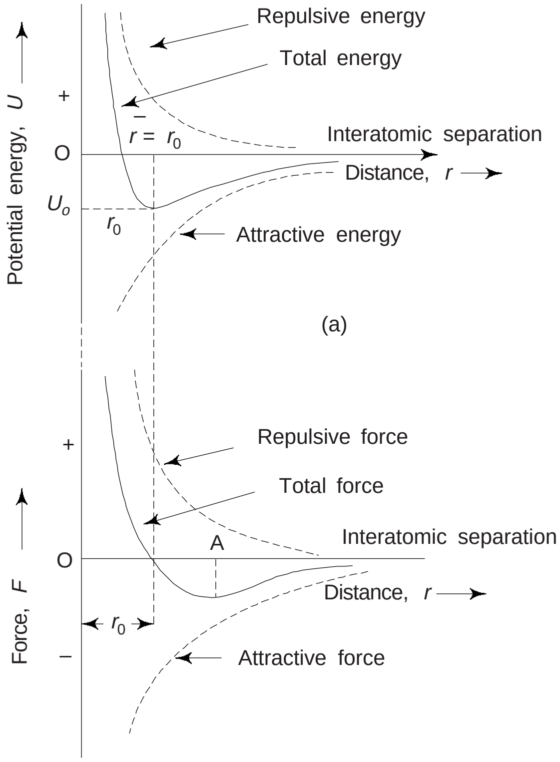


Fig. 4.2 Potential energies (a) and interatomic forces (b) versus interatomic distance, *r* in a system of two atoms

Let us consider two atoms in their ground states and infinitely far apart so that they do not interact with each other to form a solid, the potential energy is zero, since this is inversely proportional to some power of the distance of separation. Assuming that the atoms consists of moving electric charges, one of the following two things can happen as the atoms approach each other. Either atoms attract or repel each other. The potential energy due to attraction is negative, since the atoms do the work of attraction. The repulsive energy between the atoms is positive since the external work must be done to bring the atoms together and it is inversely proportional to some power of interatomic distance r. We must note that the net potential energy is sum of both terms. We may also note that the *bond energy* between atoms will tend to stay until energy equal to the bond energy is supplied. The *bond energy* is defined as the decrease in energy when atoms separated by infinite distance are brought together at equilibrium distance. The bond energy is normally described per mole basis. However, the dissociation energy will be equal to the bond energy but of opposite sign.

Since the attractive forces decrease the potential energy of the system and the repulsive forces increase it, the net energy of the system is equal to the algebraic sum of these two energies and one can express it as

$$\begin{aligned} U &= \underset{\text{(due to attraction)}}{\text{decrease in potential energy}} + \underset{\text{(due to repulsion)}}{\text{increase in potential energy}} \\ &= U_{\text{att}} + U_{\text{rep}} \\ &= -\frac{A}{r^m} + \frac{B}{r^n} \quad (n > m) \end{aligned} \tag{1}$$

where A is proportionality constant for attraction, and B for repulsion. m and n are also constants and depend upon the nature of the participating atoms. Generally $n > m$ which indicates that the increase in repulisve energy is faster than the increase in attractive energy particularly for very small values of inter-atomic distance. Eq. (1) indicates that the magnitudes of both attractive and repulsive energies increase with decrease in atomic distance. The repulsive forces are, therefore, known as *short range forces*. Obviously, the repulsive interaction between the nuclei becomes appreciable only for very small distances. r in eq. (1) is the centre to centre spacing between atoms.

The variations of the attractive energy, repulsive energy and total energy versus interatomic distance are shown in Fig. 4.2(a). In order to give a minimum potential energy condition at the equilibrium distance r_0, which signifies the formation of a stable chemical bond, m must be greater than n, so that d^2U/dr^2 is positive. Since the attractive forces in interatomic bonds are largely electrostatic, m is usually 2 as per Coulomb's law and the value of n is usually 7 to 10. The resultant force of interaction F is obtained as

$$\begin{aligned} F &= -\frac{dU}{dr} \\ &= -\frac{mA}{r^{m+1}} + \frac{nB}{r^{n+1}} \end{aligned} \tag{2}$$

This gives the net or total force between the two atoms placed at a distance r from each other. The first term on the right hand side represents the attractive force and the second term represents the repulsive force. Figure 4.2(b) shows the variations of the attractive force, repulsive force and total force with interatomic distance. At the equilibrium spacing

$$F(r) = 0 \text{ when } r = r_0$$

Thus

$$0 = -\frac{mA}{r_0^{m+1}} + \frac{nB}{r_0^{n+1}}$$

or

$$B = +A\left(\frac{m}{n}\right)\frac{r_0^{n+1}}{r_0^{m+1}} = +A\left(\frac{m}{n}\right) r_0^{n-m} \tag{3}$$

or

$$r_0^{n-m} = +\left(\frac{B}{A}\right)\frac{n}{m}$$

From Eq. (1), the energy at the equilibrium distance r_0 becomes

$$U_0 = -\frac{A}{r_0^m} + \frac{B}{r_0^n} = \left(-\frac{A}{r_0^m}\right)\left(1 - \frac{m}{n}\right) \tag{4}$$

Since $m \neq n$, $U_0 \neq 0$, it follows that, although the attractive force is equal to the repulsive force at equilibrium distance, the attractive energy is not equal to the repulsive energy.

For $n >> m$, $U_0 = -\frac{A}{r_0^m}$,

Obviously, the total energy is essentially the energy of attraction. Also, if the total energy U has to be minimum at $r = r_0$, then

$$\left.\frac{d^2U}{dr^2}\right|_{r=0} > 0$$

i.e.

$$-\frac{m(m+1)A}{r_0^{m+2}} + \frac{n(n+1)B}{r_0^{n+2}} > 0$$

Using Eq. (3), we find that this condition is satisfied only for $n > m$. It indicates that the repulsive forces should be of short range than the attractive forces.

The representation of energy by a power function of the type of Eq. (1) is, in general not quite accurate. It is, however, useful to draw some important qualitative conclusions about the bonding of atoms in solids:

(i) Cohesive Energy: The energy corresponding to the equilibrium position ($r = r_0$), denoted by $U(r_0)$ or U_0 is called the *binding energy* or the *energy of cohesion* of the molecule. This is the energy required to dissociate the two atoms of the molecule into an infinite separation. Obviously, this energy is also called the energy of *dissociation* and given by Eq. (4). This much energy is required to separate the atoms of a diatomic molecule to an infinite distance apart. This is generally of the order of a few electron volts. The cohesive energy may also be defined as the energy released when two atoms are brought close to each other at an equilibrium distance r_0. Larger the energy released, more stable the bond formed and hence more stable is the crystal structure.

When the atoms in a molecule are separated by only a few atomic diameters the repulsive forces between like charges start to assert themselves. At the equilibrium separation, the forces of attraction is just equal to the forces of repulsion and the potential energy is at a minimum (U_0). This is the bonding energy of the molecule or system and is approximately equal to the heat of dissociation of molecule.

In crystals every atom is surrounded by number of other atoms and the simple expression for attractive and repulsive energies given by Eq. (1) is not applicable. To have the exact picture of these energy terms, one must investigate their origins in detail. It further requires the knowledge of distributions of charge, particularly of valence electrons of the atoms. In certain crystals, the valence electrons are transferred from one atom to the other during bond formation. In some crystals, the sharing of electrons takes place among the neighbouring atoms while in some others the valence electrons behave as free electrons and move from one part of the crystal to another. There are also other types of electronic interactions present in the crystals. The nature of crystals formed depends upon the nature of these interactions or bonds. The strength of a bond is best measured by the energy required to break the bond. It is the amount of heat which must be supplied to vaporize the solid, and hence separate the constituent atoms. The energy required for the formation of one mole of a substance from its ions or atoms is called the *bond energy*. The melting points of the elements also found to depend on the strength of the bond, i.e., the stronger the bond the higher are the melting points. These bonds may be broadly classified into five categories. (Table 4.1).

Bond Length

The separation distance (r) between the centre to centre of two bonding atoms which gives the stable bond is called *bond length*. Obviously, greater the force of attraction between the two bonding atoms, smaller will be the bond length (r). We must note that the primary bonds are stronger than the secondary bonds. One can use the bond length for finding the diameter of ions or atoms as follows:

$$r = r_c + r_a, \text{ where } r_c \to \text{radii of a cation and } r_a \to \text{radii of anion}$$

When $r_c = r_a$, $r = 2r_c = 2r_a$ (atomic diameter of atom)

In the case of two similar atoms, the bond length is equal to the atomic diameter. We must note that the bond length of a primary bond is around 1-2 Å, and for secondary bond, it is about 2-5 Å.

4. IONIC BONDING

Perhaps *ionic* or *hetropolar* bonding which is formed by the actual transfer of electrons from one atom to the other so that each atom acquires a stable configuration similar to the nearest inert gas atoms, is the simplest type of chemical bonding to visualize since it is almost totally electrostatic in nature. Ionic bonding occurs between electropositive elements (metals, i.e., those elements on the left side of the periodic table) and the electronegative elements (non-metals; i.e. on the right hand of the periodic table). These bonds are formed mainly in inorganic compounds, e.g. sodium chloride (common salt NaCl), MgO, CuO, CrO_2, and MoF_2. In MgO the ions are doubly ionized leading to a stronger interatomic bond and hence a higher melting point (~2800°C), compared to 800°C for NaCl. Examination of the formation of ionic bonds, e.g., cupric oxide, chromous oxide and molybdenum fluoride, reveals that the metallic element need not be from Group I or II but that any metal may get ionized by losing its valence electrons.

The criterion for ionic bonding is the difference in *elect-ronegativity* (the tendency to acquire electrons). An ionic bond is truly speaking the attractive force existing between a positive and negative ion when they are brought into close proximity. These ions, of course, are formed when the ato-ms involved lose or gain electrons in order to stabilise their outer shell electron configuration. The cohesive energy of ionic crystals is quite large; it is of the order of 5 to 10 eV.

A good example of ionic crystals is the crystal of sodium chloride (NaCl). The electronic configurations of white soft metal Na and Cl atoms are as follows:

Na : $1s^2\ 2s^2\ 2p^6\ 3\ s^1$ [K (2) L (8) M (1)]
Cl : $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^5$ [K (2) L (8) M (7)]

Na atom has a low ionisation energy and hence easily loses an electron; where as Cl atom has a high electron affinity and strongly tends to acquire an electron. The reaction can be represented by

$$Na + Cl \to Na^+ + Cl^- \to NaCl.$$

Since chlorine exists as molecules, we can write the chemical reaction as

$$2Na + Cl_2 \to 2Na^+ + 2Cl^- \to 2NaCl.$$

After the transfer of an electron from 3s orbital of Na to $3p$ orbital of Cl, the electronic configurations become

Na^+ : $1s^2\ 2s^2\ 2p^6$ (same as Ne)
Cl^- : $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6$ (same as Ar)

Schematic representation of the formation of an ionic molecule of sodium chloride is shown in Fig. 4.3. Figure 4.4

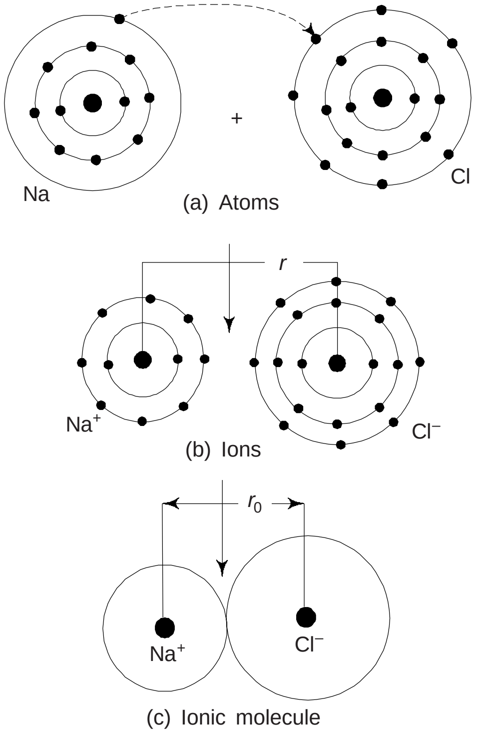


Fig. 4.3 **Schematic representation of the formation of an ionic molecule of NaCl (Common salt).**

shows the ionic crystal structures of NaCl and CsCl crystals, some other examples of ionic crystals are LiF, KCl, CsCl, Al_2O_3, etc.

The reaction for MgO in which ions are doubly ionized is represented by

$$2\ Mg + O_2 \rightarrow 2\ Mg^{++} + 2O^{--} \rightarrow 2\ MgO$$
$$Mg + Cl_2 \rightarrow Mg^{++} + Cl^{-} \rightarrow MgCl_2$$

Since ionic crystals have large binding energy, these are, in general, hard and exhibit high melting point and boiling points. At normal temperatures, ionic crystals are poor conductors of electricity but the conductivity increases with increase in temperture owing to the increased mobility of ions. Ionic crystals are generally transparent to visible light but exhibit characteristic absorption peaks in infrared region. Ionic crystals are also soluble in polar solvents, e.g. water. Ionic crystals have high melting and boiling points, high modulus of elasticity, and low coefficients of compressibility and linear expansion.

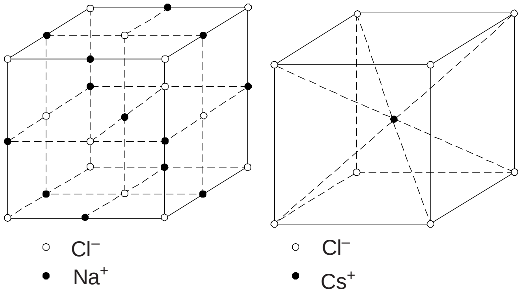


Fig. 4.4 Ionic crystal structure of Nacl and Cs Cl.

4.1 Ionic Solids

These contains atoms of different sizes which are held together by directional bonds. This group can be exemplified by crystals NaCl, CsCl (Fig. 4.4.) and FeO (Fig. 4.5).

Lattice of FeO consists of negatively charged oxygen ions and positively charged iron ions. Ionic solids cannot form close-packed structure life FCC and HCP, because the atoms are not of the same size. In these solids, the coordination number depends on the ratio of the radii of the metallic and non-metallic ions, since each ion tends to attract as many ions of the opposite sign as possible. Ions are arranged in a lattice like spheres of different diameters.

The radius of non-metallic ion is larger than that of a metallic one, because of which metallic ions fill in voids in the crystal lattice formed by ions of non-metal. In ionic solids, the coordination numbers of opposite-sign ions that surround a given ion.

The geometry of packing of spheres of different diameter suggests the following ratios of the radius of a metal, R_m (r_c), to that of non-metal, R_{nm} (r_a), and corresponding coordination numbers:

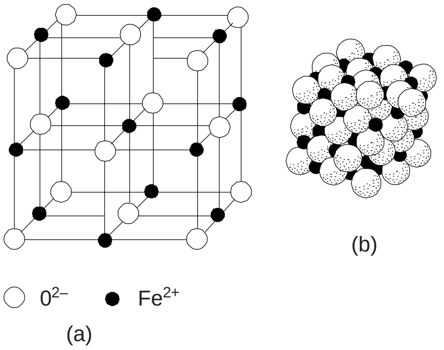


Fig. 4.5 Crystal lattice of FeO: (a) schematic (b) three dim-ensional view.

Table 4.2

C (coordination number)	8	6	4	2
R_m/R_{nm} (or r_c/r_a)	1-0.73	0.73-0.41	0.41-0.22	0.22

The ratio of ionic radii may change on heating, since the ionic radius of a non metal then increases more intensively than that of a metallic ion. This causes a change in crystal structure, i.e. a polymorphous transformation.

Most common ionic crystal structures are NaCl, CsCl, and ZnS (zinc blende) structures. The main factor which determines the structure adopted is the ratio of the ionic radii r_c/r_a ($r_c \rightarrow$ ionic radius of a cation and r_a of an anion), since like ions must not touch because of the large electrostatic repulsion. The ionic radii ratio r_c/r_a = 0.5 for NaCl, 0.92 for CsCl and 0.48 for ZnS. The coordination number for NaCl is 6, for CsCl is 8 and for ZnS is 4.

The force of attraction between ions in ionic solids is the Coulomb's force: $-e^2/4\pi\,\varepsilon_0\,r_0^2$, where r_0 is the equilibrium separation. One can use this force to calculate the total energy of a single ion in the presence of other ions. For e.g., in the NaCl structure, the central Na^+ ion is surrounded by six nearest neighbours of opposite electric charge at a distance r_0, 12 next neighbours of like charges at a distance $\sqrt{2}\ r_0$, 8 second nearest neighbours of unlike charges at a distance $\sqrt{3}\ r_0$, etc.

We can write the potential energy of a single molecule as

$$U_0 = -\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r_0^2} + \frac{B}{r_0^n}$$

Now, the total energy of the central Na^+ ion will be

$$U_0 = -\frac{6}{4\pi\varepsilon_0}\frac{e^2}{r_0} + \frac{12}{4\pi\varepsilon_0}\frac{e^2}{\sqrt{2}r_0} - \frac{8}{4\pi\varepsilon_0}\frac{e^2}{\sqrt{3}r_0} + \ldots + \frac{c}{r_0^n}$$

Where c is some constant representing the net repulsion due to all the overlapping electron clouds. We can write the potential energy as

$$U_0 = \frac{M}{4\pi\varepsilon_0}\frac{e^2}{r_0} + \frac{c}{r_0^n}$$

Where M is a constant known as the *Madelung constant*. Madelung constant is a geometric term. Knowledge of such electrostatic energies as given by the Madelung constant is of importance in the calculation of the cohesive energies of ionic crystals. Madelung constant (M) is characteristic of the lattice structure but independent of the dimensions of the lattice.

The Madelung constants for a number of common ionic crystal structures are given in Table 4.3. The calculation of Madelung constants requires some care since they are given by slowly and convergent series. In case of NaCl structure, the Madelung constant is

$$\begin{aligned} M &= \frac{6}{\pi} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{\sqrt{4}} + \frac{24}{\sqrt{5}} \ldots \\ &= 1.7475 \end{aligned}$$

Table 4.3 Madelung constants for some common ionic crystals

Crystal structure	*Madelung constant (M)*
Sodium Chloride (NaCl)	1.7476
Cesium Chloride (CsCl)	1.7627
Zinc blende, α-ZnS	1.6381
Wurtzite, β-ZnS	1.641
Fluorite CaF_2	5.0388
Cuprite, CuO_2	4.1155
Rutile TiO_2	4.816
Anatase TiO_2	4.800
Corundum Al_2O_3	25.0312

4.2 Characteristics of Ionic Solids

(i) Ionic solids are generally rigid and crystalline in nature

(ii) The bond energy of ionic crystals is close in magnitude to that of covalent crystals and exceeds the bond energy of metallic and more so, of molecular crystals. For that reason, these have high melting

and boiling points, high modulus of elasticity, and low coefficient of compressibility and linear expansion.

(iii) In ionic bonding, a metallic element loses from the outer electron shell of its atom a number of electrons equal to its valency (numerical). Obviously, an electrostatic attraction between positive and negative ions occurs.

(iv) In ionic crystal each positive ion attracts all neighbouring negative ions and vice versa, the bond itself is non-directional.

(v) Ionic solids are generally non-conductors of electricity.

(vi) Ionic solids are highly soluble in water but insoluble in organic solvents.

(vii) In many ionic crystals, a fraction of covalent bond can also be present. Ions become polarized under the action of their electromagnetic fields and produce a combined ionic-covalent bond. Polarization deforms the electron zones, so that ions lose their spherical symmetry. With a change of the fraction of covalent bond, the properties of ionic crystals are changed. In crystals formed by elements of various groups, such as $A^{I}\,B^{VII}$, $A^{II}\,B^{VI}$, $A^{III}\,B^{V}$, $A^{IV}\,B^{IV}$, the fraction of covalent bond increases from the first to the last. In compounds formed by elements of the fourth group ($A^{IV}\,B^{IV}$), the fraction of covalent bond is 90% and they are essentially covalent crystals. An increase of the fraction of the covalent bond increases the electric conductivity, so that $A^{I}\,B^{VII}$ crystals are dielectrics and $A^{IV}\,B^{IV}$ crystals, semiconductors.

Example 1 *The potential energy of a system of two atoms is given by the relation*

$$U = -\,A/r^2 + B/r^{10}$$

A stable molecule is formed with the release of 8 eV energy when the interatomic distance is 2.8 Å. Find A and B and the force needed to dissociate this molecule into atoms and the interatomic distance at which the dissociation occurs.

Solution We have the potential energy of the system

$$U = -A/r^2 + B/r^{10} \qquad (i)$$

$$\therefore \qquad F = -\frac{dU}{dr} = -\frac{2A}{r^3} + \frac{10B}{r^{11}} \qquad (ii)$$

(ii) At equilibrium distance r_0, $F = 0$. Therefore, Eq. *(ii)* gives

$$A = \frac{5B}{r_0^8} = \frac{5B}{(2.8\times10^{-10})^8} = 1.32\times10^{77}B \qquad (iii)$$

∴ Bond energy or the dissociation energy, U_o is obtained from Eq. *(i)* by putting $r = r_0$,

$$U_o = -\frac{A}{r_0^2} + \frac{B}{r_0^{10}} = -\frac{A}{r_0^2}\left(1 - \frac{B}{Ar_0^8}\right)$$

Using Eq. *(iii)*, one obtains

$$U_o = -\left(\frac{A}{r_0^2}\right)\left(1 - \frac{1}{5}\right) = -\frac{4}{5}\,\frac{A}{r_0^2}$$

Negative sign signifies the release of energy.

$$\therefore \qquad A = \left(\frac{5}{4}\right)(2.8\times10^{-10})^2\,(8.0\times1.6\times10^{-19})$$

$$= 1.256\times10^{-37}\ \text{J-m}^2$$

and $\qquad B = 9.52\times10^{-110}\ \text{J-m}^2$

The critical interatomic distance r_c is obtained from the condition

$$\left.\frac{dF}{dr}\right|_{r=r_c} = 0$$

From Eq. *(iii)* $\dfrac{6A}{r_c^4} - 110\dfrac{B}{r_c^{12}} = 0$

It yields $$r_c = (110\ B/6A)^{1/8} \tag{iv}$$

$$= \left(\frac{110 \times 9.52 \times 10^{-115}}{6 \times 1.256 \times 10^{-37}}\right)^{1/8}$$

$$= 3.3 \times 10^{-10}\ \text{m}$$

Putting $r = r_c$ in Eq. *(ii)*, the force required to dissociate the molecule is

$$F = -\left(\frac{2}{r_c^3}\right)\left(A - \frac{5B}{r_c^8}\right)$$

Using Eq. *(iv)*, it becomes

$$F = -\left(-\frac{2}{r_c^3}\right)\left[A - 5B\left(\frac{6A}{110B}\right)\right]$$

$$= \left(\frac{2}{r_c^3}\right)\left(\frac{8A}{11}\right)$$

$$= -\frac{2}{(3.3 \times 10^{-10})^3}\left(\frac{8 \times 1.256 \times 10^{-37}}{11}\right)$$

$$= -5.08 \times 10^{-9}\ \text{N.}$$

The negative sign indicates that the force existing at $r = r_c$ is the force of attraction. The force needed to dissociate the molecule is 5.08×10^{-9} N.

Example 2 *The lattice energy of KCl crystal containing N-molecules of KCl is given by* $U = -N(Mq^2/R - B/R^n)$. *Find the repulsive exponent n. Given: nearest neighbour equilibrium distance,* $R_0 = 3.14$ *Å, compressibility of KCl,* $K = 5.747 \times 10^{-11}\ m^2/N$ *and Madelung constant* $M = 1.748$.

Solution $$U = -N\left(\frac{Mq^2}{4\pi\varepsilon_0 R} - \frac{B}{R^n}\right)$$

Differentiating, we get

$$\frac{dU}{dR} = \frac{NMq^2}{4\pi\varepsilon_0 R^2} - \frac{nNB}{R^{n+1}} \tag{i}$$

For $R = R_0$, $dU/dR = 0$, this yields

$$B = \frac{Mq^2}{4\pi\varepsilon_0 n} R_0^{n-1} \tag{ii}$$

The structure of KCl is identical to that of NaCl. The compressibility is reciprocal to the bulk modulus, and hence

$$\frac{1}{K} = \left(\frac{1}{18\,NR_0}\right)\frac{d^2U}{dR^2}\bigg|_{R=R_0} \qquad \text{(iii)}$$

From Eq. *(i)*, we have

$$\frac{d^2U}{dR^2}\bigg|_{R=R_0} = -\frac{2\,NMq^2}{4\pi\varepsilon_0 R_0^3} + \frac{n(n+1)\,NB}{R_0^{n+2}}$$

$$= \frac{2\,NMq^2}{4\pi\varepsilon_0 R_0^3}(n-1) \text{ (Using Eq. } (ii))$$

Hence from Eq. (3.46), we have

$$n = 1 + \frac{18\,R_0^4\,(4\pi\varepsilon_0)}{KMq^2}$$

$$= 1 + \frac{18\times(3.147\times10^{-10})^4}{5.747\times10^{-11}\times1.748\times(1.6\times10^{-19})^2\times9\times10^9}$$

$$= 1 + 7.63 = 8.63$$

Example 3 *The force of attraction between ions of Na and Cl is 3.02 × 10^{-9} N when the two ions just touch each other. Given: ionic radius of Na^+ ion is 0.95 Å, e = 1.6 × 10^{-19} C, ε_0 = 8.854 × 10^{-12} $C^2/N - m^2$. Find the radius of Cl^- ion.* [M. Sc. (MS), B.E.]

Solution We have $F_1 = -\frac{Z_1 Z_2 e^2}{4\pi\varepsilon_0 r_2}$ (force of attraction)

$$\therefore \quad F_1 = 3.02\times10^{-9}\text{ N} = \frac{(+1)(-1)(1.6)^2\times(10^{-19})^2}{4\pi\times8.854\times10^{-12}\,r^2}$$

$$= \frac{0.023}{r^2}\times10^{-26}\text{ N}$$

$$\therefore \quad r^2 = \frac{0.023\times10^{-26}}{3.02\times10^{-9}} = 0.0076\times10^{-17}$$

Z_1 = +1 for Na^+

Z_2 = –1 for Cl^-

$e = 1.6 \times 10^{-19}$ C

$\varepsilon_0 = 8.854 \times 10^{-12}$

$$\therefore \quad r = 0.276\times10^{-9}\text{ m} = r_{Na^+} + r_{Cl^-}$$

$$\therefore \quad r_{Cl^-} = 2.76\times10^{-10} - 0.95\times10^{-10} = 1.81\times10^{-10}\text{ m}$$

$$= 1.81\text{ Å}$$

Example 4 *The ionic radii of Mg^{++} and S^{--} respectively are 0.65 Å and 1.84 Å. Calculate the force of attraction between these ions. Given: e = 1.6 × 10^{-19} C, ε_0 = 8.854 × 10^{-12} $C^2/N\text{-}m^2$* [AMIE]

Solution Force of attraction

$$F_1 = \frac{Z_1\,Z_2\,e^2}{4\pi\varepsilon_0\,r^2}$$

Z_1 = +2 for Mg^{++}

Z_2 = –2 for S^{--}

$$= \frac{2 \times (-2) \times (1.60 \times 10^{-19})^2}{4\pi \times 8.854 \times 10^{-12} \times (2.49 \times 10^{-10})^2}$$

r = radius of Mg^{++} ion + radius of S^{--}

$= (0.65 + 1.84)$ Å

$= 2.49 \times 10^{-10}$ m

$$= \frac{4 \times 2.56 \times 10^{-38}}{111.27 \times 6.2 \times 10^{-32}} \text{ N} = 0.015 \times 10^{-6} \text{ N} = 1.5 \times 10^{-8} \text{ N}$$

5. COVALENT BOND

This type of bonding is formed by an equal sharing of electrons between two neighbouring atoms each having incomplete outermost shells. The atoms do so in order to acquire a stable electronic configuration in accordance with the octet rule. Unlike ionic bonding, the atoms participating in the covalent bond have such electronic configurations that they cannot complete their octets by the actual transfer of electrons from one atom to the other. Obviously, there is no charge associated with any atom of the crystal. The majority of solids incorporating covalent bonds are also bound by either ionic or Vander Waals bonds.

A covalent bond is formed between similar or dissimilar atoms each having a deficiency of an equal number of electrons. When two atoms, each having a deficiency of one electron, come so close that their electronic shells start overlapping, the original atomic charge distributions of atoms are distorted and each atom transfers its unpaired electron to the common space between the atoms. Obviously, the common space contains a pair of electrons which belongs equally to both the atoms and serves to complete the outermost shell of each atom. This is called *sharing of electrons*. The sharing is effective if the shared electrons have opposite spins. In such a case the atoms attract each other and a covalent bond is formed. As the participating atoms in the bond have the same valence state, this bond is also called the '*valence bond*'.

An excellent example of covalent bonds is seen in the chlorine molecule. The outer shell of each chlorine atom possesses seven electrons. Obviously, each chlorine atom would like to gain an electron to form a stable octet. This can be achieved by sharing of two electrons between pairs. Chlorine atoms (Fig. 4.6) thereby producing stable diatomic molecules. The nature of sharing of the covalent bonds in molecules of hydrogen, hydrogen fluoride and water molecule are shown in Fig. 4.7.

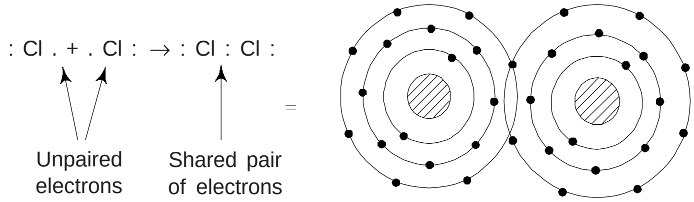


Fig. 4.6 Covalent bond between two atoms of chlorine

Covalent bonds are formed by the elements of groups IV, V, and VI, subgroup *B*. They are electronegative due to a high ionization potential and, when interacting with elements of other groups, take up valence electrons to complete their valence zone; in interaction of these elements with each other, the valence electrons are shared by the neighbouring atoms so that the valence zone is completed.

Covalent crystals are formed by carbon, silicon, germanium, antimony, bismuth, and some other elements.

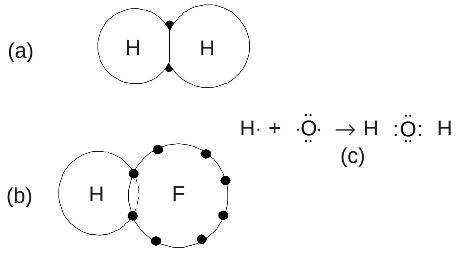


Fig. 4.7 Covalent bonds (a) between two atoms of hydrogen (b) in hydrogen fluoride (c) in water molecule

The number of atoms with shared electrons depends on the valency of an element and can be determined by the rule (8-*N*), where *N* is the valency. For example, the number of such atoms for carbon is equal to four.

In covalent crystals, atoms are arranged in the sites of the crystal lattice and the same forces are acting between them. For example, a carbon atom has four valence electrons by means of which it forms four directed bonds and comes into the interchange interaction with four neighbouring atoms. Each pair of carbon atoms can interchange their valence electrons similar to what occurs between atoms in a hydrogen molecule. The carbon atom in the ground state has the electronic configuration $1s^2\ 2s^2\ 2p^2$, i.e. it has two unpaired electrons in the outermost orbit. Since the energy difference between 2*s* and 2*p* states is small and therefore the carbon atom is known to form four covalent bonds; it is proposed that one of the electrons of carbon is transferred from 2*s* state to 2*p* state. This modifies the electronic configuration as

$$1s^2\ 2s^1\ 2\,p_x^1\ 2\,p_y^1\ 2\,p_z^1$$

Obviously, in the excited state, the *C* atom has four unpaired electrons and hence is able to make four covalent bonds with four neighbouring *C* atoms.

Carbon occurs in nature in two crystalline modifications (Fig. 4.8): diamond and graphite. The structure of diamond is a good manifestation of the directional properties of covalent bonds.

A covalent bond is characterized by directivity, since each atom can be in interchange with a definite number of neighbouring atoms. For that reason, atoms in covalent crystals are not packed closely and form crystal structure with low coordination number. For, e.g., the cubic lattice of diamond has the coordination number 4.

The directivity of interatomic bonds and loosely packed crystal structures are responsible for a low plasticity and high hardness (diamond is the hardest).

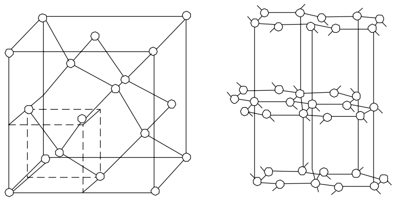

(a) (b)

Fig. 4.8 Crystal lattices of (a) diamond and (b) graphite.

The covalent bond is a strong bond. This is apparent from the binding energy of carbon in diamond which is 7.4 eV. In view of high bond energy, covalent crystals are characterized by high melting points (3540°C for diamond) and high boiling points.

The formation of completely occupied valence zones with this type of bond is the reason why covalent crystals are mostly semiconductors or even dielectrics (e.g., diamond). The high electric conductivity of graphite can be attributed to the replacement of one of the four covalent bonds by a Vander Waals bond, resulting in the appearance of free carrier of electric current.

The temperature coefficient of electric resistance of covalent crystals is negative. i.e. the electric resistance decreases on heating. Covalent crystals are formed by many compounds consisting of unlike atoms such as silicon carbide, aluminium nitride, etc.

With the help of wave mechanics, one can explain why a covalent bond tends to be the strongest in the directions when $|\psi^2|$ is maximum. We have the Schrödinger wave equation

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi+\frac{8\pi m^2}{h^2}(E-V)\psi = 0$$

where ψ is the wave function, $m \rightarrow$ mass of an electron, $h \rightarrow$ Planck's constant, $E \rightarrow$ total energy of the electron and $V \rightarrow$ the potential energy of the electron.

$|\psi^2|$ represents the probability of locating an electron. Value of $|\psi^2|$ show the extent of binding between two orbitals. $|\psi^2|$ will be maximum when electron density between two overlapping orbitals is maximum. Obviously, we can say that the greater value of $|\psi^2|$, the greater would be the overlapping of electrons, which in turn implies that greater the overlapping of two orbitals, greater would be the directional strength of the covalent bond. This is why, depending on direction, in which the value of $|\psi^2|$ is maximum, the covalent bond would also be stronger in the direction.

5.1 Characteristics of Covalent Compounds

(i) Covalent compounds are mostly gases and liquids.

(ii) They are usually electric insulators.

(iii) They are directional in nature.

(iv) They are insoluble in polar solvents like water but are soluble in non-polar solvents, e.g,, benzene, chloroform, alcohol, paraffins etc.

(v) Covalent compounds are homopolar, i.e. the valence electrons are bound to individual or pairs of atoms and electrons cannot move freely through the material as in the case of metallic bonds.

(vi) Covalent compounds are soft, rubbery elastomers, and form a variety of structural materials usually termed as plastics. The melting and boiling points of these compounds are low.

(vii) In the exceptional case of diamond covalent bonding is very strong due to the very large cohesive energy. Usually the binding energy ranges from 1 to 5 eV. The strong covalent bonding makes diamond very hard and with a high melting point.

5.2 Covalent Solids

The covalent bonds are very strong and rigid, and their directional nature have to be maintained. This is why covalent solids do not form close-packed structures. Diamond (also germanium and silicon) is the example of the simplest covalent structure, which is fairly open and empty and far from close-packed (Fig. 4.8(a)). The coordination number is 4. Diamond lattice is a complicated cubic lattice with two atoms, e.g., situated at (000) and (1/4, 1/4, 1/4) associated with each of the cubic lattice points.

The graphite is the low pressure crystal form of carbon (Fig. 4.8(b)), which differs in many respects from diamond. Graphite is a large structure in which the atoms are covalently bonded only to the nearest neighbours in the layer, the layers themselves being only weakly bonded by Vander Waals forces. In graphite, there is one electron per atom that is not employed in the covalent bonding and therefore there is an abundant supply of free electrons. This is why graphite is a good electrical conductor, whereas diamond is an almost insulator (perfect). However, the free electrons cannot cross easily from layer to layer in graphite.

6. METALLIC BOND

Metallic bond is formed by the partial sharing of valence electrons by the neighbouring atoms. They are formed by the elements of all subgroups *A* and I-III, subgroups *B*. Unlike the case of covalent bond, the sharing in metallic bond is not localized. Hence metallic bond may also be considered as delocalized or unsaturated covalent bond. Metallic bonds are electropositive. When interacting with elements of other groups, atoms in a metallic crystal can easily give off their valence electrons and change into positive ions.

When interacting with one another, the valence energy zones of atoms overlap and form a common zone with unoccupied sublevels. The valence electrons thus acquire the possibility to move freely with the zone. Obviously, valence electrons are shared in the volume of a whole crystal. Thus the valence electrons in a metal cannot be considered lost or acquired by atoms. They are shared by atoms in the volume of a crystal, unlike covalent crystals where sharing of electrons is limited to a single pair of atoms.

Due to delocalized nature of valence electrons, the metallic bond is much less directional than covalent bond. This results in a high coordination number and close packing of crystal structures of metals. High coordination numbers are typical of crystal structures with FCC and HCP lattices. A number of metals, e.g., Ni, Ag, Cu, Au, Fe_r, Pt, Al and Pb, have an FCC lattice. Many metals have an HCP lattice, but only Mg and Co among them have the axial ratio $c/a = 1.633$ corresponding to the spherical symmetry of atoms.

Deviations of the c/a ratio from 1.633 can be explained by the presence of partial covalent bonds resulting in a non-spherical symmetry of atoms. In that case, the metallic interaction in the direction of atom stretching is supplemented by covalent interaction. If non-spherical atoms are arranged in a crystal so that their major axis is along the z-axis, the c/a ratio is greater than 1.633 (Zn and Cl). If however, their minor axis is arranged along the z-axis, the c/a ratio is less than 1.633 (Be, Ti_α and Zr_α).

The formation of crystal structures with a BCC lattice is also explained by partial covalent bonds and non-spherical symmetry of atoms. In these structures, atoms are not packed closely. Crystal structures with a BCC lattice are typical of Fe_{α}, Cr, Mo, W, V, Ta, Ti_{β}, Nb, Zr_{β}, and some other metals.

Many metals and some non-metals are *polymorphous*, i.e., may have different types of crystal structures in the solid state depending on temperature. Such crystal structures are called allotropic forms, or *modifications*. They are usually designated by Greek letters α, β, γ, δ, and so on, with 'α' denoting the form existing at the lowest temperature.

The stability of allotropic modifications at definite temperatures and pressures is determined by the thermodynamic potential (free energy)

$$G = H - TS \tag{5}$$

The modification having a lower algebraic value of thermodynamic potential is more stable at a given temperature, which can be achieved either by decreasing the enthalpy H or increasing the entropy S.

In metallic crystals, closed-packed structures: HCP (C = 12) and FCC (C = 12) [Fig. 4.9] have a lower enthalpy and thus are stable at low temperatures (upto T_1). The 'looser' BCC strucutre (C = 8) has a higher entropy and therefore is stable at elevated temperatures. This explains the stability of BCC structures at elevated temperatures in many metals: Ti, Zr, Fe, U. The stability of the BCC lattice of iron at low temperatures is attributed to an increase of the electronic component of entropy.

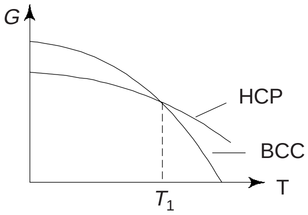


Fig. 4.9 Variations of the thermodynamic potentials of two modifications of a metal during heating.

The stability of allotropic modifications may change due to changes in the type of bond. For example, the tin modification with diamond type lattice Sn_{α}, is stable at low temperatures owing to the high energy of covalent bond, and therefore, low enthalpy. On heating, it changes to the Sn_{β} modification which has a weaker metallic bond.

Around 30 metals, e.g. Titanium, Zirconium, Tin , Uranium, Iron, Cobalt, etc. exhibit the phenomenon of *temperature polymorphism*. A high-temperature modification can be retained in a metal for a long time at 20-25°C by fast cooling, since the diffusion mobility of atoms at such temperatures is low and cannot cause rearrangement of the lattice.

Besides, some polymorphous transformations occur under the effect of pressure and temperature. For example, the graphite modification of carbon recrystallizes to diamond when heated to 2000°C under a pressure of about 10^{10} Pa. Under very high pressures, iron has a low-temperature modification with HCP lattice.

An increase of pressure at low temperatures can change less closely packed modifications into close-packed structures. As has been observed for Ge, Si and Sn_{α}, application of a high pressure can change their covalent crystals with diamond type lattice (C = 4) into metallic crystals with a tetragonal body centered lattice (C = 8).

The entropy of metallic bond is somewhat lower than that of covalent bond, because of which most metals have lower melting and boiling points and lower elastic moduli, but higher temperature coefficients of linear expansion, compared to covalent crystals.

In most cases, an increase of bond energy increases the melting point,elastic modules, an activation energy of self diffusion, but, on the contrary, decreases the coefficient of linear expansion α. These regularities have been observed experimentally and have a number of exceptions. In particular, the elastic modulus of beryllium is anomalously high, because of which beryllium alloys have an elevated stiffness. Ti and Zr exhibit too low values of elastic moduli and activation energy of self diffusion. The latter explains their low heat resistance.

By virtue of non-directivity of metallic bonds and close-packed structures, metallic crystals are more plastic and less hard than covalent crystals. Their high electric conductivity is due to the presence of free sublevels in valence energy zone.

The temperature coefficient of electric resistance of metallic crystals is positive, i.e., the electric resistance increases on heating.

The metallic bond is weaker than ionic or covalent bond. The binding energy ranges from 1 to 5 eV per bond.

Quantum mechanically, it can be shown that the wave functions of valence electrons in metals are so spread out that these electrons can no longer belong to any one nucleus in the solid but are a property of the solid as a whole. This delocalization of the valence electrons leads to a decrease in both potential and kinetic energy which is responsible for metallic bonding.

6.1 Characteristics of Metallic Crystals

(i) Metallic compounds are crystalline in nature due to the symmetrical arrangements of the positive ions in a space lattice .

(ii) Metallic compounds are good conductors of heat and electricity. They have high electrical conductivity and higher thermal conductivity.

(iii) Metallic bonds being weak, metals have a melting point moderate to high, i.e., the melting points of metallic crystals are lower than those of electrovalent crystals.

(iv) Metals are opaque to light, because the free electrons in a metal absorb light energy.

(v) Metallic crystals have good lusture and high reflectivity.

7. COMPARISON OF IONIC, COVALENT AND METALLIC BONDS

(i) Ionic and Covalent Bonds: Ionic compounds are crystalline in nature and have high hardness. They are non-directional and rigid. They have high metting and boiling points. They are poor conductors and exist in solid form only. They are soluble in water and polar solvents like ammonia but insoluble in organic solvents.

The ionic character can be calculated from the following relation

$$\% \text{ ionic character} = [1 - \exp\{(-1/4)\ (X_A - X_B)^2\}] \times 100$$

where X_A and X_B are electronegativities of elements A and B respectively, joining the ionic-covalent bond. Obviously, the ionic character of the bond increases if the electronegativity between the joining atoms increases. This type of mixed bond is obtained in several semiconducting material is, e.g. Ga As and ZnS. In ZnS, Zn belongs to II *B* group and *S* belongs to VI *A* group has higher ionic character than Ga As in which Ga and As respectively belong to III *A* and V *A*.

The electronegativities are: Ga – 1.8, As – 2.2
Zn – 1.7, S – 2.4

% Ionic character of Ga As

$$= \{1 - \exp[-1/4\ (1.8 - 2.2)^2]\} \times 100$$
$$= 3.92\ \%$$

% Ionic character of ZnS

$$= \{1 - \exp\ [-1/4\ (1.7 - 2.4)^2]\} \times 100$$
$$= 11.53\%$$

Covalent compounds are rubbery in nature and covalent bonds can be observed in gases, liquids and many solids. They have low melting and boiling points. With the exception of diamond and graphite, compounds of covalent bonded materials are usually soft, easily fusible and electrically insulators. They have open structures and covalent bonds are directional.

(ii) Ionic and Metallic Bonds: Ionic bonds result from the mutual attraction of positive and negative ions. Ionic compounds have low thermal and electrical conductivity but have fairly high melting points. Ionic bonds are stronger than metallic bonds. Ionic bond itself is non-directional.

In ionic bonding, a metallic element loses, from the outer electronic shell of its atom, a number of electrons equal to its valency (numerical). These lost electrons are transferred to the outer electronic shells of non-metallic atoms with which the metal is combining. In this way, a complete octet of electrons is left behind in metallic particle and completed in the non-metallic particle, finally both gets stable outer shells.

When a variable number of valence electrons are shared by a variable number of atoms of the same metal, a metallic bond results. In comparison to non-metals, the valence electrons in metals lie much farther from the nucleus. The valence electrons of metals move freely through the metal, i.e. they are not bound to indivdual atoms or pairs.

Metallic compound have higher thermal conductivity and moderate melting points but their bonds are weaker than ionic bonds.

(iii) *Covalent and Metallic Bonds:* Metallic bonding, due to unlimited sharing of electrons, is also called unsaturated covalent bonding. Truly speaking, in some respects, the metallic bond is similar to the covalent bond and yet quite different. In metallic bonding the valence electrons which hold the atoms together move freely through the metal, whereas in the covalent bond, two atoms are paired or shared by a particular pair of electrons. Metallic bonds are not directional unlike covalent bond. Common to both type of bonds is the sharing of electrons.

Most transition metals have mixed *metallic covalent* bonding involving *dsp* bonding orbitals. This type of bonding results in high bonding character changes from pure covalency at the top to pure metallic at the bottom. C forms pure covalent bond with other carbon atoms (diamond), Si and Ge have mixed covalent and metallic bonds. Sn and Pb form pure metallic bond.

Table 4.4 gives the comparison of all the three, i.e. ionic, covalent and metallic bonds.

Table 4.4 Comparison between ionic, covalent and metallic bonds

S.No.	*Ionic Bond*	*Covalent Bond*	*Metallic Bond*
1.	This type of bonding exists due to electrostatic force of attraction between positive and negative ions of different elements.	This type of bonding exists due to the electrostatic force of attraction between atoms which share the electron pairs to form a covalent bond.	This type of bonding exists due to electrostatic force of attraction between electron clouds and postive ions of same or different metals.
2.	This type of bonding is formed between two different elements. one of the atom loses its valence electron and other accepts it. The ions so formed attract each other and an ionic bond is formed.	This type of bonding is formed due to sharing of electron pairs between the atoms of same or different elements.	When the valence electrons detach themselves from their parent atoms and form a common pool, metallic bond is formed. The force which binds the electron cloud and positive ions of the metal forms the metallic bond.
3.	Due to their crystalline structure, ionic solids have high hardness.	Except diamond, silicon and carbide, etc., the covalent solids have low hardness.	The metallic solids are soft in nature and have crystalline structure.
4.	The ionic solids have very low thermal and electrical conductivities.	The covalent solids have low thermal and electrical conductivities.	The metallic solids have very high thermal and electrical conductivities.
5.	The ionic solids have high melting and boiling temperatures.	The covalent solids have lower melting and boiling temperatures in comparison to ionic solids.	In comparison to ionic solids, metallic solids have slightly lower and boiling temperatures.
6.	The ionic solids are not malleable and ductile.	The covalent solids are not malleable and ductile.	The metallic solids are malleable and ductile.
7.	The ionic solids are soluble in water.	The covalent solids are soluble in toluene and benzene etc.	The metallic solids are neither soluble in water nor in benzene, etc.
8.	The ionic compounds exist in the solid form only.	The covalent compounds exist in solid, liquid and gaseous form.	Except mercury, metallic compounds exists in solid form.

8. SECONDARY BONDS

Besides the three primary bonds as discussed above, there exist various secondary bonds. Secondary bonds are called *molecular bonds* and are secondary in the sense of being relatively weak in comparison with the three primary bonds. The bond energy of secondary bonds vary between 4 to 42 kJ/mole. Secondary bonds are formed in case of elements or compounds whose electron configuration is such that very little electron transfer takes place between them. Secondary bonds result from the *electrostatic attraction of dipoles*, the dipoles forming as a result of unbalanced distribution of electrons in asymmetrical molecules. Secondary bonds are also formed as a result of weak *Vander Waals forces* of attraction which exist between atoms or molecules which have their outermost shells completely filled and hence have no tendency to gain, lose or share valence electrons with other atoms or molecules in the solid. Vander Waals forces are due to the electrostatic attraction between the nucleus of one atom and the electrons of the other. The crystals resulting from this type of bonding are called *molecular crystals.* The examples of such solids are crystalline states of inert gases, such as He, Ne, A, etc. and other gases like O_2, Cl_2, CO_2, CH_4, etc. The rare gases, e.g. argon and helium, condense to form solids at sufficiently low temperatures, forming secondary bonds which result neither from the transfer of electrons (metallic, ionic) nor from the sharing of electrons (covalent). These bonds are weaker than primary bonds.

The Vander Waals forces are very weak forces. At room temperature, the thermal energy acquired by atoms or molecules is sufficient to make these forces ineffective. Thus molecular solids exist in gaseous state at room temperature. At low temperatures, however these forces dominate over the thermal forces and a significant role in the transformation of gaseous state to liquid or solid state. The energy released during the formation of Vander Waals bonds is of the order of 0.1 eV per bond only. This is why molecular solids are characterized by very low melting and boiling points, low mechanical strength and easy deformability. These solids are also poor conductors of heat and electricity. Molecular compounds are generally transparent to light and good insulators, with the exception of water (H_2O).

Secondary bonds are very crucial in determining the structure and properties of many non-metallic materials that include plastics, paraffins and graphite. The secondary bonds or mixed bonds are of the following three types:

(i) Dispersion bonds *(ii)* Dipole bonds *(iii)* Hydrogen bonds

(i) Dispersion Bonds: In a symmetrical molecule, the electrons are distributed uniformly around the nucleus and they are constantly in motion. Obviously, in a symmetrical molecule, the centres of positive charges and negative charges coincide with each other (Fig. 4.10). However, it is observed that at certain times the distribution of electrons in the molecule is not symmetrical around its nucleus. This causes in the displacement of the centres of positive and negative charges (Fig. 4.10(b)). It creates electronic imbalance of the charge which is known as *polarisation*. The nature of this polarisation is fluctuating and is known as *dispersion effect.* Due to dispersion effect, there exists a weak force of attraction between the two molecules of the same element. Due to this force a bond is formed between the two molecules which is called *dispersion bond*.

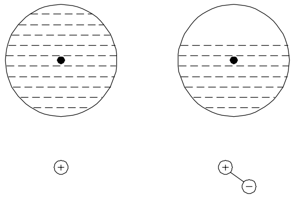


Fig. 4.10 Dispersion effects (a) same centre (b) different centres

The molecules of noble gases like argon, helium, which consists of single atom are held together by dispersion bonds, when they are solidified at very low temperature. It is this condensation which makes it evident that these are weak interatomic attractions which pull the atoms together. This is what is known as *Vander Waals forces* and bonding as *Vander Waals bonding*.

(ii) Dipole Bonds: When the outermost electrons of an atom are shared by two or more atoms, a covalent bond is formed. Sometimes there is an unequal sharing of electrons between two atoms. The unequal sharing of electrons takes place only in those substances in which one of the atoms, in a molecule, has a high affinity to attract electrons than the other. The overall effect of this unequal sharing of electrons is to create opposite charges on the parent atoms. As a result, permanent dipoles are produced. Such dipoles attract each other and a bond is formed which is known as *dipole bond*. Dipole bonds are weaker than ionic bonds but stronger than dispersion bonds.

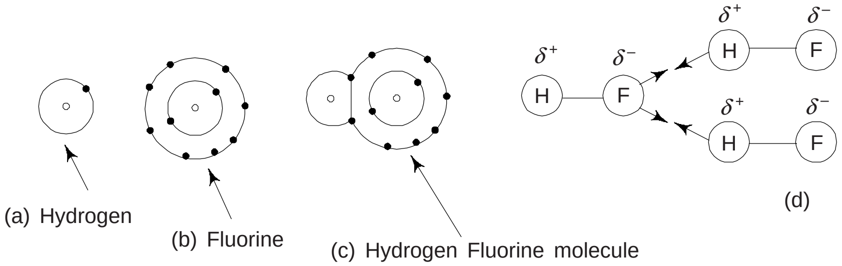


Fig. 4.11 Dipole bonds (a) Hydrogen (b) Fluorine (c) Hydrogen Fluorine molecule

Figure 4.11 illustrates the dipole bond in the case of hydrogen fluoride molecule.

The atomic number of hydrogen atom is 1 and its electronic configuration is $1s^2$. The atomic number of fluorine is 9 and its atomic configuration is $1s^2\ 2s^2\ 2p^5$. Obviously, the hydrogen atom has only one electron in its outermost energy level whereas fluorine has 7 electrons in its outermost energy level ($2s^2\ 2p^5$) as shown in Fig. 4.11. Thus the hydrogen atom requires one more electron to acquire stable configuration, whereas fluorine atom also requires one more electron to acquire stable configuration. This is why the hydrogen and fluorine atoms share a pair of electrons to acquire stable configuration (Fig. 4.11(c)). This causes to the formation of covalent bond, and the molecule so formed is known as hydrogen fluoride.

We must not that in this covalent bond, the fluorine atom has a high affinity than the hydrogen atom. This is why the shared electron pair shifts towards the fluorine atom. This shifting of electron pair produces a dipole. Figure 4.11(d) illustrates, how the resulting dipole produces a mechanism for secondary forces of attraction. Other examples are hydrogen chloride molecule, SO_2, HBr and hydrogen cyanide.

(iii) Hydrogen Bonds: This is a special type dipole bond, but it is considerably stronger. It is a special case of inter-molecular attraction produced between certain covalently bonded hydrogen atoms and one pairs of electrons of another atom. When a hydrogen atom is covalently bonded to a relatively large atom such as nitrogen, oxygen or fluorine, a powerful permanent dipole is set up. There is a positive field adjacent to the hydrogen and a negative field around the electron pair. One finds that this force is very important in plastics and in biological molecules such as DNA. A good example of hydrogen bond is water molecule. In water, the hydrogen and oxygen atoms are held together by covalent bonds in a configuration as shown in Fig. 4.12.

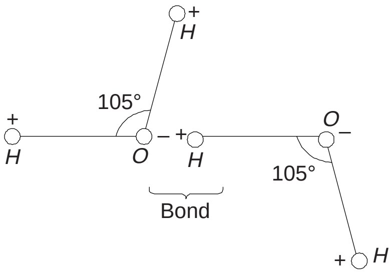


Fig. 4.12 Hydrogen bond between two water molecules

We must note that in this covalent bond, the oxygen atom has higher affinity for the shared electron pair than the hydrogen atom. Obviously, the shared electron pair shifts towards the oxygen atom. This shifting produces a dipole. Similar dipoles are formed in the other molecules as well. All such dipoles attract each other to form a liquid water.

Hydrogen bonds are directional in character. They are found in various tissues, organs, blood, skin and bones in animal life. The hydrogen bonds of the type N . . . H . . . O, which found in proteins, play an important role in determining their structures.

9. MIXED BONDS (MORE ABOUT SECONDARY BONDS)

There are many substances in which bonding between atoms does not occur as one of the primary or secondary types of ideal bonding mentioned above but, rather, as a mixture of these bondings. Quartz, glass or silica are the substances in which mixed ionic covalent bonding occurs.

Ionic and metallic bonding produces many structural defects in the lattice of materials. It is found that there is a continuous change in bonding character in a series of alloys of metals such as copper-zinc, Copper-nickel and copper-tin. Although, bonds other than metallic are present but certain metallic properties are retained.

(i) An atom with a closed electronic shell consists essentially of a positive nucleus surrounded by a spherical cloud of negative charge. If the electron clouds of two such atoms were static and undeformable then no force would exist between these atoms. However, the electron clouds result from the motion of various electrons around the nucleus. This means that an atom can on an average have no electrical dipole moment but have a rapidly fluctuating dipole moment. At any moment the centre of the negative charge distribution does not coincide with the nucleus rather fluctuates rapidly about it. This fluctuating charge on one molecule tends to interact with the fluctuating charge on a neighbouring molecule and causing a net attraction as shown in Fig. 4.13. When the gases are solidified, molecules of inert gases which consist of single atoms are held together by dispersion forces.

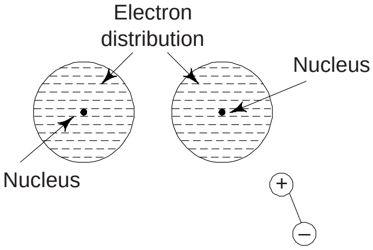


Fig. 4.13 Dispersion bond

(ii) Figure 4.14 shows that when two atoms approach each other, the rapidly fluctuating dipole moment of each atom affects the motion of the electrons in the other, and a lower energy is produced if the fluctuation occur in asympathy with one another. In figure, each atom is shown as a small electric dipole. Obviously, the centre of positive charge and the centre of the negative charge do not coincide and an electric dipole is produced. This exhibit a marked attractive force providing a mechanism for molecular bonding. Moreover, this is also due to the Vander Waals forces of attraction. We know that Vander Waals' bonds are non-directional. These bonds are formed between those molecules that have no permanent dipoles. Polar compounds HCl and PVC are the examples in which this type of binding is common.

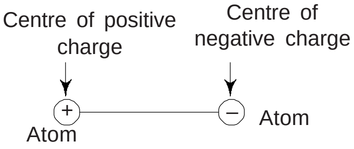


Fig. 4.14 An electric dipole

(iii) We have already remarked that hydrogen bonds are a special type of dipole bond which can be formed between atomic groups with no electrons to share. These bonds occur only when a hydrogen atom is present but they are similar to Vander Waals bonds. Hydrogen often plays a major role in organic materials and hydrogen bond frequently occurs in these materials. The unusual physical properties of ice and water are due to hydrogen bonds (Fig. 4.15). In a H_2O, i.e., water molecule, the electrons shared between the oxygen and hydrogen atoms tend to stay close to the oxygen atom than the hydrogen atoms because of the greater electronegativity of oxygen. This is why, oxygen acts as the negative end of the dipole and the hydrogen atoms acts as the positive ends. The positive end then attract the negative end of another water molecule and this way bond the molecules together. We have already mentioned earlier that hydrogen bond is also called a hydrogen bridge and is important in several biological molecules, e.g. DNA.

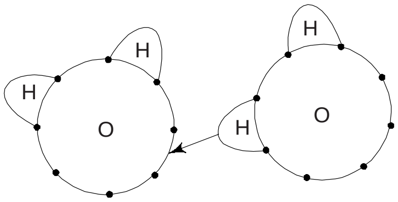


Fig. 4.15 Hydrogen bond in water

10. PROPERTIES OF SOLID MATERIALS

So far we have discussed the classification of solid materials according to their chemical bonds, i.e. ionic, metallic, covalent and molecular. The properties of various chemical bonds are summarized in Table 4.5.

In order to know about the properties of chemical bonds, one will have to confine to the strength of the bond. The strength of the bond is best measured by the energy required to break the bond, i.e. the amount of heat which must be supplied to vapourize the solid, i.e. to separate the constituent atoms. From Table 4.5, it is evident that Vander Waals and hydrogen bonds are weakest and then comes the metallic, which is followed by the ionic and covalent chemical bonds, whose strengths are almost comparable.

Melting points of substances are also indicative of bond strength. We know that melting occurs when the thermal vibration becomes so great that bonds are broken and atoms become mobile. It is observed that

Table 4.5 Properties of various chemical bonds

Properties	*Ionic bonds (1)*	*Covalent bonds (2)*	*Metallic bonds (3)*	*Molecular (Vander Waals) bonds and hydrogen (4)*
Bonding force	Electrostatic attraction of oppositely charged ions	Mutual sharing of valence electrons between atoms	Attraction between the ion cores and conduction electron gas	Electrostatic force due to oscillating dipoles, or permanent dipoles between molecules or dipoles between similar molecules
Energy of bond kJ/k mol	(500–1000) 10^3	(60–700) 10^3	(100–350) 10^3	(2.1-50)10^3
Character of bond	Non-directional	Directional	Non-directional	Directional for dipole and hydrogen bonds, non-directional for dispersion ones
Bond formation condition	When one of the atoms has smaller number of valence electrons	When atomic orbitals of two atoms overlap	For elements having small number of valence electrons	For molecules forming dipoles
Electrical conductivity	Low conductivity	Low electrical conductivity	Good conductivity	Low conductivity
Lustre	Transparent or coloured, but not highly reflective.	Transparent to highly reflective	Highly reflective	Transparent or coloured.
Colour	Wide variety of goods colours	White to grey	Usually grey with a few exceptions	Colour derives from molecule, not from crystal
Density	Intermediate	Intermediate	High	Low
Cleavage	Very good	Fairly good	Difficult except at low temperatures because of high ductility	Easy
Melting point	Intermediate (700–3300 K)	High (500–3700 K)	Intermediate (230–450)	Low(< 600 K)
Vapour state	Diatomic molecules or atoms	Atoms or occasionally small molecular aggregates	Atoms or occasionally small molecular aggregates.	Molecules
Heat of vaporization H (kcal/mole)	121–242	85–405	26–94	0.02–2.4
Electrical resistivity (Ω-m)	10^{12}–10^{20}	10–10^{20}	10^{-3}–10^{-5}	10^{12}–10^{20}
Thermal Conductivity (W-m^{-1}-k^{-1}	2–20	4–600	40–400	0.04–4
Examples	NaCl, CsCl, LiF, $CaCl_2$, Mgo, etc.	Diamond, Germanium, Si, SiC, etc.	Al, Cu, Na, Fe, etc.	Argon, Calomel, ice, solid CO_2, Ne, He, H_2, CH_4, etc.

the strength of ionic bonds increases with valency. We can easily understand why the melting point of magnesium oxide is higher as compared to sodium chloride.

One can easily determine the strength of materials and their crystal structure by knowing the bond strength.

Ionically or covalently bonded materials are non-conductors since their electrons are bound tightly in the bonds. We can understand that metals are excellent conductors because of the fluctuation of the electron cloud on application of a potential difference.

Due to shifting of electrons, metals possess ductility. Since the impinging radiant energy is easily absorbed by valence electrons, metallic bonded materials are opaque and lustrous.

11. PERIODIC TABLE AND CHEMICAL BONDING: ELECTRONEGATIVITY

We have stated earlier that the degree to which an atom can attract electrons to itself, is termed its *electronegativity*. In general, it is observed that electronegativity increases steadily across the periodic table Group I to Group VII. Moreover, electronegativity varies within a group, decreasing with increasing atomic number, except in the case of transition metals. Two atoms of similar electronegativity, form either a metallic or covalent bond, depending upon whether they can release or accept electrons. When their electronegativity is different, the bond is partially ionic and the ionic character increases with the difference in electronegativity. In general, electronegativity decreases with increase in atomic number, except in case of transition elements.

We must note that in any one row of the periodic table, (from right to left), the metallic character of the bond increases and covalent character decreases, whereas in any one column (from top to bottom), the metallic character increases or the covalent character decreases. From periodic table, we can see that Li, Na, K, Rb, Ls and Fr are metals with increasing metallic character, whereas at the other end, F, Cl, Br and I form covalent bonds of decreasing strength. Cs and Fr (at the bottom left of the periodic table) are more metallic than others and their ionization potentials is the lowest. F and Cl (at the top right-hand corner of the periodic table) are very non-metallic and exhibit the greatest electron affinity.

The bond between atoms of inert elements, between molecules of group VII, and VI and V, between molecules of unlike atoms is a weak secondary one.

We find that the transition metals have a partial covalent character and hence their melting points are high. Fe (1535°C), Ni (1453°C), Co (1495°C), V (1900°C), Nb (2415°C), Ta (2996°C), Mo (2610°C), Zr (1852°C), Ti (1668°C) and W (3410°C) are common engineering metals from the transition series, the melting point of each of these elements is shown within parenthesis.

Example 5 *The outermost electron which can easily be removed from Na atom requires +5.14 eV of energy. When this electron is transferred to a vacant position on a distant chlorine atom, it gives back –4.02 eV of energy. How much net energy is spent in the said process?* [AMIE, B.Sc. (Hons)]

Solution Net spent energy = +5.14 – 4.02 = 1.12 eV.

Example 6 *The enthalpy of fusion of ice is 6.02 kJ/mol. Estimate the fraction of hydrogen bonds that are broken when ice melts? (Given: Hydrogen bond energy = 20.5 kJ/mol)* [B.E.]

Solution There are two moles of hydrogen bonds per mole of H_2O in ice. Assuming that all the heat absorbed during melting goes into breaking the bonds, the fraction of bonds broken = $\frac{6.02}{20.5 \times 2}$ = 0.15.

We must note that bonds are continuously broken and remade in a liquid, and hence the above value should be considered as time-averaged value.

Example 7 *The enthalpy of atomization of diamond is 713 kJ/mol. The bond energy of the diamond reported is 347 kJ/mol. Explain the difference between the two.* [M.Sc. (M.S.)]

Solution We know that in diamond, each carbon atom forms four discrete bonds. As two atoms are required to form a bond, and hence the effective number of bonds per atom is 4/2 = 2. Obviously, when one mole of diamond is converted into carbon atoms in the gaseous state, 2 moles of C-C bonds are broken. Therefore, the enthalpy of atomization of diamond = 347 × 2 = 694 kJ/mol. We can see that this value is very close to 713 kJ/mol.

SUGGESTED READINGS

1. L. Pauling, *The Nature of the Chemical Bond*, Cornell Univ. Press, Ithaca, New York (1960).
2. W.L. Masterton and C.N. Hurley, *Chemistry, Principles and Reactions*, 2nd ed., Harcourt, Fort Worth, Tx (2000).
3. J.C. Kotz and P. Treichel, Jr., *Chemistry and Chemical Reactivity*, 4th ed., Saunders, Fort Worth, Tx (1999).
4. S.L. Kakani and C. Hemrajni, *Solid State Physics*, 4th ed., SultanChand, New Delhi-2 (2004).

REVIEW QUESTIONS

1. Describe the nature and origin of various forces existing between the atoms of a crystal. Explain the formation of a stable bond using the potential energy versus interatomic distance curve.

2. List the various types of bonds occurring in a crystal. Discuss any one of them in brief. [AMIE]

3. Explain electron cloud. What is the role of electron cloud in metallic bond? [Diploma]

4. List the different types of bonds occurring in a crystal. Describe the characteristics of metallic bonds. [AMIE, M.Sc. (M.S.)]

5. Describe the binding of atoms in metals. [Diploma]

6. Distinguish between ionic and metallic bonds in solids. Illustrate with examples. [AMIE]

7. Why are close-packed structures experienced mostly in metals and not in ionic and covalent bonds? [B.E.]

8. Explain the origin of metallic bonding. How does it differ from ionic bonding? Explain with suitable examples. [AMIE, Diploma]

9. How are atoms held together in a metallic bond? Explain diagrammatically. [Diploma]

10. What is hydrogen bond? How is it different from a dipole bond? Explain the role of hydrogen bond during the formation of ice?

11. Derive an expression for the binding energy of an ionic crystal and obtain the expressions for Madelung constant. Evaluate the Madelung constant for a linear ionic crystal.

12. What are various types of bonding in materials? Explain the different types of bonding and illustrate them with suitable examples. [AMIE]

13. How do the melting points of the ionic solids vary?

14. Why do inert gases get liquified and solidified at very low temperatures?

15. Define cohesive energy and determine its value for crystals of inert gases.

16. Using the concept of wave mechanics, explain why a covalent bond tends to be the strongest in the directions when $|\psi^2|$ is a maximum. [AMIE, BE]

17. Explain why carbon atoms in diamond bond covalently, while lead atoms bond metallically, even though carbon and lead have four valence electrons each. [AMIE, Diploma]

18. Explain with the help of suitable sketches the various types of bonding in crystals. [AMIE]

19. Describe the ionic bonding and giving example show that the bond energy of ionic bonding decreases as the size of electrons increases. Explain this behaviour. [M.Sc. (MS)]

20. What is a secondary bond? Explain the concept of dipole bonding. Why are some compounds non-polar even if they have dipoles?

21. Compare and contrast metallic, covalent and ionic bonds. Give examples. [AMIE]
22. What do you understand by percentage ionic factor of a covalent bonds? [B.Sc. (M.S.)]
23. Explain in which of the ionic and covalent bonds wave characteristics of electrons play important role? [M.Sc. (M.S.)]
24. Describe the essential features of the following with one example of each category: *(a)* metallic bond *(b)* ionic bond *(c)* covalent bond [AMIE]
25. Explain briefly: *(a)* covalent bonding *(b)* ionic bonding [AMIE]
26. It is observed that CCl_4 has no net dipole moment. What inference do you draw about C-Cl bond in this compound?
27. Describe compounds which show mixed ionic and covalent bonds? [M.Sc. (MS)]
28. The potential energy of a diatomic molecule is given in terms of the interatomic distance r by the expression: $U(r) = -\frac{a}{r^m} + \frac{b}{r^n}$, where symbols have usual meaning. Derive an expression for equal spacing of the atoms and hence obtain the dissociation energy. [B.E.]
29. Explain the origin of Vander Waals forces in molecular crystals. How does the binding due to Vander Waals forces differ from the binding the valence crystals. [M.Sc. (MS)]

PROBLEMS

1. The potential energy of a system of two atoms is given by

$$U = -\alpha/r^4 + \beta/r^{12}$$

Calculate the amount of energy released when the atoms form a stable bond. Also determine the bond length. $[(4\alpha^3/27\beta)^{1/2}, (3\beta/\alpha)^{1/8}]$

2. Assume the energy of two particles in the field of each other is given by the following function.

$$U = -a/r + b/r^8$$

where a and b are constants and r is the distance between the centres of the particles. Show that if particles are pulled apart, the molecules will break as soon as $r = (36b/a)^{1/7} = r_0\ (4.5)^{1/7}$.

3. The interaction energy of a system of two atoms is given by

$$U = -A/r^6 + B/r^{12}$$

The atoms form a stable bond with bond length of 3 Å and bond energy of 1.8 eV. Calculate A and B. Compute the forces required to break the molecule and the critical interatomic distance for which it occurs. [4.19×10^{-76} J-m^6, 1.53×10^{-133} J-m^{12}, 2.56×10^{-9} N, 3.33×10^{-10} m]

4. Assume that the energy of two particles in the field of each other is given by $U(r) = -\frac{\alpha}{r} + \frac{\beta}{r^8}$, where α, β are constant and r is the distance between the centres of the particles. Show that the two particles form a stable compound for $r = r_0 = (8\beta/\alpha)^{1/7}$ and that in the stable configuration, the energy attraction is 8 times energy of repulsion.

5. Consider a line of alternate positive and negative ions each carrying a charge q. The repulsive potential energy between the neighbours is given by A/r^n. Show that for a total of 2N ions, the equilibrium potential energy of the system is

$$U_0 = -(2Nq^2\, ln\ 2/r_0)\ (1 - 1/n)$$

where r_0 is the equilibrium interionic distance.

6. The potential energy U of a system of two atoms varies as a function of their distance of separation r as

$$U = -A/r^n + B/r^m$$

Show that equilibrium: *(i)* $r = r_0 = (mB/nA)^{1/(m-n)}$ *(ii)* the energy of attraction is m/n times the energy of repulsion and *(iii)* the bond energy,

$$U_0 = \frac{A}{r_0^n}\left(\frac{m-n}{m}\right)$$

7. The distance between the nearest positive and negative ions in KCl is 3.14 Å. The structure is similar to NaCl. Find the ionic bond energy in the solid in kilocalories per mole.

OBJECTIVE QUESTIONS

1. Sharing of electrons between neighbouring atoms results in
(1) metallic bond (2) ionic bond (3) covalent bond (4) none of above (3)

2. Primary bonds have energy range in kJ/mol.
(1) 10-100 (2) 10-1000 (3) 1-10 (4) 100-5000 (2)

3. The nature of binding for a crystal with alternate and evenly spaced positive and negative ions is
(1) dipole (2) ionic (3) metallic (4) covalent (2)

4. The atomic bond in NaCl is
(1) ionic (2) metallic (3) covalent (4) Vander Waals (1)

5. Thermal expansion of materials arises from
(1) thermal vibrations (2) weak bonds
(3) strong bonds (4) asymmetry of potential energy curve (4)

6. To break one H-Cl bond, 4.4 eV of energy is required. This energy is equal to
(1) 420 kJ/mol (2) 420×10^3 kJ/k-mol
(3) 420 J/k-mol (4) 42 J/k-mol (2)

7. The electron affinity of helium in kJ/mol is
(1) 4 (2) 16 (3) –16 (4) 0 (4)

8. The length of H-H bond is
(1) 1 nm (2) 0.1 nm (3) 0.037 nm (4) 0.074 nm (4)

9. The Fe-Fe bond length is 2.48 Å. The radius of iron atom is
(1) 0.62 Å (2) 1.24 Å (3) 2.48 Å (4) 3.96 Å (2)

10. Which of the following elements is a covalently bonded crystal?
(1) germanium (2) aluminium (3) lead (4) sodium chloride (1)

11. Hydrogen bonds are stronger than
(1) ionic bonds (2) metallic bonds
(3) covalent bonds (4) Vander Waals bonds (2)

12. Which of the following relation represent the potential energy of a diatomic molecule?
(1) $Ar^m + Br^n$ (2) $Ar^m - Br^n$ (3) $A/r^m - B/r^n$ (4) $-A/r^m + b/r^n$ (4)

13. The radius of anion is r_a and of cation is r_c, the bond length is
(1) $\sqrt{3}\,(r_c + r_a)$ (2) $r_a - r_c$ (3) $(r_c + r_a)$ (4) $(r_c + r_a)/2$ (3)

14. If 0.28 nm is the spacing between the nearest neighbour ions in NaCl crystal lattice, the unit cell parameter is
(1) 5.6 Å (2) 1.6 Å (3) 0.7 Å (4) 1.4 Å (1)

15. The bond energy of NaCl molecule is given by the relation:
(1) $V = e/4\pi\varepsilon_0 r_0^2$ (2) $V = -e/4\pi\varepsilon_0 r_0^2$
(3) $V = e^2/4\pi\varepsilon_0 r_0$ (4) $V = -e^2/4\pi\varepsilon_0 r_0$ (4)

16. Which of the following gives the lattice energy per k-mol of NaCl crystal? (Symbols have their usual meanings)

(1) $-\dfrac{AN_A\, e^2}{4\pi\varepsilon_0\, r_0}\left(\dfrac{n-1}{n}\right)$ (2) $\dfrac{AN_A\, r_0}{4\pi\varepsilon_0\, e}\left(\dfrac{n}{n-1}\right)$

(3) $-\dfrac{AN_A\, e^2}{4\pi\varepsilon_0\, r_0^2}\left(\dfrac{n-1}{n}\right)$ (4) $\dfrac{AN_A\, r_0}{4\pi\varepsilon_0\, e}\left(\dfrac{n}{n-1}\right)$ (1)

17. Metallic bond is not characterized by
(1) opacity (2) ductility (3) high conductivity (4) directionality (4)

18. Mixed ionic-covalent bonds are found in
(1) high strength materials (2) semiconductors
(3) heat insulators (4) none of the above (2)

19. The bond energy of C-C covalent bond in kJ//mol is
(1) 370, 800 (2) 640 (3) 1640 (4) zero (1)

20. The sublimation energy of Sodium is
(1) 1.08×10^3 kJ/k-mol (2) 10.8×10^3 kJ/k-mol
(3) 108×10^3 kJ/k-mol (4) 0.108×10^3 kJ/k-mol (3)

SHORT QUESTION ANSWERS

1. How atomic bonding in solids are considered?

Ans. These are considered in terms of attractive and repulsive forces.

2. How many types of primary bonds are?

Ans. Three: ionic, covalent and metallic.

3. Give names of secondary bonds.

Ans. Vander Waals and hydrogen bonds. These are weak in comparison to primary bonds.

4. Mention the range of energies of primary and secondary bonds.

Ans. Primary bond energies are in the range of 100-1000 kJ/mol, whereas that of secondary bonds are in the range of 1-50 kJ/mol.

5. How non directional ionic bond is developed?

Ans. For ionic bonds, electrically charged ions are formed by the transference of valence electrons from one atom to another. The forces are Coulombic.

6. What is covalent bonding?

Ans. In a covalent bonding there is sharing of valence electrons between adjacent atoms.

7. What is a metal?

Ans. A metal is an array of positive ions which are held together in a cloud of free electrons.

8. What is metallic bonding?

Ans. With metallic bonding, the valence electrons form a ‘sea of electrons’ that is uniformly dispersed around the metal ion cores and acts as a form of glue for them. The metallic bond is non directional and generally weaker than ionic and covalent bond.

9. What is the relation between melting and boiling points of materials and bond strength?

Ans. Melting and boiling points of materials increase with increasing bond strength. Strong and directional bonds result in hard and brittle solids.

10. How secondary bonds are formed?

Ans. Secondary bonds result from attractive forces between electric dipoles, of which there are two types—induced and permanent.

11. What happens when hydrogen covalently bonds to a non metallic element such as fluorine?

Ans. Highly polar molecule is formed.

5

Electron Theory of Metals

1. INTRODUCTION

There are many fundamental properties of solids, e.g. electrical and thermal conductivities, magnetic and optical properties, etc. depend upon their electronic structure. We can understand many physical properties of solids in terms of electron theory of solids. The development of the electron theory of solids, started in the beginning of the 20th century. Today, it is the basis for the classification of all solids. When free electron theory applied to metals, it explains forces of cohesion and repulsion, binding the energy levels and the behaviour of conductors and insulators and magnetic materials. According to this model, the valence electrons of the constituent atoms become conduction electrons and move about freely through the volume of the metal.

The first version of the free electron model was introduced by P. Drude in the early 1900s, with improvements soon after by H.A. Lorentz. This is now known as *Drude-Lorentz free electron theory*. The other theories are:

(i) Sommerfeld free-electron theory
(ii) Zone Theory

2. METALLIC BONDING

In chapter 4, we have already discussed about metallic bonding. In the metal, the valence electrons are free to move in different directions. The valence electrons inside the metal are called free electrons and move inside the metal, constrained only by the surface of the sample. Collisions between electrons are neglected. The binding forces in the metals are due to the electrostatic attraction between the positive ions and negative cloud or dilute gas of electrons. Such a free electron model was first introduced by P. Drude in the early 1900s, with improvements soon after by H.A. Lorentz.

We have now a modern picture of metallic bonding. According to this picture, the metallic bond is more closely related to the covalent or electron pair bond and resembles the ionic bond. Essentially, the metallic bond can be said to be an unsaturated covalent bond which allows a large number of atoms to be held together by mutual sharing of free electrons. Furthermore, the density of electrons between the atoms in metallic bonding is much lower than allowed by Pauli exclusion principle. This permits the electrons to move freely from point to point without significant increase in energy.

We have read that the bonds between atoms in solids are made of cohesive and repulsive forces which holds the atoms at definite distances from each other. It is proper to have an understanding of these forces upon which the electron structure of atoms depends. Primarily, due to the close proximity of two atoms places too many electrons into interacting locations, mutual repulsion results. The cohesive and repulsive forces gets equal when the equilibrium position is reached. We have already remarked in chapter 4 that large cohesive forces produce high melting points and fairly large elastic values of materials and higher mechanical strength.

3. DRUDE-LORENTZ THEORY

Drude, in 1900, postulated that the metals consist of positive ion cores with the valence electrons moving freely among these cores. The electrons are, however, bound to move within the metal due to electrostatic attraction between the positive ion cores and the electrons. The potential field of these ion cores, which is responsible for such an interaction, is assumed to be constant throughout the metal and the mutual repulsion among the electrons is neglected. The behaviour of free electrons moving inside the metals is considered to be similar to that of atoms or molecules in a *perfect gas*. The free electrons are, therefore, also referred to as *free electron gas* and the theory is accordingly known as Drude and Lorentz's *classical free electron theory*. The movement of electrons obeys the laws of the classical kinetic theory of gases. Lorentz in 1909, applied Maxwell-Boltzmann statistics to the electron gas with the following two assumptions:

(i) The mutual repulsion between the negatively charged electrons is negligible.
(ii) The potential field due to positive ions within the crystal can be assumed to be constant everywhere.

Since the electrons move freely inside the metals irrespective of the crystal structure, the ratio of the electrical conductivity, σ, to the thermal conductivity, k, should be constant for all metals at a constant temperature, i.e.

$$\frac{\sigma}{k} = \text{constant} \tag{1}$$

This is called the *Wiedemann-Franz law* and has been realized in practice.

This theory explained a number of properties of a metal, e.g. electrical conductivity, thermal conductivity, luster and opacity. The main drawbacks of this theory are:

(i) The theory correctly predicted the room temperature resistivity of various metals but the temperature dependence of resistivity could not be established accurately. The theory predicted that resistivity varies as $\sqrt{T}$ whereas actually it is found to be linearly with temperature.
(ii) The theory yielded incorrect magnitudes of the specific heat and paramagnetic susceptibility of metals.

The above shortcomings of Drude-Lorentz theory were removed by Sommerfeld in 1928. He applied Fermi-Dirac statistics instead of Maxwell-Boltzmann statistics. The possible electronic energy states in the potential energy box and the distribution of electrons in these states are then determined using quantum statistics.

4. SOMMERFELD FREE-ELECTRON THEORY

The basic assumptions of this theory are:

(i) The valence electrons in a metal are free.
(ii) Valence electrons in a crystal are confined to move within the boundaries of a crystal. Obviously, the electrons within the crystal have a lower potential energy than outside. We must note that the potential energy of an electron is uniform or constant within the crystal (Drude theory).
(iii) The electrons are free to move within the crystal, but are prevented from leaving the crystal boundaries by very high potential energy barriers at its surface.
(iv) The allowed energy levels of an electron, bound to a single atom are quantized.
(v) The electronic specific heat of metals is very low.

We consider the one dimensional and three dimensional cases separately.

4.1 Free Electron Gas in One-Dimensional Box

Consider an electron of mass m which is bound to move in a one dimensional crystal of length as shown in Fig. 5.1. The electron is prevented from leaving the crystal by the presence of a large potential energy

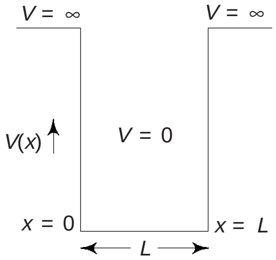


Fig. 5.1 **Potential well (one dimensional) bounded by infinite potential energy barriers**

barrier at its surfaces. The potential energy everywhere within the crystal is assumed to be constant and equal to zero. Thus we have

$$V(x) = 0 \text{ for } 0 < x < L$$

$$V(x) = \infty \text{ for } x \leq 0 \text{ and } x \geq L \quad (1)$$

Sommerfeld, in his free electron quantum theory assumed that the potential of an electron in a metal is uniform. He applied the one dimensional Schrödinger equation

$$\frac{d^2\psi_n}{dx^2} + \frac{8\pi^2 m}{h^2}(E_n - V)\,\psi_n = 0 \quad (2)$$

to calculate the total energy E_n, where ψ_n is the wave function of the electron occupying the nth state the E_n represents the kinetic energy of the electron in the nth state and V is its potential energy. The potential energy everywhere within the crystal is assumed to be constant and equal to zero. Therefore the Schrödinger equation (2) becomes

$$\frac{d^2\psi_n}{dx^2} + \frac{8\pi^2 m}{h^2} E_n\psi_n = 0 \quad (3)$$

The general solution to this equation is

$$\psi_n(x) = A \sin kx + B \cos kx \quad (4)$$

where A and B are arbitrary constants to be determined from boundary conditions. One obtains $A = \sqrt{2/L}$ and $B = 0$ and k is given by

$$k = \sqrt{\frac{4\pi m E_n}{h}} \quad (5)$$

One obtains the following expression for the allowed discrete energy values

$$E_n = \frac{h^2}{8mL^2} n^2 \quad (6)$$

where $n = 1, 2, 3 \ldots$

Obviously $E_n \propto n^2$

The number n is called the quantum number. The energy spectrum consists of discrete energy levels where the spacing between the levels is determined by the values of n and L. It decreases with increasing L. If L is of the order of a few centimeters, the energy levels form almost a continuum. But if L has atomic dimensions, the spacing between the levels becomes appreciable. The plot of E_n versus n is shown in Fig. 5.2. The energy levels and wave functions corresponding to $n = 1, 2, 3$ and 4 are shown in Fig. 5.3. Thus, we find that if the total number of electrons to be accommodated is seven, the energy levels with $n < 4$ would be occupied while the level with $n > 4$ would be empty. The topmost filled energy level at 0K is known as *Fermi level* and the energy corresponding to this level is called the *Fermi energy*, E_F.

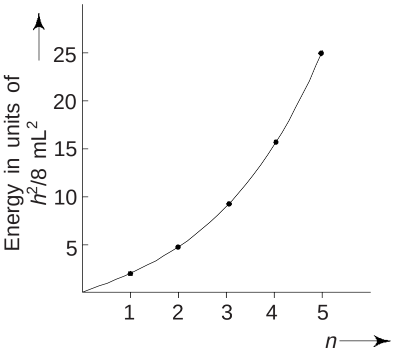


Fig. 5.2 **E_n versus n for a one-dimensional crystal**

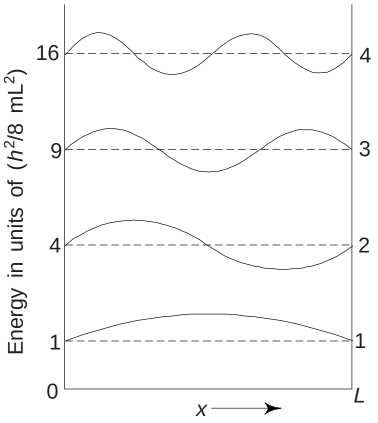


Fig. 5.3 **First four wave functions (solid lines) and the corresponding energy levels (broken lines) of an electron in a one dimensional crystal**

Free Electron Gas in Three Dimensions

The eq. (6) is the expression for energies in one dimension only. The corresponding three-dimensional case would be one in which an electron moves in all directions so that quantum numbers n_x, n_y and n_z are required corresponding to the three coordinate axes. In terms of a cubically shaped block of metal of side L, one obtains the expression for the allowed energies as

$$E_n = \frac{h^2}{8mL^2}\left(n_x^2 + n_y^2 + n_z^2\right) \tag{7}$$

$$= \frac{h^2}{8mv^{2/3}}\left(n_x^2 + n_y^2 + n_z^2\right) \qquad 7(a)$$

The integers n_x, n_y and n_z are the first three quantum numbers of an electron and $v = L^3$ is the volume of the assumed metal cube. The expression (7) gives the energies of free electrons in a metal. We must note that expressions (6) and (7) are of the same form, except for the number of integers. We must note that for various combinations of three integers n_x, n_y and n_z (e.g., 211, 121 and 112), one obtains the same energy value or level. However, each combination of integers represents a different wave function, having the same energy. Such an energy level is said to three fold degenerate.

From de Broglie relation, we have

$$\lambda = \frac{h}{mv} = \frac{2L}{n}$$

and wave number $$k = \frac{2\pi}{\lambda}$$

Using these relations, the expression (6) can be written as

$$E_n = \frac{k^2 h^2}{8m\pi^2} \tag{8}$$

we can see that the relation between energy and wave number obtained is parabolic (Fig. 5.4).

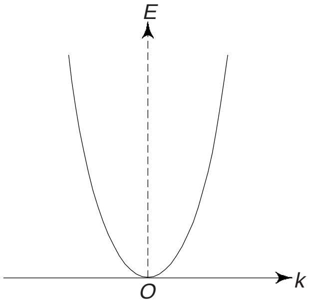


Fig. 5.4 **Parabolic relationship between energy and wave number**

5. FERMI-DIRAC DISTRIBUTION FUNCTION
(Electron Energies in a Metal)

The total values of energies of the valence electrons in a metal is given by Eq. (7). From Eq. (7) it is obvious that among energies E_1, E_2, E_3, etc., each energy value is greater than the preceding value by the same amount E_1. As stated earlier, for each different set of (n_x, n_y, n_z), there is an energy state. The number of states that have a given energy increases rapidly with increase in the number of different values of n_x, n_y, n_z. We must note that any change in these values influences the change in E by the square of their values $\left(n_x^2, n_y^2, n_z^2\right)$. This means the number of states in a metal is very large. If a plot is made for the number of states per interval of energy $N(E)$ or *density of states* V_s the total energy E, $N(E)$ increases parabolically with increasing E (Fig. 5.5).

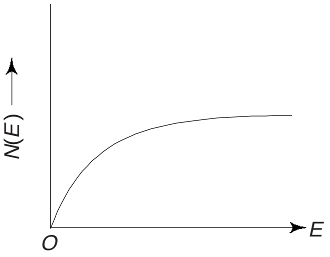


Fig. 5.5 Density of energy levels of free electrons in a solid vs energy

We know that the valence electrons have a tendency to occupy the lowest available energy states in a system. However, it is essential to consider all the electrons in the single system because of mutual interactions among all the electrons in the single system due to the mutual interactions among all the electrons forming the electron gas. This is possible only when we apply Pauli exclusion principle. We know that according to Pauli exclusion principle, only two electrons can occupy a given state, specified by the three quantum numbers (n_x, n_y, n_z), one with spin up and the other with spin down or opposite spin. If the metal is in its ground state, which occurs at absolute zero, all electrons occupy the lowest possible energy levels compatible with the exclusion principle, as indicated in Fig. 5.6. If the total number of electrons per unit volume (say n_0) is less than the total number of energy levels available in the band, the electrons will then occupy all energy states up to a maximum, designated by $E_{\max}$. The value of $E_{\max}$ depends on how many free electrons there are. The maximum energy level ($E_{\max}$) called the *Fermi level* and all quantum states in the energy levels above this level are empty (Fig. 5.6). The level at which the probability of occupation is 50% is the *Fermi level*, E_F. If we want to remove an electron from the Fermi level and take it out of the metal, some energy is required. This is called the *work function* and usually denoted by ϕ, which is equal to the energy normally measured when an electron is removed from the surface of the metal. Since thermal energies are very small compared with the Fermi energy, only a very few electrons are excited above the Fermi energy even at room temperatures. For that reason the work function is practically constant over a wide range of temperatures. We must note that at 0°K and electron in the Fermi level possesses the highest energy of the electrons in the metal and is the easiest to remove.

We now discuss the effect of temperature on the electron energy levels. It is apparent that, for temperatures greater than 0K, the Fermi level may not be the topmost filled level since some of the electrons from the filled energy levels may be excited to the higher levels. The probability that a state at a level of energy E is occupied by an electron at $T = 0°\text{K}$ is constant and equal to 1 (unity) upto the Fermi level E_F and zero above it. When the temperature of the system is raised, excited electrons move into the new energy levels (Fig. 5.7). Thus some of the levels below $E_{\max}$ would be empty while some above it would be occupied. The probability that a particular quantum state of energy E is occupied at a temperature T is given by the so called *Fermi function*.

$$f(E) = \frac{1}{\left[1 + \exp\left\{E - E_F / kT\right\}\right]} \tag{a}$$

The above also represents the change in electron energy distribution, i.e. $\int N(E) = f(E)$. Here $\int N(E)$ represents the probability that a state of energy E is occupied at temperature T. Thus at $T = 0°\text{K}$ when $E < E_F$, the exponential term in Eq. (a) becomes zero and

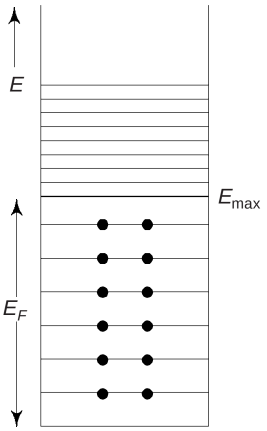


Fig. 5.6 **Filling of energy levels by electrons**

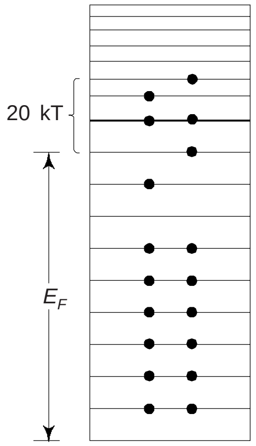


Fig. 5.7 **Distribution of free electrons among energy states (excited state)**

$$f(E) = \int N(E) = 1$$

when $E > E_F$, the exponential term becomes infinite and

$$f(E) = \int N(E) = 0$$

By putting $\int N(E) = f(E) = \frac{1}{2}$

one can easily see that $E = E_F$. Obviously, at absolute zero all states above $E = E_F$ are empty and all states below E_F are occupied (Fig. 5.8). The Fermi distribution function is a *step function.*

At any temperature one can define the Fermi level as that level for which the occupation probability is 1/2. The values of Fermi energy for few metals are given in Table 5.1.

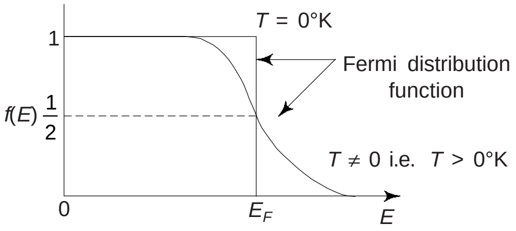


Fig. 5.8 **Probability of occupancy by an electron at various temperatures different from absolute zero**

Table 5.1 Fermi energy

Metal	*Li*	*Na*	*K*	*Rb*	*Cs*	*Cu*	*Ag*	*Mg*	*Al*
Fermi energy, E_F (eV)	4.72	3.12	2.14	1.82	1.53	4.07	5.54	7.3	11.9

6. BAND THEORY OF SOLIDS

The free electron theory of Sommerfeld was successful in explaining the various electronic and thermal properties of metals such as specific heat, paramagnetism, electrical conductivity, electron emission, etc. However, there are various other properties which could not be explained by this theory. For example, the theory could not explain why certain solids have a large number of free electrons and thus behave as good conductors while certain others have hardly any electrons and are, therefore, insulators. It also could not account for the variation of resistivity with temperature for the latter type of materials. Moreover, the properties of semiconductors also could not be explained on the basis of this theory.

The failure of the free electron theory is due to the oversimplified assumption that a conduction electron in a metal experiences a constant or zero potential due to the ion cores and hence is free to move about in a crystal; the motion being restrained only by the surface of the crystal. In fact, the potential due to ion cores is not constant and may change with position of the electron in the crystal. Some contribution to potential may also arise because of the other electrons present in the crystal. Obviously, the actual nature of potential under which an electron moves in a crystal is very complicated. The band or zone theory of solids explains this with the concept of periodic potential field by assuming that the electrons move in a region of constant potential. The electrons move in a periodic potential field provided by crystal lattice. We have already mentioned in the discussion on periodic table of elements that the atomic arrangement is periodic in all solids, i.e., within any real metal there is a periodic arrangement of positively charged ions through which the electrons move. To a reasonable approximation, the ion cores may be considered at rest and the potential experienced by an electron in a crystal is assumed to be periodic with period equal to the lattice constant a as shown in Fig. 5.9. The deep attractive wells mark the positions of the positive ions. We note that a deep potential well is located at each ion due to the Coulomb forces when the potential V varies periodically with distance x. The motion of an electron in a periodic potential yields the following conclusions:

(i) There exist allowed energy bands separated by forbidden regions or energy gaps.

(ii) The electronic energy functions $E(\vec{k})$ are periodic in wave vector $\vec{k}$.

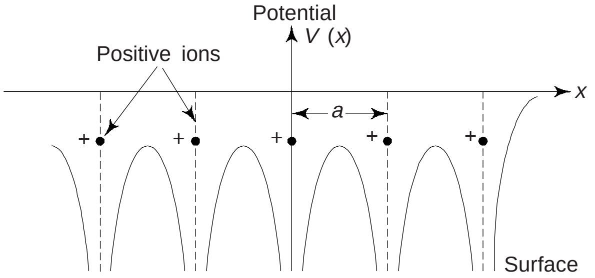


Fig. 5.9 One-dimensional representation of periodic potential experienced by an electron in a perfectly periodic crystal lattice with lattice constant a. The potential near the surface is shown on the right.

These results are in contrast with those obtained from the free electron theory (constant or zero potential case) in the sense that, in the free electron theory, E varies with k as, $E = \dfrac{4\pi^2 h^2 k^2}{2m}$ i.e., there is no upper limit to the energy and k can have discrete values.

Let us now consider the wave motion of electrons, when the wavelength of an electron travelling through the crystal is such that it will be diffracted by a particular set of atomic planar as shown in Fig. 5.10, the Bragg condition is satisfied, i.e.

$$n\lambda = 2d\,\sin\theta \tag{10}$$

where $n \rightarrow$ is any integer, $\lambda \rightarrow$ the wavelength of electrons, $d \rightarrow$ the lattice spacing and $\theta \rightarrow$ the angle of incidence of electrons. Since $K = \dfrac{2\pi}{\lambda}$, where K is the wave number. From (10), we have

$$K = \frac{n\pi}{d\sin\theta} \tag{11}$$

We must note that K is a vector in the same direction in which the electron is moving and its magnitude is proportional to the magnitude of its momentum.

According to Eq. (11), there is one series of values corresponding to the integral n for which electrons are diffracted and do not enter freely through the crystal.

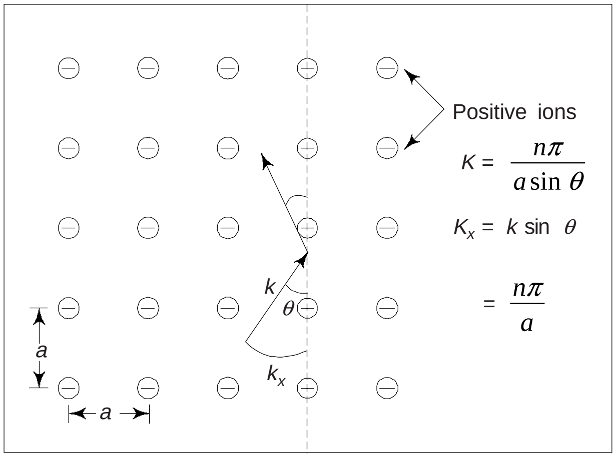


Fig. 5.10 Bragg reflection exhibiting the movement of electrons through the crystal lattice

7. BRILLOUIN ZONES

In relation (11) K represents the wave number. K is a vector in the direction in which the electron is moving and its magnitude is proportional to the magnitude of its momentum. The components of vector K are K_x, K_y and K_z along the x, y and z directions respectively, representing a *momentum, i.e. K-space*. When the Bragg condition (10) is satisfied, the moving electron suffers diffraction by the lattice plane. For this to happen, we have

$$K_x = \frac{n\pi}{d} \qquad (12)$$

for a one dimensional lattice where $n = \pm 1, \pm 2$ for the critical value of K.

Let us consider a simple cubic lattice of two dimensions. We can see that the Bragg's condition is satisfied by the (100) and (010) planes of this lattice. Further, we note that as the energy of the electrons increases, a point is reached when Bragg's condition for another set of planes (110) is satisfied. The area in wave-number space enclosed by a line corresponding to K-vectors is called the *Brillouin zone*. The range of K values between $-\pi/a$ and $+\pi/a$ constitutes the first Brillouin zone. For K between $-2\pi/a$ and $-\pi/a$ and between π/a and $2\pi/a$, is called the second Brillouin zone, and so on. We note that for the first zone, one integer is ± 1 and the other integer is 0. For the second zone, each integer is ± 1. The first two Brillouin zones for a square lattice is shown in Fig. 5.11. We can see that the K-space is divided into zones. Within each zone, when electron touches the lattice boundary, electron encounters an energy barrier preventing further increase of wave number unless some energy is available to excite the electron over the potential barrier. The potential barrier is termed as the *energy gap*. Figure 5.12 shows the energy gaps and Brillouin zones for a simple cubic lattice. Under the action of an external force, the wave number and the energy of the electron vary within the zone, as shown by the arrow.

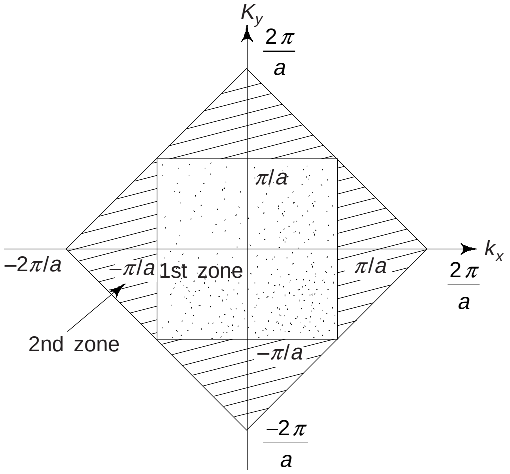


Fig. 5.11 Brillouin zones for a two dimensional crystal lattice

We must note that the different Brillouin zones for a given crystal have the same area. Obviously, the areas of

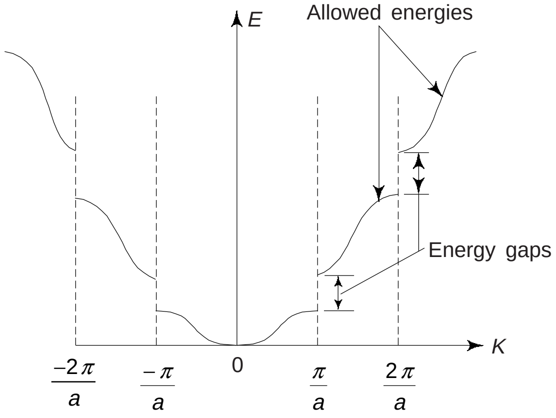


Fig. 5.12 **Energy gaps and Brillouin zones for a simple cubic lattice**

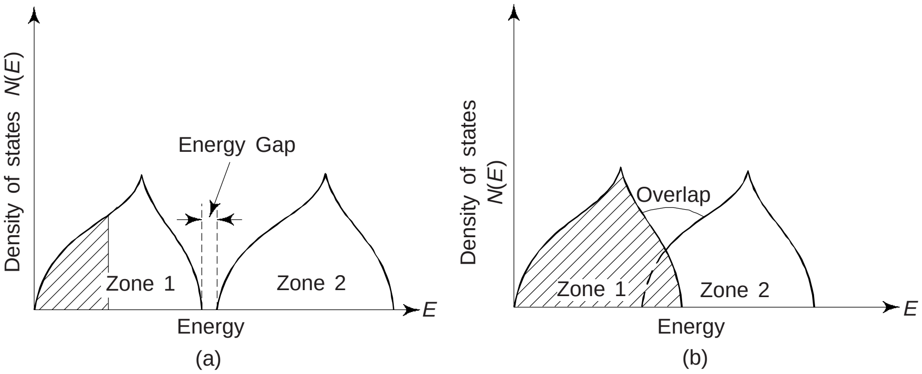
the first zone and the second zone are equal. This is very important characteristic of Brillouin zones. With the help of Brillouin zones, one can easily classify solids into conductors, insulators.

8. DISTINCTION BETWEEN CONDUCTORS, INSULATORS AND SEMICONDUCTORS

On the basis of Zone theory, the solids may be classified into the following three categories: *(i)* conductors *(ii)* Insulators and *(iii)* Semiconductors.

(i) Conductors: In order to find the electrical conductivity of a solid, it is essential to know the size of the energy gap between zones and how completely the Brillouin zones are filled. When the solid is subjected to an external field, there is a flow of electrons in the direction of the electric field. Electrons so excited are capable of moving to the second Brillouin zone and filling the vacant energy levels or states. We know that in a metal there are always unfilled states at energy levels slightly higher than the highest energy of filled states and by applying the external field, electrons can be raised.

Let us consider the case of monovalent metals such as Sodium, Potassium etc. In these monovalent metals, theoretically at the most two electrons per atom are present. Since the atoms of the metal have only a single valence electron per atom, we can easily see that only half the first zone could be filled with the available electrons. (Fig. 5.13(a)). The second Brillouin zone is at a comparatively higher energy level. Since half the energy levels are vacant, and hence any small applied electric field can accelerate electrons to move into the second zone or higher energy levels or states.



(a)

(b)

Fig. 5.13 **Brillouin zones for conductors**

We can see that conduction also occurs when there are enough electrons to fill the Brillouin zone completely, provided there is an overlapping of zones (Fig. 5.13(b)). We must note that in this case the highest energy level in the first zone is much greater than the lowest energy level in the second zone. As a result, one finds that there is an overlapping of energy levels, e.g. zones overlap in metallic crystals like silver and copper.

It is assumed that there are sufficient electrons to fill up the first zone, in the overlapping-zone system. Due to overlapping in energy levels, higher energy electrons can be moved to other levels in the second zone and there will be conduction. We note that due to transfer of electrons from the first zone to the lower portion of the second zone, both zones are partly filled with the energy levels.

Insulators

These are the solids in which there exist a large energy gap between the first zone and the second zone (Fig. 5.14), and the energies of the lowest levels in the second zone should be much higher than that of the upper levels in the first zone. Let us consider a monovalent metal that has just enough electrons to fill the first zone completely. We know that each zone is capable of holding a maximum of two electrons per atom. Obviously, the first zone contains two electrons per atom and as the energy gap between the first and second zone is wide, the charge cannot flow to the second zone in the solid. The electrons in the first zone cannot move to the higher energy levels as all the energy states are filled there and a normal electric field cannot excite them to rise to the next allowed energy level. Obviously, the electrical conduction in insulators cannot take place. It is reported that if the temperature of an insulator is raised, some electrons do jump into the second zone and therefore a small electrical conduction may take place. Rubber, bakelite, mica, etc. are few examples. Diamond is an excellent insulator. Diamond has four valence electrons per atom in each zone, just sufficient to completely fill the first zone. There is wide energy gap between zone 2 and zone 1.

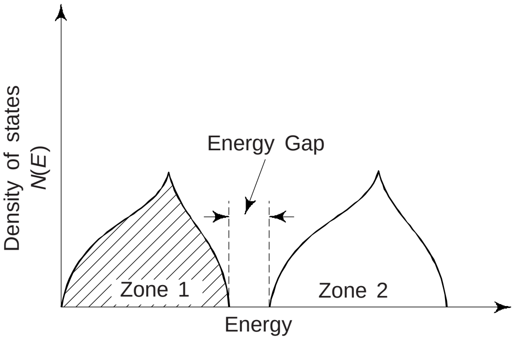


Fig. 5.14 Zones for insulators

Semiconductors

These are a group of materials in which the energy gap between the filled and unfilled zones is very small and zones do not overlap (Fig. 5.15). Due to small energy gap between the two zones, no electron can jump from the first zone to the second zone (which is empty) at absolute zero temperature (i.e., at 0°K). When temperature is raised, say, at room temperature, electrons get sufficient thermal energy which makes electron to jump across the small gap into the empty zone 2. Obviously, in semiconductors, the electrical conductivity (or flow of electron, i.e., current) increases with the increase in temperature. At 0°K, a semiconductor behaves as an insulator since the thermal energy available to cross the energy is zero. The electrical conductivity of semiconductors lies in between those of conductors and insulators.

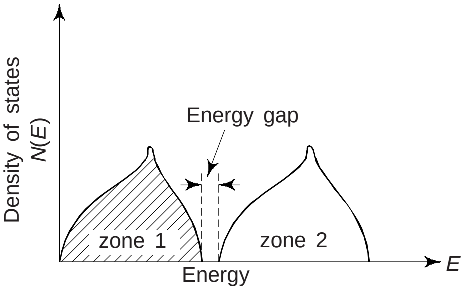


Fig. 5.15 Zones for semiconductors

9. ELECTRICAL RESISTANCE OF MATERIALS

The electrical resistance of materials is affected by the following factors:

(i) Alloying elements or impurities causes increase in lattice imperfections and also reduces the electrical conductivity in comparison to pure metal.

In general the resistivity of an alloy is given by the relation

$$\rho_{\text{alloy}} = \rho_{\text{Metal}} + x\,\rho_l$$

where

ρ_{metal} = Resistivity of the parent metal,

x = Amount of impurity added in the metal. This is expressed in atomic percent

ρ_l = Resistivity of impure metal

For example, if nickel is added to copper, the resistivity of copper goes up by 1.3 micro-ohm-cm for each atm. % addition of nickel. One can express this by the following relation

$$\rho_{Ni\text{–}Cu} = \rho_{\text{Cu}} + x \times 1.3 \text{ (Micro-ohm-cm)}$$

(ii) The electrical resistance of metallic conductors also increases with increase of temperature. The relation exhibiting the variation of a metal resistance with temperature is

$$\rho_t = \rho_{20}\,[1 + \alpha\,(t - 20)]$$

where $\rho_t \rightarrow$ Resistivity of the metal at t°C, $\rho_{20} \rightarrow$ resistivity of the metal at 20°C (i.e., at standard room temperature), and $\alpha \rightarrow$ coefficient of temperature resistance.

We must remember that the metals have positive value of temperature coefficient. The higher the temperature the wider is the thermal vibrations of atoms in the lattice. This causes in atomic spacings being less regular, i.e., decreasing the mobility of electrons in metals.

(iii) Electrical resistance of a material is affected by the nature of the material, its composition and condition of its surface.

(iv) Mechanical processes like cold working and increased strain hardening can cause a number of dislocations in a material, thereby increasing the electrical resistance.

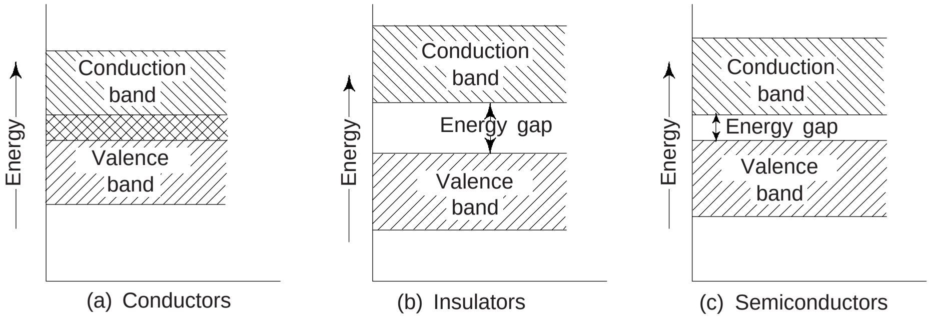
(v) It is observed that due to age hardening, electron mobility is restricted. This causes the electrical resistance of the metal increases. It is found that annealing has an opposite effect.

10. ENERGY BANDS

One of the most important reasons for the initial success of the band theory of solids was that it offered a simple explanation of markedly different electrical behaviour of solids (i.e., conductors, semiconductors and insulators).

Conductors

A crystalline solid is called a conductor or metal if the uppermost energy band is partly filled [Fig. 5.16(a)] or the uppermost filled band and the next unoccupied band overlap in energy. Here, the electrons in the uppermost band find neighbouring vacant states to move in, and thus behave as free particles. When an electric field is applied, these electrons gain energy from the field and produce an electric current, so that a metal is a good conductor of electricity. The partially filled band is called the *conduction band.* The electrons in the conduction are known as *free electrons* or *conduction electrons.* Silver, copper, aluminium, etc. are few examples of good conductors.



(a) Conductors (b) Insulators (c) Semiconductors

Fig. 5.16 Energy band structure (a) metal (b) insulator (c) semiconductor.

Insulators

In some crystalline solids, the forbidden energy gap between the uppermost filled band, called the valence band, and the lowermost empty band, called the conduction band, is very large [Fig. 5.16(b)]. In such solids at ordinary temperatures only a few electrons can acquire enough thermal energy to move from the valence band into the conduction band. Such solids are known as insulators. Since only a few free electrons are available in the conduction band, an insulator is a bad conductor of electricity. It has been observed that if the temperature of an insulator is increased, some electrons do jump to the conduction band. Therefore a small electrical conduction may take place. Most covalent solids, which are composed of atoms having an even number of valence electrons, are insulators. Rubber, bakelite, mica etc. are good examples of insulators.

Semiconductors

A material for which the width of the forbidden energy gap between the valence and the conduction band is relatively small (~1 eV) is referred as semiconductor [Fig. 5.16(c)]. Germanium and Silicon having forbidden energy gaps of 0.78 and 1.2 eV respectively at 0K are typical semiconductors. As the forbidden gap is not very wide, some of the valence electrons acquire enough thermal energy to go into the conduction band. These electrons then become free and can move about under the action of an applied electric field. The absence of an electron in the valence band is referred to as a *hole*. The holes also serve as carriers of electricity. The electrical conductivity of a semiconductor is less than that of a metal but greater than that of an insulator. The energy gaps of some insulators and semiconductors are given in Table 5.2.

Table 5.2 Energy gaps (eV)

Insulators	*Energy gap (eV)*	*Semiconductors*	*Energy gap (eV)*
Diamond	5.33	Silicon	1.14
Zinc oxide	3.2	Germanium	0.67
Silver chloride	3.2	Tellurium	0.33
Cadmium sulphide	2.42	Indium antimonide	0.23

The electrical conduction in semiconductors which we have described is called *instrinsic conductivity*. The conductivity of a semiconductor can also be enhanced by the addition of certain impurities.

11. EQUATION OF MOTION OF AN ELECTRON

According to Drude-Lorentz theory, the motion of electrons is random (when no electric field is applied). Obviously, the number of electrons in a metal moving from left to right at any time is the same as that moving from right to left. This shows that no net current flows through the metal. However, if we apply an electric field across a metal, the electrons move in the positive direction of the field and current is produced.

Let us consider that an electron within a metal moving in any direction and at any time under the applied field 'E'. Let m is the mass of an electron, v is the velocity of the electron and e be the charge on the electron. The force experienced by an electron due to the applied electric field, E is

$$F = eE \tag{13}$$

Due to this force, the electron moves with an average acceleration, d^2x/dt^2 (= dv/dt). The force with which the electron moves,

$$F = m\,\frac{dv}{dt} \tag{14}$$

from (13) and (14), we have

$$m\,\frac{dv}{dt} = eE$$

or
$$\int dv = \frac{eE}{m} \int dt$$

or
$$v = \frac{eE}{m}\ t + K \tag{15}$$

where K is constant of integration, which represents the random velocity of the electrons. The average value of random velocity must be zero, otherwise there will be a flow of current even in the absence of external field. Thus $K = 0$. We have from Eq. (15)

$$v = \frac{eE}{m}\ t \tag{16}$$

Obviously, velocity is directly proportional to time 't'. This clearly reveals that the velocity of an electron continues to increase with time till the collision does not occur. Now, if 't' is the collision time, i.e. average time between the two successive collisions, we have the average velocity of the electron

$$v = \frac{eE \cdot t}{m} \quad \text{or} \quad \frac{v}{e} = \frac{et}{m} \tag{17}$$

Equation (17) is called the equation of motion of an electron under the applied electric field. The average velocity is also called as '*drift velocity*', because the drift in electrons is due to applied field (E).

12. RESISTIVITY AND CONDUCTIVITY

The property (or ability) of a metal due to which it opposes the flow of current through it is called resistivity (ρ). The unit of resistivity is Ohm-m (Ω-m). The reciprocal of resistivity is termed as *conductivity* (σ). The unit of conductivity is Ohm^{-1}-m^{-1} or mhos/metre (mhos-m^{-1}). Thus

$$\sigma = \frac{1}{\rho} \tag{18}$$

13. CURRENT DENSITY IN METAL

Let us consider that N number of free electrons ae distributed uniformly throughout a metal or conductor of length L and cross-sectional area A (Fig. 5.17). Now, an electric field E is applied to such a conductor. The electrons travel a distance L metres in T seconds. We note that this makes the average velocity of electrons equal to $1/T$.

Now, the number of electrons passing through any area per second, $= N/T$.

$\therefore$ Total charge passing through any area per second, called current = charge on an electron × Number of electrons per second crossing any area

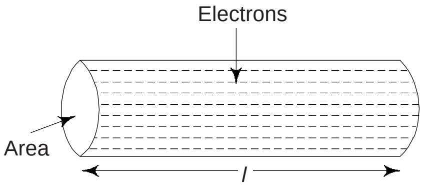


Fig. 5.17

$$= e \times \frac{N}{T} = \frac{Ne}{T}$$

Multiplying and dividing R.H.S. of the above equation by the total charge passing through any area per second, we have

$$= \frac{eN}{h}\frac{L}{T} = \frac{eNv}{h} \tag{19}$$

Obviously, the current, $I = \dfrac{eNv}{h}$ and the current per unit area, i.e. current density,

$$J = \frac{I}{A} = \frac{eNv}{LA}$$

Here LA is the volume of the conductor containing N electrons and therefore the concentration of electrons per unit volume

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

and $$J = env$$

But $$v = \frac{eEt}{m}$$

$$\therefore \qquad J = en\left(\frac{eEt}{m}\right) = \frac{e^2 nEt}{m} \tag{20}$$

In Eq. (20), e, n, t and m are all constants for any conductor. Therefore, the term e^2nt/m is a constant and is equal to conductivity of a metal, i.e.

$$\sigma = \frac{e^2 nt}{m} \tag{21}$$

Using (21), Eq. (20) reduces to

$$J = \sigma E \tag{22}$$

From (22) it is obvious that the current density in a conductor is directly proportional to the applied electric field (E). Now, substituting

$$\frac{v}{E} = \frac{et}{m}$$

We have

$$\sigma = \frac{nev}{E} \tag{23}$$

14. MOBILITY

It is observed that the average velocity of the electrons in a conductor is directly proportional to the applied electric field, i.e.,

$$v \propto E = \mu E \tag{24}$$

The constant of proportionality, μ is called mobility of the electrons. It is usually expressed in m^2/Volt-sec. or cm^2/Volt-sec. The magnitude is given by

$$\mu = \frac{et}{m} \tag{25}$$

where e is charge of an electron, $m \rightarrow$ mass of the electron and $t \rightarrow$ collision time. Now, substituting μ (= v/E) in eq. (23), one obtains

$$\sigma = \frac{nev}{E} = ne\mu \tag{26}$$

Sometimes, one requires to determine the concentration of free electrons per unit volume (n). The number of free electrons per unit time is given by

$$n = \frac{\text{Avogadro's number (N)}}{\text{Molar volume}}$$

The molar volume is given by

$$= \frac{\text{Atomic weight}}{\text{Density}}$$

$$\therefore \qquad n = \frac{N \times \text{Density}}{\text{Atomic weight}} = \frac{Nd}{\text{Atomic weight}}$$

Example 1 *Evaluate the temperature at which there is one per cent probability that a state with an energy 0.5 eV above the Fermi energy will be occupied by an electron.* [B.E.]

Solution The probability that a particular quantum state at energy E is filled, is given by

$$f(E) = \frac{1}{1 + \exp{(E - E_F)/kT}}$$

$$\therefore \qquad 0.01 = \frac{1}{1 + e^x}$$

where

$$x = \frac{E - E_F}{kT}$$

$$E - E_F = 0.5 \text{ eV}$$

$$f(E) = 1\% = 0.01$$

$$1 + e^x = 100$$

$$e^x = 99$$

or

$$x = 2.303 \log 99$$

or

$$\frac{0.5}{kT} = 2.303 \log 99$$

$$\therefore \qquad kT = \frac{0.5}{2.303 \times \log 99} = 0.109 \text{ eV} = 0.109 \times 1.6 \times 10^{-19} \text{ J}$$

$$\therefore \qquad T = \frac{0.109}{k} = \frac{0.109 \times 1.6 \times 10^{-19}}{1.38 \times 10^{-23}} = 1264 \text{ K.}$$

Example 2 *There are 10^{19} electrons/m^3, which serves as carriers in a material. The conductivity of material is 0.01 Ohm^{-1}/m. Find the drift velocity of these carriers, when 0.17 Volt is applied acorss 0.27 mm distance with the material. Given: e = 1.602 $\times 10^{-19}$ C and m = 9.1 $\times 10^{-31}$ kg.* [AMIE]

Solution

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

$n = 10^{19}/\text{m}^3$

$\sigma = 0.01 \text{ Ohm}^{-1}\text{m}^{-1}$

$V = 0.17$ Volt

$d = 0.27 \text{ mm} = 0.27 \times 10^{-2}$ m

$e = 1.602 \times 10^{-19}$ C

$m = 9.1 \times 10^{-31}$ kg

$$= \frac{0.17}{0.27 \times 10^{-3}} = 630 \text{ Volt/m}$$

Let v be the drift velocity. The conductivity

$$\sigma = \frac{nev}{E}$$

or
$$0.01 = \frac{10^9 \times 1.602 \times 10^{-19}\ \text{V}}{630} = 2.54 \times 10^{-3}\ \text{v}$$

∴
$$v = \frac{0.01}{2.54} \times 10^{-3} = 3.93\ \text{m/s}.$$

Example 3 *Find the conductivity of copper at 300 K. The collision time for electron scattering in copper at 300 K is 2 × 10^{-14} sec. Given that density of copper = 8960 kg/m^3, atomic weight of copper = 63.54 amu and mass of an electron = 9.1 × 10^{-31} kg.* [AMIE, Diploma]

Solution We know that 63.54 gms. of copper contains 6.023 × 10^{23} free electrons (i.e., Avogadro's number) as the one atom contributes one electron. The volume of 63.54 gm of copper is 8.9 cc. Obviously, the number of electrons per unit volume (cc),

Temp = 300 K
Time $t = 2 \times 10^{-14}$ s
Density of copper = 8960 kg/m^3
At. Wt. of copper = 63.54 amu
$m = 9.1 \times 10^{-31}$ kg.

$$n = \frac{6.023 \times 10^{23}}{63.54/8.9} = 8.5 \times 10^{22}$$

Number of electrons per cubic meter
$$n = 8.5 \times 10^{22} \times 10^{6} = 8.5 \times 10^{28}$$

The conductivity,
$$\sigma = \frac{e^2 nt}{m} = \frac{(1.6 \times 10^{-19})^2 \times 8.5 \times 10^{28} \times 2 \times 10^{-14}}{9.1 \times 10^{-31}}$$
$$= 4.78 \times 10^{-7}\ \text{mho/m}$$

Example 4 *The mean free time between the collisions is 10^{-14} sec. Find the mobility of electrons?* [AMIE]

Solution We have
$$\mu = \frac{et}{m}$$

$e = 1.602 \times 10^{-19}$ C
$m = 9.1 \times 10^{-31}$ kg
$t = 10^{-14}$ sec.

$$= \frac{1.602 \times 10^{-19} \times 10^{-14}}{9.1 \times 10^{-31}}$$
$$= 1.76 \times 10^{-3}\ \text{m}^2/\text{V-s}$$

Example 5 *The conductivity of silver is 6.5 × 10^{7} per Ohm per m and number of conduction electrons per m^3 is 6 × 10^{28}. Find the mobility of conduction electrons and the drift velocity in an electric field of 1 V/m. Given m = 9.1 × 10^{-31} kg and e = 1.602 × 10^{-19} C.* [M.Sc. (M.S.)]

Solution
$$\sigma = ne\mu$$

or
$$\mu = \frac{\sigma}{ne}$$

$\sigma = 6.5 \times 10^{7}\ \text{Ohm}^{-1}\text{-m}^{-1}$
$n = 6 \times 10^{23}/\text{m}^3$

$$E = 1 \text{ V/m}$$

$$e = 1.602 \times 10^{-19} \text{ C}$$

$$m = 9.1 \times 10^{-31} \text{ kg}$$

$$= \frac{6.5 \times 10^7}{6 \times 10^{23} \times 1.602 \times 10^{-19}}$$

$$= \frac{6.5 \times 10^7}{9.612 \times 10^4} = 6.76 \times 10^3 \text{ m}^2/\text{V-s}$$

Drift velocity, $v = \mu E = 6.76 \times 10^3 \times 1$ m/s

$$= 6.76 \times 10^3 \text{ m/s.}$$

Example 6 *The density of silver is 10.5 gm/cc and its atomic weight is 107.9. Assuming that each silver atom provides one conduction electron, find the number of free electrons/cc. Take conductivity of the silver as 6.8 $\times 10^7$ mhos/m. Also calculate the mobility of electrons. Given e = 1.6 $\times 10^{-19}$ C.*

Solution Let μ be the mobility of free electrons. The number of free electrons,

$$d = 10.5 \text{ gm/cc}$$

$$\text{At. Wt} = 107.9$$

$$\sigma = 6.8 \times 10^7 \text{ mhos/m}$$

$$= 6.8 \times 10^5 \text{ mhos/cm}$$

$$e = 1.6 \times 10^{-19} \text{ C}$$

$$\because \quad n = \frac{Nd}{At\ wt}$$

$$= \frac{6.023 \times 10^{23} \times 10.5}{107.9}$$

$$= 5.86 \times 10^{22}$$

$$\sigma = ne\mu$$

$$\therefore \quad 6.8 \times 10^5 = 5.86 \times 10^{22} \times 1.6 \times 10^{-19} \times \mu = 9.39 \times 10^3\ \mu$$

$$\therefore \quad \mu = \frac{6.8 \times 10^5}{9.39 \times 10^3} = 72.42 \text{ cm}^2/\text{V-s.}$$

15. MEAN FREE PATH

One can define a metal as a substance, which consists of a lattice of a positive ion cores held together by means of a loosely bound valence electrons, also called gas of electrons or delocalized electrons. We know that these electrons have a wave characteristics as they move throughout the metal. When waves travelling through a periodic structure, i.e., a structure which has uniform repetition, proceed with a minimum interruption. Any irregularity in the periodic, i.e. repetitive structure, through which the wave travels, will deflect the wave. Obviously, when an electron is travelling towards the positive electrode, a foreign or displaced atom could cause it to be reflected towards the negative electrode.

We must note that while moving towards the positive electrode, electrons continuously acquire additional momentum and hence more velocity. And when electrons move towards the negative electrode, they continuously lose momentum and hence velocity. Thus the distance between reflections and deflections determines the *net* or *drift* velocity of electrons. Obviously, *mean free path is an average* distance which an electron covers in its wavelike pattern without any reflection or deflection. Mathematically, one finds the following relation for mean free path,

$$\lambda = vt \tag{27}$$

where $\nu \rightarrow$ velocity of an electron and $t \rightarrow$ collision time or mean free time.

For metals, the velocity of an electron corresponds to that of Fermi energy (E_F) and is given by the relation

$$\nu_F = \sqrt{\frac{2W_F}{m}} \tag{28}$$

where $W_F \rightarrow$ Fermi energy in Joules (= 1.602×10^{-19} times the Fermi energy in electron volts, i.e. $1.602 \times 10^{-19}\ E_F$) and $m = 9.1 \times 10^{-31}$ kg.

Example 7 *(i) Estimate the maximum velocity of an electron in a metal in which Fermi energy has a value of 3.75 eV. Given $e = 1.602 \times 10^{-19}$ C and $m = 9.1 \times 10^{-31}$ kg. (ii) What will be the mobility of electrons when the mean free time between the collisions is 10^{-14} S?* [M.Sc.]

Solution

$$W_F = 1.602 \times 10^{-19}$$
$$E_F = 1.602 \times 10^{-19} \times 3.75$$
$$= 6 \times 10^{-19}\ \text{J}$$

(i) $$\nu_F = \sqrt{\frac{2W_F}{m}} = \sqrt{\frac{2 \times 6 \times 10^{-19}}{9.1 \times 10^{-31}}}\ \text{m/s}$$
$$= 1.76 \times 10^{6}\ \text{m/s}$$

(ii) $$\mu = \frac{et}{m} = \frac{1.602 \times 10^{-19} \times 10^{-14}}{9.1 \times 10^{-31}} = 1.76 \times 10^{-3}\ \text{m}^2/\text{V-s}$$

Example 8 *The Fermi level for Potassium is 2.1 eV. Estimate the velocity of electrons at the Fermi level. Given $e = 1.6 \times 10^{-19}$ C and $m = 9.109 \times 10^{-31}$ kg.* [B.E.]

Solution

$$E_F = 2.1\ \text{eV}$$
$$e = 1.602 \times 10^{-19}\ \text{C}$$
$$m = 9.109 \times 10^{-31}\ \text{kg}$$
$$t = 10^{14}\ \text{s}$$
$$W_F = 1.602 \times 10^{-19}\ E_F$$
$$= 1.602 \times 10^{-19} \times 2.1\ \text{J}$$
$$= 3.364 \times 10^{-19}\ \text{J}$$

The velocity of electrons,

$$\nu_F = \sqrt{\frac{2W_F}{m}} = \sqrt{\frac{2 \times 3.364 \times 10^{-19}}{9.109 \times 10^{-31}}} = 0.86 \times 10^{6}\ \text{m/s.}$$

Example 9 *Estimate the mean free path of free electrons in pure copper at 4 K. The collision time for photon scattering at 4 K is 10^{-9} s. The Fermi energy level for copper is 7 eV. Given: $e = 1.602 \times 10^{-19}$ C and $m = 9.109 \times 10^{-31}$ kg.*

Solution

$$t = 10^{-9}\ \text{s}$$
$$E_F = 7\ \text{eV}$$
$$e = 1.602 \times 10^{-19}\ \text{C}$$
$$m = 9.109 \times 10^{-31}\ \text{kg}$$
$$W_F = 1.602 \times 10^{-19}\ E_F$$
$$= 1.602 \times 10^{-19} \times 7$$
$$= 11.2 \times 10^{-9}\ \text{J}$$

Velocity of electrons,

$$v_F = \sqrt{\frac{2W_F}{m}} = \sqrt{\frac{2 \times 11.2 \times 10^{-19}}{9.109 \times 10^{-31}}} \text{ m/s}$$

$$= 1.57 \times 10^6 \text{ m/s}$$

Mean free path

$$\lambda = v_F\ t = 1.57 \times 10^6 \times 10^{-9} = 1.57 \times 10^{-3} \text{ m} = 1.57 \text{ mm}$$

Example 10 *Find the conductivity of copper at 300 K. The collision time (τ) for electron scattering 2×10^{-14} s at 300 K. Given e = 9.1 $\times 10^{-31}$ kg and m = 9.1 $\times 10^{-31}$ kg.* [B.E.]

Solution The number of electrons per m^3 of copper is obtained as

$$n = \frac{6.023 \times 10^{23} \times 8960}{0.06354}$$

$$= 8.50 \times 10^{23}/\text{m}^3$$

We have, conductivity

$$\sigma = \frac{ne^2\ \tau}{m}$$

$$= \frac{8.50 \times 10^{28} \times (1.602 \times 10^{-19})^2 \times 2 \times 10^{-14}}{9.1 \times 10^{-31}}$$

$$= 4.8 \times 10^7 \text{ Ohm}^{-1}\text{-m}^{-1}$$

16. THERMOELECTRICITY

This implies direct conversion of heat into electricity. If an electric circuit is made by connecting two dissimilar metals and the junctions of the two metals are kept at different temperatures, a current flows in the circuit in a direction depending on the temperature difference between the two junctions and the nature of metals. This effect is known as *Seeback effect* after the name of the discoverer, Seeback. The electric current is called the *thermoelectric current* and is the result of an e.m.f. in the circuit caused by the temperature difference between the junctions. This e.m.f. is termed as *thermo e.m.f.* The circuit formed by the two different metals whose junctions are at two different temperatures is known as *thermocouple* (Fig. 5.18).

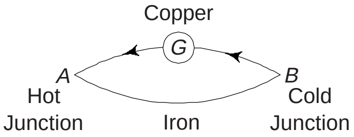


Fig. 5.18 Thermocouple

When the two junctions of a couple are *T K* and the temperature of one junction is raised by a differential *dT*, causing a thermo E.M.F. *dE* in the circuit, the thermoelectric power of the two metals at temperature *T* is defined by

$$P = \left.\frac{dE}{dT}\right|_T \tag{29}$$

The temperature of the hot junction at which the thermo e.m.f. or the thermoelectric current attains its maximum is called the neutral temperature t_n for the given couple. The temperature of the hot junction where the thermo e.m.f. or the current is zero and changes direction, is termed as the *inversion temperature* t_i. The inversion temperature is as much above the neutral temperature as the cold junction is below it

$$t_n = \frac{t_c + t_i}{2} \tag{30}$$

where t_c is the temperature of the cold junction. The variation of the thermo e.m.f. with the temperature of the hot junction is shown in Fig. 5.19. The nature of the curve is the same for any pair of metals in the Seeback series (Bi, Ni, Co, Pd, Pt, U, Cu, Mn, Ti, Hg, Pb, Sn, Cr, Mo, Rh, Ir, Au, Ag, Zn, W, Cd, Fe, As, Sb, Te); the numerical values will however will be different for different couples. In most of the cases, the curve of Fig. 5.19 represent a parabola to a first approximation.

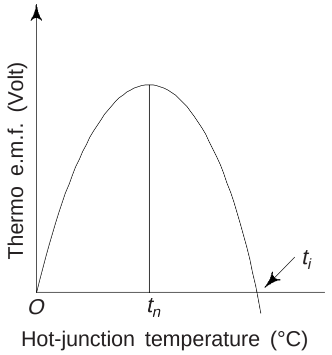


Fig. 5.19 Variation of the thermo e.m.f. with the temperature of the hot junction

17. ORIGIN OF THE THERMOELECTRIC EFFECT

One can explain the origin of the thermoelectric effect with the help of the free electron theory. A conductor possesses a large number of free electrons, the concentration of which is different for different metals. When the wires of two different metals are placed in contact with each other, the electrons diffuse one metallic wire to another metallic wire because of the concentration difference (i.e. gradient). As a result of this, one of the metallic wire becomes positively charged and the other negatively charged. This results in a setting up of a potential difference across the contact. This is termed as *contact potential.* We must note that the contact potential is strongly affected by the temperature. When one of the junctions of the thermocouple is heated (keeping the other junction cold), the potential difference set up at the hot junction is comparatively more than that of the cold junction. As a result of this, a net e.m.f. is produced in the thermocouple and called as thermo e.m.f. The thermo e.m.f. is the cause of thermoelectric current.

18. MAGNITUDE AND DIRECTION OF THERMO E.M.F.

It is observed that the magnitude and direction of the thermo e.m.f. produced in a thermo couple depends on the following factors:

(i) Nature of the metals forming the thermocouple: Seeback made a series of metals, known as *Seeback series*, such that if a couple is constructed with any two metals in this series, the thermocurrent flows across the hot junction from the metal appearing earlier in the series to the one appearing later. Seeback series is as follows: Bi, Ni, Co, Pd, U, Cu, Mn, Ti, Hg, Pb, Sn, Cr, Mo, Rh, Ir, Au, Ag, Zn, W, Cd, Fe, As, Sb, Te.

The greater the separation of the two metals in the series, the higher the thermo e.m.f. for a given temperature difference between the junctions. Obviously, the thermo e.m.f. for Ni-Fe couple is greater than for Cu-Fe couple. The direction of the current will be from a metal occurring earlier in this series to a metal occurring later in the series through the cold junction. For, e.g., in Cu-Fe thermo couple, the current flows from Cu to Fe through the cold junction and from Fe to Cu through the hot junction (Fig. 5.20).

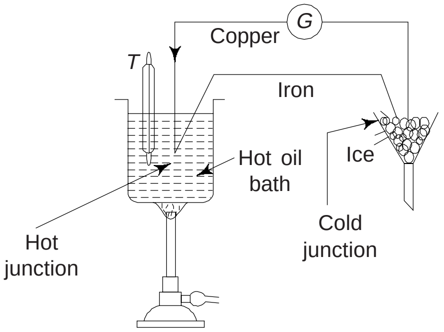


Fig. 5.20 Cu-Fe thermocouple

The magnitude of the thermo e.m.f. is of the order of a few micro-volts per degree temperature difference between the two junctions. Magnitude of thermo e.m.f. depends upon how far the metals are separated in Seeback series. As a general rule more the separation, greater will be the magnitude of thermo e.m.f. The magnitude of thermo e.m.f. for a difference of 100°C temperature is found to be about 1.3 mV for Cu-Fe thermocouple and about 8 mV for antimony-bismuth thermocouple.

(ii) Temperature Difference Between Two Junctions of the Thermocouple: Let us consider a Cu-Fe thermocouple as shown in Fig. 5.20. One of the junctions of the thermocouple is kept hot by immersing in oil bath and heated with a burner. The other junction is kept cold by immersing it in powdered ice (Fig. 5.20). One can record the temperature (T) of the hot junction by a thermometer placed in the hot oil bath. The magnitude of thermoelectric current in the circuit is measured with the help of a galvanometer (G) connected in the circuit.

The temperature of the hot junction (t_n) can be varied by adjusting the amount of heat supplied by the burner. The temperature of the cold junction in t_c (which is 0°C). As the temperature of hot junction is raised above 0°C, the deflection of the galvanometer is directly proportional to the thermoelectric current and hence the thermo e.m.f. As the temperature of the hot junction is further increased, a state is reached at which the thermo e.m.f. becomes maximum, is called a neutral temperature (t_n). We must note that for a given thermocouple, the neutral temperature has a fixed value and does not depend upon the temperature of cold junction. The value of t_n for Cu-Fe thermocouple is 270°C.

As the temperature of the hot junction is further increased, the magnitude of thermo e.m.f. decreases, and ultimately, it becomes zero. Now, if the temperature of the hot junction is further increased, the e.m.f. is produced in the reverse direction. As stated earlier, the temperature of the hot junction, at which the direction of thermo e.m.f. reverses its direction is called the temperature of inversion (t_i). We have already stated that the temperature of inversion is as much above the neutral temperature as the neutral temperature is above the temperature of the cold junction, i.e.

$$t_n - t_c = t_i - t_n$$

or

$$t_n = \frac{t_i + t_c}{2}$$

Obviously, the neutral temperature is the mean of the temperature of inversion (t_i) and temperature of cold junction (t_c).

19. USES OF THERMOCOUPLES

The well-known practical application of the thermoelectric effect is the measurement of temperature. The choice of thermocouple wires depends on the range of temperature involved. For temperatures between 1000°C and 1700°C, platinum coupled with a platinum rhodium alloy is used since these metals do not melt or show an inversion effect at such high temperatures. For temperature differences of one or two hundred degrees, a Cu-Fe couple can be used. It cannot be used at higher temperatures since the neutral temperature is 275°C. A copper-constantan thermocouple is more useful because it offers a large e.m.f. for the same temperature difference. The general requirement is a steady measurable variation of thermo e.m.f. with temperature avoiding neutral and inversion points. The major advantages of the thermocouples are: *(i)* they have a small heat capacity and high conductivity so that their response to the temperatures is very quick, *(ii)* they can be placed at points inaccessible to ordinary temperatures, *(iii)* they can measure temperatures as accurately as the e.m.f. and *(iv)* the thermocouples can be constructed with temperature-resistant metals, so that the temperatures which could not be investigated with ordinary materials like glass can be studied.

The other uses of thermocouples are: *(i)* Thermopile *(ii)* Thermogalvanometer *(iii)* Thermocouple pyrometer, etc.

SUGGESTED READINGS

1. F. Seitz, *Modern Theory of Solids*, McGraw Hill, New York (1940).
2. C. Kittel, *Introduction To Solid State Physics*, 7th ed. Wiley, New York (1996).
3. S.L. Kakani and C. Hemrajani, *Solid State Physics*, 4th ed., Sultan Chand, New Delhi (2004).

REVIEW QUESTIONS

1. Describe the free electron gas model of metals.
2. Write the expression for the Fermi-Dirac distribution function and discuss its behaviour with change in temperature.
3. Explain the physical basis of classification of solids into conductors, semiconductors and insulators. [AMIE]
4. What are the effects of electron energy levels in metals? [AMIE]
5. Discuss the factors which affect the electrical resistance of materials. [AMIE]
6. Explain the difference between conductors, semiconductors and insulators. Give two examples of each category. [AMIE, B.Sc.]
7. Discuss briefly the Zone theory of solids. [AMIE]
8. Draw the energy band diagram of an insulator. [AMIE]
9. Explain the formation of energy bands in solids or on the energy band diagram for metal, semiconductor and insulator. [AMIE]
10. Distinguish between conductors, semiconductors and insulators. [AMIE]
11. Discuss the following:
  (a) electron energies in metals
  (b) zones in conductors and insulators
12. Compare the salient features of the band theory vis-a-vis the free electron theory. Discuss the successes and failures of these theories.
13. State the Bloch theorem and discuss the importance in the band theory.
14. What is thermoelectricity?
15. What are the uses of thermocouples?
16. Why is a Cu-Fe thermocouple not used to measure temperatures above 250°C?
17. What are Brillouin zones? How are they related to the energy levels of an electron in a metal? How does zone theory explain the conducting or insulating character of a material?
18. What are Brillouin zones? Illustrate your answer by constructing two Brillouin zones for a square lattice.

PROBLEMS

1. The atomic radius of silver which crystallizes in FCC structure is 1.4. Calculate the Fermi energy of silver at 0*K*. [5.7 eV]
2. Show that the wavelength associated with an electron having an energy equal to the Fermi energy is given by $\lambda_F = 2(\pi/3n)^{1/3}$.
3. Estimate the fraction of electrons excited about the Fermi level at room temperature for Na and Cu. [0.8%, 0.36%]
4. A copper wire has a resistivity of 1.8×10^{-8} Ω-m at 300°K. Assuming copper is very pure, estimate the resistivity at 700°C and the percentage change in the resistivity from room temperature to 700°C. [5.84×10^{-8} Ω-m, and 22.4%]

OBJECTIVE QUESTIONS

1. If the mobility of electrons in a metal increases, the resistivity
(1) decreases (2) increases
(3) first decreases and then increases
(4) first increases and then decreases (1)

2. Mobility of the electron is
(1) reciprocal of electrical conductivity
(2) flow of electron per unit electric field
(3) average electron drift velocity per unit electric field
(4) none of the above (3)

3. Most widely used conducting materials are
(1) gold and silver (2) copper and aluminium
(3) tungsten and platinum (4) germanium and silk (2)

4. The average drift velocity v of electrons in a metal is related to the electric field E and collision time, τ as

(1) $v = \frac{eE\tau}{m}$ (2) $v = \frac{m}{\sqrt{eE\tau}}$ (3) $v = \sqrt{\frac{eE\tau}{m}}$ (4) $v = \left(\frac{eE\tau}{m}\right)^{2/3}$ (1)

5. The temperature dependence of the classical expression for electrical resistivity of a metal is

(1) $\rho \infty \sqrt{T}$ (2) $\rho \infty T^{-1}$ (3) $\rho \infty T^{-2}$ (4) $\rho \infty T^{2}$ (1)

6. If e, μ and n represent the charge, mobility and concentration of electrons respectively, then the electrical conductivity of the metal is given by

(1) ne (2) $ne\mu$ (3) $\frac{\mu e}{n}$ (4) $n/\mu e$ (2)

7. The Fermi energy of Silver at 0K is 5 eV, the mean energy of electron in silver at 0K is
(1) 3 eV (2) 1.5 eV (3) 4.5 eV (4) 6 eV (1)

8. If the Fermi energy of a metal is 1.4 eV, the Fermi temperature of the metal is approximately
(1) 1.6×10^8 K (2) 1.6×10^6 K (3) 1.6×10^4 K (4) 1.6×10^2 K (3)

9. Which of the following metals has the lowest temperature coefficient of resistance?
(1) Silver (2) Gold (3) Copper (4) Aluminium (2)

10. At very high temperatures, the mean free path and collision time of an electron in a metal are proportional to
(1) $1/T^2$ (2) $1/T$ (3) T^2 (4) independent of temperature (2)

SHORT QUESTION ANSWERS

1. What are metallic conductors?

Ans. Metallic conductors obey Ohm's law, which states that the current in the steady state is proportional to the electric field strength.

2. What are the characteristics of metallic materials?

Ans. Electrical resistivity increases with temperature, impurity content and plastic deformation. The contribution of each to the total resistivity is additive. The resistivity of an impure specimen is given by Mathiessen's rule.

$$\rho = \rho_0 + \rho\,(T)$$

where ρ_0 is a constant for the impure specimen and $\rho(T)$ is the temperature dependent resistivity of the pure specimen.

3. How the resistivity of a metal varies with temperature at low temperatures?

Ans. At low temperatures, the resistivity (ρ) is proportional to the fifth power of absolute temperature, i.e., $\rho \propto T^5$.

4. What is magnetic resonance?

Ans. The conductivity of a metal varies in the presence of magnetic field. This effect is known as magnetic resonance.

5. What do you know about electric and thermal conductivities of metals?

Ans. Metals have high electric and thermal conductivities.

6. What is Wiedemann-Franz law?

Ans. The ratio of electrical to thermal conductivities of a metal is directly proportional to the absolute temperature.

7. Where pure metals are preferred?

Ans. Pure metals are chosen for high conductivity and large temperature coefficient of resistance.

6

Photoelectric Effect

1. INTRODUCTION

We have stated that the characteristic of light is its wave-corpuscle (wave-particle) dualism. On the one hand, it possesses wave properties, manifested in the phenomena of interference, diffraction and polarization; on the other hand, light is a stream of particles, called *photons*, which have zero rest mass and travel at a velocity equal to that of light in a vacuum. The energy (E) and momentum (P) of a photon for the corresponding electromagnetic wave of frequency ν and wavelength λ in a vacuum are:

$$E = h\nu = \frac{hc}{\lambda} \quad \text{and} \quad p = \frac{h\nu}{c} = \frac{h}{\lambda} \tag{1}$$

where h is Planck's constant.

We must note that the wave properties of light predominate at low frequencies; the corpuscular properties predominate at high frequencies.

Photons are produced (emitted) in the processes of transition of atoms, molecules, ions and atomic nuclei from the excited state to a state of less energy, as well as in the acceleration and deceleration of charged particles, and in the decay and annihilation of particles.

The photoelectric effect (photoeffect) is the process of emission of electrons from a metal surface when it is illuminated by high frequency electromagnetic radiation. The effect was discovered by Heinrich Hertz in 1887. When he observed that the passage of an electric current through a gas discharge tube was made much easier when ultraviolet light was allowed to fall on the cathode. It was observed that a plate of metal such as zinc when exposed to ultraviolet radiation became positively charged which showed that it had lost negative charges from its surface. Hallwachs undertaken further experimental study in 1888. Lenard in 1899 showed that carriers of electricity emitted from metal surface, under the action of high frequency radiations were electrons. This phenomenon is known as *photoelectric effect* and the emitted electrons are known as *photoelectrons*. The visible and ultraviolet regions of the electromagnetic spectrum are most often involved, although the infrared and X-ray regions are also of interest.

Metals like Zinc, Cadmium, Selenium, etc. are sensitive only to ultraviolet light; whereas alkali metals like sodium, potassium, etc. are sensitive to visible light. Various devices based on this effect are photo sensitive devices like *photo tubes* or *electric* eyes, photoconducting devices like automatic door opening and closing devices and photovoltaic cells like solar batteries or solar cells, etc.

2. EXPERIMENTAL ARRANGEMENT TO STUDY THE PHOTO ELECTRIC EFFECT

A simple experimental arrangement for the study of photoelectric effect is shown in Fig. 6.1.

The emitting surface, i.e. photosensitive plate A is mounted opposite a metal plate B in a highly evacuated glass tube C. The two plates A and B form two electrodes, to which a variable potential difference can be applied. The evacuated glass tube is fitted with quartz window D (if the cell is to operate in ultraviolet

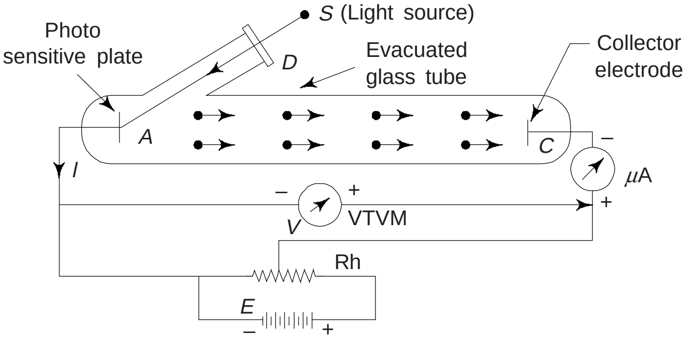


Fig. 6.1 Experimental arrangement to study the photoelectric effect

light), whereas if it is to be used in visible light it is made of common glass, through which ultraviolet light or visible light (as the case may be) from the source *S* is transmitted and allowed to fall on the plate *A*.

When the applied potential difference between the plates *A* and *C* is such that *A* is at negative potential with respect to *C*, the photoelectrons emitted from *A* are accelerated towards plate *C*.

The resulting photoelectric current *I* flowing in the circuit is measured by the micro ammeter (*m* A) and the accelerating potential difference *V* is measured by vacuum tube voltmeter. The variations in the current value versus collector potential *V* are shown in Fig. 6.2.

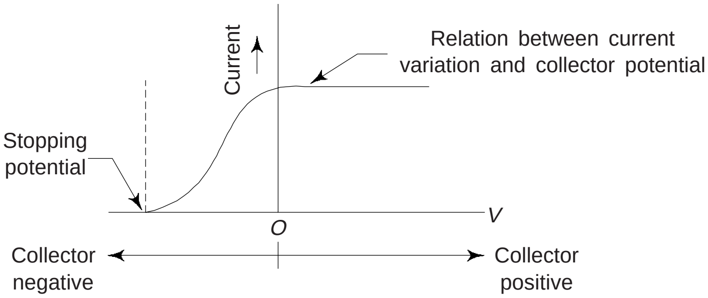


Fig. 6.2

A beam of incident light passing through the window of a photoelectric cell acts like a switch that completes the electric circuit. When the incident light beam strikes the photosensitive metal surface *A*, there is a flow of electrons to the collector electrode (*C*). This causes to flow electric current in the circuit. When the intensity of the light increases, there is a corresponding increase in the number of photoelectrons, and the current in the circuit therefore increases. When we stop the incident light, the photoelectric action also stops and the current ceases. We must note that if the metal film is positively charged, the cell becomes inactive to light since the electrons attempting to leave the plate are held back by electrostatic repulsion. It is found that the photoelectric current depend on the following factors:

(i) The frequency or wavelength of the incident light or radiation
(ii) The intensity of incident light or radiation
(iii) The potential difference between the electrodes, and
(iv) The nature or the photosensitivity of the coated metal

The important experimental observations are as follows:

(i) Effect of Frequency on the Photoelectric Current: The collector plate *B* is made sufficiently positive w.r.t. the emitter photosensitive plate *A*. The surface of plate *A* is illuminated with monochromatic light of

different frequencies. It is observed that the photoelectric current is produced only when the frequency ν of the incident light is greater than a certain minimum value ν_0. This minimum value of the frequency is called the *threshold frequency* for the given surface.

(ii) Effect of the Intensity of the Incident Light: To study the effect of the intensity of the incident light, the collector plate B is made sufficiently positive with respect to the emitter plate A. By keeping the frequency of the incident light and the potential difference V constant, the photoelectric current I is measured for the various intensities of the incident light. Figure 6.3 shows the variation of photoelectric current with the intensity. Obviously, the photoelectric current is *proportional to the intensity of incident light*.

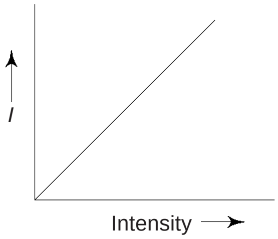


Fig. 6.3 Variation of the photoelectric current (*I*) with the intensity of the incident light

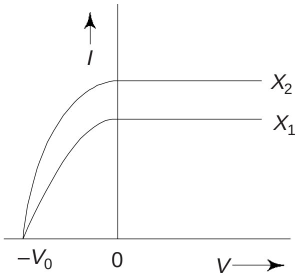


Fig. 6.4 Variation of photoelectric current with the potential difference

(iii) Effect of Potential Difference: Keeping the frequency and the intensity X, of the incident light constant the potential difference between the electrodes is varied. Starting with a small positive potential, it is reduced to zero and then made negative with the help of a reversing key. The variation of photoelectric current with the potential difference is as shown in Fig. 6.4. When the collector plate B is positive, the photoelectric current remains constant. When the collector plate is made more and more negative, the photoelectric current goes on decreasing until it is stopped entirely. The retarding potential that stops the photo current is called the stopping potential. If the intensity of the incident light is increased to X_2, then initial photo current for potential V_0 is found to be the same for the light of the same frequency.

The above experimental observation is explained by assuming that when light of a certain frequency $\nu(\nu > \nu_o)$ is incident on a emitting surface, photoelectrons with kinetic energies ranging from zero to a certain maximum value are emitted from the surface. The photoelectric which are emitted with the maximum kinetic energy are stopped by the stopping potential V_o volts. Obviously, the work done by the retarding potential on the faster electrons must be equal to their kinetic energy, i.e.,

$$eV_o = \frac{1}{2} m\, v_{\max}^2 \tag{2}$$

where m is the mass of the electron, e is the electronic charge and $v_{\max}$ is the maximum velocity of emission of the electrons.

(iv) Effect of Frequency on the Stopping Potential: To study the effect of the frequency of the incident light on the stopping potential, the intensities of light of different frequencies are adjusted to produce the same maximum (I_m) value of the photoelectric current when collector B is positive. The potential of collector B is then reduced in steps and made zero; it is then made more and more negative with the help of the reversing key. For the given photocathode, the graphs obtained are as shown in Fig. 6.5. From Fig. 6.5

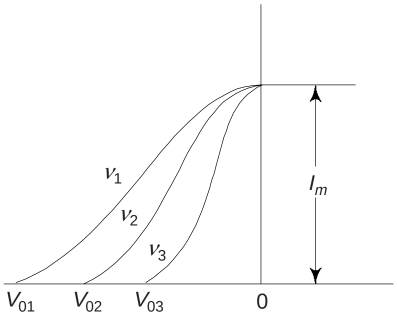


Fig. 6.5

it is evident that the stopping potential increases as the frequency of the incident light is increased. Thus if the frequencies are in the increasing order, i.e. $\nu_1 > \nu_2 > \nu_3$ then the corresponding stopping potentials are also in the increasing order, i.e. $V_{01} > V_{02} > V_{03}$.

(v) Effect of Frequency on the Maximum Kinetic Energy: The magnitudes of the stopping potential at different frequencies are converted into the maximum kinetic energies. A plot of maximum kinetic energy against the frequency is as shown in Fig. 6.6. A straight line is obtained. The straight line *(a)* meets the frequency axis at a certain frequency ν_0. This reveals that ν_0 is the minimum frequency for which the maximum kinetic energy is zero. This minimum value of frequency, is called the threshold frequency for the emitting surface.

Thus the threshold frequency for a given emitting surface defined as minimum frequency of incident light, below which no photo-electrons are emitted, however intense the incident light may be.

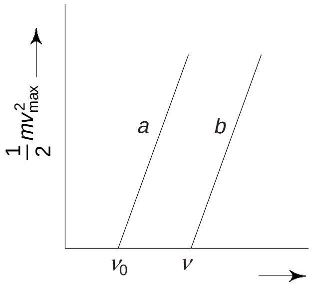


Fig. 6.6

From Fig. 6.6 it is evident for another emitting surface, a straight line *(b)* parallel to *(a)* is obtained.

On the basis of above experimental observations, we can summarize the following important characteristics of the photoelectric effect:

(i) For a given photosensitive surface, there is a minimum frequency ν_0 of incident radiation, below which there is no emission of photoelectrons. This frequency (ν_0) is called the *threshold frequency* for the surface and its value depends on the material and the nature of the emitting surface.

(ii) For a given frequency ν ($\nu > \nu_0$) the number of photoelectrons emitted per second from a given surface, under a constant accelerating potential difference is directly proportional to the *intensity* of the incident radiation.

(iii) When radiation of a given frequency ν ($\nu > \nu_0$) is incident on a given surface, photoelectrons with kinetic energies ranging from zero to a certain maximum value, are emitted from the surface. The stopping potential for the photoelectrons and hence the maximum kinetic energy is independent of the intensity of the radiation.

(iv) The maximum kinetic energy $\left(\frac{1}{2} m v_{max}^2 = e V_0\right)$ of the photoelectrons emitted from a given surface increases linearly with the frequency of the incident radiation. The straight line so obtained for the given surface meets the frequency axis at the threshold frequency ν_0 for the surface. The slope of the straight line is the same for all emitting surfaces and is equal to Planck's constant h ($= 6.6 \times 10^{-34}$ J-sec).

(v) The emission of photoelectrons from photosensitive surfaces is an instantaneous process. As soon as radiation of frequency greater than the threshold frequency is incident on a given photosensitive surface, there is emission of photoelectrons.

3. FAILURE OF CLASSICAL THEORY

The classical theory which assume that light is propagated as electromagnetic waves, fails to explain the above experimental observations on electromagnetic waves. The draw-backs of classical theory are:

(i) According to the classical theory, if the intensity of illumination of incident radiation were increased, electrons would be emitted with greater energies. This is contrary to the experimental observation, i.e. on increasing the intensity of incident radiation the photoelectrons are ejected with the same kinetic energy.

(ii) The existence of threshold frequency for a given material cannot be explained on the basis of wave theory of radiation.

(iii) The instantaneous ejection of electrons from the metal surface also cannot be explained from classical theory. If the illumination is very faint, the classical theory indicates that time of several minutes

would have to elapse before a single electron is released and this is contrary to the above experimental observation.

Einstein applied Planck's quantum theory to explain the photoelectric effect and in 1905 he obtained the photoelectric equation.

4. EINSTEIN'S PHOTOELECTRIC EQUATION

Einstein's photoelectric theory is based on the following two equations:

(i) A radiation of frequency ν consists of a stream of discrete particles, called photons. Energy of each photon or quanta is $h\nu$, where h is Planck's constant. The photons move through space with the speed of light.

(ii) When a quanta of energy $h\nu$, i.e. a photon is incident on a metal surface, the entire energy of the photon is absorbed by a single electron without any time lag. The probability of absorbing two or more photons at the same time by a metal surface is negligible.

Suppose that radiation of frequency ν is incident on the surface of a metal in vacuum (Fig. 6.7). The entire energy of a photon, i.e. $h\nu$ is absorbed by an electron of the metal. We must note that the incident radiation penetrates several atomic diameters into the metal, and electrons at different depths absorb this amount of energy from the incident photons. The energy $h\nu$ absorbed by the electron in the metal is used in the following two ways:

(i) A part of the photon energy $h\nu$ is used up by the electron to do a certain minimum amount of work to overcome attractive forces of positive ions of the metal. This minimum amount of energy is called the *photoelectric work function* and usually denoted by W of the metal. The work function (W) depends on nature of the emitting surface.

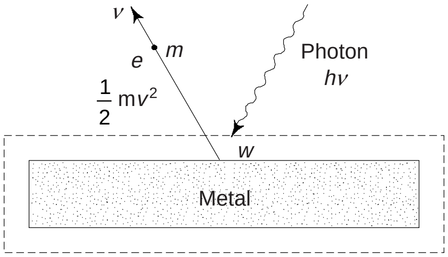


Fig. 6.7

(ii) The remaining energy ($h\nu - W$) appears as the kinetic energy of the photoelectron. If the electron with this kinetic energy comes from the surface, it does not suffer any collision with another electron. In this situation, it will be emitted from the surface into vacuum with this energy as maximum kinetic energy $\frac{1}{2}mv_{\max}^2$, where m is the mass of the electron and $V_{\max}$ its maximum speed for the given frequency. Thus for such an electron, we have

$$\frac{1}{2}mv_{\max}^2 = h\nu - w \tag{3}$$

The above equation is called *Einstein's photoelectric equation*. Suppose the frequency of the incident radiation is reduced to a certain value ν_o such that an electron is just emitted from the surface with zero kinetic energy. Then we have when $\nu = \nu_o$, $\frac{1}{2}mv_{\max}^2 = 0$. Then from eq. (3), we have

$$O = h\nu_o - W$$

or

$$W = h\nu_o \tag{4}$$

Substituting for $w = h\nu_o$ from (4) in (3) we have

$$\frac{1}{2}mv_{\max}^2 = h\nu - h\nu_o = h(\nu - \nu_o) \tag{5}$$

Eq. (5) is another form of Einstein's photoelectric equation. The frequency ν_o is called the *threshold frequency*. On the basis of Einstein's theory, one can define the threshold frequency as follows:

It is the frequency of incident photon which has just sufficient energy to liberate an electron with zero kinetic energy from the emitting metal surface.

It is worthwhile to mention that the energy of light photon incident on the surface of the metal is not divided, but the whole energy of a photon is used up by the same electron in ejecting from the parent atom and imparting kinetic energy.

Explanation of the characteristics of the photoelectric effect from Einstein's equation:

(i) Equation (5) shows that the graph of $\frac{1}{2}mv_{\max}^2$ against ν is a straight line for an emitting surface. For all emitting surfaces the straight lines will have the same slope equal to Planck's constant h.

(ii) When the intensity of incident radiation is increased, then the increased number of photons in the incident radiation will cause the emission of more photoelectrons per second. Obviously, the photoelectric current is proportional to the incident of the incident radiation. The increase of the intensity does not affect the maximum kinetic energy.

(iii) From Eqs. (4) and (5), we have

(a) For $\nu < \nu_o$, no photoelectrons can be emitted.

(b) For $\nu = \nu_o$, photoelectrons are just emitted with zero kinetic energy.

(c) For $\nu > \nu_o$, photoelectrons are emitted with kinetic energies ranging from zero to a maximum value.

(d) For $\nu > \nu_o$, and the intensity of incident light is increased, the number of photoelectrons emitted per second will increase.

Some recent important observations about the photoelectric effect are:

(i) Only a small fraction (~5%) of the incident photon succeeds in ejecting photoelectrons from photometal while rest of the photons are absorbed by the system as a whole and generate thermal energy.

(ii) Photoelectric effect is also observed for isolated atoms in the form of a gas , e.g. Na, K vapour and this process is called photo-ionization.

(iii) The energy required for ejecting the electrons may also be provided by heating the metal, which results in the thermionic emission of the electrons.

(iv) The advent of lasers capable of emitting coherent radiation at high power levels, two or multi photon photo-emission has now been observed. Theory predicts that the double quantum photo current should be proportional to the square of the power of the incident radiation as opposed to the nearly linear relationship that holds for the single quantum photoeffect. In a multi photon photoelectric effect, the kinetic energy of the photoelectron is given by

$$\frac{1}{2}mv^2 = Nh\nu - eV_s \tag{6}$$

and the critical frequency is $\frac{eV_s}{Nh}$ which is smaller than the corresponding frequency for single photon process by a factor of $1/N$.

In the analysis of the photoelectric effect, we have considered a photon as a wave packet of energy, while not making any statements for the quantization of momentum. In fact, since a significant amount of momentum was carried away by the metal, conservation of momentum could not be used for the photoelectron system. A photon as a particle must have both a quantum of energy as well as a quantum of momentum. A. H. compton in 1921 demonstrated in the scattering of X-rays by electrons that a photon have both a quantum of energy as well as a quantum of momentum. We will discuss it later on.

5. MILLIKAN'S VERIFICATION OF EINSTEIN'S EQUATION

From equations (2) and (5), we have

$$eV_o = h\nu - h\nu_o$$

$$\therefore \qquad V_o = \frac{h}{e}\nu - \frac{h}{e}\nu_o \tag{7}$$

The above equation shows that the stopping potential V_o for photoelectrons from a given surface varies linearly with the frequency ν of the incident light. This equation was used by Millikan in 1916 for determining Planck's constants.

The experimental arrangement used by Millikan is shown in Fig. 6.8. Three different alkali metal surfaces are mounted on a drum *D* suitably mounted inside an evacuated chamber. The drum *D* can be rotated from outside so that the surface of any one metal is first scraped clean by sharp knife edge *K* and then brought before the quartz window *W*. A beam of monochromatic light passing through window *W* falls on cleaned surface *C*. The emitted photoelectrons are collected by an electrode *A* which may be maintained at positive or negative potential w.r.t. *C* (or drum *D*) by means of a potential divider potential. The millimeter (mA) measures this photoelectric current.

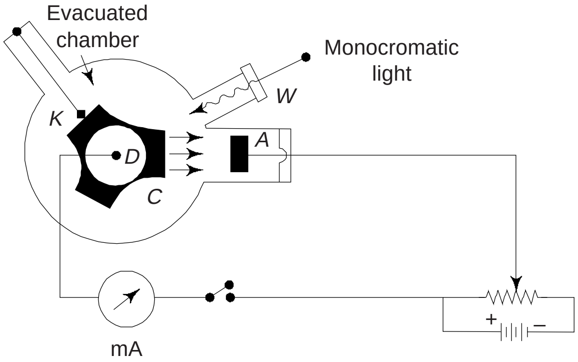


Fig. 6.8

Working. The anode *A* is first given a negative potential w.r.t. *C*, the electrons emitted from *C* are retarded. If the potential is small, only the slower electrons are pulled back and the current falls. With the increase in negative potential, more and more faster electrons are pulled back. Ultimately at a certain potential V_o called the stopping potential, even the fastest electrons fail to reach *A*. At this stage the photoelectric current is zero. Obviously, the kinetic energy of fastest electrons just reduces to zero when they move across the retarding potential V_o. Thus one obtains

$$\frac{1}{2} m v_{\max}^2 = eV_o \tag{8}$$

Obviously, *the stopping potential multiplied by electronic charge gives the maximum kinetic energy of the photoelectrons*.

Millikan plotted a graph between V_o and the frequency of the incident light over a wide range of frequencies and obtained a straight line as shown in Fig. 6.9. Parallel lines were obtained by Millikan for other photo-metals.

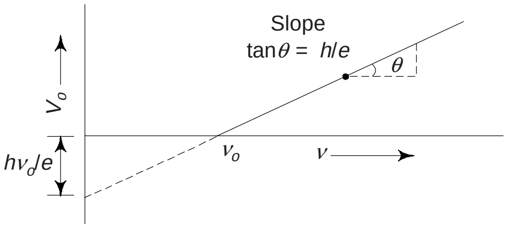


Fig. 6.9

Now, according to Einstein's equations, we have

$$\left(\frac{1}{2} m v^2\right)_{\max} = h(\nu - \nu_o)$$

But from Eq. (8),

$$\left(\frac{1}{2} m v^2\right)_{\max} = eV_o$$

$$\therefore \qquad eV_o = h(\nu - \nu_o)$$

or
$$V_o = \left(\frac{h}{e}\right)\nu - \left(\frac{h}{e}\right)\nu_0 \quad (9)$$

Since h and e are constants and ν_o is constant for a given surface, Eq. (9) indicates that the graph between the stopping potential V_o and frequency of light ν must be a straight line. This is actually the case, as found by Millikan. Obviously, Einstein's equation is of correct form.

Determination of Planck's Constant (h)

Equation (9) shows that the slope of the straight-line graph is h/e, and its intercept on the potential axis is $-h\nu_o/e$. Hence by measuring the slope of the straight-line obtained and, using the known value of e, the value of h can be obtained. The currently accepted value of h is 6.625×10^{-34} Js.

Determination of Work Function

Equation (9) shows that when $V_o = 0$; $\nu = \nu_o$. That is, the intercept of the straight line on the ν-axis gives the threshold frequency ν_o for the particular metal surface. This when multiplied by h, gives the work function $h\nu_o$.

6. PHOTOELECTRIC CELLS

A device which converts radiant energy into electric energy is called a *photo cell.* Photo cells are generally of three types:

(i) Photo-emissive cells *(ii)* Photo-voltaic cells and *(iii)* Photo-conductive cells.

(i) Photo-emissive cells or Phototubes: The emission of electrons in photo-emissive cells is caused by the energy of light beam striking a metallic surface having a low work function. This sensitive surface is enclosed in a vacuum, a gas-filled envelope or quartz and the emitted electrons are collected by a positive anode. These cells are of two types:

(a) Vacuum photo-emissive cell or phototube
(b) Gas-filled photo-emissive cell or gas phototube

(a) Vacuum Photo-emissive Cell The cathode surface in this phototube is treated with a low work function material and a small wire is used (Fig. 6.10). It consists essentially of two electrodes, a cathode C and an anode A. The cathode is semicircular in appearance and made of a metal plate with a coating of a photosensitive material such as Cs and acts as the emitter of the electrons. The anode, acting as the collector, has the shape of a straight wire or a loop so that it does not obstruct the incident light. The anode collects the electrons emitted when the light strikes the light-sensitive cathode. Anode voltages are large enough to achieve current saturations, i.e., all electrons emitted by voltages over about 40 V reach the anode.

To avoid damage of the cathode, the photo current is limited to about 10 μA per lumen of the incident radiation. The chief *advantages* of this type of photo cells are that *(i)* there is no time-lag between the incident light and the emission of photoelectrons and (ii) the photocurrent is proportional to the intensity of light. The variation of current with p. d. (the characteristic curves of the photo cell) is linear (Fig. 6.11), the saturation current being proportional to the intensity. This type of photo cell is extremely accurate in response and hence is used in television and photometry.

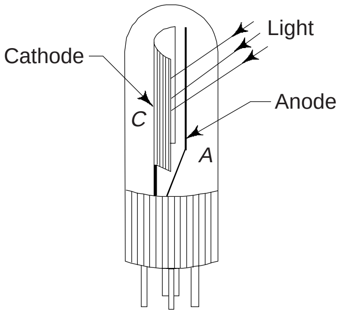


Fig. 6.10 Vacuum tube photo emissive cell

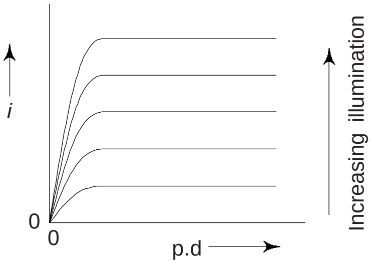


Fig. 6.11 Characteristic curve of photo cell

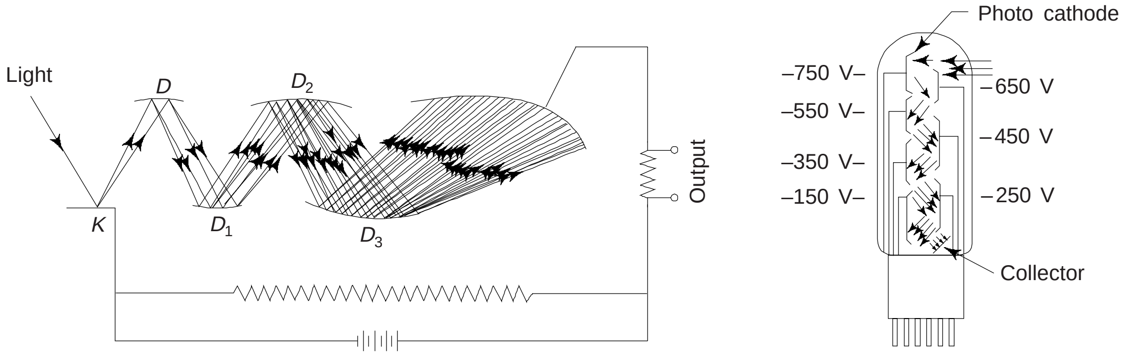
(b) Gas Phototube The current output and sensitivity for a given luminous flux of vacuum photo cell can be increased by introducing an inert gas such as argon or neon at low pressure (few mm of mercury) inside the tube. Ionization by collision of electrons with gas molecules increase the current five to ten times. This effect is also known as *thermionic effect*. This type of photo cells are called as *gas filled photo cells*. When the p. d. between the electrodes exceeds the ionization potential of the gas, the emitted photoelectrons ionise the gas atoms, and thereby a large current flows. The p. d. however should not exceed 100 V for otherwise ions would damage the cathode by an excessive bombardment.

The gas-filled photo cell cannot be used to register very rapid changes of illumination, as it takes only 10^{-4} s for the ions to recombine and the current cannot follow the changes in such a short time. The current thus cannot maintain proportionality with changing intensity of light. However, the gas-filled photo cells could be usefully employed in a film projector to follow the changes in light intensity from the sound track.

When photo-tubes are used at extremely low light levels, the output of the tube may be fraction of an ampere. For many practical applications, e.g. for the reproduction of sound recorded on film in television, etc., it is necessary to amplify the current output of phototubes. These currents can be directly amplified in a device known as *photomultiplier tube* or *electron multiplier*. These tubes are based on the principle of secondary emission. When an highly energetic electron strikes the surface of a metallic plate, it usually causes the emission of other electrons from the surface. The electron striking the surface is called ‘primary electron’ while those emitted from the surface are called ‘*secondary electrons*’. This phenomenon is known as ‘secondary electron emission. By special treatment of the surface anything upto 4 or even more secondary electrons can be emitted for each primary electron.

Photomultiplier Tube: Based on secondary electron emission, this is used for very high amplification of very weak signals.

It consists of a long evacuated tube with a photosensitive cathode K and an anode A (Fig. 6.12). Between K and A, a number of subsidiary anodes D_1, D_2, D_3, . . ., called the *dynodes*, are fitted. The dynodes are coated with a material that allows a copious emission of secondary electrons.



(a) (b)

Fig. 6.12 Photomultiplier tube

The voltages on the dynodes are arranged to increase successively in positive steps of about 200 V.

When a weak signal of light falls on the cathode K, one or more photoelectrons are emitted from K. The photoelectrons are directed by a suitable electrostatic field to the dynode D_1 where a number of secondary electrons are emitted for each (primary) photoelectron. These, in turn, are pulled to dynode 2 where the electrons are further multiplied by secondary emission. This process is repeated at as many as 10 or 12 dynodes, and the highly amplified signal in the form of photo current finally emerges at the anode A.

If there are n-dynodes, each of which has a secondary emission ratio of m, then the number of electrons N obtained at the anode is given by

$$N = N_o\, m^n \tag{10}$$

Where N_0 is the initial number of photoelectrons produced at the cathode. Thus, for e.g., if there are 10 dynodes ($n = 10$) each with a ratio 4 ($m = 4$), then

$$\frac{N}{N_o} = 4^{10} \approx 10^6,$$

i.e., the signal is amplified about a million times. Obviously, the photomultiplier tube can detect very much weaker intensities than the human eye or the photo cell. It is usually attached to a scintillation counter in order to detect very weak X-rays, γ-rays, etc., in spectrophotometers to measure the intensities of spectral lines and in TV pick up devices.

There are two types of photomultiplie actually in use: the *American type* called RCA and the *english type* called EMI.

(ii) Photovoltaic Cells: It is a different type of photocell which develops a small e.m.f. and is therefore called photovoltaic cell. The junction of semiconductors and metals exhibit the photovoltaic effect. This cell is also called as the *barrier layer cell*.

In this type of photocell, the radiant energy striking the photocell surface generates an e.m.f. This e.m.f. can be applied to external circuit. One can design the external circuit in such a way that the external current flows is a linear function of the light intensity. This type of photocell consists of oxides or compounds coated on metallic plates. These cells can be classified as:

(a) Copper-copper oxide photovoltaic cells.
(b) Iron-selenium photovoltaic cells.

Copper-copper photovoltaic cells are either back-wall or front-wall cells. In the back-wall type cell, the semiconductor cuprous oxide is deposited on copper-based plate. The barrier, i.e. the blocking layer existing between the copper-based plate and the cuprous oxide is the point of origin of photoelectrons. *A* semi-transparent metallic layer is placed over the oxide layer. The copper-based plate constitutes one electrode and a thin copper ring in contact with the semi-transparent metallic layer constitutes another electrode. When illuminated, the photoelectrons travel from the cuprous oxide to the copper, i.e. from semiconductor to the metal through the barrier layer as shown in Fig. 6.13(a). Since the barrier layer is remote from the incident light, this cell is known as a back-wall cell.

The cell is called a *front-wall cell*, when the barrier layer lies between the top electrode and cuprous oxide. The direction of the current is from the semiconducting cuprous oxide to the metal electrode through the barrier layer. Obviously, the direction of electron flow in the external circuit of this cell is opposite to that of the black-wall cell (Fig. 6.13(b)).

Both, the front-wall and back-wall cells are usually called as *photon cells*. In comparison to the back-wall cell, the sensitivity of front-wall cell is greater because light reaches the barrier without being partially absorbed by the ineffective cuprous oxide.

In Fig. 6.13(c), the iron-selenium front-wall cell is shown. This cell consists of a base plate of iron. In this cell a thin layer of iron selenide is deposited on the iron-based plate and then iron-selenide layer is covered with a silver layer. The base plate of this cell is connected to a copper ring and forms one electrode. Another copper ring in contact with the silver layer works as the second electrode of the cell.

A photovoltaic effect is produced when radiation is absorbed near a *P-N* junction is a semiconductor. Excellent photovoltaic cells have been made with silicon, germanium or indium antimonide as the semiconductor. Silicon photovoltaic cells are used to power radio-transmitters in artificial satellites. Photovoltaic cells are small and portable, and does not require external battery. Its sensitivity is much greater than that of a photoemissive cell. Following are the main characteristics of photovoltaic cells:

(i) Like photoconductive or photoemissive cells, no external voltage is required in photovoltaic cells.
(ii) Amplication directly by vacuum tube is not possible because a small dc voltage is generated.
(iii) Due to high-barrier-layer capacities, the response falls rapidly with modulated light.
(iv) One can increase the voltage sensitivity by connecting cells in series.

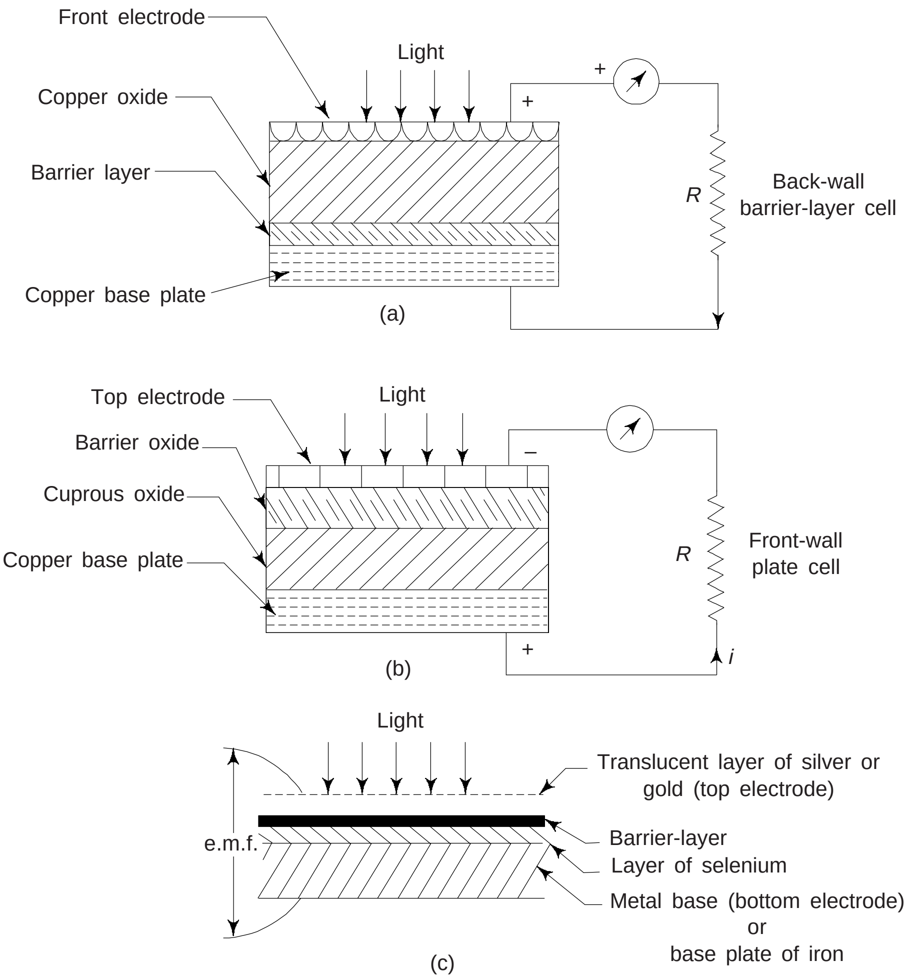


Fig. 6.13 Photovoltaic cells

(v) These cells are highly sensitive to high temperatures. The selenide cell is damaged if operated above 65°C temperature.

This type of cells are used in devices like portable exposure-meters, direct reading illumination meters, solar batteries, solar cells and other monitoring operations in various industries.

Solar Cell or Battery

When we want that the photovoltaic cell may serve as an source of electric energy supply, we consider them as solar cells. They function as a battery to supply voltage, but the sun light serve as the source of energy instead of the chemicals of dry batteries. No doubt, extended period of good weather are a requirement.

The solar cell is a *p-n* junction diode that transform sunlight to electricity with a large conversion efficiency. Solar cells consist of carefully refined wafer-thin strips of silicon, about the size of razor blades, into the faces of which an impurity has been diffused in the form of a microscopic layer near the surface. These formed strips constitute a *p-n* junction, which serves as a solar cell. When solar light is allowed to fall on this *p-n* junction, the energy in sunlight moves electrons to one face and holes to the other, thus creating polarity to the solar battery. The current varies as the intensity of sunlight and continues to flow as long as the *p-n* junction is exposed to sunlight.

Solar cells have long life, relatively high efficiency and more power developed than in other photovoltaic devices.

Solar batteries serve as long-duration power supplies in satellites, space vehicles, remote repeater stations and wire communication circuits.

(iii) Photoconductive Cell: Photoconductivity is a general property of semiconducting solids and consists in the change of resistance of a semiconducting material when radiation is incident upon it. The resistivity of semiconducting materials like selenium, lead sulphide, cadmium sulphide, etc. decreases when irradiated. Obviously, these materials have dark-resistance and low-irradiated resistance. The semiconducting compounds like lead sulphide, lead selenide and lead telluride, when properly prepared in the form of thin films, respond well to the infra red and are now widely used. It is reported that, for photo-emission to occur, relatively large photon energies were necessary, and, as a result, presently there are no known photoemitters for wave lengths beyond the near infrared region of the spectrum. However, in semiconductors, relatively small photon energies are capable of creating electron-hole pairs internally, and thus increasing the carrier concentrations and conductivity of the material. One can use this photo-induced conductivity for light detection from the ultraviolet through the infrared region.

When the energy gap of the semiconductor is in the right range, the electrons in the valence band can be excited into the conduction band by visible light. The incident light creates additional holes and electrons, which in turn lead to an increase in the current in an approximate external circuit. We know that the current is a direct measure of the light intensity and the devices used for detecting and measuring the light energy are called photo conductors.

The simplest form of a photoconductive cell using selenium is shown in Fig. 6.14. We note that there are two electrodes provided with the semiconductor material. When the photo cell is dark, i.e. unilluminated, its resistance is always high and hence current through the circuit is low. As soon as the cell is illuminated, the current through the circuit will become large due to a decrease in its resistance. In order to get a large ratio of dark to light resistance, the shape of the semiconductor is properly fabricated. Usually the resistance between the terminals falls from about 10^7 ohm in the dark to about 10^6 ohm in bright light.

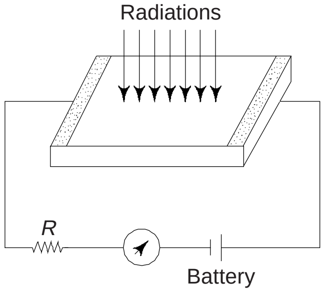


Fig. 6.14 Photoconductive cell

The another commonly used photoconductive cadmium cell is shown in Fig. 6.15. Due to its particular shape, this cell provides very high dark-to-light ratio as well as a maximum response at 5000 Å. The two electrodes of the cell are generally extended in the interdigital pattern so that one could increase the contact area.

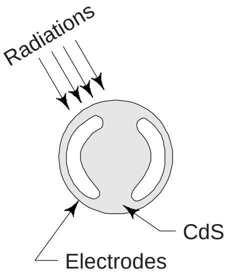


Fig. 6.15 Cadmium sulphide photoconductive cell

These cells are generally used for detecting ships and aircrafts by the radiations given out by their exhausts or fuel gases, as well as for telecommunication by modulated infrared light. When used alongwith suitable amplifiers the device finds use in burglar alarms. The device, however, is sluggish. It takes some time to fully respond to illumination. Further one change of illumination affects its response to the next change.

A circuit for automatic door opening is shown in Fig. 6.14. From figure, it is obvious that a photoconductor is arranged with a light beam falling on it continuously. When an obstacle, i.e. an object approaches the door, this light beam is interrupted, reducing the current flow in the photoconductor which then activates the circuit to open the door. One can also use this cell for low-speed counting or sorting work, in burglar or fire alarms and street lights can be automatically switched on when the sun goes down. Photoconductive devices with an invisible infrared beam can be used for detection of a distant sky object.

7. USES OF PHOTOELECTRIC CELLS

The photoelectric cells are widely used in various fields. Some of the important uses are:

(i) In reproduction of sound in motion films

(ii) In television and cathode ray tubes

(iii) In transmission of pictures to distant places (phototelegraphy)

(iv) In photoelectric counters for counting objects

(v) In automatic light switches for switching on and switching off the street light and in other automatic timing devices

(vi) In burglar alarms to detect thieves and in the fire alarms to indicate the outbreak of fire.

(vii) In determination of temperature of stars

(viii) As illumination-meters in photography, as exposure-meters in photography and as day-light recorders in meterology.

(ix) As photoelectric sorter for sorting out objects of different shades.

(x) In photometry

(xi) In fire alarms

(xii) Cell system in hotels

(xiii) Solar batteries

(xiv) Remote sensing in defence work

(xv) Automatic relays

(xvi) Contrast and brightness control in televisions

(xvii) Sorting machines

(xviii) Photostate work or photocopiers

(xix) In the field of communication photocells are used to transmit static images in phototelephones on infra-red rays without using wires. Photo cells are used in astro compass which finds use in aeroplanes for autopilot. In the military field photocells are used for signal communication, for location of invisible rays and in self-aiming projectiles.

Example 1 *The wavelength of the photoelectric threshold of tungsten is 2300 Å. Determine the kinetic energy of electrons ejected from the surface by ultraviolet light of wavelength 1800 Å.*

[$h = 6.62 \times 10^{-34}$ J-s, $c = 3 \times 10^8$ m/s]

Solution Work function, $W = h\nu_0 = hc/\lambda_0$

$$= \frac{6.62 \times 10^{-34} \times 3 \times 10^8}{2300 \times 10^{-10}}$$

$$= 0.86 \times 10^{-18} \text{ J}$$

The energy of the incident radiation is

$$h\nu = \frac{hc}{\lambda} = \frac{6.62 \times 10^{-34} \times 3 \times 10^8}{1800 \times 10^{-10}} \text{ J}$$

$$= 1.10 \times 10^{-18} \text{ J}$$

Kinetic energy of the ejected electron E is given by

$$E = h\nu - W = (1.10 - 0.86) \times 10^{-18} \text{ J}$$

$$= \frac{0.24 \times 10^{-18}}{1.6 \times 10^{-19}} \text{ eV} = 1.5 \text{ eV}.$$

Example 2 *The work function of sodium is 2.3 eV. Calculate the threshold frequency and the corresponding wavelength.* [AMIE, B.E.]

Solution $W = h\nu_0 = 2.3 \text{ eV}$

$$\nu_0 = \frac{W}{h} = \frac{2.3 \times 1.602 \times 10^{-19}}{6.625 \times 10^{-34}} \text{ Hz}$$

Now $$W = h\nu_0 = \frac{hc}{\lambda_0} = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{\lambda_0} \text{ J}$$

(λ_0 in meters)

$$= \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{1.602 \times 10^{-19}\ \lambda_0} \text{ eV}$$

(λ_0 in Å)

$$= \frac{12400}{\lambda_0} \text{ eV}$$

The work function of sodium is 2.3 eV

or $$2.3 \text{ eV} = \frac{12400}{\lambda_0}$$

∴ $$\lambda_0 = \frac{12400}{2.3} = 5390 \text{ Å}$$

Example 3 *Electrons are emitted with zero velocity from a certain metal surface when it is exposed to radiations of 6800 Å. Calculate the threshold frequency and work function of the metal.*

[AMIE, Diploma]

Solution $$h\nu = W + \frac{1}{2} mv^2$$

$$= h\nu_0 + \frac{1}{2} mv^2$$

As $v = 0,\ \frac{1}{2} mv^2 = 0, \quad \therefore \quad h\nu = h\nu_0$

or $$\nu = \nu_0 = \frac{c}{\lambda} = \frac{3 \times 10^8}{6800 \times 10^{-10}} = 4.416 \times 10^{14} \text{ Hz}$$

Work function $W = h\nu_0 = 6.625 \times 10^{-34} \times 4.416 \times 10^{14}$
$= 2.918 \times 10^{-10}$ J

Example 4 *A blue lamp emit light of a mean wavelength of 4500 Å. The lamp is rated at 150 W and 8% of energy appears as emitted light. How many photons are emitted by lamp per second?*

[AMIE, Diploma]

Solution Light energy emitted/sec $= 150 \times \frac{8}{100}$ J $= 12$ J

$$\nu = \frac{c}{\lambda}$$

Now $$h\nu = h\frac{c}{\lambda}$$

$c = 3 \times 10^8$ m/s
$h = 6.625 \times 10^{-34}$ J/s
$\lambda = 10^{-10}$ m

$$= \frac{6.625 \times 10^{-34} \times 3 \times 10^{8}}{4500 \times 10^{-10}}$$

$$= 44 \times 10^{-20} \text{ J}$$

Now, the number of photons emitted by the lamp per second

$$= \frac{12}{44 \times 10^{-20}} = 26.98 \times 10^{8} \simeq 27 \times 10^{8}$$

Example 5 *Determine the region of electron spectrum which liberates photo-electrons from potassium. Given that the work function of potassium is 2.24 eV.*

Solution The photoelectric equation is

$$h\nu = \frac{1}{2} mv^2 + W$$

When electrons are just liberated, $v = 0$ and the corresponding frequency is

$$\frac{W}{h} = \frac{2.24 \times 1.6 \times 10^{-19}}{6.6 \times 10^{-34}} \text{ sec}^{-1} = 5.43 \times 10^{14} \text{ Hz.}$$

But $$\lambda = \frac{c}{\nu} = \frac{3 \times 10^{8}}{5.43 \times 10^{14}} \text{ m} = 5.520 \times 10^{-7} \text{ m} = 5520 \text{ Å}$$

Obviously, all wavelengths shorter than 5520 Å can liberate electrons from potassium.

Example 6 *A radio receiver operates at a frequency of 440 kHz and a power of 10 kW. Show that it emits about 34 $\times 10^{30}$ photons per second.*

Solution Power of radio receiver = 10 kW = 10,000 W

= 10,000 J/s

= 10^4 J/s

Energy of each photon

$$E = h\nu = 6.625 \times 10^{-34} \times 440 \times 10^{3} \text{ J}$$

$$= 291.28 \times 10^{-30} \text{ J}$$

Number of photons emitted/sec $= \frac{\text{Power}}{\text{Energy}} = \frac{10^4}{291.28 \times 10^{-34}}$

$$= 34.3 \times 10^{30}$$

$$\simeq 34 \times 10^{30}$$

Example 7 *The work functions of tungsten and barium are 4.52 and 2.5 eV respectively. Calculate the wavelength of light which can just eject electrons from tungsten and from barrium. Which metal would you select for a photocell for use with visible light?*

Solution Threshold wavelength $\lambda_0 = \frac{c}{\nu_0} = \frac{ch}{W}$

For tungsten, $\phi = 4.52 \text{ eV} = 4.52 \times 1.6 \times 10^{-19}$ J

$$\therefore \quad \lambda_0 = \frac{3 \times 10^{8} \times 6.62 \times 10^{-34}}{4.52 \times 1.6 \times 10^{-19}} = 2.746 \times 10^{-7} \, m$$

$$= 2746 \text{ Å}$$

For barium,

$$\lambda_0 = \frac{4.52}{2.5} \times 2746 \text{ Å} = 4965 \text{ Å}$$

SUGGESTED READINGS

1. A. Beiser, *Concepts of Modern Physics*, 5th ed., McGraw Hill (1996).
2. S.L. Kakani, *Modern Physics*, Viva Books Pub. New Delhi (In press).

REVIEW QUESTIONS

1. What is photoelectric effect? Write Einstein's photoelectric equation.
2. What is photoelectric effect? Explain the characteristics of the photoelectric effect.
3. Describe briefly an arrangement for studying the photoelectric phenomenon. Point out the important features of the experimental results. How did Einstein explain them?
4. Derive Einstein's photo electric equation and show that the frequency of the incident radiation affects the photoelectric emission.
5. What is photoelectric effect? Explain the laws of photoelectric emission.
6. How would you explain terms related to photo electric effect?
- *(i)* Threshold frequency
- *(ii)* Stopping potential
- *(iii)* Photoelectric work function
- *(iv)* Planck's quantum theory of radiation
- *(v)* Critical wavelength

7. What are the fundamental laws of photoelectric emission? Write Einstein's photoelectric equation.
8. What is a photo cell? Describe the different types of photoelectric cells? Mention the important applications of these cells.
9. Describe in brief the emission of photocells and effect of temperature on their emission. [AMIE]
10. What is photoelectric effect? Describe an experiment to verify this law. [AMIE]
11. Describe some of the devices operating on the photoelectric effect. [AMIE]
12. State and explain Einstein's photoelectric equation. [AMIE]
13. State Einstein's law for the photoelectric effect. Describe an experiment to verify this law. [AMIE]
14. State and explain Einstein's photoelectric equation. [AMIE]
15. Describe Millikan's experiment for the verification of Einstein's photoelectric effect.
16. Write short explanatory notes on
- *(a)* Solar batteries
- *(b)* Photomultiplier tube
- *(c)* Phototubes
- *(d)* Photo sensitive tubes
- *(e)* Photoelectric work function
- *(f)* Einstein's Photoelectric equation
- *(g)* Photovaltaic cell

PROBLEMS

1. Determine the region of the electron spectrum which liberates photo-electrons from potassium. Given that the work function of potassium is 2.24 eV. [All wavelengths shorter than 5520 Å]
2. Light of wavelength 2000 Å falls on an aluminium surface (work function of aluminum 4.2 eV). Calculate: *(i)* the kinetic energy of fastest and slowest emitted photo-electrons *(ii)* stopping potential and *(iii)* cut off wavelength for aluminium. [*(i)* 2 eV, 0 eV, *(ii)* 2 V *(iii)* 3000 Å]

3. The work function of a particular emitter is 2 eV and light of wavelength 3000 Å is used to cause emission. Find the stopping potential and the velocity of the most energetic materials. [2.14 V, 8.68 × 10^5 m/s]
4. The work function of potassium is 1.9 V. If radiation of wavelength 4.5 × 10^{-7} m be incident on it, what would be *(i)* the threshold wavelength *(ii)* the maximum kinetic energy of ejected electrons and *(iii)* the minimum retarding potential? [*(i)* 6.53 × 10^{-7} m *(ii)* 1.37 × 10^{-19} J *(iii)* 0.86 V]
5. The work function for zinc is 3.6 eV. The threshold frequency for the metal is 9 × 10^{14} Hz. Find the value of Planck's constant. [h = 6.4 × 10^{-34} J-s]

OBJECTIVE QUESTIONS

1. The work function for aluminium is 4.2 eV. The cut off wavelength for the photoelectric effect for the surface is
(1) 2955 Å (2) 3100 Å
(3) 4200 Å (4) 1500 Å (1)
2. The work function for aluminium surface is 4.2 eV and that for sodium surface is 2 eV. The two metals were illuminated with appropriate radiations so as to cause photo-emission. Then
(1) both aluminium and sodium will have threshold frequency
(2) the threshold frequency of aluminium will be more than that of sodium
(3) the threshold frequency of aluminium will be less than that of the sodium. (3)
(4) none of the above
3. X-rays of frequency ν are used to irradiate sodium and copper surface in two separate experiments and the stopping potential determined. Then
(1) the stopping potential is more for copper than sodium
(2) the stopping potential is more for sodium than for copper
(3) the stopping potential is same for sodium and copper
(4) none of the above
4. Light of two different frequencies, whose photons have energies 1 and 2.5 eV respectively successively illuminate a metal whose work function is 0.5 eV. The ratio of maximum speeds of the emitted electrons will be
(1) 1 : 1 (2) 1 : 2 (3) 1 : 4 (4) 1 : 5 (2)
5. The work function of a photoelectric material is 3.3 eV. The threshold frequency will be equal to
(1) 4 × 10^{14} Hz (2) 5 × 10^{20} Hz (3) 8 × 10^{10} Hz (4) 8 × 10^{14} Hz (4)
6. Light of wavelength 4000 Å is incident on a metal plate whose work function is 2 eV. The maximum kinetic energy of the emitted photoelectron would be
(1) 0.5 eV (2) 1.5 eV (3) 20 eV (4) 1.1 eV (4)
7. In photoelectric effect the number of electrons emitted is proportional to
(1) work function of cathode (2) velocity of incident beam
(3) frequency of incident beam (4) intensity of incident beam (4)
8. What is the order of time interval between the falling of radiation on a metallic surface and the ejection of photoelectron from it?
(1) 10^{-9} s (2) 10^{-6} s (3) 10^{-4} s (4) 10^{-2} s (1)
9. The threshold wavelength for photoelectric emission from a material is 5200 Å. Photoelectrons will be emitted when this material is illuminated with monochromatic radiation from a
(1) 50 W infrared lamp (2) 1W infrared lamp
(3) 50 W ultraviolet lamp (4) 1W ultraviolet lamp (3, 4)

SHORT QUESTION-ANSWERS

1. How the kinetic energy of the photoelectrons depends on the frequency and intensity of the incident radiation?

Ans. The kinetic energy of the emitted photoelectrons changes only with a change in the frequency of the incident radiation. A change in intensity only changes the number of photoelectrons emitted.

2. What is the condition for photoemission from a photosensitive surface?

Ans. $h\nu > W$, i.e. the work function of the photosensitive surface must be less than incident photon energy.

3. What is photoelectric effect?

Ans. This is the process of interaction of electromagnetic radiation with matter as a result of which the energy of photons is transmitted to the electrons of the substance.

7

Diffusion in Solids

1. INTRODUCTION

Diffusion refers to the transport of atoms through a crystalline or glassy solid. Many processes occurring in metals and alloys, especially at elevated temperatures, are associated with self-diffusion or diffusion. Diffusion processes play a crucial role in many solid-state phenomena and in the kinetics of microstructural changes during metallurgical processing and applications; typical examples include phase transformations, nucleation, recrystallization, oxidation, creep, sintering, ionic conductivity, and intermixing in thin film devices. Direct technological uses of diffusion include solid electrolytes for advanced battery and fuel cell applications, semiconductor chip and microcircuit fabrication and surface hardening of steels through carburization. The knowledge of diffusion phenomenon is essential for the introduction of a very small concentration of an impurity in a solid state device.

2. TYPES OF DIFFUSION

(i) Self Diffusion: It is the transition of a thermally excited atom from a site of crystal lattice to an adjacent site or interstice.

(ii) Inter Diffusion: This is observed in binary metal alloys such as the Cu-Ni system.

(iii) Volume Diffusion: This type of diffusion is caused due to atomic movement in bulk in materials.

(iv) Grain Boundary Diffusion: This type of diffusion is caused due to atomic movement along the grain boundaries alone.

(v) Surface Diffusion: This type of diffusion is caused due to atomic movement along the surface of a phase.

3. DIFFUSION MECHANISMS

Diffusion is the transfer of unlike atoms which is accompanied with a change of concentration of the components in certain zones of an alloy. Various mechanisms have been proposed to explain the processes of diffusion. Almost all of these mechanisms are based on the vibrational energy of atoms in a solid. Direct-interchange, cyclic, interstitial, vacancy etc. are the common diffusion mechanisms. Actually, however, the most probable mechanism of diffusion is that in which the magnitude of energy barrier (activation energy) to be overcome by moving atoms is the lowest. Activation energy depends on the forces of interatomic bonds and crystal lattice defects which facilitate diffusion transfer (the activation energy at grain boundaries is only one half of that in the bulk of a grain). For metal atoms, the vacancy mechanism of diffusion is the most probable and for elements with a small atomic radius (H, N and C), the interstitial mechanism. Now, we will study these mechanisms.

(i) Vacancy Mechanism: This mechanism is a very dominant process for diffusion in FCC, BCC and HCP metals and solid solution alloy. The activation energy for this process comprises the energy required to create a vacancy and that required to move it. In a pure solid, the diffusion by this mechanism is shown

in Fig. 7.1(a). Diffusion by the vacancy mechanism can occur by atoms moving into adjacent sites that are vacant. In a pure solid, during diffusion by this mechanism, the atoms surrounding the vacant site shift their equilibrium positions to adjust for the change in binding that accompanies the removal of a metal ion and its valency electron. We can assume that the vacancies move through the lattice and produce random shifts of atoms from one lattice position to another as a result of atom jumping. Concentration changes takes place due to diffusion over a period of time. We must note that vacancies are continually being created and destroyed at the surface, grain boundaries and suitable interior positions, e.g. dislocations. Obviously, the rate of diffusion increases rapidly with increasing temperature.

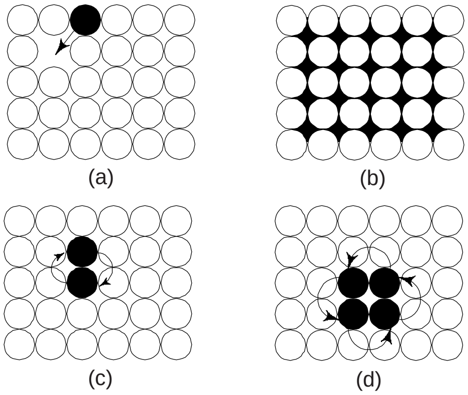


Fig. 7.1 Various diffusion mechanisms
(a) Vacancy mechanism
(b) Interstitial mechanisms
(c) Two atoms interchange mechanisms
(d) Four atoms interchange mechanisms

If a solid is composed of a single element, i.e. pure metal, the movement of thermally excited atom from a site of the crystal lattice to an adjacent site or interstice is called *self diffusion* because the moving atom and the solid are the same chemical-element. The self-diffusion in metals in which atoms of the metal itself migrate in a random fashion throughout the lattice occurs mainly through this mechanism.

We know that copper and nickel are mutually soluble in all proportions in solid state and form substitutional solid solutions, e.g., plating of nickel on copper. For atomic diffusion, the vacancy mechanism is shown in Fig. 7.2.

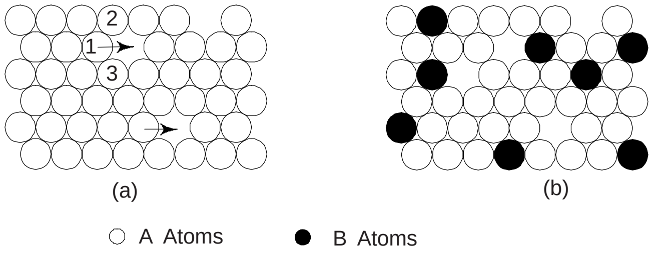


Fig. 7.2 Vacancy mechanism for atomic diffusion
(a) pure solid solution, and
(b) substitutional solid solutions

(ii) The Interstitial Mechanism: The interstitial mechanism where an atom changes positions using an interstitial site does not usually occur in metals for self-diffusion but is favoured when interstitial impurities are present because of the low activation energy.

When a solid is composed of two or more elements whose atomic radii differ significantly, interstitial solutions may occur. The large size atoms occupy lattice sites where as the smaller size atoms fit into the voids (called as interstices) created by the large atoms. We can see that the diffusion mechanism in this case is similar to vacancy diffusion except that the interstitial atoms stay on interstitial sites (Fig. 7.1(b)). We must note that an activation energy is associated with interstitial diffusion because, to arrive at the vacant site, it must squeeze past neighbouring atoms with energy supplied by the vibrational energy of the moving atoms. Obviously, interstitial diffusion is a thermally activated process. The interstitial mechanism process is simpler since the presence of vacancies is not required for the solute atom to move. This mechanism is vital for the following cases:

(a) The presence of very small atoms in the interstices of the lattice affect to a great extent the mechanical properties of metals.

(b) At low temperatures, oxygen, hydrogen and nitrogen can be diffused in metals easily.

(iii) Interchange Mechanism: In this type of mechanism, the atoms exchange places through rotation about a mid point. The activation energy for the process is very high and hence this mechanism is highly unlikely in most systems.

Two or more adjacent atoms jump past each other and exchange positions, but the number of sites remains constant (Fig. 7.1(c) and (d)). This interchange may be two-atom or four-atom (Zenner ring) for BCC. Due to the displacement of atoms surrounding the jumping pairs, interchange mechanism results in severe local distortion. For jumping of atoms in this case, much more energy is required. In this mechanism, a number of diffusion couples of different compositions are produced, which are objectionable. This is also termed as *Kirkendall's effect*.

Kirkendall was the first person to show the inequality of diffusion. By using an α-brass/copper couple, Kirkendall showed that Zn atoms diffused out of brass into Cu more rapidly than Cu atoms diffused into brass. Due to a net loss of Zn atoms, voids can be observed in brass.

From theoretical point of view, Kirkendall's effect is very important in diffusion. We may note that the practical importance of this effect is in metal cladding, sintering and deformation of metals (creep).

4. DIFFUSION COEFFICIENT: FICK'S LAWS OF DIFFUSION

Diffusion can be treated as the mass flow process by which atoms (or molecules) change their positions relative to their neighbours in a given phase under the influence of thermal energy and a gradient. The gradient can be a concentration gradient; an electric or magnetic field gradient or a stress gradient. We shall consider mass flow under concentration gradients only. We know that thermal energy is necessary for mass flow, as the atoms have to jump from site to site during diffusion. The thermal energy is in the form of the vibrations of atoms about their mean positions in the solid.

The classical laws of diffusion are Fick's laws which hold true for weak solutions and systems with a low concentration gradient of the diffusing substance, dc/dx (= $c_2 - c_1/x_2 - x_1$), slope of concentration gradient.

(i) Fick's First Law: This law describes the rate at which diffusion occurs. This law states that

$$dn = -D\,\frac{dc}{dx}\,a\,dt \tag{7.1}$$

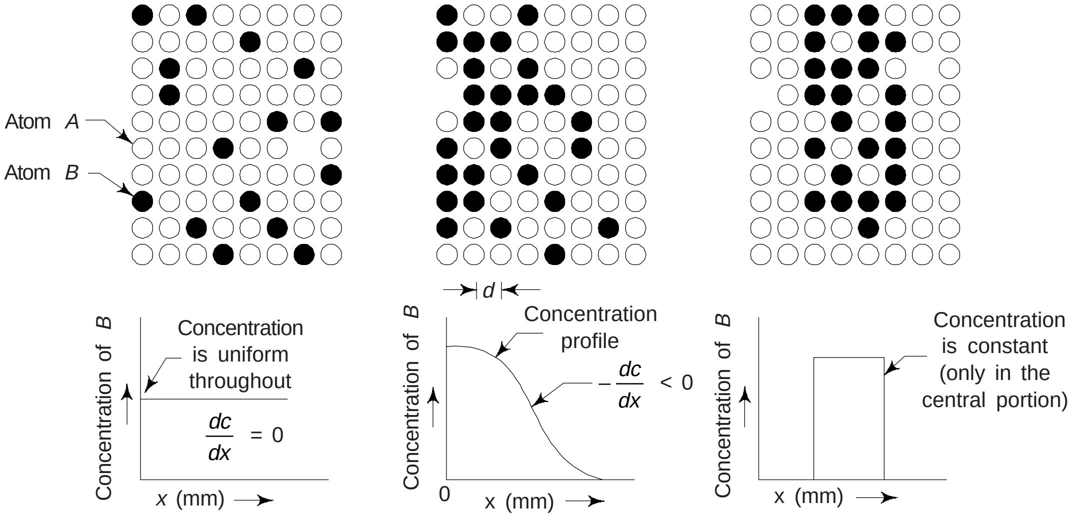
i.e. the quantity dn of a substance diffusing at constant temperature per unit time t through unit surface area a is proportional to the concentration gradient dc/dx and the coefficient of diffusion (or diffusivity) D (m^2/s). The 'minus' sign implies that diffusion occurs in the reverse direction to concentration gradient vector, i.e. from the zone with a higher concentration to that with a lower concentration of the diffusing element.

The equation (7.1) becomes:

$$\frac{dn}{dt} = -D\,\frac{dc}{dx}\,a$$

$$\therefore \qquad J = -\frac{1}{a}\frac{dn}{dt} = -D\frac{dc}{dx} \tag{7.2}$$

where J is the flux or the number of atoms moving from unit area of one plane to unit area of another per unit time, i.e. flux J is flow per unit cross sectional area per unit time. Obviously, J is proportional to the concentration gradient. The negative sign implies that flow occurs down the concentration gradient. Variation of concentration with x is shown in Fig. 7.3. We can see that a large negative slope corresponds to a high diffusion rate. In accordance with Fick's law (first), the B atoms will diffuse from the left side. We further note that the net migration of B atoms to the right side means that the concentration will decrease on the left side of the solid and increase on the right as diffusion progress.



(a) (b) (c)

Fig. 7.3 Model for illustration of diffusion: Fick's first law. We note that the concentration of *B* atoms in the direction indicates the concentration profile

This law can be used to describe flow under *steady state conditions*. We find that it is identical in form to Fourier's law for heat flow under a constant temperature gradient and Ohm's law for current flow under a constant electric field gradient. We may see that under steady state flow, the flux is independent of time and remains the same at any cross-sectional plane along the diffusion direction.

Diffusion coefficient (diffusivity) for a few selected solute solvent systems is given in Table 7.1.

Table 7.1 Diffusion coefficient (diffusivity) of a few solute solvent systems

Solute	*Solvent (host structure)*	*Diffusion coefficient (m^2/s)*	
		500°C	*1000°C*
C	FCC Fe	(5×10^{-15})	3×10^{-11}
C	BCC Fe	10^{-12}	(2×10^{-9})
Fe	FCC Fe	(2×10^{-23})	2×10^{-16}
Fe	BCC Fe	10^{-20}	(3×10^{-14})
Ni	FCC Fe	10^{-23}	2×10^{-16}
Mn	FCC Fe	(3×10^{-24})	10^{-16}
Zn	Cu	4×10^{-18}	5×10^{-13}
Cu	Al	4×10^{-14}	10^{-10}
Cu	Cu	10^{-18}	2×10^{-13}
C	HCP Ti	3×10^{-16}	2×10^{-11}

Parentheses indicate that the phase is metastable

(ii) Fick's Second Law: This is an extension of Fick's first law to *non steady flow*. Fick's first law allows the calculation of the instantaneous mass flow rate (flux) past any plane in a solid but provides no information about the time dependence of the concentration. However, commonly available situations with engineering materials are non-steady. The concentration of solute atom changes at any point with respect to time in non-steady diffusion.

If the concentration gradient varies in time and the diffusion coefficient is taken to be independent of concentration, the diffusion process is described by Fick's second law which can be derived from the first law:

$$\frac{dc}{dt} = \frac{d}{dx}\left[D\left(\frac{dc}{dx}\right)\right] \tag{7.3}$$

Equation (7.3) is Fick's second law for unidirectional flow under non steady conditions. A solution of Eq. (7.3) is given by

$$c\,(x,\,t) = \frac{A}{\sqrt{Dt}}\ \exp\,[-(x^2/4Dt)] \tag{7.3a}$$

where A is constant. Let us consider the example of self diffusion of radioactive nickel atoms in a non-radioactive nickel specimen. Equation 7.3(a) indicates that the concentration at $x = 0$ falls with time as $t^{-1/2}$ and as time increases the radioactive atoms penetrate deeper in the metal block [Fig. 7.3(a)]. At time t_1 the concentration of radioactive atoms at $x = 0$ is $c_1 = A/(Dt_1)^{1/2}$. At a distance $x_1 = 2\,(Dt_1)^{1/2}$ the concentration falls to 1/e of c_1. At time t_2, the concentration at $x = 0$ is $c_2 = A/(Dt_2)^{1/2}$ and this falls to 1/e and $x_2 = 2\,(Dt_2)^{1/2}$. These results are in agreement with experiments.

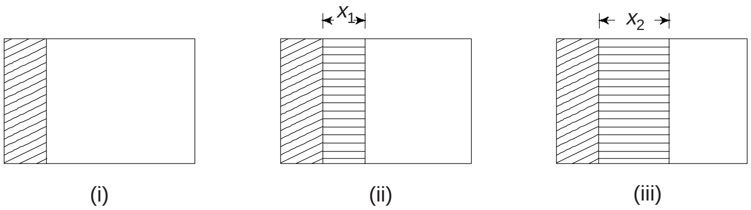


Fig. 7.3 (a) **The radioactive sheet of Nickel (shown by shaded section) is kept in contact with a block of nonradioactive nickel. Radioactive atoms diffuse from the sheet to the bulk metal and can be detected as a function of time. In figure, the diffusion of atoms is shown (i) for $t = 0$, (ii) for t_1, and (iii) t_2 with $t_2 > t_1$**

If D is independent of concentration, Eq. (7.3) simplifies to

$$\frac{dc}{dt} = D\ \frac{d^2 c}{dx^2} \tag{7.4}$$

Even though D may vary with concentration, solutions to the differential Eq. (7.4) are quite commonly used for practical problems, because of their relative simplicity. The solution to Eq. (7.4) for unidirectional diffusion from one medium to another across a common interface is of the general form

$$c(x,\,t) = A - B\ \mathrm{erf}\ (x/2\sqrt{Dt}\,) \tag{7.4a}$$

where A and B are constants to be determined from the initial and boundary conditions of a particular problem. The diffusion direction x is perpendicular to the common interface. The origin for x is at the interface. The two media are taken to be semi-infinite, i.e., only one end of each of them, which is the interface, is defined. The other two ends are at an infinite distance. The initial uniform concentrations of the diffusing species in the two media are different, with an abrupt change in concentration at the interface. erf in Eq. 7.4(a) stands for '*error function*', which is

$$\operatorname{erf} \frac{x}{2\sqrt{Dt}} = \frac{2}{\sqrt{\pi}} \int_{0}^{x/2\sqrt{Dt}} \exp(-\eta^2)\, d\eta \tag{7.4(a)}$$

η is an integration variable, that gets deleted as the limits of the integral are substituted. The lower limit of the integral is always zero, while the upper limit of the integral is the quantity, whose error function is to be determined. $2\sqrt{\pi}$ is a normalization factor.

The diffusion coefficient D(m^2/s) determines the rate of diffusion at a concentration gradient equal to unity. It depends on the composition of alloy, size of grains, and temperature.

Solutions to Fick's equations exist for a wide variety of boundary conditions, thus permitting an evaluation of D from c as a function of x and t.

A schematic illustration of the time dependence of diffusion is shown in Fig. 7.4. The curve corresponding to the concentration profile at a given instant of time t_1 is marked by t_1. We can see from Fig. 7.4 that at a later time t_2, the concentration profile has changed. We can easily see that this change in concentration profile is due to the diffusion of B atoms that has occurred in the time interval $t_2 - t_1$. The concentration profile at a still later time t_3 is marked by t_3. Due to diffusion, B atoms are trying to get distributed uniformly throughout the solid salution. From Fig. 7.4 it is evident that the concentration gradient becoming less negative as time increases. Obviously, the diffusion rate becomes slower as the diffusion process progress.

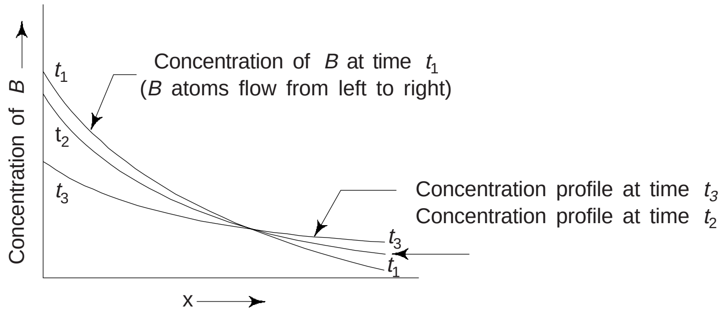


Fig. 7.4 Time dependence of diffusion (Fick's second law)

5. DEPENDENCE OF DIFFUSION COEFFICIENT ON TEMPERATURE

The diffusion coefficient D (m^2/s) determines the rate of diffusion at a concentration gradient equal to unity. It depends on the composition of alloy, size of grains, and temperature.

The dependence of diffusion coefficient on temperature in a certain temperature range is described by *Arrhenius exponential relatioship*.

$$D = D_0 \exp(-Q/RT) \tag{7.5}$$

Where D_0 is a preexponential (frequency) factor depending on bond forces between atoms of crystal lattice; Q is the activation energy of diffusion; where $Q = Q_v + Q_m$, Q_v and Q_m are the activation energies for the formation and motion of vacancies respectively, the experimental value of Q for the diffusion of carbon in α-Fe is 20.1 k cal/mole and that of D_0 is 2×10^{-6} m^2/s, and R is the gas constant.

6. FACTORS AFFECTING DIFFUSION COEFFICIENT (D)

We have mentioned that diffusion coefficient is affected by concentration. However, this effect is small compared to the effect of temperature. While discussing diffusion mechanisms, we have assumed that atoms jumped from one lattice position to another. The rate at which atoms jumped mainly depends on their vibrational frequency, the crystal structure, activation energy and temperature. We may note that the activation energy depends on the energies of the bonds which are formed between the jumping atoms and their neighbours. Further, the bond energies depend on concentration.

Later, Fick's laws have been modified for describing diffusion processes in complex alloys, in particular by introducing the concepts of gradient in chemical potential and the gradient of thermodynamic activity instead of concentration gradient.

7. SELF DIFFUSION

It is the transition of a thermally excited atom from a site of the crystal lattice to an adjacent site or interstice. If the solid is composed of a single element (say pure copper), the movement of the atom is called self diffusion because in this case the moving atom and the solid are the same chemical element.

The process of self diffusion is very important for annealing and creep. Use of radioactive tracers have found to be quite useful in determining self diffusion coefficient.

8. INTER-DIFFUSION

It is contrary to self diffusion and takes place in binary metallic alloys, e.g. the Cu-Ni system. If nickel had been plated onto the surface of copper, then atomic diffusion would bring about nickel homogenization within the copper, after a sufficient time, at elevated temperatures.

9. EXPERIMENTAL DETERMINATION OF *D* USING A DIFFUSION COUPLE

Using a diffusion couple, we can determine the diffusion coefficient *D*. A diffusion couple consists of two solids (say metal *A* and metal *B*), so that each solid tries to diffuse into the other. A diffusion couple, shown in Fig. 7.5 consists of two long metal bars *A* and *B* welded face to face, the concentration of diffusing species is one, c_2, being higher than that in the other, c_1. Diffusion takes place across the common face. In Fig. 7.5, the concentration is shown along the Y-axis; the origin of diffusion direction *x* is at the common interface. The rate of diffusion of *A* into *B* and *B* into *A* for any given value of *x* can be easily determined by Fick's first law.

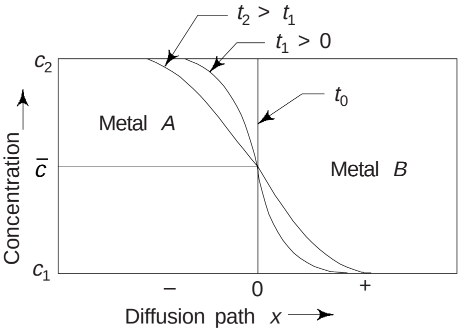


Fig. 7.5 The diffusion couple set up for nonsteady-state flow

The initial conditions for the diffusion couple set up are:

$$c(x,\ 0) = \begin{cases} c_1 \ x > 0 \\ c_2 \ x < 0 \end{cases} \qquad (7.6)$$

The constants in Eq. 7.4 (a) are obtained as

$$A = \frac{c_1 + c_2}{2} \qquad (7.7)$$

$$B = \frac{c_2 - c_1}{2} \qquad (7.8)$$

The thermal energy at room temperature is usually insufficient to cause an appreciable rate of diffusion. If the diffusion couple is heated to a high temperature, say, near the melting point, sufficient thermal energy becomes available and the diffusion species start to move from metal bar *B* to metal bar *A*. The concentration-distance profiles at different lengths of diffusion time are as shown in Fig. 7.5. Time t_0 is the instant at which diffusion begins and we note that concentration profiles show a step change at the contact surface of *A* and *B*. *A* and *B* diffuse into each other and the concentration profiles change with time as shown by curves for t_1 and $t_2 (> t_1)$. The concentration at the junction is independent of time and is $\bar{c} = (c_1 + c_2)/2$, where $\bar{c}$ is the average concentration. The concentration of either *A* or *B*, as a function of *x* and *t* can be found from the solution of Fick's second law as applied to the diffusion couple. With *x* and *t* known, *D* can be computed. Using Eq. (7.5) *D* can also be determined as a function of temperature.

10. DIFFUSION WITH CONSTANT CONCENTRATION (CASE HARDENING)

A process in which one element (usually in gaseous form) is diffused into another (a solid form) and the diffusion remains limited to a small region near the surface is called *case hardening*. As a result, the properties of this region changes. Generally, it is observed that the surface regions become harder and brittle whereas the core remains ductile.

Case hardening process is governed by Fick's second law. Diffusion of gas atoms into the solid takes place by one of the diffusion mechanism. It is found that this diffusion mechanism is usually the interstitial mechanism. It is reported that the depth to which the atoms of gas penetrate increases with time and thus the depth of interstitial alloy increases as diffusion progresses.

We now consider an example of diffusion of nitrogen into iron, which is usually termed as nitriding. Nitrogen can be dissolved in the interstitial sites of an iron crystal and the resulting Fe-N alloy is stronger, harder and more brittle than the original iron. In accordance with the concentration profiles, the interstitial atoms inhibit dislocation motion and the nitrogen concentration would be highest near the surface.

Surface hardening of steel objects such as gears is frequently done by *carburizing* or *nitriding*. The carburizing process consists of diffusing carbon into the surface layers of the steel object. In pack carburizing the object is packed in solid carbon powder, whereas in gas carburizing, an atmosphere of methane gas that is reach in carbon surrounds the object to the carburized. The reaction which takes place at the steel surface is as follows:

$$CH_4\ (g) \rightarrow 2H_2\ (g) + C\ (\text{steel})$$

When the steel object is annealed at an elevated temperature in the carburizing medium, carbon diffuses into the steel from the surface under a concentration gradient. We may note that surface hardening improves the wear resistance of components such as gears, without impairing the bulk mechanical properties, e.g. toughness.

The opposite of *carburization* is called *decarburization*. Here, the carbon is lost from the surface layers of the steel, due to an oxidizing atmosphere that reacts with carbon to produce CO or CO_2. The fatigue resistance of steel is lowered due to this and, therefore, it should be avoided by using a protective atmosphere during the heat treatment of steel.

11. THE KIRKENDALL EFFECT

When we consider in a binary solution of *A* and *B*, the rates at which *A* and *B* diffuse are not necessarily equal. It is observed that, usually, the lower melting component diffuses much faster than the other. This lead to certain effects which are interesting and first observed by Kirkendall.

Inert markers, i.e. thin rods of a high melting point substance which is insoluble in the diffusion matrix, are placed at the weld joint of the couple, prior to the diffusion of anneal. It is found that these markers shift during the anneal in the same direction as the slower moving species. The extent of this shift is reported to be proportional to the square root of the diffusion time. This type of movement reveals that the net mass flow due to the difference in diffusivities is being compensated by a bulk flow of matter in the opposite direction within the diffusion zone. Obviously, lattice planes are created on one side of the diffusion zone, while they are destroyed on the other side of the diffusion zone, and the resulting bulk flow carries the markers along. We must note that the bulk flow occurs relative to the ends of the diffusion couple. It is interesting to note that it is quite a different phenomenon from the diffusion process itself. In several cases, one observes porosity on the lower-melting component side, indicating that the bulk flow does not fully compensate for the difference in diffusivities of the two species.

To understand the Kirkendall effect, one may consider the analogy of gaseous interdiffusion. Let us consider that hydrogen and argon at the same pressure be kept in two chambers interconnected through a tube and a frictionless piston in the tube separates the gases. When an orifice in the piston is opened, the gases interdiffuse. Obviously, the lighter hydrogen will diffuse faster, resulting in a pressure difference that will tend to shift the piston in the same direction as the slower diffusion argon is moving.

12. DIFFUSION IN OXIDES AND IONIC CRYSTALS

Diffusion in oxides and other crystalline compounds that have ionic or partial ionic bonds can occur by diffusion mechanisms. In the present case, the diffusion process is somewhat complicated due to the requirement of electrical neutrality. Vacancies must be formed in such a way that the crystal does not acquire charge. This means that the vacancies be formed in pairs, i.e. with each pair consisting of one cation (+ve charge) and the other anion (–ve charge) vacancy.

We may note that Schottky and Frankel defects help diffusion in ionic crystals. In the Frankel diffusion mechanisms, the cation (+ve) interstitial carries the flux, whereas in the Schottky diffusion mechanism, the cation vacancy carries the diffusion flux. The cations diffuse through cation vacancies and anions through anion vacancies but the activation energies are not the same for both processes.

It is reported that the diffusion in ionic crystals and oxides is very sensitive to the concentration of impurities as they affect the number of vacancies. Let us consider the addition of a small quantity of cadmium as an impurity to an NaCl crystal. We know that the valence of Na is +1 and that of Cd as +2. Obviously, a single Cd ion would have to replace two Na ions in crystal in order to maintain electrical neutrality and thereby generating a cation vacancy. This means that the addition of Cd cations will increase the diffusion rate.

Metallic oxides often contain impurities and the rate of oxygen diffusion through an oxide is often increased by the presence of impurities which occupy anion sites.

Diffusion in oxides and ionic crystals is quite useful in fabricating parts from high temperature ceramics by the powder metallurgy technique. Due to diffusion, the powered particles would bond together in the sintering process.

13. SURFACE DIFFUSION

Diffusion takes place along the surfaces of a solid and in the *grain boundaries* of a polycrystal as well as through the volume of a material. Atoms at the surface of a solid form fewer bonds than atoms at the interior of a solid and hence one can expect lower activation to surface diffusion than volume diffusion. During the solidification process, migration of atoms takes place and this is termed as solid diffusion (Fig. 7.6).

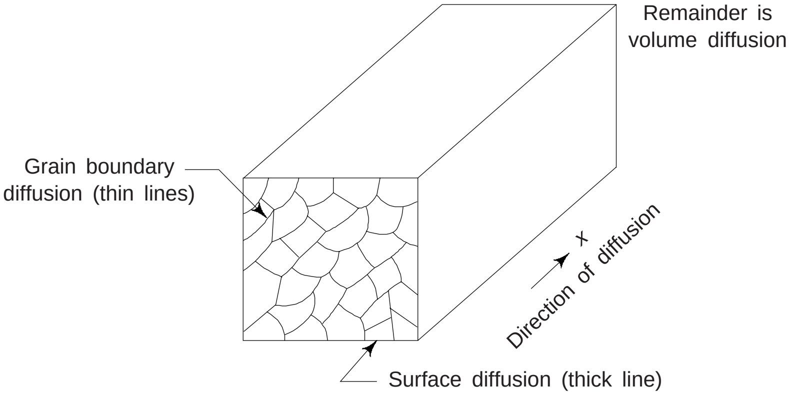
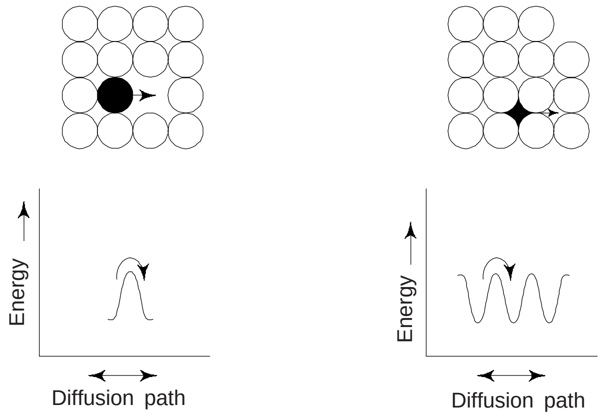


Fig. 7.6 Areas of volume, grain boundaries and surface diffusion

We must note that the atoms in regions of grain boundaries are not bonded as tightly as interior atoms. This is why, they diffuse more readily. As regards surface and boundary diffusions, these are produced along paths other than those of crystal lattice diffusion.

14. ACTIVATION ENERGY OF DIFFUSION

We have already mentioned that atomic movement is responsible for the diffusion process. Obviously, some energy is required by the atoms for each movement or jumping from one lattice position to another lattice position. To overcome this energy barrier, the energy required by the atom is called the activation energy of diffusion (Fig. 7.7).



(a) (b)

Fig. 7.7 Activation energy for diffusion (a) vacancy mechanism (b) interstitial mechanism

The energy is required to pull the atom away from its nearest atoms in the vacancy mechanism. Energy is also required to force the atom into closer contact with neighbouring atoms as it moves along them in interstitial diffusion. If the normal inter-atomic distance is either increases or decreases, addition energy is required. We may note that the activation energy depends on the size of the atom, i.e. it varies with the size of the atom, strength of bond and the type of the diffusion mechanism. It is reported that the activation energy required is high for large-sized atoms, strongly bonded materials, e.g. corrundum and tungsten carbide (since interstitial diffusion requires more energy than the vacancy mechanisms).

15. APPLICATIONS OF DIFFUSION

Diffusion processes are the basis of crystallization, recrystallization, phase transformation and saturation of the surface of alloys by other elements. Few important applications of diffusion are:

(i) Oxidation of metals
(ii) Doping of semiconductors
(iii) Joining of materials by diffusion bonding, e.g. welding, soldering, galvanizing, brazing, and metal cladding
(iv) Production of strong bodies by sintering, i.e. powder metallurgy
(v) Surface treatment of steels, e.g. case hardening
(vi) Important in heat treatment, e.g. homogenising treatment of castings, recovery, recrystallization and precipitation of phases.
(vii) Diffusion is fundamental to phase changes, e.g. γ to α-iron.

Now, we may discuss few applications in some detail. A common example of solid state diffusion is surface hardening of steel, commonly used for gears and shafts. Steel parts made in low carbon steel are brought in contact with hydrocarbon gas like methane (CH_4) in a furnace atmosphere at about 927°C temperature. The carbon from CH_4 diffuses into surface of steel part and thereby carbon concentration increases on the surface. Due to this, the hardness of the surface increases. We may note that the percentage of carbon diffuses in the surface increases with the exposure time. The concentration of carbon is high near the surface and reduces with increasing depth (Fig. 7.8).

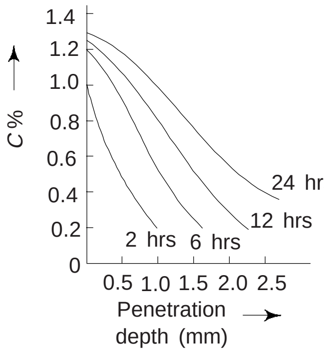


Fig. 7.8 *C* gradient in 1022 steel carburized in 1.6% CH_4, 20% CO and 4% H

Modern electronic integrated circuits uses very thin wafers of silicon-diffused with impurities. In doing so the surface of silicon wafer surface is exposed to vapour of appropriate impurity when wafer surface is kept at a temperature of about 1100°C in quartz tube furnace. In case if a part of the surface is required to be kept free from doping, i.e. free from doping, then it can be covered while being treated. We note that this process is very much similar to the one used for surface hardening of machine parts and the concentration of impurities from surface reduces almost in the same way as shown in Fig. 7.8.

Example 1 *Steel gear, having carbon content of 0.2% is to be gas carburized to achieve carbon content of 0.96% at the surface and 0.4% at 0.5 mm depth from the surface. If the process is to be carried out at 927°C, find the time required for carburization. Given, diffusion coefficient of carbon in given steel =* $1.28 \times 10^{-11} m^2/s$. [B.E., M.Sc (M.S.)]

Solution We have
$$\frac{dc}{dt} = D\frac{d^2c}{dx^2} \tag{1}$$

The solution of Eq. (1) can be written as

$$\frac{c_s - c_x}{c_s - c_0} = \text{erf}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{2}$$

Substituting the proper values in Eq. (2), we hve

$$\frac{0.9-0.4}{0.9-0.2} = \text{erf}\left(\frac{0.5\times10^{-3}}{2\sqrt{12.8\times10^{-12}\,t}}\right)$$

or
$$0.7143 = \text{erf}\left(\frac{69.88}{\sqrt{t}}\right)$$

Let
$$z = \frac{69.88}{\sqrt{t}}$$

$\therefore$
$$\text{erf } z = 0.7143 \tag{3}$$

$x = 0\ .5$ mm $= 0.5 \times 10^{-3}$ m
$D = 1.28 \times 10^{-11}$ m^2/s
t = ?
$c_s \rightarrow$ Surface concentration of diffusion element in the surface = 0.9%
$c_o \rightarrow$ Initial uniform concentration of the element in the solid = 0.2%
$c_x \rightarrow$ Concentration of the diffusing element at a distance x from the surface = 0.4%

From table, we find the erf z lies between 0.7112 and 0.7421 for which z lies between 0.75 and 0.80. One can find the number z by interpolation.

$$\frac{0.7143-0.7112}{0.7421-0.7112} = \frac{z-0.75}{0.80-0.75}$$

or
$$z = 0.755$$

From (3),
$$\frac{(69.88)^2}{(0.755)^2} = t$$

$\therefore$
$$t = 8566.35 = 142.8 \text{ min.}$$

Table 7.1 Error function

z	*erf z*	*z*	*erf z*
0.000	0.0000	0.85	0.7707
0.025	0.0282	0.90	0.7970
0.05	0.0564	0.95	0.8209
0.10	0.1125	1.0	0.8427

(*Contd.*)

Table 7.1 (*Contd.*)

z	*erf z*	*z*	*erf z*
0.15	0.1680	1.1	0.8802
0.20	0.2227	1.2	0.9103
0.25	0.2763	1.3	0.9340
0.30	0.3268	1.4	0.9523
0.35	0.3794	1.5	0.9661
0.40	0.4284	1.6	0.9763
0.45	0.4755	1.7	0.9838
0.50	0.5205	1.8	0.9891
0.55	0.5633	1.9	0.9928
0.60	0.6039	2.0	0.9953
0.65	0.6420	2.2	0.9981
0.70	0.6778	2.4	0.9993
0.75	0.7112	2.6	0.9998
0.80	0.7421	2.8	0.9999

Example 2 *To obtain a p-type semiconductor, boron (an element from third column) is doped in pure silicon. This doped is carried through a* B_2O_3 *vapour phase of partial pressure equal to 1.5* N/m^2. *This atmosphere is equivalent to a surface concentration of* 3×10^{26} *boron atoms per* m^3. *Show that the time required to get a boron content of* 10^{23} *atoms per* m^3 *at a depth of 2* μ*m is 3845 s. Given: the doping temperature = 1100°C and* D_B *in Si at this temperature =* 4×10^{-17} m^2/s. [B.E., B.Sc. (Hons: M.S.)]

Solution We have

$$c(x, t) = A - B \text{ erf } (x / 2\sqrt{Dt})$$

Using appropriate values, we obtain

$$\text{erf } (x/2\sqrt{Dt}) = 0.99967$$

Using table 7.1, we obtain

$$x/2\sqrt{Dt} = 2.55$$

$c_1 = 0$
$c_s = 3 \times 10^{26}$ atoms
$c(x) = 10^{23}$ atoms
$x = 2 \times 10^{-6}$ m
$D = 4 \times 10^{-17}$ m^2/s

$$\therefore \qquad t = \frac{(2)^2 \times 10^{-12}}{2 \times (2.55)^2 \times 4 \times 10^{-17}} = 3845 \text{ sec}$$

Approximate calculations can be made using $x = \sqrt{Dt}$. c_1 is negligible compared to c_2 and for the diffusion couple set up, this approximation yields the concentration at x as $c(x) = 0.24\ c_2$.

Example 3 *The diffusion coefficients of carbon in* α*-Ti were determined at following temperatures.*

Temperature (°C)	*D* (m^2/s)
736	2×10^{-13}
782	5×10^{-13}
835	1.3×10^{-12}

(a) Assuming that Arhenius type rate process equation is valid, find the constant of the equation and activation energy, and (b) determine the diffusion coefficient at 500°C. [B.E., M.Sc., (M.S.)]

Solution We have

$$D = D_0 \exp(-\Delta E/kT) \qquad (1)$$

Taking log of both the sides of Eq. (1), one obtains

$$\text{Log } D = \text{Log } D_0 - \frac{\Delta E}{kT} \qquad (2)$$

Plotting D as a function of $1/T$, a straight line is obtained which have an intercept of log D_0 on log D axis and a slope $\Delta E/k$ (Fig. 7.9)

$D \rightarrow$ Coefficient of diffusion (m^2/s)
$D_0 \rightarrow$ Constant of equation (m^2/s)
ΔE = Activation energy (J)
k = 1.38×10^{-23} J/k

Now, $T_1 = 736 + 273 = 1009$ K
$D_1 = 2 \times 10^{-13}$ m^2/s
and $T_2 = 782 \times 273 = 1055$ K
$D_2 = 5 \times 10^{-13}$ m^2/s

Substituting in Eq. (2), one obtains

$$\log (2 \times 10^{-13}) = \log D_0 - \frac{\Delta E \times 10^{23}}{1.38 \times 1009}$$

$$\log (5 \times 10^{-13}) = \log D_0 - \frac{\Delta E \times 10^{23}}{1.38 \times 1055} \qquad (3)$$

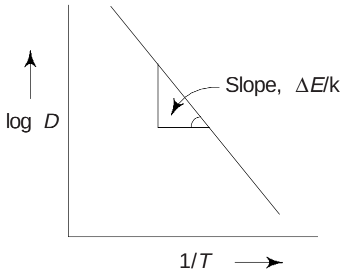


Fig. 7.9 log *D* versus 1/*T*

Subtracting the above, one obtains

$$\log \frac{2}{5} = \frac{\Delta E}{1.38} \times 10^{20} (0.948 - 0.991)$$

or
$$-0.9163 = \frac{\Delta E}{1.38} \times 10^{20} (-0.043)$$

or
$$\Delta E = \frac{0.9163 \times 1.38}{0.043} \times 10^{-20} \text{ J}$$
$$= 2.941 \times 10^{-19} \text{ J} \qquad (4)$$

Using this value of ΔE in first of (3), one obtains

$$2 \times 10^{-3} = D_0 \exp \left[\frac{-2.941 \times 10^{-19} \times 10^{23}}{1.38 \times 1009} \right]$$
$$= D_0 \exp (-21.12)$$
$$= 6.71 \times 10^{-10} \; D_0$$

$\therefore$
$$D_0 = \frac{2 \times 10^{-13}}{6.71 \times 10^{-10}} = 2.98 \times 10^{-4} \text{ m}^2\text{/s}$$

We can check the correctness of the solution from third observation:

$T_3 = 825 + 273 = 1108$ K

$\therefore$
$$D_0 \exp \left(-\frac{\Delta E}{kT} \right) = 2.98 \times 10^{-4} \exp \left[\frac{-2.941 \times 10^{-10}}{1.38 \times 10^{-23} \times 1108} \right]$$
$$= 298 \times 10^{-4} \times 4 \times 10^{-9} = 1.32 \times 10^{-12} \text{ m}^2\text{/s}$$

i.e. $D_3 = 1.32 \times 10^{-12}$ m^2/s

We note that the observed value of D_3 is 1.3×10^{-12} which is almost equal to the calculated value. Thus the value of D_0 and ΔE at (5) and (4) respectively are correct.

Now, using $T_4 = 500 + 273 = 773$ K, one obtains,

$$D_4 = 2.98 \times 10^{-4} \exp\left[-\frac{2.94\times10^{-19}}{1.38\times10^{-23}\times773}\right] \times 1.063 \times 10^{-12} = 2.98 \times 10^{-4}$$

or Diffusion coefficient at 500°C = 3.17×10^{-16} m^2/s

Example 4 *The diffusion coefficient for copper in aluminium at 500 and 600°C are 4.8×10^{-14} and 5.3×10^{-3} m^2/s respectively. Determine the approximate time at 500°C that will produce the same diffusion result (in terms of concentration of copper at some specific point in aluminium) as 10^{-h} heat treatment at 600°C.* [B.E.]

Solution The composition in both diffusion situations will be at the same position, i.e. x is also a constant. Thus

$$Dt = \text{constant}$$

Now,
$$D_{500}\, t_{500} = D_{600}\, t_{600}$$

or
$$t_{500} = \frac{D_{600}}{D_{500}}\; t_{500}$$

$$= \frac{(5.3\times10^{-13}\ \text{m}^2/\text{s})\,(10\,\text{h})}{(4.8\times10^{-14}\ \text{m}^2/\text{s})}$$

$$= 110.4\ \text{h}.$$

SUGGESTED READINGS

1. P.G. Shewmon, *Diffusion in Solids*, McGraw Hill, New York (1963).
2. J. Crank and G.S. Park, *Diffusion in Polymers*, Academic Press, New York (1968).
3. M. Glicksman, *Diffusion in Solids*, Wiley Interscience, New York (2000).
4. R.J. Borg and G.J. Dienes (Eds.), *An Introduction to Solid State Diffusion,* Academic Press, San Diego (1988).

REVIEW QUESTIONS

1. Explain the importance of diffusion in engineering.

2. State and explain Fick's laws of diffusion. How diffusion coefficient depends upon temperature?

3. Explain the activation energy of diffusion. [A.M.I.E.]

4. Explain the process of diffusion in
   (i) ionic crystals and oxides
   (ii) Grain boundaries and oxides [B.E.]

5. Explain briefly the following
   (i) Diffusion in metals and alloys
   (ii) Self-diffusion
   (iii) Diffusion coefficients
   (iv) Diffusion couple
   (v) Diffusion with constant concentration
   (vi) Kirkendall effect
   (vii) Carburization and decarburization of steel

6. How Fick's second law can be used to determine D?

7. Explain mechanisms of diffusion.

8. How will the diffusivity of NaCl change, when it is doped with *(i)* KCl and *(ii)* $CaCl_2$? Explain your answers. [B.E.]

9. Compare the diffusivities of hydrogen, nitrogen ad nickel in iron at 300 K and explain the difference between the three values. [B.E., M.Sc, (M.S.)]

PROBLEMS

1. The diffusion rate of A in B was studied at 500°C and 850°C. It was reported that, for the same diffusion time, the depths of penetration x_1 and x_2 in the two experiments were in the ratio of 1: 4. Show that the activation energy of A in B is 57 kJ/mol. [B.E., M.Sc, (M.S.)]

2. An amount Q of a dopant is deposited on the surface of a silicon substrate. During a subsequent anneal without the dopant in the atmosphere, the concentration c of the dopant as a function of depth x and time t is given by

$$c = [Q/\sqrt{\pi Dt}\,] \exp(-x^2/4Dt)$$

Show that this is solution of Fick's second law, where D is independent of concentration. [B.E.]

3. The activation barrier for jump is 100 kJ/mol. Calculate the rate at which a vacancy jumps in copper at 20°C. [1.5×10^{-5}/s] [B.E.]

OBJECTIVE QUESTIONS

1. The unit of diffusion coefficient is
(1) $m^2\,s^1$ (2) $m^2\,s^{-1}$ (3) $m^2\,s^{-2}$ (4) $m^2\,s^2$ (2)

2. Substitutional diffusion becomes possible if
(1) the temperature of solid solution is sufficiently high.
(2) the activation energy for vacancy formation is low.
(3) atoms are significantly different in size.
(4) vacancies are present in a substitutional solid solution. (4)

3. The unit of flux J is
(1) moles $m^{-2}s^{-1}$ or atoms $m^{-2}\,s^{-1}$ (2) atoms $m^2\,s^{-1}$
(3) moles $m^{-3}\,s^{-1}$ (4) moles $m^{-1}s^{-1}$ (1)

4. For metal atoms, the most probable mechanism of diffusion is
(1) Vacancy mechanism (2) Interstitial mechanism
(3) Direct interchange mechanism (4) None of the above (1)

5. The dependence of diffusion coefficient on temperature in a certain temperature range is described by
(1) $D = D_0 \exp(Q/T)$ (2) $D = D_0 \exp(-Q/RT)$
(3) $D = D_0 \exp(-Q^2/RT)$ (4) $D = D_0 \exp(Q^4/RT)$ (2)

6. The fastest diffusion coefficient in Fe is
(1) C (2) W (3) Ni (4) H (4)

7. Fick's first law of diffusion is applicable under
(1) steady state conditions of mass flow
(2) non-steady state conditions
(3) steady as well as non-steady state conditions
(4) none of the above (1)

8. In the presence of impurities and deviations from stoichiometry, the diffusion rate of ions in ionic crystals
(1) enhances (2) decreases
(3) remains unaffected (4) first decreases and then enhances (1)

9. In comparison to lattice diffusion, the activation energy for diffusion along surfaces and grain boundaries is

(1) higher (2) lower

(3) almost negligible (4) infinite (2)

10. Interstitial diffusion is much ___________ than substitutional diffusion. (faster)

11. Substitutional solutes ___________ by interchanging positions with vacancies. (diffuse)

12. Solutions to Fick's second law are available for given set of initial and ___________ conditions. (boundary)

13. The solution to the diffusion couple set up can be used to determine D ___________. (experimentally)

14. Diffusion is the transfer of ___________ atoms which is accompanied with a change of concentration of the components in certain zones of an ___________. (unlike, alloy)

15. The error function of ∞ is ___________. (1)

16. The error function of 0 is ___________. (0)

17. If the diffusion jump distance is 1.5 Å, the theoretical value of D_0 in m^2/s is approximately ___________. (2.25×10^{-7})

18. For the same diffusion time, the depth of diffusion penetration at 500°C and 850°C is in the ratio of 1:6. The activation energy for diffusion in kJ/mol is ___________. (74)

19. The unit of ratio σ/D $(= ne^2z/kT)$ is ___________. $(c^2m^{-3}J^{-1})$

20. If $D_0 = 0.4 \times 10^{-4}$ m^2/s and $Q = 100$ kJ/mol, to double the depth of penetration, the initial temperature of 900°C should be increased to ___________. (1083°C)

SHORT QUESTION-ANSWERS

1. What do you understand by solid state diffusion?

Ans. This is a means of mass transport within solid materials by stepwise atomic motion.

2. What is self diffusion?

Ans. This refers to the migration of host atoms; for impurity atoms, the term interdiffusion is used. Two mechanisms are possible: vacancy and interstitial. For a given host metal, interstitial atomic species generally diffuse more rapidly.

3. What is steady state diffusion?

Ans. The concentration profile of the diffusing species is time dependent, and the flux or rate is proportional to the negative of the concentration gradient. Fick's first law is applicable.

4. How the mathematics for nonsteady state are described?

Ans. Under nonsteady state conditions, the flux changes with time and position along the diffusion direction. Fick's second law, a partial differential equation describes the nonsteady flow. Solutions to Fick's second law are available for a given set of initial and boundary conditions.

5. What can be inferred from the magnitude of the diffusion coefficient?

Ans. This is indicative of the rate of atomic motion being strongly dependent on and increasing exponentially with increasing temperature.

6. What is the effect of the presence of impurities and deviations from stoichiometry on the diffusion rate of ions in ionic crystals?

Ans. This enhance the diffusion rate of ions in ionic crystals by several order of magnitude.

7. What is activation energy for diffusion?

Ans. The activation energy for diffusion along surfaces and grain boundaries is lower than that for lattice diffusion. Diffusion along these special paths is effective at low temperatures or for very small particle or grain sizes, which a high boundary area to volume ratio.

8

Mechanical Properties of Materials and Mechanical Tests

1. INTRODUCTION

In determining the fabrication and possible practical applications, the mechanical properties of materials, their strength, rigidity and ductility are of vital importance. The important mechanical properties of materials are: *elasticity, plasticity, strength, ductility, hardness, brittleness, toughness, stiffness, resilience, malleability, fatigue, creep*, etc. The complete specifications of mechanical properties and composition of various materials have been standardised by BIS. Materials exhibit a wide range of mechanical properties ranging, e.g., from the ductility of pure copper to the hardness of diamond and most surprising elastic behaviour of rubber. There are several materials behave quite differently when stressed in different ways, e.g. steel and wood are stronger in tension whereas cast iron, cement and bricks are much stronger in compression. Obviously, stresses can produce a shape change and may also cause a material to break or fracture. For materials which have to withstand high temperatures, the melting point is also of importance.

The combination of high yield strength and good fracture toughness or ductility makes steel an excellent structural material. Modern high strength low alloy (HSLA) steels have a very fine grain size around 10 micro-meters which provides both high strength and good crack growth resistance or fracture toughness.

In addition to mechanical properties of materials, the following properties are also important for an engineer, to enable him in selecting suitable metals for various jobs:

(i) Physical Properties: These properties of materials include shape, size, colour, lusters, specific gravity, porosity, structure, finish, etc.

(ii) Technological Properties: We may note that all the technical properties of a metal are essentially its mechanical properties, which include properties like malleability, machinability, weldability, formability or workability, castability, etc.

(iii) Thermal Properties: Specific heat, thermal conductivity, thermal expansion, latent heat, thermal stresses, thermal shock, etc. fall under thermal properties of materials.

(iv) Electrical Properties: These include conductivity, resistivity, relative capacity, dielectric strength, etc.

(v) Chemical Properties: These properties include atomic weight, equivalent weight molecular weight, atomic number, acidity, alkalinity, chemical composition, corrosion, etc.

In this chapter, we shall discuss mechanical and technological properties. Other properties will be discussed in subsequent chapters. Initially, we present a brief review of some definitions of common mechanical properties.

2. COMMON TERMS

(i) Isotropy: A body is said to be isotropic if its physical properties are not dependent upon the direction in the body along which they are measured, e.g. a polycrystalline material in which the grains or crystals are randomly oriented behaves isotropically, i.e. its properties are independent of direction. Aluminium, steels and cast irons have random distribution of crystals. When a solid material is subjected to a force and if it withstand applied force, then such a material is of interest to engineers from design point of view. There are many solid materials which behave elastically when subjected to a load or force. By this it is meant that the solid material deforms when subjected to force, i.e. loaded, but returns to its original condition when the force is removed. A body displaying isotropy has only one refractive index, one dielectric constant and so on. R. Hooke studied the elastic behaviour of metals and formulated a law, now known as Hooke's law. When a specimen of length l and cross-sectional area A is loaded with a tensile strength P, the length increases by an amount δl, which is termed as *elongation*. On further increasing the load, the elongation also increases upto a certain limit (Fig. 8.1).

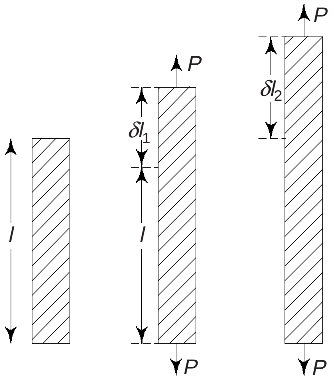


Fig. 8.1 Effect of axial tensile load on a specimen

(ii) Anisotropy: The quality of variation of a physical property with the direction in a body along which the property is measured, e.g., the resistivity of certain single crystals measured with the electric field along a particular crystallographic direction may be higher than along directions perpendicular to it. Thus such crystals are anisotropic with respect to resistivity. Obviously, anisotropy is a state having different properties in different directions. Various composite materials, aggregates of polycrystals with a preferred orientation, wood, laminated plastics, etc. are few examples showing high anisotropy.

(iii) Elasticity: It is the property of a material which enables it to regain its original shape and size after deformation within the *elastic limit*. However, in nature no material is perfectly elastic, i.e., a certain limit exists for every material beyond which it will not be able to regain its original shape and size. This limit is termed as *elastic limit*. Materials with high elastic limit are called as more elastic than the others. This property is always desirable in metals used in machine tools and other structural constituents. Steel and rubber are amongst the common examples of materials having elasticity. Elasticity of solids has its origin in the existence and stability of interatomic and intermolecular bondings, i.e. rubber and steel.

(iv) Plasticity: It is the ability of material to be permanently deformed (without fracture) even after the load is removed. To some extent all materials are plastic. This property of a material, which is extreme opposite of elasticity, is of importance in deciding manufacturing processes like forming, shaping, extruding operations etc. Metals possess more plasticity at high temperatures. Usually, plasticity of a material increases with increase in temperature and this is important in deciding the further operations.

The plasticity of a metal depends upon its nature and the environmental conditions, i.e., whether the metal is shaped red hot or in cold. Materials such as clay, lead are plastic at room temperature. Steel is plastic at bright red hot and hence shaped when hot. This property finds its use in forming, shaping and extruding operations of metals.

Plastic deformation takes place only after the elastic range has been exceeded. A general expression of plastic action would involve the time rate of strain since in the plastic state materials can deform under constant sustained stress. It would also involve the concept of limit of deformation before rupture. Evidences of plastic action in structural materials are called *yield*, *plastic flow* and *creep*.

As stated above, plasticity is important in forming, shaping and extruding operations. Some meterials are shaped cold, e.g. the deep drawing of the sheets, several materials particularly metals are shaped hot, e.g. the rolling of the structural steel shapes and the forging of certain material parts. Clay, lead etc. are the materials, which are plastic at room temperatures. Figure 8.2 shows elastic and plastic region of materials in a graph applied load vs. strain.

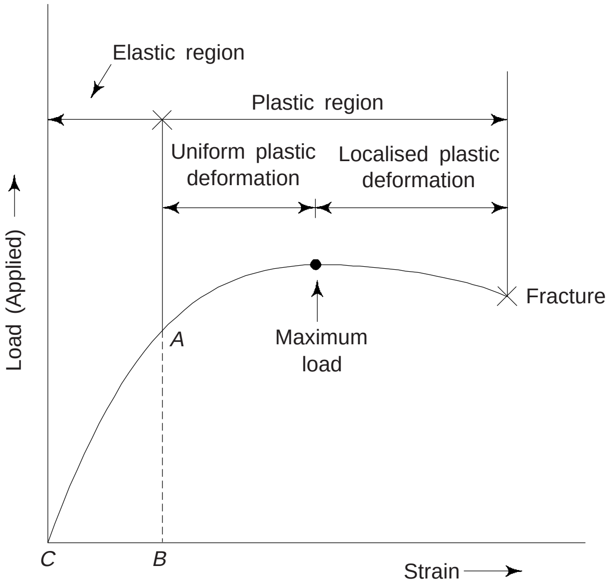


Fig. 8.2 Elastic and plastic regions of materials

(v) Ductility: It is defined as the property of a metal by virtue of which it can be drawn into wires or elongated before rupture takes place. It is the deformation produced in a material at the breaking point and measured by the percentage of elongation and the percentage of reduction in area before rupture of test piece. Its value is expressed as elongation, i.e., percentage elongation is most widely used to measure ductility. The term percentage elongation is the maximum increase in the length expressed as percentage of original length. Mathematically, one can express percentage elongation as

$$\text{Percentage elongation} = \frac{\text{Increase in length}}{\text{Original length}} \times 100$$

$$= [(l_f - l_o)/l_o] \times 100$$

Similarly, the term percentage reduction of cross-sectional area is the maximum decrease in cross-sectional area. Mathematically, one can express the percentage reduction in cross-sectional area

$$= \frac{\text{Decrease in crass-sectional area}}{\text{Original cross-section area}} \times 100$$

$$= [(A_o - A_f)/A_o] \times 100$$

In the above relations l and A represent the length and area of cross-section respectively, f and o are respectively the suffixes to denote final and original values.

Ductility commonly referred to in tensile test, which is strain at fracture. The unit of Ductility is same as that of strain. As stated above, the valuable information about ductility of a material is obtained from the form of test curves and by the percentage elongation and percentage reduction in the area of test piece at the neck. Ductility is a measure of the amount of permanent deformation that has occurred when the material reaches its breaking point.

We can see that brittle materials, e.g., cast irons show little or no plastic deformation before fracture, i.e. they are not ductile. A little consideration shows that a metal with a good percentage of elongation or reduction in cross-sectional area explains its high ductility. Metals with more than 15% elongation are considered as ductile. Metals with 5 to 15% elongation are considered of intermediate ductility. However, the metals with less than 5% elongation are considered as brittle ones. Brittle materials such as cast irons show little or no plastic deformation before fracture, i.e. they are not ductile. Copper is a ductile material, show considerable plastic flow due to high ductility before fracture. One can draw wires due to this property. Ductility of glass is high when hot and hence drawn in various shapes in hot condition. The order of ductility for few common metals is as under:

1. Gold, 2. Platinum, 3. Silver, 4. Iron, 5. Copper, 6. Aluminium, 7. Nickel, 8. Zinc, 9. Tin and 10. Lead.

Ductility is an important property of a material which governs its ability to be deformed in processes, e.g. drawing, rolling and forging. Adequate ductility ensures that the material during these processes will not fracture. There is an associated property by virtue of which sheets can be rolled from material is called '*malleability*'.

We must note that for any given material, the strength and ductility are inversely proportional to each other. Any treatment which increases strength, decreases the ductility. Ductility and strength, both are appreciably affected by temperature.

(vi) Strength: It may be defined as the capacity of material by virtue of which it can withstands or support an external force or load with rupture. It is expressed as force per unit area of cross-section. This is most important property of a metal, which plays a decisive role in designing various structures and components.

A material has to withstand different types of load, e.g. tensile, compressive and shear load. Accordingly, one may broadly classify strength into following two categories:

(a) Depending upon the value of stress, the strengths of a metal may be elastic or plastic.

(b) Depending upon the nature of stress, the strengths of a metal may be tensile, compressive, shear, bending and torsional. Now, we study all these types of strengths.

Ultimate strength: It is the load required to fracture a unit cross-section of material.

Elastic Strength: It is the value of strength corresponding to transition from elastic to plastic range, i.e. when material changes its behaviour from elastic range to plastic range. One cannot measure it accurately. This is known as ideal stress value and used to define elastic strength of a material.

Plastic strength: It is the value of strength of the material which corresponds to plastic range and rupture. It is also termed as ultimate strength. In actual practice, a specimen is subjected to a stress which is always less than the working stress. The ratio of ultimate stress to the working stress of a metal is termed as *factor of safety* or factor of ignorance. This greatly depends upon the nature of loads or stresses. Usually, the following values of factor of safety are generally kept for various loads.

Dead load	4 or 5
Live load	6
Alternating kind of load	8 to 12
Shock loading	12 to 15

One adopts the lower value of factor of safety by ensuring the metal to be without any defect, which is done through 'non destructive' tests. In determining the reliability of the design, the factor of safety is of great importance.

Tensile Strength: It is the maximum tensile stress which a material is capable of developing when subjected to loading upto rupture. Mathematically,

$$\text{Tensile stress} = \frac{\text{Maximum tensile load}}{\text{Original cross-sectional area}}$$

The tensile stress is expressed in N/mm^2 or MN/m^2. In actual practice, a given specimen is always subjected to a tensile stress less than the working tensile stress. Tensile strength is obtained from the following relation

$$\text{Tensile strength} = \frac{\text{Maximum Force in kg}}{\text{Original area in sq. cm}}$$

Tensile strengths and tensile modulus of few selected materials are given in Table 8.1

Table 8.1 Tensile strength and tensile modulus of few selected materials

Material	*Tensile strength* (N/mm^2)	*Tensile Modulus* ($\times 10^3$ N/mm^2)
Copper alloys	200-1200	110
Grey C.I.	150-350	100-150
Aluminium alloys	100-550	70
Magnesium	150-350	45
Zinc alloys	200-350	100
Mild steel	350-500	200
Martensitic S.S.	450-1300	200
Ferrite S.S.	500-600	200
Nickel alloys	400-1600	200
Polyethylene high density	20-40	0.9
Polyvinylchloride	40-60	2.7
Nylon-6	70-90	2.2
Polystyrene	30-70	3.3

Tensile strength is the ultimate strength in tension and corresponds to the maximum load in a tension test. It is measured by the highest point on the conventional stress-strain curve. This strength provides the basic design information on the material's acceptance in engineering tests. In ductile materials the load drops after the ultimate load because of necking and this indicates the plastic instability. For working stresses, the ultimate tensile strength is a logical basis in brittle materials. We may note that like yield strength, tensile strength is used with a factor of safety.

Compressive Strength

For a metal, the compressive strength is the value of load applied to break it off by crushing. Mathematically,

$$\text{Compressive Stress} = \frac{\text{Maximum compressive load}}{\text{Original cross-sectional area}}$$

Compressive stress is also expressed in N/mm^2 or MN/m^2. In actual practice, a given specimen is also subjected to a compressive stress less than the working compressive stress.

Shear Strength

The shear strength of a metal is the value of load applied tangentially to shear it off across the resisting section. Mathematically, we can express,

$$\text{Shear Stress} = \frac{\text{Maximum tangential load}}{\text{Original cross-sectional area}}$$

It is also expressed in N/mm^2 or MN/m^2. We must note that in actual practice, the given specimen is also subjected to shear stress less than the working shear stress.

Bending Strength

For a metal, it is that value of load which can break the metal off by bending it across the resisting section. Mathematically,

$$\text{Bending Stress} = \frac{\text{Maximum bending load}}{\text{Original cross-sectional area}}$$

This is also expressed in N/mm^2 or MN/m^2. In practice, the given specimen is also subjected to bending stress less than the working bending stress.

Torsional Strength

For a metal, the torsional strength is that value of load applied to break the metal by twisting across the resisting section. Mathematically,

$$\text{Torsional stress} = \frac{\text{Maximum twisting load}}{\text{Original cross-sectional area}}$$

This is also expressed in N/mm^2 or MN/m^2. In practice, the given specimen is also subjected to torsional stress less than the working torsional stress.

(vii) Stress and Strain: Stress is defined as the force per unit area and strain as the fractional change in length, area or volume. Obviously, this is the resistance of the body to deformation due to the application of external force.

Stress describes the intensity of a force that acts on a unit area. Its unit is N/mm^2 or N/m^2 which is called Pascal in SI and denoted by Pa. When the force acting over an area is uniformly distributed, we have

$$\text{Stress} = \frac{\text{Force}}{\text{Area}}$$

In several cases, such uniformily distributed loads are not present and therefore stress is non-uniform. This is why, the stress is always referred to a point and in a body the stress varies from point to point over any section. If P is the total load acting on the original cross-sectional area A_o, then normal stress,

$$\sigma = \frac{P}{A_0} \tag{1}$$

Obviously, stress is the intensity of internal force. The stress is said to be *normal* if load P is normal to the surface and *tangential* or shearing, if load is tangential to this surface. The normal or direct (tensile or compressive) stress is produced over a section when force is acting normal to the section. If the force is acting away from the section, the stress is tensile, if it is acting towards the section the section is compressive.

In general, the stress at any point will have six components and its nature is different than that of force and area both. In fact, stress at a point is a *tensor quantity* and needs the following specifications for complete specification: *(i)* magnitude *(ii)* plane passing through the point, on which stress is being defined and *(iii)* the direction in which stress is acting.

Strain is the deformation produced per unit length of a body due to the effect of stress on it. It is the ratio of the change in length of the specimen to its original length. If L is the original length of the sample and l is the change in length, then

longitudinal strain,
$$\varepsilon = \frac{l}{L} \tag{2}$$

Strain is simply a ratio and has no unit and it is a dimensionless quantity. Depending upon the type of load, strain can be *lateral* strain or *shear* strain.

As there are different types of stresses, there are different types of strains, e.g. *(i) tensile strain (ii) compressive strain*, *(iii) shear or transverse strain* and *(iv) volumetric strain.* The strain associated with the change in length is called the *elongation strain* (l/L). Similarly $\Delta V/V$ is the volume strain where V is the volume. When there is a change in shape and no change in volume, corresponding strain is called *shear strain*. The shear strain is measured by the angle. The behaviour of a material within the elastic limit is the same under compression as under tension.

Corresponding to *elastic and plastic* properties of materials, we have two classes of strain:

(a) *Elastic Strain:* It is the change in dimension of a body when it is subjected to a load. This is reversible phenomenon, i.e., elastic strain disappears after the applied load is removed. This is proportional to the stress applied.

(b) *Plastic Strain:* This is the permanent change in the body when subjected to a load. The change remains even after the applied load is removed.

The amount of elongation, expressed as a percentage of the original gauge length is called as the percentage elongation:

$$\text{Percentage elongation} = \frac{\text{Final length} - \text{Initial length}}{\text{Initial length}} \times 100$$

(viii) Hooke's Law: In 1678, Robert Hooke, for the first time stated that *within elastic limits, stress is proportional to strain*, i.e.

$$\text{Stress} = \text{a constant} \times \text{strain}$$

The ratio of stress to strain is a constant characteristic of a material, and this proportionality constant is called *modulus of the material*. It differs from material to material and for different nature of stresses. When the stress applied is *tensile or compressive*, the constant is called *Young's modulus* of elasticity. The slope of stress-strain diagram upto the limit of proportionality is called *Young's Modulus of elasticity* (Y or E) (Fig. 8.2).

$$\therefore \quad Y \text{ or } E = \frac{\text{Stress}}{\text{Strain}} = \frac{\sigma}{\varepsilon} = \text{constant} \tag{3}$$

When the shear stress (τ) and strain (γ) are used, it is called *modulus of rigidity* (G). It is given by

$$\tau = G\gamma \tag{4}$$

G is also called shear modulus. G and E are related as

$$G = \frac{E}{2}(1 + \nu) \tag{5}$$

where ν is Poisson's ratio.

For volumetric stresses and strains, the constant is called bulk modulus (K). The relation between K and E is

$$K = \frac{E}{3(1-2\nu)} \tag{6}$$

In Table 8.2 we list the values of Y, G and ν (Poisson's ratio) of few polycrystalline solids at room temperature. Nature of the bond and melting point, T_m (°C) are also given. We must note that the single crystal values can differ widely from polycrystalline solids, depending on the direction.

Table 8.2 Young's modulus (Y or E), shear modulus (G), poisson ratio (ν) of some polycrystalline solids. Melting point (T_m) and type of bond are also given

Material	*Y or E* ($\times 10^{10}$ N/m^2)	*G* ($\times 10^{10}$ N/m^2)	∂	T_m(°C)	*Bond*
Aluminium	7.0	2.4	0.34	660	Metallic
α-Fe	20.7	8.3	0.29	1538	Metallic
Cu	12	4.5	0.34	1084.5	Metallic
Ag	7.2	2.9	0.38	960.5	Metallic
Pb	1.6	0.6	0.45	327.4	Metallic
Tungsten	41	15.2	0.30	3370	Metallic
Si	11	—	—	1420	Covalent
Alumina	40	—	0.23	2050	Covalent + ionic
Magnesia	31	—	0.19	2900	Covalent + ionic
Polystyrene	0.3	—	0.33	100	Vander Waal

Poisson's Ratio (ν)

When a specimen is stressed by a uniaxial force, it is strained in the direction of the force and also in a direction perpendicular to the direction of the force. The strain in the direction of force is called longitudinal strain and that perpendicular to it as lateral strain. The relationship existing between the lateral strain and axial strain is called Poisson's ratio (ν):

$$\text{Poisson's ratio } (\nu) = \frac{\text{Lateral strain}}{\text{Longitudinal strain}} = \frac{l}{m} = \nu \text{ (constant)}$$

Poisson's ratio is an important elastic constant and its value is different for different materials. For most engineering structural materials its value is usually between 0.3 and 0.6. The relation between E, G and ν is

$$\nu = \frac{E}{2G} - 1 \tag{7}$$

Stress-Strain Relation

The stress and strain relation is commonly shown by means of a stress-strain diagram. These diagrams are obtained by drawing a graph or curve from the data obtained in a tensile test, in which an increasing tensile stress is applied to a specimen (Fig. 8.1). There are resulting changes in length which can be observed and recorded by strain measuring devices. Stress-strain for different engineering materials are shown in Fig. 8.3.

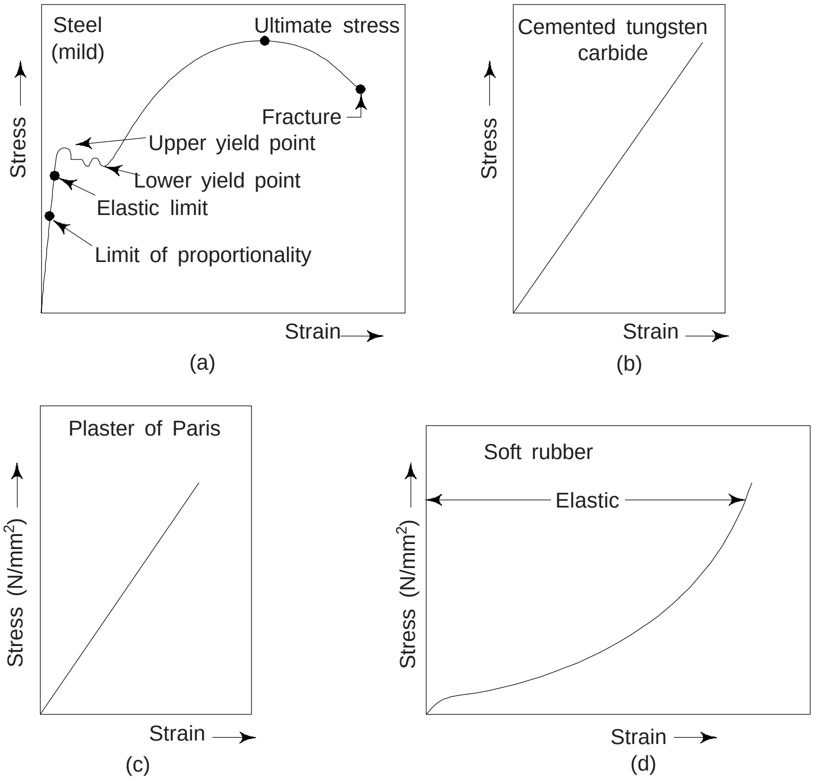


Fig. 8.3 Stress-strain curve for different engineering materials

In the case of ductile materials, at the beginning of the test, the material extends elastically. The strain (both longitudinal and lateral) at first increases proportionally to the stress and the sample or specimen returns to its original length on removal of the stress. The limit of proportionality (stress $\propto$ strain) is the stage upto which the specimen, i.e., material obeys Hooke's law perfectly (Fig. 8.3(a)).

On further increasing the applied stress, i.e., beyond the elastic limit, it produces plastic deformation so that a permanent extension remains even after the removal of the applied load, i.e. stress. The resultant strain, in this stage begins to increase more quickly than the corresponding stress and continues to increase till the *yield point* is reached. We must note that at the yield point the material suddenly stretches.

The ratio of applied load to original cross-sectional area is called the *normal stress* and this continues to increase with elongation, due to work hardening or strain hardening, until the tensile stress is maximum. This is the value of stress at maximum load and one can calculate it by dividing the maximum load by the original cross-sectional area. This stress is called *ultimate tensile stress* (Fig. 8.3(a)).

From Fig. 8.3(a) it is evident that at a certain value of load the strain continues at slow rate without any further stress or loading. This phenomenon of slow extension increasing with time, at constant stress, is termed *creep*. A neck begins to develop at this point, along the length of the specimen and further plastic deformation is localized within the neck. The cross-sectional area decreases in proportion to the increasing length during elastic elongation. We must note that the volume of the test bar, i.e. specimen remains constant. Figure 8.3(a) is a stress-strain diagram for mild steel. This diagram clearly shows the limit of proportionality, elastic limit, yield point, ultimate tensile stress and fracture stress at the breaking points. We note that this diagram shows a well-defined yield point.

Poorly defined yield point as in the case of brittle materials is shown in Fig 8.4. For the determination of the yield strength in such materials, following the general practice, one has to draw a straight line parallel to the elastic portion of the stress-strain curve at a predetermined strain ordinate value (say 0.1%). The point at which this line intersects the stress vs. strain curve is the yield point at off-set and called the yield strength at 0.1% or 0.2% of set strain.

In case of hard steels and non-ferrous metals stress is specified corresponding to a definite amount of permanent elongation. This stress is termed as the proof stress. We must note that the proof stress is applied for 15 seconds and, when removed, the specimen should not lengthen permanently beyond 0.1%. The method of finding the proof stress from the stress-strain curve is shown in Fig. 8.4.

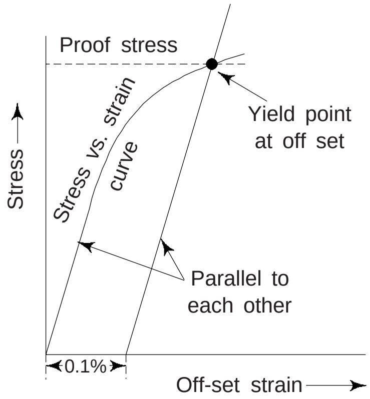


Fig. 8.4 Stress vs. strain curve for a brittle material

Stress vs. strain curves also help to explain the properties of ductile materials. We find that, greater the angle of inclination of the line of stress vs. strain proportionality to the ordinates, the more elastic is that metal. A higher yield point reveals greater hardness of the metal. A higher value of the maximum stress point shows that the metal is a stronger one. Similarly, the toughness and brittleness of metal are indicated by the distance from the ordinates of the breaking stress or load point. The metal is more brittle when the distance is shorter. Stress vs. strain curves for ferrous and non-ferrous materials are shown in Fig. 8.5. We find that brittle materials show little or no permanent deformation prior to fracture. Some metals and magnesium oxide exhibit brittle behaviour (Fig. 8.5). Prior to fracture, the small elongation reveals that the material gives no indication of impending fracture and brittle fracture usually occurs rapidly. It is reported that this is often accompanied by loud noise.

The values of *offset strain* (ε_o) have been standardized for different materials. ε_o equal to 0.2% or 0.35% is often used. The yield strength determined by the offset method is always described as yield strength for an offset ε_o. Commonly used offset values are given in Table 8.3

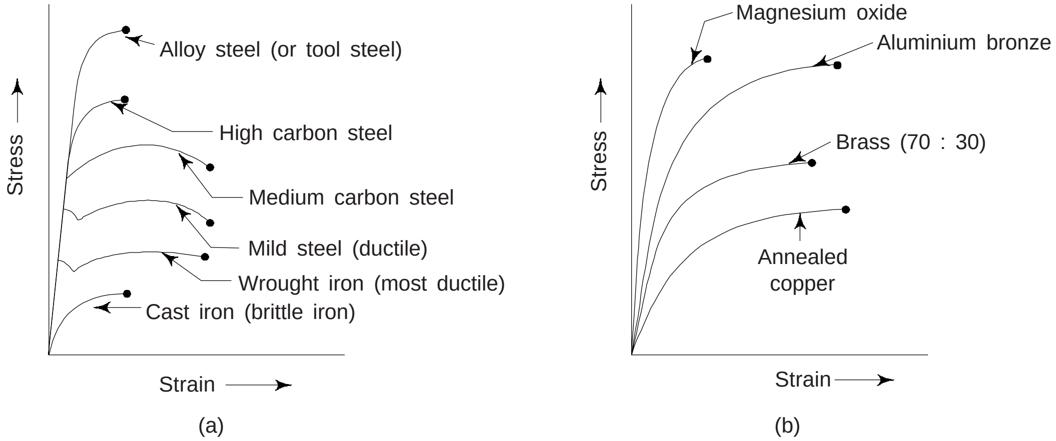


Fig. 8.5 Stress vs. strain curves (a) ferrous metals (b) non-ferrous metals

Table 8.3 Commonly used offset values (ASTM, *E* 6.35 *T*)

Material	*Stress (type)*	*Offset strain (ε_o%)*	*Equivalent strain*
Aluminium alloys	Tension and compression	0.20	0.0020
Brass and Bronze	Tension and compression	0.35	0.0035
Concrete	Tension	0.02	0.0002
Cast iron	Compression	0.05	0.0005
Steel	Tension or compression	0.20	0.0020
Wood	Compression	0.05	0.0005

Engineering and True Stress-Strain Diagrams

Stress and strain defined in the aforesaid manner are usually termed as conventional, nominal or engineering stress and strain. Truly speaking they are simply a measure of the load deformation on a standard specimen. In order to study the plastic flow of materials it is preferred to use the concepts of true stress and true strain. In tensile test for any engineering material one region of the specimen begins to deform much quicker than the rest as the test progresses. The reduction in area therefore raises a question, which area should be considered for calculating stress, the original area or the actual area of the specimen at any instant of the load. When we calculate the stress on the basis of the original area, it is called the *engineering* or *nominal stress*. If we calculate the stress based upon the instantaneous area at any instant of load it is then termed as *true stress*. If we use the original length to calculate the strain, then it is called the *engineering strain*. Now, we have

$$\text{True stress } (\sigma_T) = \frac{\text{Instantaneous load}}{\text{Instantaneous cross-sectional area}}$$

$$= \frac{P}{A_i} \tag{8}$$

where A_i is the actual area of the cross-section corresponding to load P. Similarly, we have

$$\text{True strain } (\varepsilon_T) = \int_{l_o}^{l_i} \frac{dl}{l} = \log\left(\frac{l_i}{l_o}\right) \tag{9}$$

where dl is the infinitesimal elongation, l_i is the instantaneous length and l_o is the original length. The true strain may also be expressed as:

$$\varepsilon' = \log_e (1 + \varepsilon) \tag{10}$$

The true stress and strain can also be determined from load and diameter measurements as

$$\sigma' = \frac{\sigma}{\pi d_i^2 \big/ A} \tag{11}$$

and

$$\varepsilon' = \log_e \frac{A_o}{A_i} = 2 \log_e \left(\frac{d_o}{d_i} \right) \tag{12}$$

where d_o and d_i are original and instantaneous diameter respectively.

Giving consideration to the mechanism of work hardening which is the interaction of mobile dislocations with each other, one will always expect the true stress-true strain curve as continually rising characteristic. Figure 8.6 shows both the engineering and true stress true-strain diagrams for copper. The ultimate strength of the engineering stress-strain curve corresponds to the beginning of necking. We can see from Fig. 8.6 that the rapid area reduction that accompanies necking causes the curve to fall. Obviously, this does not happen with the true curve. After beginning of necking, the strain is not homogeneous, when one part of the specimen is elongating more than the rest, it is not meaningful to use the entire length of the specimen for determining deformation. One can express the true strain in terms of area by substituting in the equation

$$\varepsilon_{\text{True}} = \varepsilon' = \log \frac{A_o}{A_i} = \frac{\text{Original area}}{\text{Instantaneous area}}$$

A comparison of stress-strain diagrams for aluminium alloy, using conventional and true values is shown in Fig. 8.7. We note that until the maximum load is reached, the points on the true stress-strain curve are always to the left of those on the conventional curve. After this point, we observe that because of the high local strains in the neck, the curve for true strain overtakes that for conventional strain and at the breaking points is far to the right.

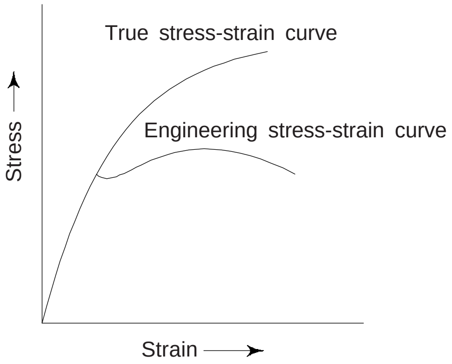


Fig. 8.6 **Engineering and true stress-strain diagrams for copper**

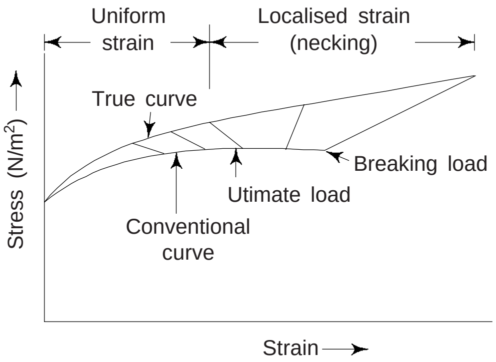


Fig. 8.7 **Typical conventional and true stress-strain curves for aluminium**

Both conventional and true stress and strain diagrams find application in engineering practice. Most of the diagrams used for practical purposes are based on engineering stress versus engineering strain. Conventional stress-strain diagrams are usually used in the elastic range, while true stress-strain diagrams are used in the plastic range. True stress-strain diagrams are required occasionally. Both types of diagrams have merits and demerits and the decision as to which should be used in any given analysis depends on several factors. However, in the specification of mechanical properties for design conventional stress-strain diagrams are largely used except for certain specific properties.

If we plot true stress against true stress a curve as depicted in Fig. 8.8 is obtained. Obviously, this relationship is non-linear and cannot be easily represented as elastic stress-strain relationship. Experimental results suggests following relationship between σ' and ε' as

$$\sigma' = k\, \varepsilon'^{\,n} \tag{13}$$

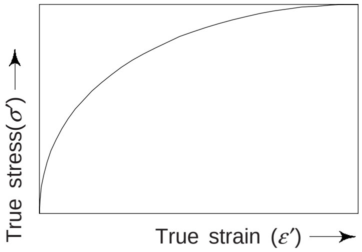


Fig. 8.8 True stress vs. true strain curve

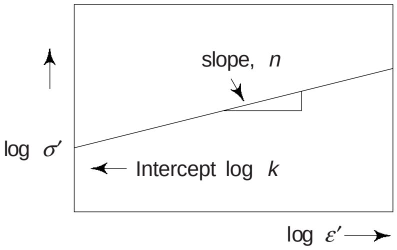


Fig. 8.9 log ε' vs log σ' curve

where k and n are constants for a material. k is called *strength coefficient* and n is *strain hardening exponent.*

Taking log of both sides of Eq. (13), we have

$$\log \sigma' = \log k + n \log \varepsilon' \tag{14}$$

The plot of the above relationship is a straight line on log σ' vs log ε' graph (Fig. 8.9). The slope of the line is n while intercept on log σ' axis is log k and thus one can determine k and n.

One can easily show that the condition for necking to begin is that the true strain becomes equal to strain hardening exponent, i.e.

$$n = \varepsilon'$$

Let us prove it. We have, the occurrence of maximum load condition as

$$dp = 0$$

or $$d(\sigma' A) = 0$$

or $$\sigma' dA + A\, d\sigma' = 0$$

Thus, we get $$\frac{d\sigma'}{\sigma'} + \frac{dA}{A} = 0 \qquad (i)$$

Also from volume constancy, AL = const.

or $$d(AL) = 0$$

or $$LdA + ADL = 0$$

or $$\frac{dL}{L} = -\frac{dA}{A} \qquad (ii)$$

But $$\frac{dL}{L} = d\varepsilon' = -\frac{dA}{A}$$

∴ $$\frac{d\sigma'}{\sigma'} = d\varepsilon'$$

or $$\frac{d\sigma'}{d\varepsilon'} = \sigma'$$

or $$\frac{d}{d\varepsilon'}(k\varepsilon'^n) = \sigma'$$

or $$kn\, \varepsilon'^{n-1} = k\varepsilon'^n$$

or $$n = \varepsilon'$$

i.e., the condition for necking to begin is that the true strain becomes equal to strain hardening exponent. Obviously, if necking is to be avoided in any mechanical process, say drawing, strain should be kept below the magnitude of strain hardening exponent.

(ix) Brittleness: It may be defined as the property of a metal by virtue of which it will fracture without any appreciable deformation. This property is just opposite to the ductility of a metal. Few examples of brittle

materials are cast iron, glass and concrete. This property of metals find its importance for design of machine tools, which are subjected to sudden loads. Metals with less than 5% elongation are known to be brittle ones.

(x) Toughness: It may be defined as the property of a metal by virtue of which it can absorb maximum energy before fracture takes place. It is the measurement of ultimate energy strength of material and is expressed as work units/unit volume, i.e. kg fm/m^3. Toughness is also calculated in terms of area under stress-strain curve (Fig. 8.9(A)). We may note that the toughness is maximum under Fig. 8.7(a) (c). We may note that the total area under stress-strain curve is the work expanded in deforming one m^3 of material until it fractures. This work is also termed as *modulus of toughness*. Tenacity and hardness of a metal are the measures of toughness. The toughness is approximately measured as the product of strength and elongation and these measurements are used for comparison purposes and termed as *merit number*.

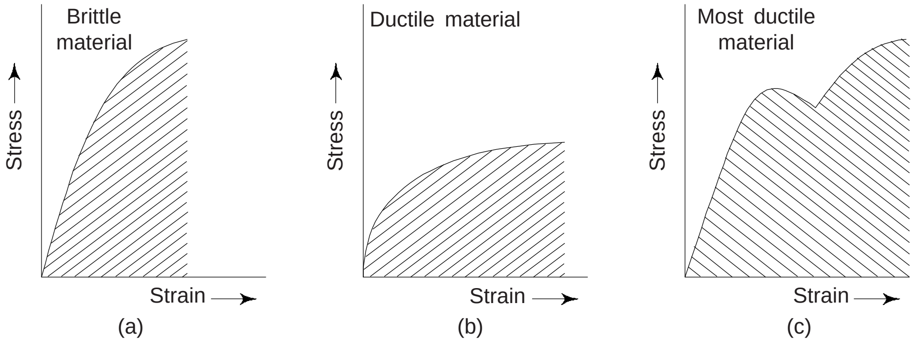


Fig. 8.9(a) Stress-strain curves

Toughness is the property of materials which enables a material to be twisted, bent or stretched under a high stress before rupture. The value of toughness falls with the rise in temperature.

Toughness is highly desirable property for structural and mechanical parts which have to withstand shock and vibration. Its importance lies in the selection of a material where the load increases beyond the elastic limit or yield point, e.g. power press punch and pneumatic hummer etc. Amongst the tough materials, few are: brass, mild steel, manganese and wrought. A ductile material with same strength as non-ductile material requires more energy for breaking and hence is more tough.

(xi) Stiffness: This may be defined as the property of a metal by virtue of which it resists deformation. Modulus of rigidity is the measure of stiffness. The term flexibility is quite opposite of stiffness. The materials which suffer less deformation under load have high degree of stiffness. We may note that the greater the stress required to produce a given strain, the stiffer is the material.

The stiffness of a metal is of great importance while selecting it for a number or a component of a machine or a structure. To make it clear, let us consider an example of two beams one of steel and other of aluminium. Both the beams are strong enough to carry the given load but aluminium beam will sag or deflect more than the mild steel beam. Obviously, the steel beam is stiffer than the aluminium beam. Stiffness is also used in graduating spring balance and spring controlled measuring instruments. Stiffness or flexibility is the function of dimensions, shape and characteristics of the material.

(xii) Resilience: This may be defined as the property of a metal by virtue of which it stores energy and resists shocks or impacts. It is measured by the amount of energy absorbed per unit volume, in stressing a material upto elastic limit. This property is of great importance in the selection of a material used for various types of springs.

The maximum energy which can be stored in a body upto elastic limit is termed as *proof resilience*. Proof resilience per unit volume is termed as *modulus of resilience*. Obviously, the energy stored per unit volume at elastic limit is modulus of resilience.

The annealed copper is note used for springs due to its very low elastic limit. However, the cold-worked copper has much high elastic limit (and resilience) and therefore it used for springs. This shows that resilience is associated with high elastic limit. We may note that resilience is of importance for materials required to bear shocks and vibrations.

(xiii) Endurance: This is defined as the property of a metal by virtue of which it can withstand varying stresses (same or opposite nature). The maximum value of stress, which can be applied for an indefinite times without causing its failure, is termed as its *endurance* limit. For ordinarily steel, the endurance limit is about half the tensile strength.

This property of a metal is of great importance in the design and production of parts in a reciprocating machines and components subjected to vibrations. It is always desirable to keep the working stress of a material well within the elastic limit.

3. ATOMIC MODEL OF ELASTIC BEHAVIOUR

We have seen in chapter 4 that the potential energy of a pair of atoms and interatomic forces between them is a function of their distance of separation r. A general expression for the potential energy U is

$$U = -\frac{A}{r^n} + \frac{B}{r^m} \tag{15}$$

where A, B, n and m are constants and $m > n$. The negative term on the right of Eq. (15) is the attractive energy term. The potential energy term arises due to the mutual repulsion between the atoms at the close distances of approach. The net interatomic force F is

$$F = \frac{dU}{dr} \tag{16}$$

The potential energy, U is a minimum and the net interatomic force is at r_o, i.e. the equilibrium distance of separation.

In order to decrease or increase the interatomic distance of separation starting from the equilibrium value r_o to r, we have to apply an external force F_{app} *equal in magnitude but opposite in sign* to F at r,

$$F_{\text{app}} = -F = \frac{dU}{dr} \tag{17}$$

From Fig. 8.10, to decrease the interatomic distance from r_o to r', a compressive force equal in magnitude to F' have to be applied. Further, to increase r_o to r'', a tensile force equal in magnitude to F'' have to be applied. From Fig. 8.10 it is evident that a tangent to the force-distance curve drawn at r_o practically coincides with the curve over a small range of displacement on either side of r_o. We can see that the strains in the elastic region for both brittle and ductile materials lie in the range of 0.001 to 0.005 and are, therefore, within this small range of displacement. Obviously, the negative of the slope of the force-distance curve at r_o is proportional to the Young's modulus, Y of a material. Further, the curvature of the potential energy curve at r_o is proportional to Y. Thus

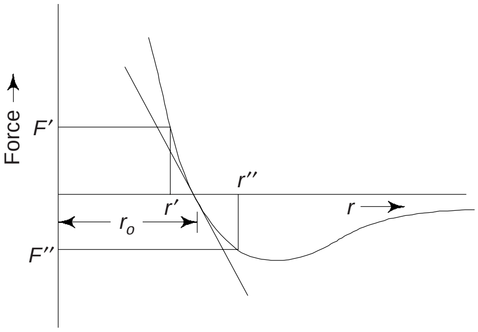


Fig. 8.10 The interatomic force-distance curve

$$Y = -\frac{dF}{dr} = \frac{d^2U}{dr^2} \tag{18}$$

We may note that materials with strong bonds have a deep potential energy well with a sharp curvature. Obviously, strong bonding results in large values for the elastic modulus. On the otherhand, the shallow potential well of the weekly bonded materials result in small values for the modulus Y of Li, Be, B and C (dia) in a row of the periodic table increase in that order

Element	*Li*	*Be*	*B*	*C (dia)*
Z (Atomic number)	3	4	5	6
Y (Young's Modulus) (GN/m^2)	11.5	289	440	1140

As we go down any given column of the periodic table, the covalent nature of the bonding decreases. We find that the elastic modulus shows a corresponding decrease. The elements of fourth column exhibit this trend.

Element	*C (dia)*	*Si*	*Ge*	*Sn*	*Pb*
Z	6	14	32	50	82
Y (GN/m^2)	1140	103	99	52	16

In a crystal, the interatomic distance, r varies with a corresponding variation in bond strength. This gives rise to *elastic anisotropy,* i.e., the elastic properties become a function of the crystal direction. We may note that the anisotropy is particularly evident in those materials which have two kinds of bonds. For, e.g., Y of graphite in a direction of crystal parallel to the sheets is 950 GN/m^2, which is much larger than that averaged over all directions, which is about 8 GN/m^2.

We may note that the transition metals have Y much higher than those of the alkali metals, as a result of the partial covalent character of their bonds. Metals of the first transition series have Y in the range of 200 GN/m^2. Y for metals of the second and the third transition series have higher values, reaching upto 600 GN/m^2.

The shear modulus G is related to the Young's modulus, Y through the Poisson's ration ν, as

$$G = \frac{Y}{2(1+\nu)} \tag{19}$$

4. MODULUS (Y) AS A PARAMETER OF DESIGN

The elastic modulus, Y is an important parameter in mechanical design. Materials with high *stiffness* and hence high modulus (Y) are often required. The stiffness of a material is its ability to resist elastic deformation or deflection on loading. It is proportional to the elastic modulus, Y for identical shapes. The value of Y for covalently bonded elements is very high, e.g. for diamond, Y = 1140 GN/m^2. However, they are not suitable for use in engineering purposes, due to high cost, brittleness and non availability. Brittle materials may have a high modulus, but they are not suitable as structural materials because they cannot withstand accidental overloading during service and may fail in a catastrophic manner.

Ductile elements are found suitable for structural components as these may withstand accidental overloading without catastrophic failure, e.g. metals. It is reported that among the metals, the elements of the first transition series offer a good compromise of adequate ductility and a moderately high Y, in the range of 200 GN/m^2. Although the metals of the second and third transition series have an even higher Y but they have the disadvantage of high density. However, by suitable alloying, Y of metals can be increased. We must remember that Y being a structure insensitive property, it can be increased only in proportion to the concentration of higher modulus solute that is added. Y for some ionic solids are listed below.

Material	NaCl	MgO	Al_2O_3	TiC	Silica glass
Y (GN/m^2)	37	310	402	308	70

Even though Y of some of ionic solids are quite high, but they also suffer from the lack of ductility like covalent solids.

Polymers, inspite of their plasticity not found suitable for applications where high stiffness is required, polymers have low Y, as the chains are bonded together by secondary bonds. The value of Y is dependent on the nature of the secondary bonding (hydrogen or Vander Waals bonding).

An attempt has been made to increase the stiffness, without the disadvantages of brittleness in *composite materials.* Boron has low density and found suitable for light weight applications for air borne structures. Y for boron is 440 GN/m^2, which is highest for elements, but it is brittle. Boron can be used as a reinforcing fibre for ductile matrix such as aluminium.

Anelastic Behaviour

Recoverable deformation that takes place as a function of time is termed an-elastic deformation. Due to some relaxation process within the material, the elastic deformation of the material continues even after the application of the load. On removal of the load, some part of the elastic deformation is recovered only as a function of time, with the reversal of the relaxation process.

Relaxation Processes

There are several relaxation proceses take place within a material in response to an externally applied stress. When the time scale of a *relaxation process* is *too fast* or *too slow* compared to the time interval over which the stress is applied, we found that the stress-strain relationship is essentially *independent of time.* However, if the time scale of the process is comparable to the time interval of stress application, the stress-strain relationship is *dependent of time* and results in *anelastic behaviour*.

Relaxation processes in materials have a wide range of time spectrum from 10^{-13} s for atomic vibrations to more than 10^6 s for the viscous flow of grain boundaries. The relaxation time for a given process is strongly dependent on temperature and decreases with increasing temperature in an exponential fashion.

Viscoelastic Behaviour

This is found in those materials which respond to an applied stress by both recoverable and permanent deformations, which are time dependent. Non-crystalline organic polymers exhibit this behaviour. Time-dependent permanent deformation is termed as viscous flow. We may note that it is analogous to the creep phenomenon in crystalline materials.

5. FUNDAMENTAL MECHANICAL PROPERTIES

The following are the fundamental mechanical properties of materials:

(i) Tensile Strength *(ii)* Hardness *(iii)* Impact Strength *(iv)* Fatigue and *(v)* Creep

We have already discussed about Tensile strength in section 2.

Hardness

This is the property of a material (metal) by virtue of it is able to resist abration, indentation (or penetration) and scratching by harder bodies. It is the resistance of a material to permanent deformation of the surface. In other words, one can define it as the resistance of the metal to penetration by an indentor. The hardness of a surface of the material is, of course, a direct result of interatomic forces acting on the surface of the material. We must note that hardness is not a fundamental property of a material, however, but rather a combined effect of compressive, elastic and plastic properties relative to the mode of penetration, shape of penetrator, etc. Hardness seems to bear a fairly constant relationship to the tensile strength of a given material and thus it can be used as a practical non-destructive test for an approximate idea of the value of that property and the state of the metal near the surface.

There are several methods of hardness testing, depending either on the direct thrust of some form of penetrator into the metal surface, or on the ploughing of the surface as a stylus is drawn across it under a controlled load, or on the measurement of elastic rebound of an impacting hammer which possessing known energy. Measurements of hardness are the easiest to make and are widely used for industrial design and in research. All hardness tests are made on the surface or close to it. We may note that in mechanical tests the bulk of material is involved.

Brinell hardness number is the hardness index, calculated by pressing a hardened steel ball (indenter) into test specimen under standeridies load. Brinell, Rockwell and Vickers hardness tests are used to determine hardness of metallic materials, to check quality level of products, for uniformity of samples of metals, for uniformity of results of heat treatment.

The Rockwell hardness is another index which is widely used by engineers. This index number is measured by the depth of penetration by a small indenter. By selecting different loads and shapes of indenter, different Rockwell scales have been developed. The value of Brinell hardness number is related to tensile strength and for two materials this is as shown in Fig. 8.11.

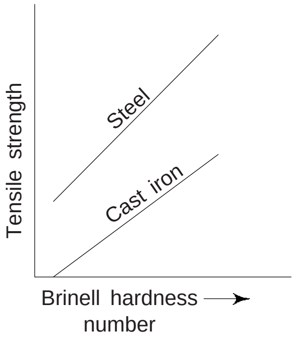


Fig. 8.11 Tensile strength versus Brinell hardness curves

The mechanism of indentation in all indentation tests is that when the indenter is pressed into the surface under a static load, a large amount of plastic deformation takes place. The material thus deformed flows out in all directions. Sometimes the material in contact with the indenter produces a ridge around the impression as a result of plastic flow. Large amounts of plastic deformation are reported to be accompanied by large amounts of transient creep which vary with the material and time of testing. We may note that the transient creep takes place rapidly at first and more slowly as it approaches its maximum. The time required for reaching maximum deformation for hard materials is short (~ few seconds) and for soft materials the time required to produce the desired indentation is quite long (~ few minutes). For minerals, hardness is expressed as resistance to abrasion which one mineral offers to another. Moh's scale has been developed for comparison of hardness. In order of increasing hardness, the following ten minerals are arranged: Talc, Gypsum, Calcite, Flourite, Apatite, Orothoclase, Quartz, Topaz, Corundum and Diamond.

Sometimes, hardness is expressed in terms of macro-hardness and micro-hardness. The macro hardness of material relates to its resistance to larger volume displacements in plastic deformation, whereas micro-hardness is the hardness of the materials in microscopically small volumes e.g. in grain boundaries.

Hardness of materials is of importance for dies and punches, limit guages, cutting tools, bearing surfaces etc. Softness of a material is opposite extreme of hardness. On heating all materials become soft.

(iii) Impact Strength: The resistance of a material to fracture under dynamic loads is characterized by what is called the impact strength. Impact strength of a material is a complex characteristic which takes into account both the toughness and strength of a material. In S.I. units the impact strength is expressed in Mega Newtons per m^2 (MN/m^2). It is defined as the specific work required to fracture a test specimen with a stress concentrator (notch) in the mid when broken by a single blow of striker in pendulum-type impact testing machine:

$KC = K/S_o$ (where K is the work of the fracture and S_o is the cross-sectional area of the specimen at the notch).

Impact strength is indicative of the toughness of the material, i.e. the ability of the material to absorb energy during plastic deformation. Obviously, brittleness of a material is an inverse function of its impact strength. Course grain structures and precipitation of brittle layers at the grain boundaries do not appreciably change the mechanical properties in static tension, but substantially reduce the impact strength.

Impact strength is affected by the rate of loading, temperature and presence of stress raisers in the material. It is also affected by variations in heat treatment, alloy content, sulphur and phosphorus content of the material.

Impact strength is determined by using the notch-bar impact tests on a pendulum-type impact testing machine. These results are used considerably in some industries to know shock-absorbing property of the material under the given variations. These results further help to study the effect of stress concentration and high velocity load application. Impact strength of a material varies with certain factors:

(i) Impact strength increases if the dimensions of the specimen are increased.
(ii) To some extent the velocity of impact also affects impact strength.
(iii) When the sharpness of the notch increases, the impact strength of the material required to cause failure decreases.
(iv) An indication about the type of the fracture that is likely to occur, i.e. ductile, brittle or ductile or brittle transition is provided by the temperature of the specimen under test.
(v) The angle of notch also improves impact strength after certain values.

Impact strength is designated as *KCU*, *KCV* or *KCT*, where the first two letters are the symbol of impact strength and the letter *U*, *V* or *T* indicates the shape of the stress concentrator (notch): U-shaped with the radius of the notch r_n = 1 mm, V-shaped with r_n = 0.25 mm or T-shaped notch with a fatigue crack at its tip. The *KCU* index is the principal criterion of impact strength, whereas *KCV* and *KCT* indices are used only in special cases.

(iv) Fatigue: It is the failure of a material by fracture when subjected to a cyclic stress. Fatigue can occur at a stress whose amplitude is much smaller than the static load required to produce fracture. The maximum stress that a material can withstand without failure for a specific large number of cycles of stress is termed its *fatigue* or *endurance limit*. Fatigue is distinguished by three main features: *(i)* loss of strength *(ii)* loss of ductility and *(iii)* increased uncertainty in strength and service life.

Engineering materials are often subjected to fluctuating loads while in service. Few examples of components which are subjected to fluctuating loads or alternate stresses are: *(i)* aircraft wings subjected to turbulent air *(ii)* leaf springs bent to and fro *(iii)* connecting rods pushed and pulled in piston engines and *(iv)* some parts of compressors, pumps and turbines, etc. subjected to repeated loading and vibration. If a metal wire is bent to and fro several times, it ultimately breaks. Rotating and vibrating parts of machines as in aeroplanes are liable to undergo fatigue and cause accidents. About 80% of failure in engineering components takes place due to fatigue failure. Steel have generally a fatigue limit which is normally 0.4 to 0.5 times the tensile strength of the material. Due to the effect of inhomogeneities under repeated loading, fatigue results in brittle fracture with no gross deformation at the fracture. Obviously, fatigue fractures occur without any warning. They result in brittle fracture.

In all materials including metals, plastics, rubber and concrete, fatigue is an important form of behaviour. In addition to the above mentioned situations there are many other situations where fatigue failure will be very harmful. Under following conditions, the fatigue fracture progresses rapidly: *(i)* maximum tensile stress of sufficiently high value *(ii)* large vibrations or fluctuations in applied stress *(iii)* large number of cycles of applied stress and *(iv)* other variables which may change the conditions of failure such as stress concentration, over loading, corrosion, residual stresses, etc. Due to difficulty of recognizing fatigue conditions, fatigue failures comprise percentage of failures occurring in engineering. The point at which the curve flattens out is termed as the *fatigue limit* and is well below the normal yield stress. The fatigue strength is usually defined as the stress that produces failure in a given number of cycles usually 10^7.

To avoid stress concentrations, rough surfaces and tensile residual stresses, fatigue specimens must be carefully prepared.

Fatigue properties of a material are affected by several factors, e.g.:

(i) Corrosion: The effect of corrosion is to reduce the number of cycles required to reach for every stress amplitude. We can protect the steel against salt corrosion by chromium or zinc-plating.

(ii) Surface Finish: Scratches, dents, identification marks can act as stress raisers and so reduce the fatigue properties. It is reported that shot peening a surface produces surface compressive residual stresses and improves the fatigue performance. Electro-plating produces tensile residual stresses and have a detrimental effect on the fatigue properties.

(iii) Temperature: As a consequence of oxidation or corrosion of the metal surface increasing, increase in temperature can lead to a reduction in fatigue properties.

(iv) Micro Structure of an Alloy: Composition of an alloy and its grain size can affect its fatigue properties. In comparison to coarse grained steel, finer grain size steels have higher strength. In addition to higher strength, fine grain size also results in better resistance to cracking, better machine finish and improved plastic deformation. In comparison to fine grains, coarse grained steels are less tough and have a greater tendency for distortion. Addition of lead and sulphur in steel increases its machinability and can act as nuclei for fatigue and so reduce fatigue properties.

(v) Residual Stresses: Residual stresses are produced by fabrication and finishing processes. Case hardening of steels by carburising results in compressive residual stresses, on surface it improves the fatigue. Several machining processes produce tensile residual stresses, which impair the properties.

(vi) Heat Treatment: This reduces residual stresses within a metal. By producing compressive residual stresses in surfaces, case hardening improves fatigue properties. However hardening and tempering treatments reduce surface compressive stresses and so adversely affect fatigue properties.

(vii) Stress Concentrations: These are caused by sudden changes in cross-section, keyways, holes, or sharp corners can more easily lead to a fatigue failure. Even a small hole lowers fatigue-limit by 30%. To study the effect of stress raisers on fatigue, a specimen containing *V*-notch or circular notch is prepared. When specimen is loaded, the notch has the following effects: *(i)* a taxial state of stress is produced *(ii)* stress gradient is set up from the root of the notch to the center of the specimen *(iii)* there is stress concentration at the notch. A crack is developed due to stress concentration at the root.

Stress Cycles

Fatigue is caused by fluctuating stresses. Figure 8.12 shows the different arrangements of fatigue loadings and corresponding stress cycles. Figure 8.12(a) shows the simplest type of load as the alternating stress, where the stress amplitude is equal to the maximum stress and the average or mean stress is zero. This stress cycle is of sinusoidal for *m*. Figure 8.12(b) shows the repeated stress cycle. This type of loading produces alternating stresses, superimposed on a steady stress the maximum stress is equal to the sum of the mean stress and stress amplitude. The maximum and minimum values can be considered as tensile and compressive stresses about the axis shown by dashed lines (---). Figure 8.12(c) shows a complicated stress cycle which is periodic and unpredictable, e.g. air craft wings.

Fatigue Fracture

This results from the presence of fatigue cracks, usually initiated by cyclic stresses, at surface imperfections, e.g. machine markings and slip steps. Although the initial stress concentration associated with these cracks are too low to cause brittle fracture, however they may be sufficient to cause slow growth of the cracks into the interior. Eventually the cracks may become sufficiently deep and therefore the stress concentration exceeds the fracture strength and sudden failure occurs.

It is reported that the extent of the crack propagation process depends upon the brittleness of the material under test. In brittle materials the crack grows to a critical size from which it propagates right through the structure rapidly, whereas with ductile materials the crack keeps growing until the area left cannot support the load and an almost ductile fracture suddenly occurs.

Fatigue Failure

One can recognize *fatigue failures* by the appearance of fracture. Fatigue failure has a number of specific features compared with failure under static loads:

(i) It occurs at lower stresses than the failure at static loads, i.e., lower than the *yield strength* or ultimate strength.

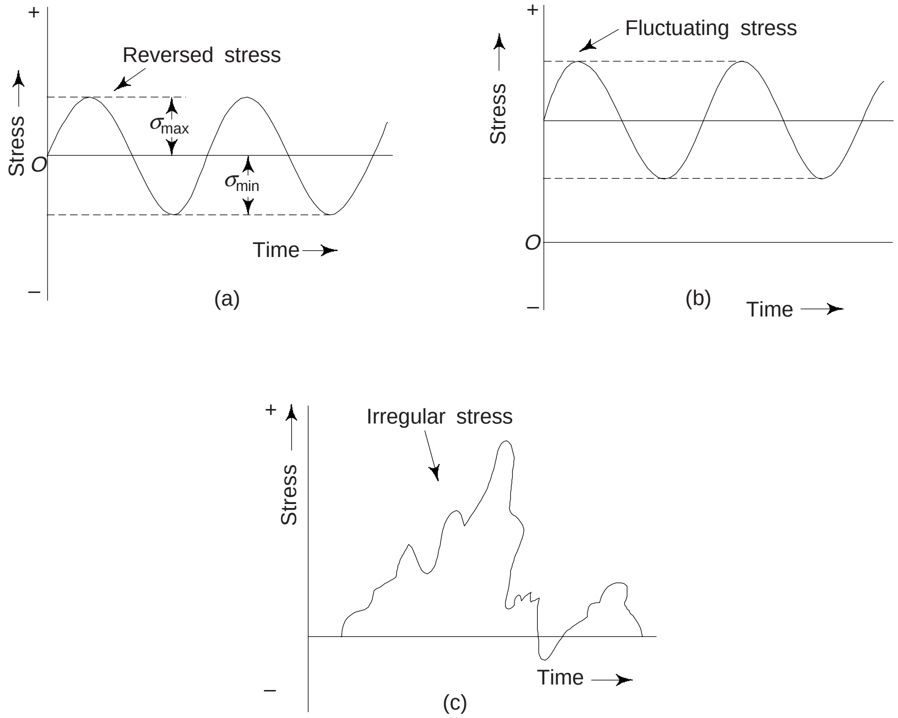


Fig. 8.12 Stress cycles for testing of fatigue

(ii) Failure starts on the surface (or near it) locally, in places of stress (strain) concentration Local stress concentrators are formed by surface defects appearing on cyclic loading or notches as traces of surface treatment or the effect of the surrounding medium.

(iii) Failure occurs in a number of stages; accumulation of defects in the material; nucleation of fatigue cracks; gradual propagation and joining of some cracks into single main crack; and rapid final destruction.

(iv) Failure has the typical structure which reflects the sequence of fatigue processes. A failure usually has the initial zone of destruction (the zone of nucleation of micro cracks), the fatigue zone, and the final failure zone (Fig. 8.13). The initial zone of failure is usually near the surface and has small size and smooth surface. The fatigue zone is the zone where a fatigue crack gradually develops. It has typical concentric ripple lines which are an evidence of jumpwise propagation of fatigue cracks. The fatigue zone develops until the increasing stresses in the gradually diminishing actual section attain a level at which instantaneous destruction takes place and forms the zone of final failure.

The main basic reasons for taking place of fatigue failures are:

(i) Surface imperfections like machining marks and surface irregularities.

(ii) Stress concentrations like notches, keyways, screw threads and matching under-cuts.

(iii) At low temperature the fatigue strength is high and decreases gradually with rise in temperature.

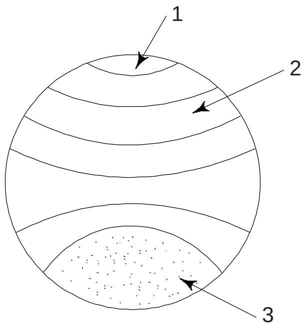


Fig. 8.13 Fatigue fracture
1-Zone of crack nucleation
2-Fatigue zone
3-Final fracture

(iv) Fatigue strength reduces by corroding environments. Following surface treatments like polishing, coatings, carburizing, nitriding, etc., their effect can be reduced.

Mechanism of Fatigue Failure

This is associated with the development and accumulation of microplastic deformations in the surface layer and is based on dislocation movement. The possibility of dislocation movement at stresses below the yield limit is due to the anisotropy and random orientation of crystals in the metal structure. At rather low mean stresses, the stresses appearing in some crystals may be sufficient for movement of weakly 'locked' dislocations. Besides thin surface layers (to a depth of 1-2 grains) have a typically low stress for operation of Frank-Read dislocations sources. For this reasons, microplastic deformations and damage of thin surface layers can be observed in mild (annealed) metals already at an early stage of loading (1-5% of the total number of cycles to failure). Microplastic deformations manifest themselves in the form of slip lines in the surface, with the density of these lines increasing as the number of cycles increases. As dislocations emerge to the surface, their damage is enhanced and appears as steps. Slip lines become wider, change to slip bands, and gradually degenerate into extrusions and intrusions (Fig. 8.14). An extrusion is a bulge and intrusion, a depression of slip band. Extrusions and intrusions (i.e. ridges and recesses) form a rough relief of the surface. Recesses are places of concentration of strain, and therefore, of vacancies and dislocations. A high local density of vacancies and dislocations can result in the formation of voids and loose places which merge together to form submicrocracks. The development and joining of submicrocracks in turn leads to the formation of microcracks.

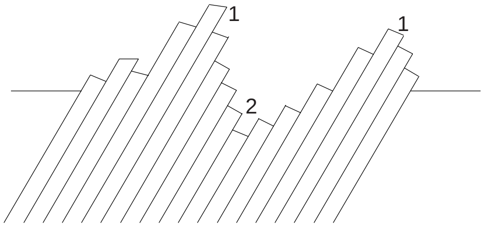


Fig. 8.14 Arrangement of slip lines in (1) extrusions (2) intrusions

Fatigue are often referred to as "progressive fractures". We have already remarked that fatigue is a result of cumulative process involving slip. High temperature increases the mobility of atoms, facilitating greater slip and deformation before fracture.

Highly localized stress is also developed at abrupt changes in cross-section, at the base of surface scratches, at the root of a screw thread, at the edge of small inclusions of foreign substances, and at a minute blow hole or similar internal defects. We may see that these are typical conditions which give rise to the susceptibility to failure by fatigue.

Design For Fatigue

To secure satisfactory fatigue life it is necessary for one to observe the following points in design procedure:

(i) To avoid stress concentration by eliminating sharp recesses and severe stress raiser, *modification of design.*
(ii) Precise control of the surface finish by avoiding damage to surface by rough machining, punching, stamping, shearing, etc.
(iii) Control of corrosion and erosion or chemical attack in service and to prevent of surface decarburization during processing or heat treatment.
(iv) Surface treatment of metal.

We have already mentioned that fatigue is essentially a surface initiated phenomenon and the conditions of the surface is particularly important. Moreover, fatigue occurs by slip, any structural condition that can inhibit slip would be useful for long fatigue life.

Fatigue Tests

Machines which may be used for making fatigue tests under cycles of repeated or reversed stress are usually classified according to the type of stress produced:

(i) Machines for cycles of axial stress (tension, compression)
(ii) Machines for cycles of flextural (bending) stress.
(iii) Machines for cycles of torsional shearing stress.
(iv) Universal machines of axial, flextural or torsional shearing stresses or their combinations.

Rotating Beam Fatigue Test

The most common test loading for fatigue testing is pure bending, because it is easy to apply. A circular specimen is gripped in collets (Fig. 8.15). Figure shows the test specimen and main features of the testing machine. The test machine has high speed electrical motor with a speed of 1000 rpm. Next to the motor there is a large bearing, which relieves the motor from large bending moment which is applied to the specimen. A collet is attached with a rotating lever arm which is further connected with a small bearing. A force is applied on the bearing, which causes the specimen to bend. The upper surface of the specimen will be under tension, whereas the lower surface in compression. Since the specimen is rotated by a motor, the surface of specimen is subjected to alternate tensile and compressive stresses. The revolution counter reader provided with the machine records the number of cycles applied and when the specimen breaks then the counter automatically disengages. Truly, the disengaging device is designed to stop the testing machine itself.

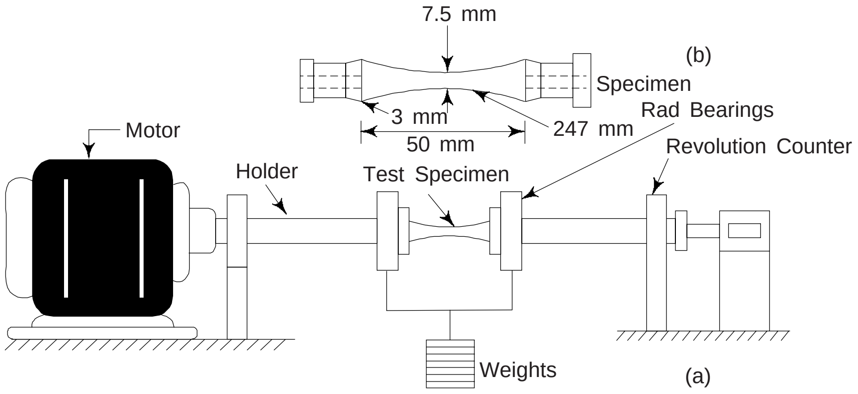


Fig. 8.15 (a) Rotating beam fatigue test Machine (b) fatigue test specimen in bent state

In order to obtain the fatigue limit, i.e. endurance limit of a metal, it is necessary to prepare a number of specimens which are representative of that metal. The first specimen of metal is tested at a relatively high stress so that the failure will occur at a small number of applications of the stress. The other specimens of the metal are tested relatively at a lower stress than the previous one. With the decrease in stress value, the life of specimen increased. The number of repetition required to produce rupture, i.e., fracture increases as the stress decreases. Specimen with stress below the endurance limit will not rupture. The life of the specimen is expressed in number of cycles required upto failure at maximum applied force. The results of fatigue tests are commonly plotted on diagrams in which values of stress are plotted as ordinates and values of cycles of stress for rupture as abscissae. The curve is called *S-N* diagram, where *S*-stands for stress and *N* for number of cycles (Fig. 8.16). These diagrams are drawn using semi-logarithmic plotting, i.e. plotting '*N*' on logarithmic scale.

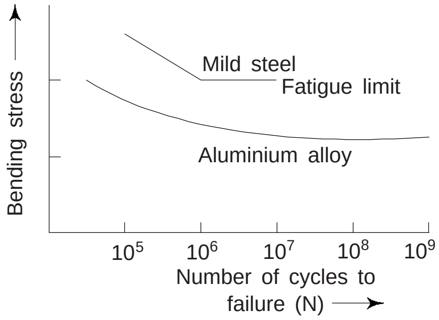


Fig. 8.16 S-N Diagram

The form of *S-N* curve is of much significance to engineers. For metals, e.g. mild steel and titanium the curve becomes horizontal at certain stress. This stress value is termed as *fatigue limit* or *endurance limit*. Below this stress value, the specimen does not fail or fracture, i.e. the material will not fail even after infinite number of stress cycles. Obviously, this shows that if the material is loaded to a stress below the fatigue limit, it will not fail, no matter how many times the stress is applied.

For all ferrous metals tested, and for most non-ferrous metals, these *S-N* diagrams become almost horizontal for values of N ranging from 1×10^6 to 5×10^7 cycles, thus indicating a well-defined endurance limit. We must note that the number of cycles are arbitrarily taken.

Wohler's Fatigue Test

Figure 8.17 shows a diagrammatic sketch of Wohler fatigue testing machine. In this machine, the specimen in the form of cantilever forms the extension of a shaft which is driven by an electric motor. Through a ball bearing, dead loading is applied to the specimen. When the machine is in action i.e. it runs, the specimen rotates and the fibres of the specimen are subjected to reversed stresses. In some instances the specimen is tapered or a two point loading is applied to obtain a uniform surface stress over a considerable length of specimen.

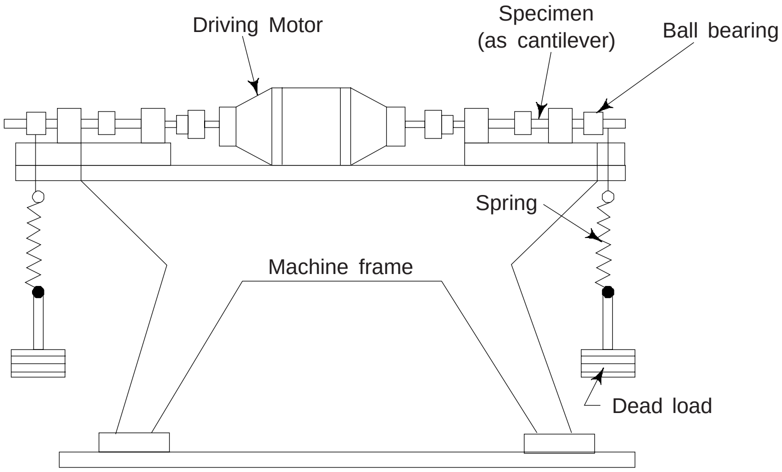


Fig. 8.17 Wohler fatigue testing machine

To cause failure the number of cycles vary with applied stress. When stress is higher, fewer are the cycles required for causing the fracture. Obviously, a stress is reached below which fracture would not take place within the limits of a standard test and this is termed as 'endurance limit'. The length of such a standard test depends on the material being tested and types of loading. Usually it is of the order of 500,000 cycles for very hard steels, 5,000,000 for soft steels; 10,000,000 for cast steel and cast iron; and for non-ferrous metals and alloys from 1,000,000 to about 50,000,000. If the fracture does not take place within these limits, then it is understood that it will not take place at all.

There are certain well defined characteristics for fatigue failures of metallic materials. The fractured surface frequently exhibit two distinct zones. One can find the cause of the failure by careful examination of such a failure. There is a smooth part usually showing concentric markings starting from a nucleus stress raisers, and rougher part often presenting crystalline faces.

Applicability of Fatigue Tests

Till today, numerous tests of various types have been successfully employed in developing a fairly adequate data on endurance limits of several metals. These data appear to be reasonably reliable and seem to form a satisfactory basis for design. However, results obtained from fatigue tests on small specimens cannot be applied to the design of complicated and built-up units. In order to obtain direct information on the behaviour of such parts or units under repeated loadings, fatigue tests have to be made on full sized specimen, e.g. axis, and rivetted, bolted and welded joints. We must note that a fatigue test is generally unsuitable as an inspection test, or quality control test, owing to the time and effort needed to collect the data.

Theories of Fatigue

There are several theories of fatigue, we will concentrate only on few theories:

(i) Orowan Theory: This theory considers the metal to contain small, weak regions. Slip occurs easily along these regions. The metal also contains impurities, i.e. inclusions, which act as notches and cause stress concentration (around inclusion). In elastic matrix, these regions are treated as plastic regions, which experience increase in stress even when the repeated applied stress is constant. This increase in stress causes plastic strain in the weak region. A crack is formed, when plastic strain in the weak region exceeds beyond certain value. Until a large crack is formed, this process is repeated again and again.

(ii) Fatigue Limit Theory: There are some metals which have well defined fatigue limit while other metals do not exhibit fatigue limit. The metals which undergo strain aging have well defined fatigue limits. The experiment is performed on a specimen in which total nitrogen and carbon content was decreased. Obviously, the specimen with lesser amount of nitrogen and carbon, will have lesser tendency for strain energy. Moreover, *S-N* curve will not show fatigue limit for this specimen. Heat treated steels exhibit definite fatigue limit. This reveals that localized strain aging affects the fatigue properties of the specimen to large extent.

(iii) Wood's Theory: The strain direction in fatigue is reversed again and again. The slip produced by fatigue consist of slip bands, which are the slip movements of the order 10^{-9} cm in length and 10^{-7} to 10^{-6} cm in height. When there is a strain in one single direction, the steps that appear are simple (Fig. 8.17 (a)) and when the loading is cyclic, the slip bands tend to group as notches or ridges (Fig. 8.17(b)). The notch acts as a starting point for fatigue crack and the notch with a very small root radius, act as stress raiser (stress concentration point) and crack slowly propagates through the material, and after some time a stage reaches when the remaining portion of cross-section is too small to bear the maximum load. Sudden failure occurs at this point. The final fracture is usually of cleavage type.

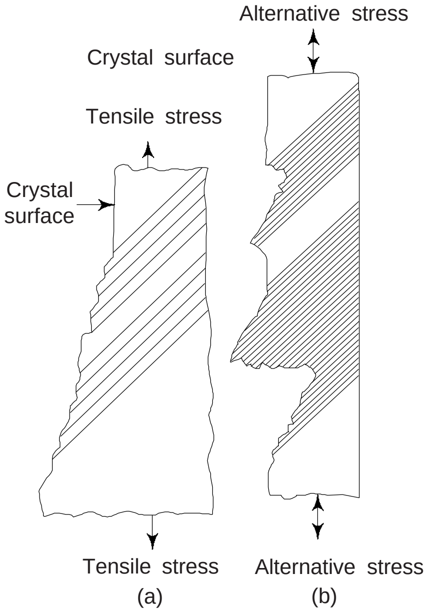


Fig. 8.17 **(a) Steps produced due to strain (in single direction) (b) When strain is cyclic, notch and edge formation in early stages on fatigue**

Dislocation Movement Theory

There is movement of dislocation during fatigue. A localised deformation occurs in the slip bands due to this movement of dislocation. This deformation is called extrusions (rides), i.e., some metal is expelled out of its surface. The extrusions are generally accompanied by intrusion (inverse of extrusion), and these extrusions and intrusions are responsible for crack initiation. It is reported that such extrusions and intrusions, i.e., disturbances appear as early as 1/10 of the whole life of specimen and they have the height of 10^{-6} to 10^{-7} cm.

Corrosion Fatigue

The destruction of the metal under the combined effect of fatigue (cyclic stress or load) and corrosion (electrochemical attack of the medium) is called corrosion fatigue. The endurance limit of metals decreases in corrosion media. Corrosion alone produces pits on the surface of metal. These pits act as stress risers and reduce fatigue strength. The reduction in fatigue strength is much more and crack propagates at much faster rate, when corrosive attack and loading occur simultaneously.

There are some materials which show definite fatigue limit when fatigue tested in air, i.e. absence of corrosion. These materials do not show any fatigue limit when tested in corrosive medium. Obviously, fatigue strength goes on reducing in corrosive medium.

Fatigue properties can be greatly improved by careful design, reduction in stress concentration and proper use of residual compressive stresses. To some extent, fatigue properties can also be improved by controlling certain metallurgical variables.

(v) Creep: It is essentially a slow rise of plastic deformation under the action of stresses below the yield strength of the material. A typical curve of deformation versus loading time is shown in Fig. 8.18. A creep curve usually consists of three portions corresponding to particular stages of creep. The first stage is the stage of what is called unstable, or transient creep (also called initial or primary creep). It features gradual decrease of the deformation rate to a definite constant value. The second stage is the stage of steady-state, or secondary creep and is characterized by a constant deformation rate. At the third stage (tertiary, or accelerated creep), the deformation rate increases upto failure. The third stage is, as a rule, short and should be avoided, since quick failure of parts is inevitable at this stage.

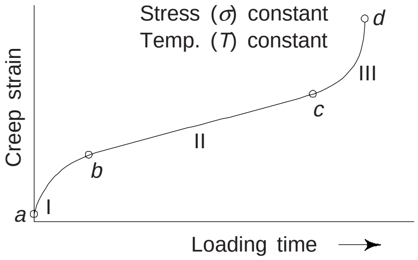


Fig. 8.18 Creep curve at constant temperature and stress: I-transient creep stage, II-steady stage creep stage and III-fracture stage

A physical explanation of the three stages of creep is as follows:

Primary Creep: This stage is mainly due to dislocation movement. The creep rate decreases with time and the effect of work hardening is more than that of the recovery process.

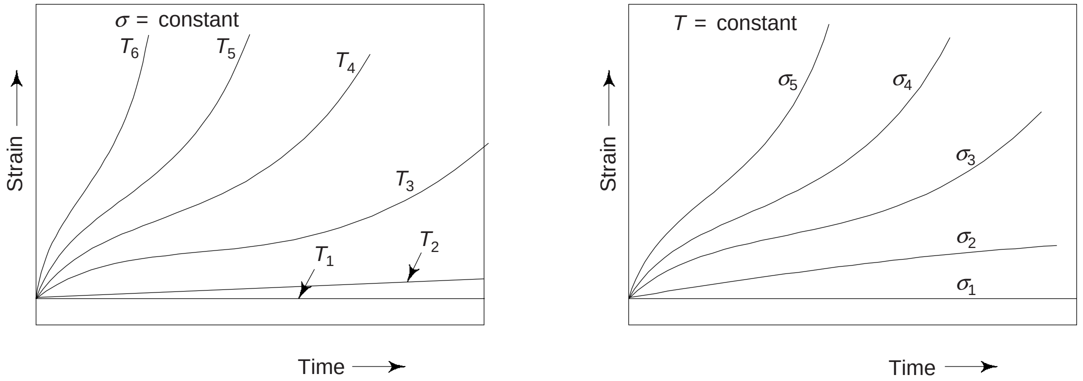
Secondary Creep: The rates of work hardening and recovery during this stage are equal, so the material creeps at a steady rate (minimum creep rate). Depending upon the state level and temperature, steady state creep may be essentially viscous or plastic in character. Structural observations reveals that polygonization is an important recovery process during secondary creep.

Tertiary Creep: Creep rate increases with time until fracture occurs in this stage. Tertiary creep can occur due to necking of the specimen or grain boundary sliding at high temperatures and this continues until specimen fractures.

The development of each stage of creep depends on the temperature and stress. For the same stress, an increase of temperature shortens the time of the second stage and accelerates failure (Fig. 8.19(a)). An increase of stress at the constant temperature has a similar effect (Fig. 8.19(b)). One can obtain the maximum operating temperatures and stresses from creep data plotted in a suitable form. The phenomenon of creep is observable in metals, ionic and covalent crystals, the amorphous materials such as glasses and polymers. In polymers, the phenomenon of creep is important at room temperature, in alloys at 100°C, and in steel above 300°C.

Metals, generally, exhibit creep at high temperatures, whereas plastics, rubbers and similar amorphous materials are very temperature-sensitive to creep. A material subjected to a constant tensile load at an elevated temperature will *creep* and undergo a time dependent deformation.

The temperature and time dependence of creep deformation indicates that it is a *thermally activated process*. There are several atomic processes known to be responsible for creep in crystalline materials.



(a) (b)

Fig. 8.19 Creep curves depending on (a) temperature (b) stress:
$T_1 < T_2 < T_3 < T_4 < T_5 < T_6$;
$\sigma_1 < \sigma_2 < \sigma_3 < \sigma_4 < \sigma_5$

The yield strength which is determined in short term tests, cannot be the criterion of high-temperature strength, since it does not consider the behaviour of a material in long-term loading. The actual criteria of high-temperature strength are the *creep limit* (creep-strength) and *long term strength*.

The *creep limit* is the stress at which a material can be formed by a definite magnitude during a given time at a given temperature. The designation of creep limit includes the temperature, the deformation, and the time in which this deformation appears. For example, $\sigma^{550}_{1/100000}$ = 100 MPa implies that a plastic deformation of 1% will appear in the material at 550°C under a stress of 100 MPa in 100000 hours.

The *long-term strength* is the stress that causes failure of the material in a definite time at the specified temperature. The designation of long term strength contains the temperature and time to failure. For example, σ^{600}_{10000} = 130 MPa denotes that the material can withstand a stress of 130 MPa at a temperature of 600°C for 10000 hours. The long-term strength characterizes the softening of a material that occurs in the process of long-term loading at elevated temperatures. The long-time strength is always lower than the ultimate strength which is determined in short-time tests at the same temperature.

Materials are required in many applications, where they are required to sustain study loads for long periods, e.g. in turbine blades, pressure vessels for high temperature chemical processes, aircraft, lead coverings (sheaths) on telephone cables, furnace parts and timber beams in roofs of buildings. The material, under such conditions may continue to deform until its usefulness is seriously impaired. Such time-dependent deformations of a structure can grow large and even result in final fracture without any increase in load. If under any conditions, deformation continues when the load (stress) is constant, this additional deformation, which is called as creep becomes important. Creep becomes important in *(i)* Rockets and missiles, supersonic jets, *(ii)* Nuclear reactor field, *(iii)* Gas turbines working at high temperatures, *(iv)* Steam and chemical plants operating at 450°-550°C, *(v)* Soft metals used at about room temperature, e.g. lead pipes, etc.

To a certain extent, most materials creep at all temperatures, although engineering materials, e.g. steel, aluminium and copper creep very little at room temperature. We may note that high temperatures lead to rapid creep which is often accompanied by microstructural changes. In high-temperature creep resistance applications generally precipitation-hardened alloys are used because in addition to acting as barriers to slip and grain boundary sliding, the fine dispersion of particles apparently restricts the rate of recovery.

Creep Curve Equations

One can consider creep curve to consist of a combination of two different creep processes, which occur after the sudden strain. These are: *(i)* transient creep having decreasing creep rate and *(ii)* viscous creep having constant rate. One can write the equation of curve with the help of following empirical equation,

$$l_t = l_o\ (1 + \beta t^{1/3})\ \exp\ (kt)$$

where, $l_t \rightarrow$ length of the specimen at time t, $l_o \rightarrow$ length of specimen after sudden creep and β and k are constants. Setting $k = 0$, for transient creep, we have

$$l_t = l_o\ (1 + \beta t^{1/3})$$

Now, change in length/unit time $= \dfrac{dl_t}{dt} = \dfrac{1}{3} l_o \beta\, t^{-2/3}$

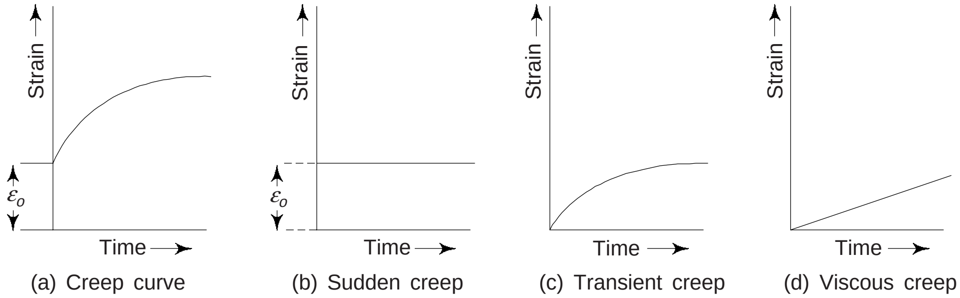
For viscous creep, setting $\beta = 0$, we have

$$l_t = l_o \exp\ (kt)$$

Change in length/unit time $= \dfrac{dl_t}{dt} = kl_o \exp\ (kt) = kl_t$

or $k = \dfrac{1}{l_t} \dfrac{dl_t}{dt} = \dfrac{\text{strain}}{\text{unit time}} = \text{strain rate}$

Strain/unit time, i.e. k proceeds at constant rate. Obviously, the slope of viscous is constant. Creeps curves for sudden, transient and viscous creep are shown in Figure 8.20.



(a) Creep curve (b) Sudden creep (c) Transient creep (d) Viscous creep

Fig. 8.20 Creep curve analysis. $\varepsilon_o \rightarrow$ instantaneous strain

Types of Creep

The creep may be classified depending on the temperature as *(i)* the logarithmic creep *(ii)* the recovery creep, and *(iii)* the diffusion creep. At low temperature the creep rate usually decreases with time and one obtains a logarithmic creep curve. At higher temperatures in the range of 0.5 to 0.7 of T_mK (melting point temperature on the absolute scale), the influence of work-hardening is weakened and there is a possibility of mechanical recovery. As a result, the creep rate does not decrease as rapidly as at low temperature. This creep, occurring at high temperature in many engineering materials particularly during the hot working of metals falls under this category. Beyond 0.7 T_mK, i.e. at very high temperature, the creep is primarily influenced by diffusion and load (stress) applied has little effect. This creep is termed as diffusion creep or plastic creep.

Factors Affecting Creep

Creep is much more affected by grain size, microstructure and previous strain history, e.g. cold work, etc. and creep is also extremely structure sensitive.

The major factor in creep is grain size. Usually, coarse-grained materials exhibit better creep resistance than fine-grained ones, since the latter have a great amount of grain boundary materials and grain boundaries behave as a quasi-viscous material with a high tendency to flow at elevated temperatures. This is why that single crystals show a higher creep resistance than polycrystalline materials. For example, tungsten filaments of electrical bulbs are made of single crystals and that is why they can withstand very high temperatures.

The thermal stability of the microstructure of alloys and its resistance to oxidation at high temperatures is another important factor. An annealed specimen for having greater thermal stabilities is far superior in its creep resistance to a quenched steel for its poor thermal stability.

It is reported that prior cold working has also a strong effect on creep which is accelerated in some metals during recrystallisation following cold work. This accelerated creep is followed by a period of relatively slow creep in the recrystallised metal.

Pure metals with high melting points and a compact atomic structure generally exhibit good creep resistance at high temperatures. By alloying the pure metals with suitable elements, the creep resistance can be increased considerably.

The manufacturing method/process of a steel also influences its creep resistance. From the data available so far, the electric-arc-steel shows better creep resistance than that produced by the open-hearth process, whereas induction-furnace steel shows better results than those shown by electric-arc furnace.

Mechanism of Creep

There are many mechanisms of creep which have been proposed. We have already remarked that creep is a thermally activated process. Some significant mechanisms that play vital roles during the creep process are:

(i) Dislocation climb *(ii)* Vacancy Diffusion and *(iii)* Grain Boundary Sliding

An appreciable atomic movement at high temperatures causes the dislocations to climb up or down. In response to the applied stress, the diffusion rate of vacancy may produce a motion, by a simple climb of edge dislocation. Obviously, edge dislocations are piled up by the obstacles in the glide plane and the rate of creep is governed by the rate of escape of dislocations past obstacles. Random thermal vibrations help such atomic movement.

Another mechanism of creep is called diffusional creep or diffusion of vacancies. The diffusion of vacancies controls the creep rate, but the mechanism does not involve the climb of edge dislocations. In response to the applied stress, the vacancies move from surfaces of the specimen transverse to the stress axis to the surfaces that are parallel to the stress axis (Fig. 8.21(b)). Over a period of time, this movement would elongate the specimen in the direction of the stress axis and contract it in the transverse direction resulting in creep.

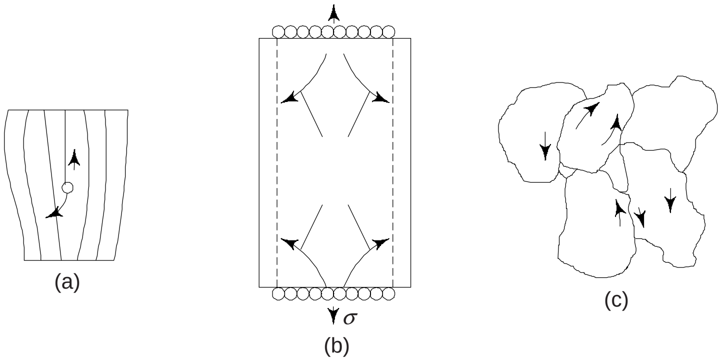


Fig. 8.21 Mechanism of creep

The third mechanism of creep is the sliding of grain boundaries, i.e. sliding of *neighbouring grains* with respect to the boundary that separates them (Fig. 8.21(c)). As compared to individual grains, grain boundaries lose their strength at a lower temperature. This effect arises from the non-crystalline structure of the grain boundaries and grain boundaries play a major role in the creep of polycrystals at high temperatures as they slide past each other or create vacancies. At higher temperatures, ductile metals begin to lose their ability to strain-harden and become viscous to facilitate the sliding of grain boundaries. We may note that at temperatures above 0.5 T_m, the viscosity of the grain boundaries is small enough for them to behave like a very viscous liquid separating the neighbouring grains and allowing them to slide against each other. At low temperatures, grain boundaries do not flow viscously, but provide effective obstacles to dislocation motion. Obviously, the grain boundaries facilitate the deformation process by sliding at high temperatures, whereas at low temperatures, they increase the yield strength by stopping the dislocations.

Creep Resistance Materials

At elevated temperatures, materials used for machine parts and structural components must be creep resistant. Materials used for turbine blades must be creep resistant, i.e., they should not deform by creep during service. Creep can result in seizing the blades with the turbine casing.

The materials to be used at high temperatures must have a high melting point. The creep becomes significant at temperatures greater than 0.4 T_m. Refractory oxides, e.g. MgO and Al_2O_3 have high melting points and, as such, are most suitable for use at high temperature. Where only compressive stresses are encountered, their brittleness limits their use to applications. The discovery of tougher ceramics, e.g. Si_3N_4 (silicon nitride) indicate that selected parts of a heat engine such as piston rings and cylinder heads can be produced from ceramic materials.

Under more versatile conditions, metals and alloys can be used. Most creep resistance alloys consist of a base metal of a fairly high melting point. In aircraft industry, aluminium alloys are mostly used. Aluminium alloy containing 4% copper, 2% nickel and 1.5% magnesium is quite useful group of creep-resisting light alloys called as *Y*-alloys. To develop creep resistance alloys there should be *(i)* increased resistance of grains and grain boundaries to flow and *(ii)* minimum softening of recovery effects.

For better mechanical properties in a low temperature application, where creep is not important, a fine grained material is desirable. However, for high temperature applications, fine grained materials are to be avoided, as grain boundary sliding can add to creep deformation. We must note that grain boundary sliding is not a problem in a single crystal as there are no grain boundaries in a single crystal. A less expensive way of minimizing grain boundary sliding is to orient the grains in such a way that the boundary sliding does not result in cavities at grain junctions. These cavities, if present, can coalesce together and cause *creep fracture.* It has been observed that creep will not occur to any measurable extent below a certain temperature and this temperature is different for different materials, as mentioned below:

Metal	Temperature (°C) (at which creep begins)
Copper	180
Lead	–70
Aluminium	40
Pure iron	330
Tungsten	940

Some of the creep resistant materials are:

(i) Low alloy ferric steels: Iron containing 4% Cr, M and V, deriving their creep resistance from carbide precipitates, good upto 650°C.

(ii) High alloy ferrite steels: 304, 316, 321 stainless steel-iron containing a solid solution (mainly Ni and Cr) and precipitates, good upto 600°C.

(iii) Ni based super alloys: Alloys of nickel containing a solid solution (mainly Cr, W and Co) and precipitates good upto 950°C.

(iv) Refractory oxides and carbides: Like alumina (Al_2O_3), glass ceramics based on SiO_2, silicon carbide (SiC), silicon nitride (Si_3N_4) and sialon-alloys of Si_3N_4 and Al_2O_3 good upto 1300°C.

Materials used in petrochemical, fertilizer industry for furnace tubes. High temperature reforming alloys are used for high temperature service. Percentage composition for some of these alloys is given below in Table 8.4.

Table 8.4 Percentage composition of reformer tube

Alloy	*C*	*Cr*	*Ni*	*Nb*	*Remarks*
HK 40	0.4	25	20	—	—
IN 519	0.3	25	24	1.5 max	—
HP Mod	0.4	25	35	1.5 max	—
H 39 WM	0.4	25	35	0.7-1.5	
SH 24T	0.4	25	24	1	Titanium and Zirconium (additional)
Manaurite XM	0.4	23-27	32-35	1.5	Zirconium (additional)

Creep Fracture

(i) It may take place at higher temperatures or longer stress times, e.g. in ductile materials. Fracture may take place after a nearly uniform deformation without forming a neck and this behaviour is usually called as thermal action.

(ii) Under moderate temperature and high stresses it takes shorter times. The nature of this type of fracture is exactly similar to that of ductile material forming a neck before fracture.

(iii) For long time durations at low stresses, fracture is similar to brittle fracture because of negligible deformation produced in the process.

(iv) Metals sometimes fracture with very little plastic elongation at high temperatures or after long periods of loading. Grain boundary shearing becomes important under these conditions. Because of their irregular shape, the movement of whole grains relative to each other causes cracks to open between them because of their irregular shape. When one crack becomes larger it spreads slowly across the member until fracture takes place, which is brittle in character.

Creep Test

This requires the measurement of four variables: stress, strain, temperature and time. The objective of this testing is to determine the continuing change in the deformation of materials at elevated temperatures when stresses below the yield point.

A creep specimen is prepared, which is usually the same as the specimen of tensile testing. At one end of the specimen a platinum wire is spot welded and to the other end a platinum tube is spot welded. The platinum wire-slides inside the tube and the reference marks are put on the both, wire and tube, which we may observe through the telescope at the middle. There is a scale inside the telescope and elongation can be measured.

The specimen is loaded in the furnace along the pullies (Fig. 8.22). With the help of the wire, the specimen is made to fix on the ground of the furnace. As soon as the load is applied, the specimen come under the tension. To measure the temperature, thermocouples are fixed along both the elongated ends of the specimen.

First, a load is applied and initial or instantaneous elongation is measured with the help of telescope. For a constant load, all that is needed is a dead weight and a system of levers to multiply it to the required load. The temperature of the furnace is allowed to increase upto a specified temperature. With the help of the

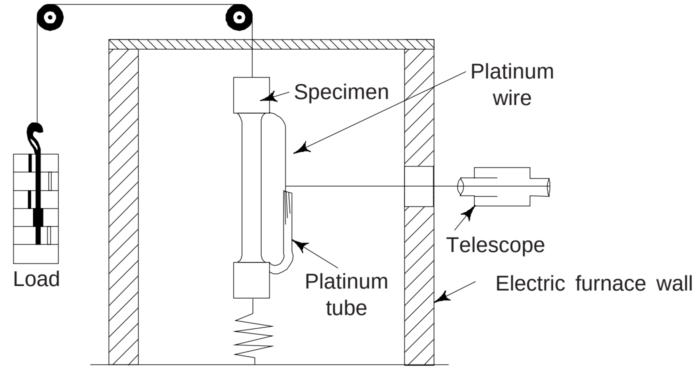


Fig. 8.22 Creep test

telescope, the increase in length of the specimen is measured along the platinum wire and tube. For testing temperature upto 600°C, temperature must be controlled within 3°C. For the testing temperature within the range 600-800°C, the accuracy of ± 4°C must be maintained whereas between 800-1000°C, the accuracy must be ± 6°C.

The strain in the specimen is measured by a strain gauge or by an optical extensiometer. At each temperature under different loads (stresses), four or five specimens are tested. For each specimen, elongation versus time curves are plotted.

The aim of the creep test to determine limiting creep stress. The limiting creep stress is that stress which will just not break the specimen with applied stress for an infinite period of time at a specific constant temperature. Creep test takes very long time, i.e. about 10,000 hours.

Creep test results are of use in the design of machine parts, furnace parts, boilers, internal combustion engine, turbine blades which are exposed to elevated temperatures.

STRESS-RUPTURE TEST

In stress-rupture test, creep test is extended upto failure. One can calculate the following from this test:

(i) time required to cause failure at a given stress and at constant temperature, and

(ii) minimum creep rate by measuring elongation as a function of time.

In order to have higher creep rates, the test is carried at higher loads. We must remember that in creep tests, lesser load is applied to avoid stage III. Usually creep tests are carried out up to 1000 hours only. In creep test, the strain is lesser, i.e. upto 0.5% while in stress rupture test the strain is about 50%. Stresses are high in stress rupture test and thus the structural changes take place in a short time. This test is used to test the strength of an alloy material for jet engine applications.

The graph between rupture time and stress is plotted (Fig. 8.23). We note that the graph is mostly a straight line. The change is due to structural changes taking place in the material, i.e. due to recrystallization, grain growth, graphitization, oxidation, spherodation, etc.

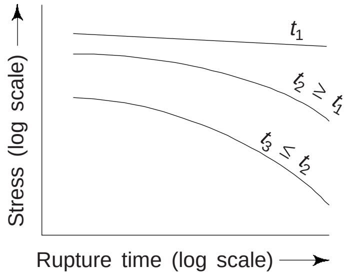


Fig. 8.23 Stress vs rupture time

Applicability of Creep Tests

The only satisfactory guide to the performance of metals for high temperature service appear to be the creep tests at high temperatures. The creep tests are inherently long time tests, but the test periods may be short in comparison with the actual service periods of the structure, therefore, extrapolation of creep test data has to be made with judgement. Although creep tests require too much time to be used as acceptance tests, but they are the basis of data to be used in design.

To obtain creep resistance at high temperatures, one should observe the following general rules:

(i) Metal with a high melting point be used

(ii) Use FCC metals with low stacking fault energy. For metal with same melting point, the FCC structure is usually superior to BCC structure in high temperature applications.

(iii) Long range order structure be used. Its behaviour is similar to extended dislocations in FCC lattices.

(iv) Stable precipitants should be used. This will prevent dislocations from excessive gliding by blocking the glide planes.

(v) Use of those elements be made which precipitate during service time.

(vi) One of the factors affecting the extent of creep is the grain size. Generally coarse grained structure reveals a better creep resistance than fine grained structure at elevated temperature.

Creep Under the Action of Several Mechanisms

It is very rare that creep deformation takes place only by one mechanism only. Usually creep deformation takes place when more than one mechanism is operating. Various different mechanisms under certain conditions control the creep rate. For simplicity purpose, let us assume that creep in a metal is dependent only on two mechanisms and net creep rate varies with temperature under constant stress. Let the following two equations represents the creep rate of two simple mechanisms

$$\dot{\varepsilon}_1 = A_1 \exp [-x_1/\text{kT}]$$
$$\dot{\varepsilon}_2 = A_2 \exp [-x_2/\text{kT}]$$

where A_1 and A_2 are constants at constant stress, ε_1 and ε_2 are creep rates and x_1 and x_2 are activation energies.

Two creep mechanisms, which are acting simultaneously at one time, can be described as *(i)* they are independent of each other and act in parallel and *(ii)* they act in series, the first mechanism operates first and second mechanism afterwards. The net strain rate $\dot{\varepsilon}_n$ in the first case is the sum of the individual strain rates, i.e.

$$\dot{\varepsilon}_n = \dot{\varepsilon}_1 + \dot{\varepsilon}_2 \tag{20}$$

where ε_1 and ε_2 are creep rates respectively for first and second mechanisms. Now, the net creep rate in the second case is equal to the sum of total strains divided by total time needed for both mechanisms to function, i.e.

$$\dot{\varepsilon}_n = \frac{\varepsilon_1 + \varepsilon_2}{t_1 + t_2} \tag{21}$$

where ε_1 and ε_2 are the respective strains corresponding to respective mechanisms, t_1 and t_2 are time intervals between operations of respective mechanisms. Equation (21) clearly reveals that when creep mechanisms are in series, the one with slower rate or longer delay period controls the creep rate, whereas Eq. (1) shows that when mechanisms work in parallel the faster rate controls the creep rate.

6. FACTORS AFFECTING MECHANICAL PROPERTIES

The following factors affect the mechanical properties of a material:

(a) Grain Size This has an important effect on the mechanical properties of metals. In comparison to coarse grained metals, fine grained steels offer better resistance to cracking, produce fine surface finishes and can be shaped better. Fine grained steels have higher strength but lower ductility and therefore, fine grain is specified for structural applications (Fig. 8.24). Coarse grained steels have higher hardenability and better forging properties.

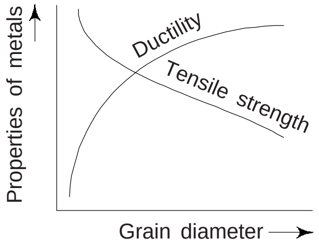


Fig. 8.24 **Effect of grain size on ductility and tensile strength of metals**

(b) Temperature It has been observed that the temperature affects the tenacity, elastic limit, Young's modulus and other properties of steel and wrought iron as follows:

(i) With rise in temperature, yield stress, toughness and ultimate tensile strength decreases, whereas Young's modulus decreases.
(ii) At higher temperatures (~ 450°C) embrittlement of steel and creep occurs. It increases with time and thereby reducing the toughness.
(iii) With rise in temperature stiffness and fracture stress of many metals decrease.
(iv) Several metals and large number of alloys exhibit the phenomenon of superconductivity (infinite conductivity and perfect diamagnetism) near absolute zero temperature.
(v) For unnotched mild steels, the elongation and reduction in cross-sectional area is satisfactory upto 130°C. It goes down considerably below that temperature. Elongation rises with further rise in temperature.
(vi) Below 100°C, non-ferrous metals show better properties than ferrous metals.
(vii) With the rise in temperature, there is also rise in thermal vibrations of atoms causing changes in structural properties.

(c) Exposure to Atmosphere When exposed to atmosphere, most metals get oxidised. Due to oxidation of metal surface, a film is formed. The presence of chloride in marine atmosphere increases the conductivity of the liquid layer in contact with the metal surface. The resistivity of metals decreases due to the presence of moisture, sulphur dioxide, hydrogen sulphide and other corrosive environments. Corrosive action is pronounced and varies with the humidity of the air. Obviously, care should be taken while selecting the materials. When exposed to atmospheric conditions, metals like chromium, S.S., aluminium and nickel show exceptional resistance.

(d) Heat Treatment By heat treatment, one can improve mechanical properties like tensile strength, hardness, toughness, and wear resistance. Heat treatment is generally carried out to

(i) Improve machinability
(ii) Improve resistance to corrosion
(iii) Improve chemical, magnetic, electrical and thermal properties
(iv) Improve mechanical properties, e.g. ductility, hardens tensile strength, shock resistance etc.
(v) Relieve internal stresses induced during cold and hot working
(vi) Modification of the structure and grain size

(e) Carbon Content Compared to the other ferrous and non-ferrous materials, carbon steels are widely used engineering materials. There are many types of carbon steels, each designed by the metallurgist, to have particular properties required by the engineer. Carbon steel is basically an alloy of iron and carbon containing manganese (1.0% max); silicon (0.5% max); sulphur (0.04% max) and phosphorus (0.05% max). The last two elements are introduced as impurities by the raw materials used during manufacture of steel and kept as minimum as possible. Silicon is residue from the steel making process. Carbon and manganese have a significant influence on the mechanical properties and uses of steel. Carbon steel is widely used due to following reasons:

(i) it is comparatively cheap.
(ii) its properties can be improved further by heat treatment.
(iii) it has wide range of mechanical properties, which can be further improved by the addition of alloying elements and proper control of carbon content.
(iv) it is readily available in different size and shapes.
(v) one can easily shape into any form by bending, drawing or forging.
(vi) one can machined and weld it easily.
(vii) it consumes less energy during production.
(viii) one can remelt and recycle it again. About 40% of todays production of carbon-steel is by recycling.

7. MECHANICAL TESTS

Engineering materials are tested by engineers to ensure that the materials supplied are sound and safe and also for the following reasons:

(i) to check chemical composition
(ii) to determine suitability of material for particular applications
(iii) to assess numerically, the fundamental mechanical properties, e.g. ductility, malleability, toughness, etc.
(iv) to determine the data such as stress values, to set up specification upon which engineer can base his design
(v) to determine surface and sub-surface defects in raw materials or processed parts.

The following tests are performed on engineering materials for the determination of their mechanical properties:

(i) Tensile test *(ii)* Impact test *(iii)* Hardness test *(iv)* Fatigue test *(v)* Creep test and *(vi)* torsion test

Mechanical tests may be grouped into following two classes:

(i) Destructive Tests: In this type of testing, the specimen or the component is destroyed and cannot be reused. The tests are conducted on similar specimens and under the same conditions. Tensile test, hardness test, impact test, fatigue test, creep test, etc. are examples of destructive test.

(ii) Non-Destructive Tests: The specimen is not destroyed and can be used after the test. Examples of non-destructive tests are the magnetic dust method, penetrating liquid method, ultrasonic test and radiography. All these non-destructive tests are used to detect various types of flaws on the surface or internal inclusions of impurities.

(a) Destructive Test

(i) Tensile Test: This test is used to find out the effect of shearing forces on the specimen which is subjected to a twisting couple and one of the widely used mechanical tests. This test helps to determine tensile properties, e.g. limit of proportionality, elastic limit, yield point or yield strength, maximum tensile strength, breaking strength, percentage elongation, percentage reduction in area and modulus of elasticity. The tensile test involves an axial load being applied to specimens of circular or rectangular cross-section (Fig. 8.25). Typical dimensions of the specimen as per BS-18 are summarized in Table 8.5.

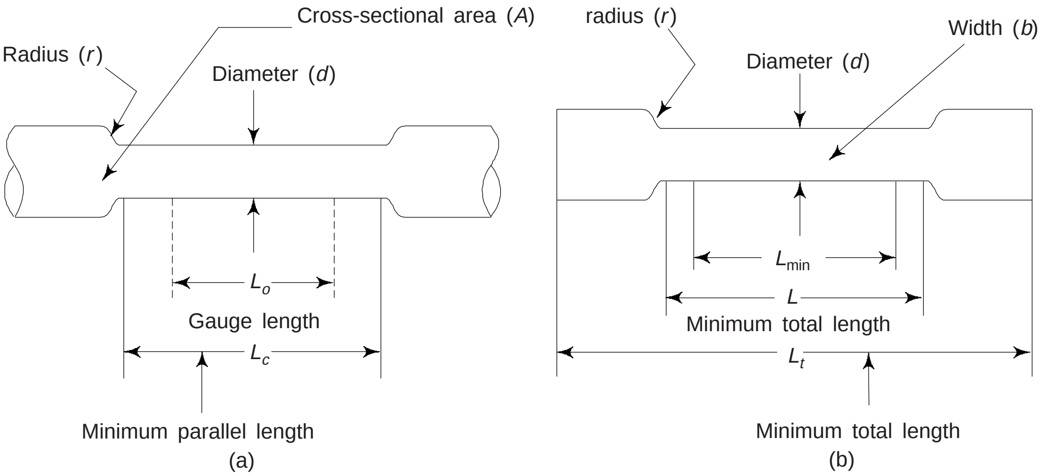


Fig. 8.25 Standard tensile test specimen: (a) round (b) rectangular

Table 8.5 Typical dimensions for tensile test specimens as per BS-18

Specimen	d	b	L_o	L_c	L_t	r
Round	15.96	—	80	88	—	15
Round	11.28	—	56	62	—	10
Flat	—	25	100	125	300	25
Flat	—	12.5	50	63	200	25

Here L_o → Width of test piece, L_o → Gauge length,
L_t → Minimum total length, L_c → Minimum parallel length
d → Diameter of specimen, r → Radius at shoulder

Generally the gauge length (L_o) is related to the cross-sectional area A by the following relation

$$L_o = 5.65\sqrt{A}$$

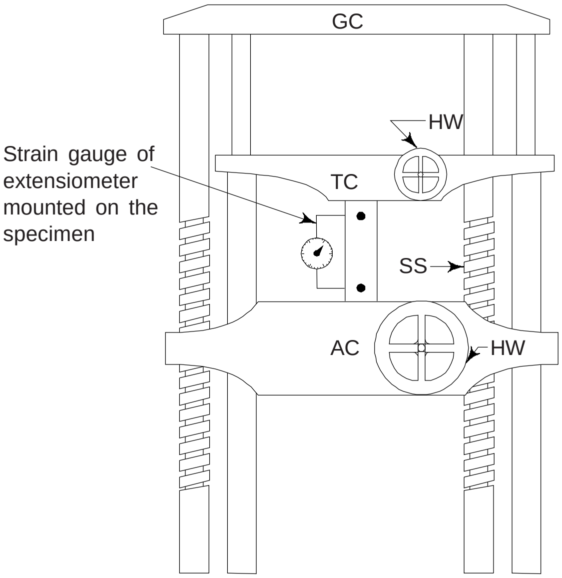
With circular cross-section, the relationship works out to be L_o = 5 d (approximately). The value of ultimate tensile strength is affected by the following: *(i)* quality of surface finish *(ii)* speed of testing and *(iii)* dimensional accuracy of the specimen.

Obviously, it is of prime importance to follow the above factors closely while conducting tensile tests. The dimensions of a standardised specimen can be

Diameter (d) = 12 mm
Gauge length = 5 d for short test pieces and
10 d for long specimens (d-diameter)
End fillet radius = 3 mm
Parallel length = 5.5 d

One end of the specimen is fastened to the frame of the machine by means of grips or clamps, while the other end is similarly fixed to the movable cross head (Fig. 8.26). A steadily increasing load is applied to the specimen usually hydraulic load is applied. The magnitude of the load is measured by the load measuring unit. By attaching an extensiometer or gauge to the specimen, elongation is measured. The oil is supplied by the rotating piston pump to hydraulic pump. The oil causes the main piston to rise upwards. This results in upward movement of movable crosshead, thus increasing the length of the specimen, i.e., deforming the specimen. The tensile test thus gives elongation as a function of load. One can calculate stress and strain from load elongation data and a graph of stress versus strain can be made of the material of the specimen. There are several machines which can automatically record stress strain curves for materials under tests.

The different values of load and elongation at different intervals are recorded during the test. To prevent any damage to the extensiometer, the extensiometer should be removed from the gauge length of the specimen just before the pointer on the load scale in the load-measuring unit of the testing



GC—Guide cross head
HW—Hand wheel
AC—Adjustable cross head
TC—Tension cross head
SS—Square screw

Fig. 8.26 Tensile Testing Machine showing the specimen in tension

machine stops moving in the forward direction. This stage or point is indicative of yield point where the internal structure of the specimen starts yielding. On further increasing the load and the maximum travel of the load-scale pointer will indicate the *ultimate load*. After this stage the pointer moves in the reverse direction to indicate necking. The pointer finally stops at a point with a noise to indicate fracture.

After fracture the two pieces of the broken specimen are placed as shown in Fig. 8.27, as if they are fixed together and the distance between the two gauge marks is measured. This is the final length of the specimen. Similarly, at the place of fracture, the average diameter is calculated and the area of fracture is calculated. One can easily calculate the various tensile properties from these measurements and other observations from the tensile test data as follows:

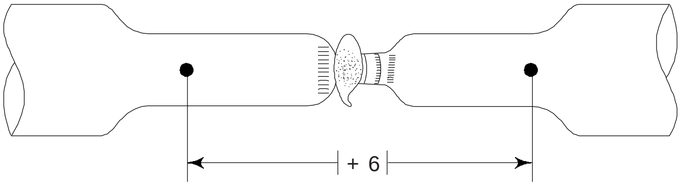


Fig. 8.27 After fracture, increased length and reduced diameter

Elastic Limit: The highest stress that the specimen can withstand without experiencing a permanent strain on removal of the load. This is given by

$$\text{Elastic limit} = \frac{\text{Load within elastic limit}}{\text{Original area of specimen}}$$

Proportional Limit: The stress at which the stress vs. strain graph deviates from the straight line law to a curve is called proportional limit.

Yield Strength or Stress: The stress at which the specimen starts yielding without any increase in load is called yield strength or stress, i.e.

$$\text{Yield Strength} = \frac{\text{Load at yield point}}{\text{Original area of the specimen at gauge length}}$$

Ultimate Tensile Strength or Stress

The maximum travel of the pointer on the dial of the load-measure range divided by the original area of the specimen is called ultimate tensile stress, i.e.

$$\text{Ultimate tensile stress} = \frac{\text{Ultimate load}}{\text{Original area of the specimen}}$$

Percentage elongation: It is given by

$$\text{Percentage elongation} = \frac{\text{Final length} - \text{Original length}}{\text{Original length}} \times 100$$

Percentage reduction in area: It is given by

$$\text{\% reduction in area} = \frac{\text{Original area} - \text{Final area}}{\text{Original area}} \times 100$$

Young's Modulus of elasticity (E or Y)

$$E \text{ or } Y = \frac{\text{Stress at any point within elastic limit}}{\text{Strain at that point}}$$

Breaking stress: The stress at the breaking point of the specimen at the tensile test is called breaking stress

$$\text{Breaking stress} = \frac{\text{Breaking load}}{\text{Original area}}$$

(ii) Hardness Test: This is essentially the resistance of the surface of a material to deformation. We may note that there is no absolute scale for hardness. The commonly used tests for hardness are:

(a) Brinell hardness test
(b) Vickers hardness test
(c) Rockwell hardness test
(d) Rebound hardness test
(e) Scratch test

Hardness tests are commonly used in industry and research as they provide an easy and reliable way for evaluating the effect of hot and cold working processes and heat treatments, upon the basic properties of metals. We may note that while performing the hardness tests plastic deformation is produced in the specimen. The factors which affect plastic deformation also affect the hardness tests. The information conveyed by a hardness test depends upon the definition of hardness upon which the test is based. Some of these are listed below:

(i) resistance to permanent indentation under static or dynamic loads-*indentation hardness*,
(ii) energy absorption under *impact loads-rebound hardness*,
(iii) resistance to *scratching-scratch hardness*,
(iv) resistance to abrasion-*wear hardness*,
(v) resistance to cutting or drilling *hardness* concerning machinability.

Therse are about 30 methods of testing hardness, we will discuss 5 most important methods stated above.

Brinell Hardness Test

This test is used: *(i)* to determine hardness of metallic materials, *(ii)* to check the quality of the product, *(iii)* for uniformity of samples of metals, and *(iv)* for uniformity of results of heat treatment. We may note that this test is not suitable for very hard or very thin materials.

In this test a standard hardened steel ball of diameter D is pressed for about 10-15 seconds into the surface of the specimen by a gradually applied load P (Fig. 8.28). The diameter of the indentation is measured after the load and ball are removed. The impression of the steel ball (indenter) so obtained, is measured to nearest 0.2 mm with the help of microscope.

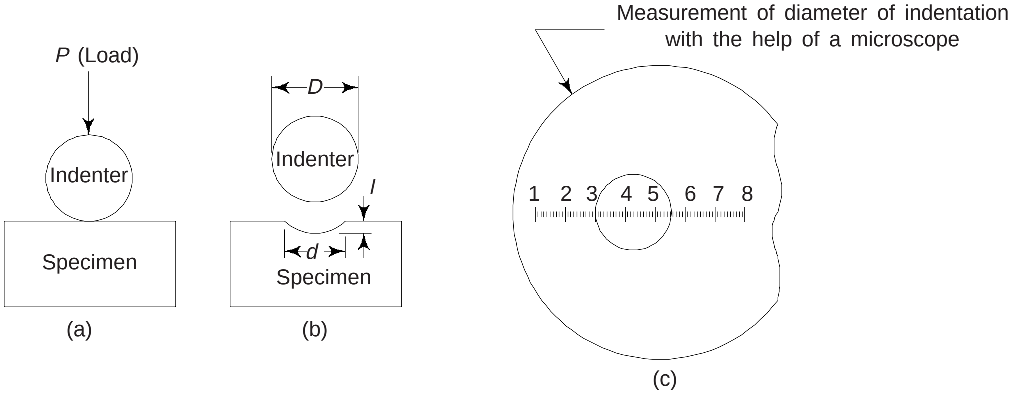


Fig. 8.28 Brinell hardness test

For curved surfaces, where the radius of curvature is not less than 25 mm, the diameter of the impression is taken as the average of maximum and minimum diameter. The Brinell Hardness Number (BHN) is obtained by dividing the applied force by the spherical surface area of the indentation, i.e.

$$\text{BHN} = \frac{\text{Applied load in kg}}{\text{Area of impression or indentation of steel ball in m}^2}$$

$$= \frac{P}{A}$$

where A is area of contact between ball and indentation in $(\text{mm})^2$. The surface contact is a spherical segment, whose area is, $A = \pi Dt$, where $t \to$ depth of impression, $d \to$ diameter of impression measured with a microscope, and $D \to$ diameter of ball. From Fig. 8.29, we have

$$OB = \sqrt{OA^2 - AB^2}$$

$$= \sqrt{\left(\frac{D}{2}\right)^2 - \left(\frac{d}{2}\right)^2}$$

Also $$t = OB = OC - OB = \frac{D}{2} - \sqrt{\left(\frac{D}{2}\right)^2 - \left(\frac{d}{2}\right)^2}$$

$$= \frac{1}{2}\left(D - \sqrt{D^2 - d^2}\right)$$

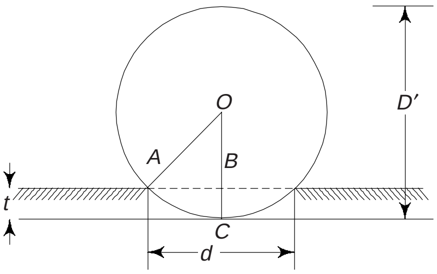


Fig. 8.29 Ball indenter

Now, $$A = \pi Dt = \frac{\pi D}{2}\left[D - \sqrt{D^2 - d^2}\right]$$

$\therefore$ $$\text{BHN} = \frac{2P}{\pi D\left[D - \sqrt{D^2 - d^2}\right]} \tag{22}$$

The BHN number is expressed in kilogram per square meter. The BHN number for some metals are:

BHN number	*Metal*
Lead	4-8
Zinc	25-40
Cast iron	115-200
Nickel	130-160
H.S.S.	450-700

The diameter (D) of the ball, 1, 2, 5 or 10 mm (generally it is 10 mm) and the applied load (P) are chosen, such that the same value of P/D 1, 5, 10 or 30 will give the same hardness number. The thickness of the material being tested should be at least 10 times the depth of the indentation.

Steel ball used in Brinell hardness testing should be polished and free from surface defects. For testing steel and cast iron, 3000 kg load is applied, whereas a load of 500 kg is used for brass and soft alloys. The thickness of the specimen should be 10 × depth of indentation. The duration of time for the Brinell hardness test is between 10 and 30 seconds, which depends upon the metal being tested. Diameter of indentation should be measured in two directions normal to each other, under microscope. The mean of these two diameters is taken as diameter of indentation (d). Tungsten carbide balls should be used, so that there should not be any permanent change in ball diameter.

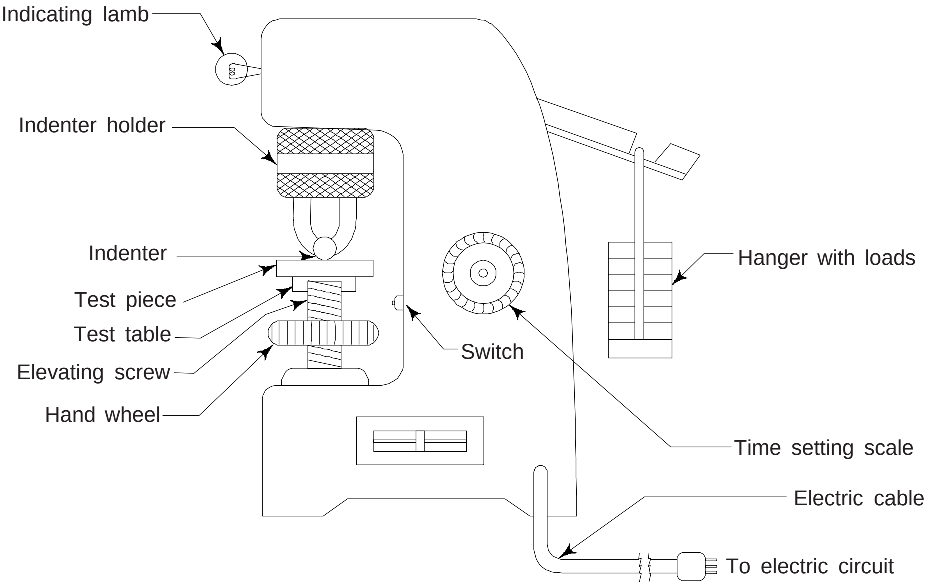


Fig. 8.30 Brinell hardness testing machine

Figure 8.30 shows a Brinell hardness tester. The procedure for testing the hardness of a sample is as follows:

(i) Place the test sample on the top of the test table and raise it with the elevating screw, till the test sample just touches the ball.
(ii) Apply the desired load (about 30 × diameter of the ball in mm) either mechanically or by oil pressure.
(iii) The steel ball during this period moves to the position of the sample and makes an impression or indentation.
(iv) Measure the indentation diameter at two places, either on the screen provided with the machine or by coinciding the two points of a reading microscope (Fig. 8.28(c)).
(v) Using relation (22), one can calculate BHN after substituting the values of P, D and d.

To some extent BHN is an indication of the tensile strength of the metal. One can correlate them by multiplying the BHN by a constant which depends upon the character of the metal. The limitations of this method are:

(i) the test machine is very heavy.
(ii) the area of indentation is quite large that it affects the surface quality. This is why, sometimes it is considered as a destructive test.
(iii) the thickness of the test sample also limits its use, e.g. thin sheets will bulge or be destroyed during the test.
(iv) for very hard materials, the test results are unreliable. The ball gets flattened on hard surfaces.
(v) one faces difficulty in measuring the indentation diameter accurately.

This test is widely used to examine the effects of heat treatment and cold working in order to control the heat treatment process.

Vickers' Hardness Test

This is similar to Brinell Hardness test, but in this method the drawback of the flattening of the steel ball in testing harder materials is eliminated. It uses a similar relationship and most of the errors and limitations of Brinell hardness test are eliminated. In this method of hardness testing a diamond square based pyramid indentor with 136° angle between opposite faces is used. The load varies from 5 kg to 120 kg in increments of 5 kg. Similar to Brinell and Rockwell hardness measuring methods, this method also uses the indentation produced by the indentor (diamond pyramid). The indentor gives geometrically similar impressions under different loads.

A piston and a dash pot of oil is used for controlling the rate and duration of loading. The specimen is placed on the anvil, which is then raised to indentor. Load is applied and then removed. The value of Vicker's hardness can be obtained by the following relation

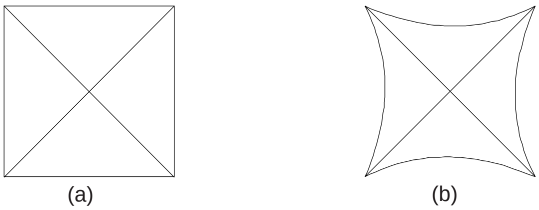
$$\text{VPN} = \text{DPN} = \frac{\text{Load applied (in kg)}}{\text{Area of indentation (in mm}^2\text{)}}$$

where VPN = Vicker's Pyramid number and DPN = Diamond Pyramid number
Let P is the load applied, d is the average length of two diagonals, in mm, and θ is the angle between opposite faces of diamond pyramid (136°). Then

$$\text{VPN} = \text{DPN} = \frac{2P \sin \theta/2}{d^2} \tag{23}$$

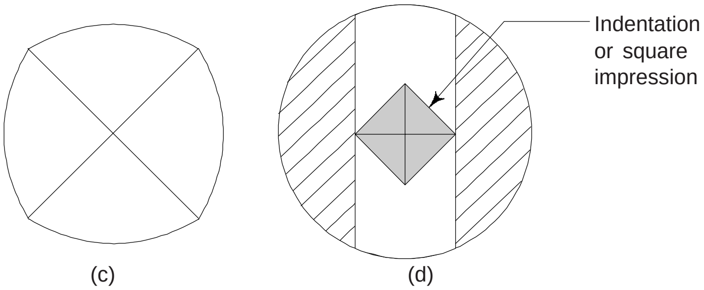
This method is used for the determination of hardness of very thin and very hard materials. This method also facilitates the ease of measurement of a diagonal of the indentation area (Fig. 8.31), as compared to circular dimensions, which are difficult to measure. This method is rapid, accurate and suitable for metals as thin as 0.15 mm. The indentor is capable of giving geometrically similar impression with different loads. Obviously, the hardness number is independent of the load applied. Some typical values of VPN are:

Material	*VPN Value (Kg/mm^2)*
Diamond	8400
Steel	210
Aluminium	22
Lead	12
Tungsten carbide	2100



(a) Perfect indentation

(b) Pin-cushion indentation due to sinking



(c) Barreled indentation due to ridging

(d) Setting of microscope on indentation area

Fig. 8.31 Diamond pyramid indentations

The values of Brinell and Vickers hardness are practically the same upto 300. We must note that Vicker's test can be carried out accurately on polished surfaces but does not give accurate results when used for rough surfaces.

Knoop's Hardness Test

This method is applicable to extremely thin metals, plated surfaces, very hard and brittle materials, carburised and nitrided surfaces and whenever applied load must be kept below 3600 gm.

Load is applied by a diamond pyramid indentor with long but short in depth diagonal in ratio 7 : 1 (shallow indentation), longitudinal angle 172° - 30′, traverse angle is 130°. In this test, the diamond pyramid is indented on the specimen and length of diagonal is measured under microscope. The Knoop hardness number (KHN) is calculated from the following relation:

$$\text{KHN} = \frac{\text{Applied load (in kg)}}{\text{Uncovered projected area}}$$

$$= \frac{P}{A} = \frac{P}{L^2 C}$$

where P is applied load in kg, A is uncovered projected area of indentation in mm^2, L is the length of long diagonal in mm and C is a constant related to the length of projected area for each indentor supplied by manufacture.

One can calculate the Knoop hardness from the projected area of the impression or indentation rather than the area of contact as in the case of indentation tests.

We must note that micro-hardness require extra care in all stages of testing. Good polishing of the surface is required, i.e. the surface of the specimen must be prepared carefully.

Rockwell Hardness Test

This is very widely used test because the testing machine is simple to operate and does not require microscope for measuring indentation diameter. This method is free from personal errors and also faster. Surface polishing of the specimen is also not required. As compared to Brinell hardness testers, this test requires much smaller penetrators (steel balls as well as cone-shaped diamond penetrators) and loads.

In this test, the dial gives a direct reading of hardness. There are two scales on a Rockwell testing machine: B-scale and C-scale. In general, there are nine scales on the dial A to H and K, but B-scale and C-scale are most widely used. The B-scale is used to record the hardness of soft metals and its range is R_B 0 to R_B 100. Steel ball indentors are commonly used for the B-scale. To test materials of hardness greater than R_B 100, e.g. cast irons, C-scale is used. For measuring hardness, a diamond-cone indentor is used. The usual range of this machine is R_C 20 to R_C 70. The working of machine is as follows:

(i) The test specimen is placed on the machine's test table.

(ii) The test piece is raised up against the steel ball indentor till the needle on the dial reads zero. This is achieved by applying minor load of 10 kg on the test sample by the machine and ensuring that the specimen is seated properly.

(iii) Now, major load of 100 kg (for B-scale) or 150 kg (for C-scale) is applied by pressing the crank provided for this purpose.

(iv) Withdraw the major load and lower the test piece.

(v) The permanent increase in depth penetration '*e*' is measured under specific conditions. The measurement unit for '*e*' is 0.002 mm from which a number, known as Rockwell hardness number is defined.

1.58 mm ball is used for soft materials and diamond cone indentor is used for hard materials. The angle of cone is 120° and radius of spherical tip is 0.2 mm.

$$\text{Rockwell B number} = \frac{130 - \text{depth of indentation}}{0.002}$$

$$\text{Rockwell C number} = \frac{100 - \text{depth of indentation}}{0.002}$$

B-scale is standardised and is most widely used. B-scale and C-scale differ by 30 hardness number. The Rockwell hardness is denoted by R.H. For example, 40 RHB indicates Rockwell hardness of 40 measured on B-scale. Similarly 80 RHC indicates Rockwell hardness of 80 measured on C-scale. The following precautions must be taken during testing.

(i) The test piece should be flat and preferably machine polished. It should be clean, dust free and free from oil and scale, etc.

(ii) The thickness of the test piece must be at least ten times the depth of indentation so that a mark is not produced on the other side of the test piece.

(iii) Care should be taken that successive impressions not be superimposed on each other and they are also not close to each other, i.e. distance between them should be at least four times the diameter of the indentation apart.

(iv) The testing machine should be protected from vibrations and dust and for this it should be placed on a rigid support and covered suitably.

(v) The test specimen, work table and indentor should be kept very clean.

Superficial Rockwell Testers

This is a special type of testing machine, in which small loads can be used. The value of the hardness of the specimen obtained is rated the superficial Rockwell hardness number.

Rebound Hardness Test

This testing machine is sometimes also known as shore-scleroscope. In this testing machine, hardness of the test specimen measurements are made by dropping a hard object, a small diamond pointed hammer, on the surface and observing the height of rebound. As it falls from a height, its potential energy is transformed into kinetic energy which is passed on the surface when it strikes. The diamond pointed hammer rebounds due to the elastic strain energy of the surface and damping capacity stiffness of the material. The height of the rebound is considered as the index of hardness. In actual testing procedure a 1/12 Oz hammer is allowed to fall from a height of 25 cm in a glass tube. The glass tube is graduated in 140 equal parts to measure the height of rebound. This testing machine is hard and can be used to test the hardness of sheet rolls.

Moh's Hardness Test or Scratch Hardness Test

This test is based on scratch resistance of the material and mostly applied to minerals and non-metallic elements.

in Moh's hardness scale there are ten degrees or steps, which consists of 10 standard minerals arranged in the order of their ability to be scratched. These hardness numbers with corresponding minerals of the original scale are:

1. Talc, 2. Gypsum, 3. Calcite, 4. Flouspar, 5. Apatite, 6. Orthoclase, 7. Quartz, 8. Topaz, 9. Sapphire or corundum, 10. Diamond.

Diamond, the hardest known material, is assigned the hardness number of 10 Mohs, whereas the softest mineral in the scale, i.e. talc is assigned the hardness number of 1 Mohs. Most hard metals fall in the Mohs hardness range of 4 to 8. If any material is having its hardness number according to this scale as '6' then it means that the material is just in a position to scratch apatite and is just to be scratched by quartz.

Moh's scale for most engineering applications is not quantitative enough to distinguish clearly between metals of similar hardness. Also, the measurement of hardness by scratching is difficult to interpret and standardise. Therefore, new hardness scales based on other methods, e.g. indentation hardness and dynamic or rebound hardness for measuring hardness of metal specimen are of major interest.

Turner's Sclerometer Test

This test is just similar to Moh's scale test. A weighed 90° cone diamond is drawn, once forward, and once backward over the smooth surface of the material to be tested to produce a scratch of 0.1 mm width. The weight in gms on the cone which is just able to produce that scratch is designated as the number of hardness of this test.

Comparison of Various Hardness Tests

While interpreting the hardness results, the surface condition of test specimen is of great significance. In most of the cases the surface is harder than the interior of material which produces a hardness gradient. Surface layers of various thickness in which hardness gradient may be much more developed due to fabricated parts. Hard layers are produced due to case hardening and cold working. When penetration of indentor is small and the gradient is not steep, one can take the hardness number as reference to the hardness of each surface of material. *(i)* The diamond pyramid test gives the best results as the depth can be kept very small by selecting small but proper loads. *(ii)* The hardness number indicates the hardness of inner material, if the penetration of indentor is deep as compared with the thickness of the layer. *(iii)* There will be no reliable indication of hardness either of the surface or of the inner material if the penetration of the indentor is nearly the same.

Impact Tests

The capacity of a material to resist or absorb shock energy before it fractures is called its impact strength. The fracture of a material can be either brittle or ductile. Less energy is absorbed with a brittle fracture than with a ductile fracture. Usually the term toughness is used to describe the ability of a material to withstand shock loads, the tougher the material the more it is able to withstand such loads without breaking. Toughness takes into account both the strength and ductility of material. Several engineering materials have to withstand impact or suddenly applied loads while in service. We may note that impact strengths are lower as compared to strength achieved under slowly applied loads. Impact strength in S.I. units is expressed in Mega Newtons per square metre (MN/m^2).

To study the behaviour of engineering materials under dynamic load, impact testing is done. The impact test can be considered as an indicator of toughness, i.e. it provides the relative toughness of the material.

In all types of impact test, a notched specimen is used, since unnotched specimen do not always reveal the susceptibility (sensitivity) of a metal to brittle structure. Some normally ductile materials have a tendency to behave in a brittle manner in the presence of notches and this is termed as *notch sensitivity*. This property of the material depends on its response to change strain rates, triaxiality and temperature. There are some materials, which are notch sensitive, e.g. carbon, steels and plastics, whereas, there are also other materials which are not notch sensitive, e.g. FCC metals. One can visualize the effect of notch sensitivity in terms of the *ductile-brittle transition* curve of fracture energy versus temperature. On the energy scale of an impact testing machine, fracture energy can be measured in kg-m.

Impact strength is affected by the rate of loading, temperature, and presence of stress raisers in the material. It is also affected by variations in heat treatment, alloy content, sulphur and phosphorus content of the material. Impact strength tests are used considerably in some industries to know shock-absorbing property of the material under the given variations. In impact tests by high velocity loading and the introduction of a notch to create triaxiality and stress concentration, a high strain rate is provided. Impact tests are usually conducted on pendulum impact testing machine. There are two standard test methods: *(i)* Charpy test and *(ii)* Izod test-for notched bar impact testing.

Impact Testing Machine

The standard machine for impact testing is of the pendulum type (Fig. 8.32). A notch is cut in the specimen, i.e. a standard test piece which struck under impact conditions by a heavy weight attached to the end of the pendulum. The test piece, i.e. specimen is held in an anvil (vice) and is broken by a single blow of the pendulum weight or hammer which is allowed to fall from a fixed starting point of a known height. The pendulum swings on after breaking the specimen and the height to which pendulum rises on the other side is measured. Obviously, the energy absorbed in breaking the specimen may be determined and if this energy is low, the specimen is brittle. One can also note the impact energy required to break the specimen from the scale provided on the impact testing machine. When one releases the pendulum from the position of maximum height or maximum energy, the pointer on the scale also moves along with the pendulum and stops at a particular position to exhibit the energy absorbed in breaking and energy still left unutilised. Most test machines are constructed in such a manner that both types of tests can be used with only minor adjustments.

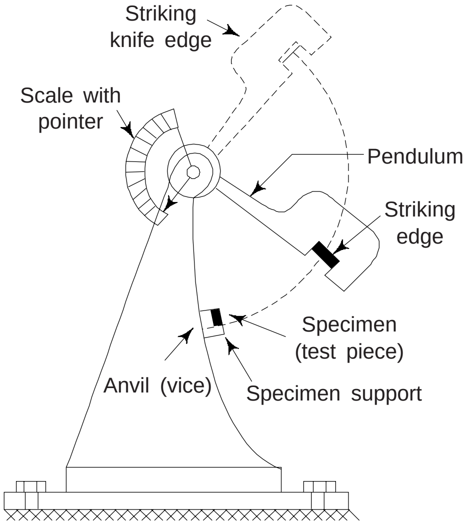


Fig. 8.32 Standard machine for impact testing

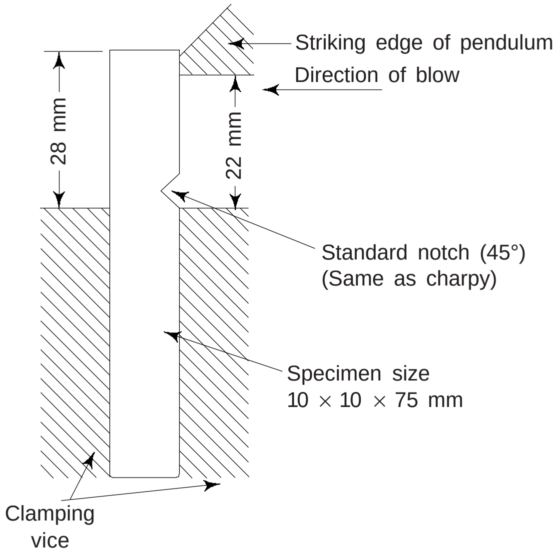


Fig. 8.33 Cantilever beam of Izod test

Izod Test

This test is carried out on a cantilever test piece of 10 mm × 10 mm section and 75 mm length (Fig. 8.33). It has a *V*-notch 2 mm deep and the angle of the notch is 45°. The pendulum hammer strikes near its end. The test piece is placed vertically between the supports. The energy required to rupture the test piece (specimen) can be calculated.

Charpy Test

This is a most common test. In this test, a bar of material to be tested is 10 mm × 10 mm cross section × 55 mm length. It has a V-notch 2 mm deep of 45° included angle and a root radius 0.25 mm. The specimen bar is placed on the supports or anvil as a simply supported beam between the two supports of the machine (Fig. 8.34).

The pendulum, having weight at one end, is raised to certain height, from where it is released. The blow of the pendulum hammer is from a direction opposite to the notch section and rupture the specimen. In its upward swing the pendulum carries the friction pointer over a semi-circular scale graduated in degrees or Kgf-m. The pointer reads the reading of the impact strength. The energy required to rupture the specimen is a function of the angle of rise.

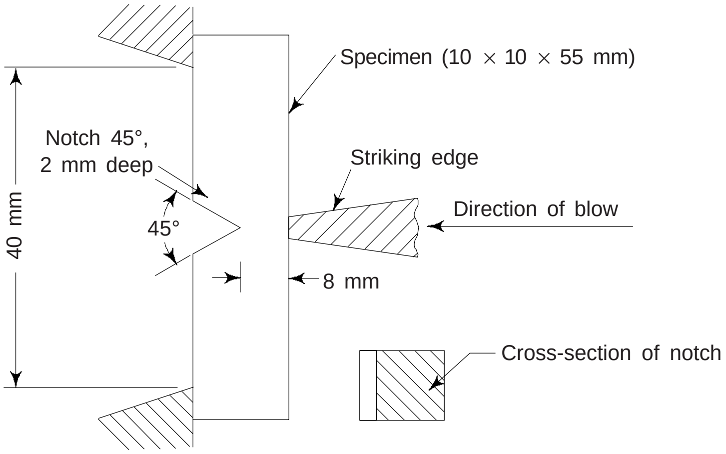


Fig. 8.34 Specimen for Charpy test

The Charpy test has two following advantages over the Izod test: *(i)* Placing the specimen on the test machine facilitates even low-temperature tests without the risk of changing temperature during the period of clamping in Izod test and *(ii)* it is not required to clamp the specimen as in the Izod test and moreover it is free from compressive stresses around the notch which are produced by vice (or anvil).

One can calculate the energy used in breaking or fracturing the specimen in both the Charpy test and Izod test as follows (Fig. 8.35).

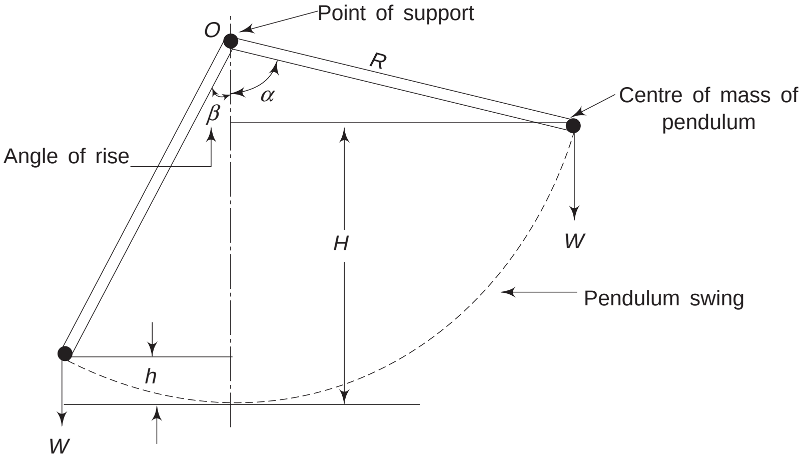


Fig. 8.35 Impact Test

Let us consider any position of pendulum in the beginning of its motion.

Let W → weight of the pendulum

H → height of fall of centre of gravity of the pendulum

h → height of rise of centre of gravity of the pendulum

R → length of the pendulum i.e. the distance from C.G. of pendulum to the axis from point of support O.

α → angle of fall

and β → angle of rise

Initial potential energy $= WH = WR\,(1 - \cos\ \alpha)$

Potential energy after rupture $= Wh = WR\,(1 - \cos\ \beta)$

Energy required to rupture the specimen

$$= WH - Wh = W\,(H - h)$$
$$= WR\,(\cos\ \beta - \cos\ \alpha)$$
$$= WR\cos\ \beta, \text{ when } \alpha = 90°$$

Typical Charpy *V* Impact strength test values for some of the metals and plastics at 0°C are summarized in Table 8.6.

Table 8.6 Charpy *V* Impact strength of some metals and plastics

Material		*Charpy V Impact strength*
Copper, (oxygen free) *HC*	annealed	34 J
Aluminium (commercial)	annealed	30 J
Cupronickel (70% Cu, 30% Ni)	annealed	157 J
Magnesium Alloy (3% Al, 1% Zn)	annealed	8 J
Nickel Alloy	annealed	290 J
Grey Cast Iron	annealed	3 J
Malleable Cast Iron black heart	annealed	15 J
Carbon Steel 0.2% C as rolled	annealed	50 J
Austenitic S.S.	annealed	217 J
Polythene high density	annealed	30 kJ/m^{2*}
PVC (unplasticised)	annealed	2 kJ/m^{2*}
ABS		25 kJ/m^{2*}

*Notch tip radius 0.25 mm, depth 2.75 mm

Fatigue Test: See section *(iv)*

Creep Test: See section *(v)*

8. NON-DESTRUCTIVE TESTING (NDT)

While studying various mechanical tests in previous sections, we have noted the effects of cracks and flaws. These should be detected at the early stages and the component replaced otherwise disaster will result. One can detect all microscopic flaws by NDT. NDT is the method of detection and measurement of properties or condition of materials, structures, machines without damaging or destroying their operational capabilities. Examples of NDT are: magnetic dust method, penetrating liquid method, ultrasonic test and radiography. All NDTs are used to detect various types of flaws on the surface of material or internal inclusions of impurities and these techniques are also very useful during preventive maintenance and repair. There are few techniques which do not require any special apparatus and are quite simple to handle and only a moderate skill being required. Some of the applications of NDTs are detecting: *(i)* surface cracks *(ii)* material composition *(iii)* internal inclusions *(iv)* internal voids and discontinuities and *(v)* condition of internal stresses. Now, we describe the various NDTs.

Magnetic Dust Method

This method is restricted to magnetic materials e.g. iron, nickel, cobalt, etc. This test is based on the principle that if there is a flaw in the magnetic material through which a magnetic field is allowed to pass, the lines of magnetic force or flux will be distorted near the flaw and lines of magnetic flux will be uniform for magnetic materials which are defect free.

This test is performed by magnetising the substance and then immersing the substance (test piece) in a bath of kerosene oil containing iron oxide powder. One can also use the coloured powder. If a crack or void lies across the path of the magnetic flux, each side of the crack or void becomes a magnetic pole which attracts iron powder. The accumulation of iron dust on the crack portion of the sample reveal the crack. This test can detect both internal and external defects. One can detect the cracks caused by quenching, fatigue failure in welding, blow holes in castings and grinding operations by this method. *Magnaflux* is the equipment used for this test.

Liquid-Penetration Test

This test is employed for detection of small defects which are very small to detect with the naked eye. This test is used to detect surface cracks or flaws in non-ferrous metals. This test employs a visible colour-contrast dye penetrant technique for the detection of open surface flaws in metallic and non-metallic objects. The penetrants are applied by spraying over the surface of material to be inspected. The excess penetrant is then washed or cleaned. Absorbent powder is then applied to absorb the penetrants in the cracks, voids which reveals the flaws. This test reveals flaws such as shrinkage cracks, porosity, fatigue cracks, grinding cracks, forging cracks, seams, heat treatment cracks and leaks etc., on castings, weldings, machined parts, cutting tools, pipes and tubes.

If the fluorescent penetrant is used, the developed surface must be examined under ultra violet light to see the presence of defects. This technique is used for non-porous and non-absorbent materials. Care may be taken to clean the surface so that it is free from dust, scale, etc. to have better results. Penetrants are highly toxic and flammable and hence proper precautions should be taken both during use and of storage of penetrants.

Ultrasonic Test

High frequency ultrasonic (sound) waves are applied to the test piece by a Piezoelectric crystal. If the test piece is free from cracks, or flawless, then it reflects ultrasonic waves without distortion. If there are any flaws in the specimen, the time taken by the ultrasonic waves will be less as the reflection of these waves will be from flaw points and not from the bottom of the specimen. Cathode ray oscilloscope (CRO) is used to receive the sound signals, whose time base circuit is connected to it. Knowing the time interval between the transmission of the sound pulse and the reception of the echo signal, we can calculate the depth of the crack. This test is a very fast method of inspection and often used to test aerospace components and automobiles. This test is generally used to detect internal cracks like shrinkage cavities, hot tears, zones of corrosion and non-metallic inclusions.

Radiography

This includes X-rays, gamma rays and radio-isotope welds. This method is used to check internal cracks, defects in materials and welds. Nowadays, radiography techniques are finding more extensive applications in the field of physical metallurgy and in the treatment of various diseases.

Rays are absorbed by the materials through which they are passed in the proportion of their density. The rays, after passing through the components, show a picture on a fluorescent screen or on a photographic plate. The cracks, blow holes and cavities appear lighter, whereas inclusions of impurities appear darker than the metal component. Developed photographic film show lighter and darker areas to represent the radiograph of defects in the component.

In X-ray photography, X-rays after allowing through the blow hole in a casting, will be absorbed to a lesser extent than X-rays which allowed to pass through sound metal. X-ray absorption measurement can be used to detect the fault in the specimen. X-rays are useful only for small thickness materials as their penetration power is less than that of gamma rays. Obviously, gamma rays are useful for greater thickness specimens.

Gamma rays have shorter wavelengths and are more penetrating than X-rays. The source for gamma radiations is usually the radioactive isotope of cobalt enclosed in a special container or capsule. Gamma rays radiography give better results for thicker materials and unlike X-rays, one can inspect at a time a number of components. We know that gamma rays do not ionise gases and also remain undeflected by magnetic fields. Now a days cheap radioisotopes are available and this test can be performed in a very short time and therefore this method is becoming more popular. However, there are some limitations of this method, e.g. handling of radioisotopes and precautions required.

The use of neutron-radiography or radioactive isotope method for studying metals and alloys depends on the characteristic decay in unit time. This method have high sensitivity and accuracy and can be used to find voids, inclusions and concentrations in explosives, polymers, lubricant papers and various space and defence equipments.

9. FRACTURE

It is defined as the separation of a specimen into two or more parts by an applied stress. Fracture can be classified as either *brittle* or *ductile*. The former occurs after little or no plastic deformation, whereas the latter occurs after extensive plastic deformation. A comparison of these are made in Table 8.7.

Table 8.7 Ductile and brittle fracture

Ductile fracture	*Brittle fracture*
• Material fractures after plastic deformation and slow propagation of crack	• Materials fractures with very little or no plastic deformation, e.g. in a china clay, glass etc.
• Fractured surfaces are dull or fibrous in appearance	• Fractured surfaces are crystalline in appearance
• Percentage elongation is about 30% prior to fracture occurs	• Percentage elongation is about 0.5% or almost nil prior to fracture occurs
• There is reduction in cross-sectional area of the specimen	• There is virtually no change in the cross-sectional area
• Fracture takes place after necking with little sound	• Fracture occurs rapidly often accompanied by a loud noise

Fracture is caused by physical and chemical forces and takes place in two stages: *(i)* crack initiation, i.e. initial formation of a crack and *(ii)* crack propagation, i.e. spreading of crack.

Metals exhibit different types of fractures, which depend upon *(i)* type of material *(ii)* rate of stressing, i.e. loading *(iii)* state of stress and *(iv)* temperature. The main types of fractures shown by metals are: *(i)* brittle fracture *(ii)* shear fracture *(iii)* cleavage fracture and *(iv)* ductile fracture (Fig. 8.36).

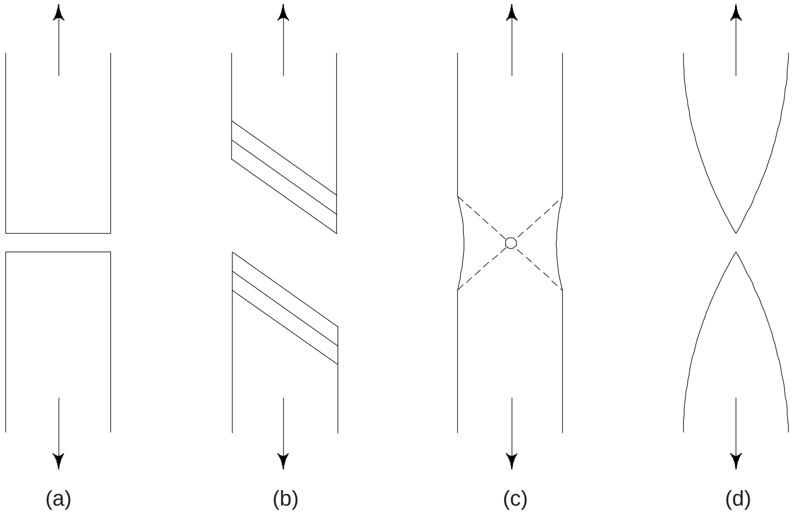


Fig. 8.36 Types of fractures observed in metals subjected to uniaxial tension (a) brittle fracture (b) shear fracture (c) ductile fracture (d) completely ductile fracture

Fracture is usually undesirable in engineering applications. We may note that flaws such as surface cracks, lower the stress for brittle fracture whereas line defects (dislocations) are responsible for initiating ductile fractures. Different types of fracture are shown in Fig. 8.37.

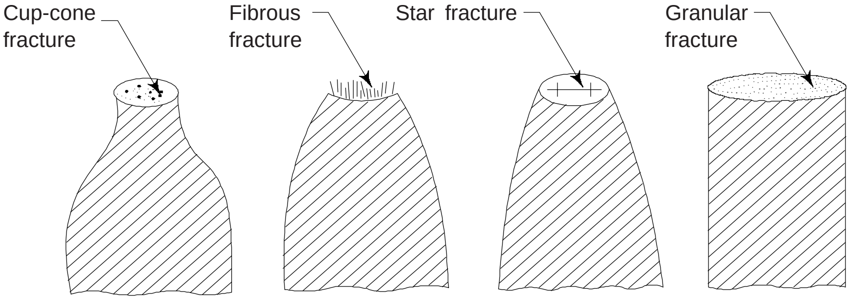


Fig. 8.37 Different types of fracture

Brittle Fracture

The word 'brittle' is associated with a minimum of plastic deformation, i.e. with a brittle fracture the material fractures with very rapid propagation of crack with very little or no plastic deformation like a china cup. The salient features of brittle fracture are:

(i) Brittle fracture occurs when a small crack in a material grows and the movement of crack involves very little plastic deformation of the metal adjacent to the crack. Growth continues until fracture occurs (crack propagation).

(ii) At the surface of a material, the atoms do not have as many neighbours as those in the interior of a solid and therefore they form fewer bonds. Obviously, surface atoms are at a higher energy than a plane of interior atoms. Brittle fracture contains in destroying the interatomic bonds by normal stresses.

(iii) Brittle fracture in metals is characterised by a rapid rate of crack propagation with minimum energy of absorption, with no gross deformation and very little microdeformation. Adjacent parts of the metal are separated by stresses normal to the fracture surface.

(iv) This does not produce plastic deformation and therefore requires less energy than a ductile failure where energy is introduced in the process of forming dislocations and other imperfections within the crystal.

(v) Brittle fracture occurs along crystal planes with fewer atomic bonds, i.e. characteristic crystallographic planes called as cleavage planes. The fracture is termed as cleavage fracture.

(vi) Brittle fracture occurs at or below the elastic limit of a material.

(vii) Normally brittle fracture follows the grain boundaries which can be identified by their granular and shiny look. In some instances this type of fracture can be caused by grain-boundary films of hard-brittle second phase, like that formed by bismuth in copper.

(viii) The tendency for brittle fracture increases with decreasing temperature, increasing strain rate and stress concentration conditions usually produced by a notch.

Brittle fracture is to be avoided at all cost, because it is very dangerous and occurs without warning and usually produces disastrous consequences. Brittle fractures are of practical importance due to the failures of pressure vessels, bridges, pipe lines, hulls of ships, etc.

Ductile Fracture

This signifies large plastic deformation, and occurs after extensive plastic deformation prior to and during the propagation of the crack. This requires considerable energy which is absorbed in forming dislocations and other imperfections (defects) in metals.

In a ductile fracture, there are three successive events involved: *(i)* test sample or specimen begins necking and minute cavities form in the necked region. The plastic deformation is concentrated in this region and indicates that the formation of cavities is closely linked to plastic deformation, hence to the dislocation movement, thereby taking the longest time in the fracture process, *(ii)* the cavities coalesce and form minute crack at the centre of the test specimen, and *(iii)* the crack propagates outwardly to the surface of the specimen by a shear separation in a direction 45° to the tensile axis, resulting in a familiar "cup and cone" type fracture.

The various stages involved in ductile fracture are shown in Fig. 8.38. Necking begins at the points of plastic instability where the increase in strength due to strain-hardening fails to compensate for the decrease in cross-sectional area. This occurs at maximum load. The formation of a neck introduces a triaxial state of stress in the region. A hydrostatic component of tension acts along the axis of the specimen at the centre of the necked region. Many fine cavities form in this region (Fig. 8.38(c)), and under continued straining these grow and coalesce into a central crack. This crack grows in a direction perpendicular to the axis of the specimen until it approaches the surface of the specimen. It then propagates to the surface of the specimen in a direction roughly 45° to the tensile axis to form the cone part of the fracture (Fig. 8.38(e)).

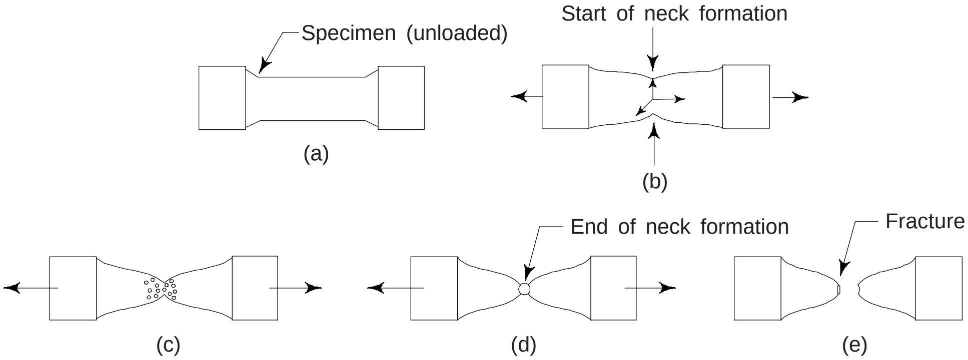


Fig. 8.38 Various successive stages in the ductile fracture of a specimen (tensile test)

One can also explain the fracture of ductile materials in terms of work-hardening coupled with crack nucleation and growth. The initial cavities are often observed to form at foreign inclusions where gliding dislocations can pile up and produce sufficient stress so that void or microcrack is formed. Let us consider a test specimen which is subjected to a slowly increasing load. Material begins to work harden, when the elastic limit is exceeded. The permanent elongation increases with increasing the load and the cross-sectional area decreases simultaneously. The associated decrease in area leads to the formation of a neck in the test specimen as stated earlier. Due to high dislocation density of the necked region and the material being subjected to a complex stress, i.e., no longer a simple tensile stress, the dislocations are separated from each other because due to repulsive interatomic forces. The dislocations come closer together with the increase of the resolved shear stress on the slip plane. The cracks are formed due to high shear stress and the presence of low angle grain boundaries. Once a crack is formed then it can grow or elongate by means of dislocations which slip. We may note that crack propagation for this mechanism is along the slip plane and these cracks coalesce. Obviously, one crack grows at the expense of others and finally crack growth results in failure.

Crack nucleation at a slip plane obstacle due to dislocation pile up is shown in Fig. 8.39 and crack nucleation at low angle grain boundaries is shown in Fig. 8.40. Figures 8.39 and 8.40 together explains the ductile fracture mechanisms.

Ductile fracture has been studied much less extensively than brittle fracture, as it is considered to be a much less serious problem. An important characteristic of ductile fracture is that it occurs through a slow

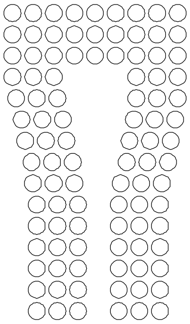


Fig. 8.39 Crack nucleation at a slip plane obstacle due to dislocation pile up

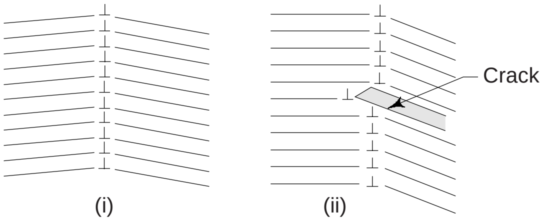


Fig. 8.40 Crack nucleation at low angle grain boundaries

tearing of the metal with the expenditure of considerable energy. We may note that unlike brittle fracture, a completely ductile fracture presents a rough dirty-grey surface.

Griffith Theory of Brittle Fracture

The stress at which a material fractures is lower than the value of the ideal breaking stress calculated from the atomic bond strength. Clearly, the fracture strength of real materials is far below the theoretical maximum value for an ideal solid. Griffith was first to explain this discrepancy.

Griffith postulated that in a true brittle material there are always present many microcracks which act to concentrate the stress at the tips. The crack could come from a number of sources, e.g. as a collection of dislocations, as flaw occurred during solidification or a surface scratch.

In order to explain the mechanism of ideal brittle fracture, let us consider the stress distribution in the vicinity of a crack and the conditions under which it propagates with a constant velocity. Lengthening of a crack causes an increase in surface area of the crack and, obviously, the surface energy of the specimen is increased. Moreover, there is also compensating release of energy. With the crack becoming longer, the material on both sides of crack can no longer store elastic energy since the tensile stress cannot be transmitted across the crack. This means, an increase in crack length causes the release of elastic energy. Griffith remarked that when the elastic energy released by extending a crack was equal to the surface energy required for crack extension, the crack would grow.

Now, we consider a crack of an elliptical cross-section in a rectangular specimen (Fig. 8.41). Let

σ → Tensile stress applied perpendicular to the length 2 C.
C → Half length of the crack
P → Radius of curvature at its tips.

The expression for maximum stress at the tip of the crack is given by

$$\sigma_m = 2\sigma\sqrt{\frac{C}{P}} \tag{24}$$

A certain amount of energy is always stored in a material, before the propagation of crack, is known as elastic strain energy. When the crack is to propagate through the material, the release of the stored energy must be sufficient to provide the energy associated with the two surfaces produced. The expression for total elastic strain energy for unit volume is

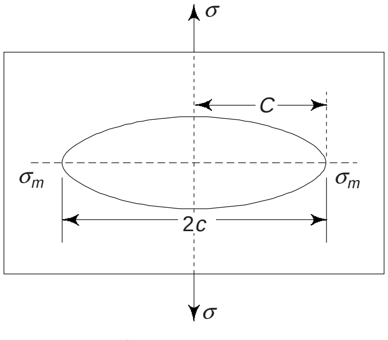


Fig. 8.41 Griffith's crack

$$U_E = \frac{\sigma^2}{2E} \times \text{Area} \times \text{Width}$$

$$= -\frac{\sigma^2}{2E} \times \pi C^2$$

In fact, the strain field should be integrated from infinity to the surface of crack, which gives the elastic energy available per unit width as

$$U_E = -\frac{\sigma^2 \pi C^2}{E}$$

where E is the Young's modulus of elasticity. The negative sign indicates that the elastic strain energy stored in the material is released as the crack formation takes place.

Now, if γ is the surface energy per unit area in J/m^2, then the surface energy due to the presence of crack of length 2 C is given by

$$U_S = (2\gamma C) \times 2 = 4\ C\gamma$$

We multiply the above expression by 2 because there are two surfaces. Applying Griffith's criterion means that the change in surface energy with crack length must just equal the change in elastic energy, i.e.

$$\frac{dU_E}{dC} = \frac{dU_S}{dC}$$

$$\frac{d}{dC}\left(\frac{\pi C^2\ \sigma^2}{E}\right) = \frac{d}{dC}\ (4\gamma C)$$

or

$$\sigma = \sqrt{\frac{2\gamma E}{\pi C}} \tag{25}$$

Relation (25) shows that the stress necessary to cause the brittle fracture (i.e., the critical value of stress), varies inversely as the square root of the crack length. This means the tensile strength of a completely brittle material is determined by the length of the largest crack existing before loading. Relation (25) is known as the *Griffth's equation.*

The above analysis is valid only for a perfect brittle material like glass. In the case of crystalline materials, the above theory has to be modified somewhat to take into account the small amount of plastic deformation that occurs. We may note that basically, the fracture mechanism is the same, although the critical crack depth is somewhat larger. The theory can be applied for metals, with certain modifications such as inclusion of plastic work (p) required to extend crack wall and crack extension force (G). The modified Griffith's equation reads as

$$\sigma_g = \left[\frac{2E\,(\gamma + p)}{\pi C}\right]^{1/2} \tag{26}$$

Equation (26) is similar to Eq. (25) but there is new interpretation of energy. p can be increased by suitable changes in the micro structure of material. When the value of p is increased, the higher rate of stress can be applied with flaws of given size or flaws of larger size can be tolerated at a given applied stress.

Ductile to Brittle Transition

This is commonly observed in BCC metals and almost missing in most of the FCC metals. This transition is observed at low temperatures, extremely high rates of strain or notching the material. This is very important when selecting materials for engineering purposes. The notched bar impact test for metals can be used to determine the temperature over which the transition from ductile to brittle takes place. Such a temperature is termed as transition temperature.

One can explain the ductile to brittle transition with the help of Fig. 8.42. Figure shows the plot of brittle fracture stress (σ_f) and the yield stress (σ_y) as a function of temperature or strain rate. We note that the curve

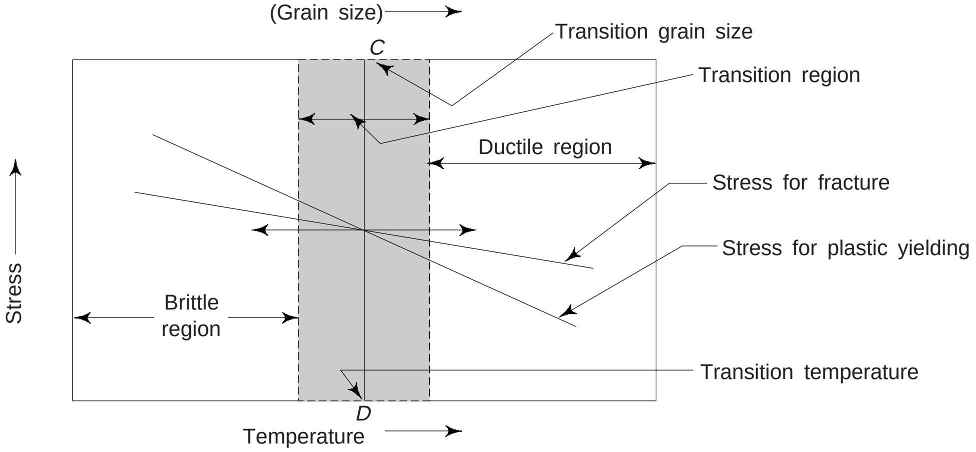


Fig. 8.42 Ductile to brittle transitions as a function of grain size and temperature

for brittle fracture stress rises slightly to the left because the surface energy increases as temperature decreases. We note a strong temperature dependence in the yield stress curve as in BCC metals and metal oxide ceramics. From figure it is clear that the two curves intersect and a vertical line is drawn at the point of intersection, which is called the *ductile-brittle transition temperature.* Now, if a material is stressed at a temperature or strain rate which is to the right side of line *CD*, it will reach its yield point prior it reaches the brittle fracture stress and will undergo some plastic deformation prior to fracture. However, applying a stress under conditions which lie left of the line *CD* will result in brittle fracture. Obviously, at all temperatures, below the transition temperature, the fracture stress is smaller than that of the yield stress. This reveals that fracture stress may be controlled by the yield stress. As the applied stress reaches a value equal to the yield stress, the crack is nucleated at the intersection of the slip planes and propagates rapidly. The temperature range over which the rapid changes takes place is termed as the *transition region*.

We may note that there is no ductile to brittle transition, if the curves of yield stress and brittle fracture do not intersect. It is interesting to note that the yield stress curves for FCC materials generally lie below the brittle fracture stress curve and FCC materials do not experience this transition.

The yield stress as well as fracture stress is a function of grain size and these stresses increase with the decreasing grain size. Obviously, the fine-grained metals have a lower transition temperature as compared to the course grained metals.

For mild steel, the consumption of energy in an impact test as a function of temperature is shown in Fig. 8.43. By fast loading, one can achieve a high strain rate in impact testing machines. We know that increasing the strain rate is equivalent to lowering the temperature. This means the materials which are ductile when strained slowly at a given temperature will behave in a brittle manner when subjected to a high strain rate.

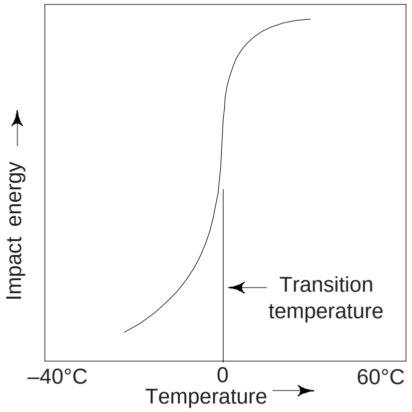


Fig. 8.43 For mild steel, the variation of impact energy as a function of temperature

We may note that the ductile to brittle transition is quite dangerous from a design point of view.

Method For Prevention of Fracture Failure

Griffith's theory shows that surface cracks are cause of more concern than internal cracks and cause brittle fracture. The surface crack of length *C* has the same effect as internal crack of length 2 C. It is reported that fatigue cracks often propagate from the surface, inwards. Methods against fracture failure by crack propagation and to make surface cracks ineffective are as follows:

(i) To counteract the tensile stress causing cracks to propagate, introduce compressive stresses on the surface. One can achieve this by heating, e.g. silicate glass to its softening temperature and then cooling it rapidly to produce contraction or introduce compressive stresses. Due to this the glass becomes stronger and is used in producing tempered glass. Such a glass is commonly used in the windshields of automobiles.

(ii) Better surface finish by polishing the surface offers better protection as it removes some of the cracks from the surface.

(iii) Avoid sharp corners and notches in the parts to eliminate points of stress concentration. One may take care of this at design stage itself.

(iv) Avoid impurities in the base metal as well as during welding process as inclusions will produce brittle phases.

(v) Stress applied should be parallel to the direction of elongation in a good design.

(vi) One should avoid failure along grain boundaries by proper distribution of solute atoms in metals to prevent the formation of brittle phases at grain boundaries. This can be achieved by preventing the formation of films of impurities along the gain boundaries.

(vii) Fine grained polycrystalline materials offer good protection against brittle fracture.

(viii) Use of composites/reinforced materials which provide protection against tensile loads, e.g. reinforced cement conceretie (RCC) and reinforced plastics. During their use, if we apply accidently a tensile load, there is no major danger as the crack formed cannot propagate due to the presence of a compressive matrix.

Important Technological Properties of Metals

Those properties which are highly desirable in the fabrication, formation and shaping of materials for various technological operations or processes are known as technological properties of metals. The important technological properties of metals are:

(i) Weldability: It is defined as the capacity of metal to be welded or joined under the fabrication conditions imposed, into a specific, suitable designed structure, and to perform the intended service satisfactorily. A metal has good weldability, provided: *(i)* it can be welded readily in order to make satisfactory fabricated structure, and *(ii)* expensive service and complicated techniques are not required to produce sound joints.

Weldability is of great importance for fabrication of metals into various structures. Factors affecting the weldability of a metal are:

(i) The composition of a metal *(ii)* Welding technique (iii) Heating and cooling effects *(iv)* Filter materials *(v)* Shielding atmosphere, i.e. affinity for oxygen (oxidation) and other gases before and at welding temperature *(vi)* Stability of micro-constituents upto welding temperature *(vii)* Proper heat treatment prior and after deposition of metal *(viii)* Strength of metal at elevated temperature *(ix)* Design, selection and service condition.

One can know weldability of a metal by determining the metal's behaviour under fusion and cooling, by crack and notch sensitivity, or by comparison of the heating and cooling effects which take place at the joint of the metal with the metal of known weldability.

(ii) Machinability: This is defined as the ease with which a given material can be machined by cutting tools in various operations, e.g. machining, drilling, boring, milling, etc. This property of the material enables it to be machined by cutting tools. One can compare machinability of different materials by their machinability index defined as below:

$$\text{Machinability index} = \frac{\begin{array}{c}\text{Cutting speed of metals for}\\ \text{20 minutes tool life}\end{array}}{\begin{array}{c}\text{Cutting speed of a standard}\\ \text{steel for 20 minutes tool life}\end{array}}$$

Machinability depends on the following factors:

Chemical Composition of Metals to be Cut

By addition of small amounts of alloying elements, e.g., lead, manganese and sulphur increases the machinability, whereas high alloy contents, presence of abrasive inclusions and more than 0.6% carbon content decrease machinability.

Microstructure of the Material to be Cut

Undistorted fine grains and uniform microstructures with very low carbon contents can help to increase machinability of a material, whereas large distorted non-uniform structures with high carbon contents affect adversely machinability of a material.

Mechanical Properties of the Materials

It is found that the machinability of soft materials is good whereas high ductile and very hard materials are difficult to machine.

Cutting Conditions of Materials

To compare machinability, different materials are cut with same tools, feeds, speeds, depth of cuts etc. Obviously, one compares the power required, tools life surface, finish etc. If power required is less the machinability is said to be high or good and tool life is more for same surface finish. The machinability of metals is affected by heat treatment processes. Annealing and normalizing processes usually help make metal parts soft and increase their machinability, whereas quenching and hardening processes lower their machinability. Malleable iron has the highest machinability as compared to other ferrous materials. Addition of small percentage of certain elements improves machinability of various steels, copper and its alloys.

The *machinability rating* of few materials is as follows:

Material	Machinability Rating (%)
Nickel steels	40-50
Carbon steels	40-60
Cast iron	50-80
Aluminium	300-200

Alloys of magnesium, aluminium and zinc have relatively excellent machinability.

The following points need consideration which determining the machinability in any production process:

- Type of work and type and shape of cutting tool.
- Life and speed of cutting tool and coefficient of friction.
- Hardness of the cutting tool and cutting temperature.
- Power required and strength of cutting forces.
- Quality of surface finish and feed, depth and velocity of the cut (metal removal rate)
- Composition and microstructure of the material to be cut.
- Type of coolant used.
- Form and size of chips.

Methods for Improving Machinability

One can improve the machinability of a metal by addition of alloying elements, e.g. lead and sulphur. Mild steel has higher and better machinability than carbon steel, high speed steel and alloy steel. In comparison to white cast iron, the grey iron has very good machinability. The grey cast iron contains carbon in free form as well as graphite flakes which cause discontinuity in ferrite. This causes the breaking of chips while machining. Cast iron contains free carbide which is very hard and hence it is difficult to machine. The following other methods can also be used to improve machinability of a material: *(i)* heat treatment *(ii)* use of proper type of coolant *(iii)* making use of quality tools *(iv)* improving microstructure of the material *(v)* hot working of hot alloys.

(iii) Castability: It is the ease with which a metal can be cast into various forms is known as castability. It is the oldest process of forming metals. It consists of introducing the molten metal into a cavity or the mould of the required form and allowing the metal to solidify. The shape of the casting is that of product required which is then further processed by machining. The mould or cavity into which the molten metal is poured is made of some heat resisting material. Castability depends on the following factors:

- *Rate of Solidification:* The ease with which a metal continues to flow after it has been poured in the mould cavity depends upon pouring temperature and gate and riser designs, etc. Some metals, as grey iron are very fluid and thin sections, i.e. intricate shapes of complicated castings can be casted easily.
- *Gas Porostity:* In the molten state, certain metals are affected by oxygen and nitrogen. During the cooling of castings these gases are trapped, resulting in the formation of small holes and voids in the castings.
- *Shrinkage:* On changing from a molten to a solid state, there is the reduction in volume of a metal and this is termed as shrinkage. We may note that the amount of shrinkage produced during casting varies from metal to metal and one will have to take into consideration of some shrinkage allowance in the design of a pattern and mould-making. To increase and decrease the shrinkage rate, one will have to add few elements to the molten metal.
- *Segregation:* The non-uniform distribution of constituents in a metal is termed as segregation. Segregation is usually a concentration of certain constituents and of impurities arising during freezing and generally persisting throughout subsequent operations.

Some of the casting processes are:

(a) Sand Casting This is the most convenient and easiest method of giving desired shape and size to the metal. It has high permeability and resistance to high temperatures. The merits and demerits of this method are:

- Less costly due to low cost of sand mould
- Casting has good hardness but poor ductility
- Casting has low density and poor strength
- Casting has poor finish and several defects, e.g. defects like blow holes, cracks, shrinkages, etc.
- There are limitations to surface finishing and dimensional accuracy
- The method is unsuitable for thinner sections of materials

(b) Centrifugal Casting In this method the molten metal is poured in the mould and allowed to solidify while the mould is revolving. Obviously, the molten metal is subjected to centrifugal force due to which it flows in mould cavities and results in the production of high density castings.

This method is quite suitable for mass production as well as for thin parts of high strength. The merits and demerits of this method are:

- Good surface finish is produced
- Castings of high density are produced
- Mass production with minimum rejection is possible with this method
- High initial cost and skilled labour required
- Method is unsuitable for complex castings
- Process is limited to cylindrical parts

(c) Die Casting The liquid metal is forced into a mould under pressure and at a relatively high velocity, and allowed to solidify with pressure maintained. This casting can be done either by gravity or by pressure die casting process. The pressure casting has the advantage that the metal can be forced into all parts of mould cavity and thus very complex shapes with better dimensional accuracy can be produced. These castings when properly designed and produced in first class dies compare favourably with many machined products. Rapid casting rates are possible, making this an inexpensive method; furthermore, a single set of dies may be used for thousands of dies. However, the speed with which the dies is filled often requires much skill in design to avoid porosity and similar faults. However, this technique lends itself only to relatively small pieces and to alloys of zinc, aluminium, and magnesium, which have low melting temperatures. This technique is employed for toys, chasis for radio sets, camperabodies, gears, bearing liners, engine-cylinder-heads, lead and tin base white metals.

(d) Investment (lost-wax) Casting In case of investment casting, the pattern is made from a wax or plastic that has a low melting temperature. Around the wax or plastic pattern is poured a fluid slurry, which sets up to form a solid mould or investment, plaster of paris is usually used. The mould is then heated, such that the wax or plastic pattern metls and is burned out, leaving behind a mould cavity of desired shape. After the mould is cooled, the casting is removed by breaking the mould. This technique is an expensive process for large scale production. This technique is employed when high dimensional accuracy, reproduction of fine detail, and an excellent finish are desired, e.g., in jewellery and dental crowns and inlays. Also, blades for gas turbines and jet engine impellers are investment cast. Obviously, this is also suitable for complex castings.

(f) Continuous Castings At the conclusion of extraction processes, several metals are solidified by casting into large ingot moulds. Normally, the ingots are subjected to a primary hot-rolling operation, the product of which is a flat sheet or slab; these are more convenient shapes as starting points for subsequent secondary metal-forming operations, i.e. forging, extrusion, drawing. One may combine these casting and rolling steps by a continuous casting (or strand casting) process. Using continuous casting, the refined and molten metal is cast directly into a continuous strand which may have either a rectangular or circular cross-section; solidification occurs in a water-cooled die having the desired cross-sectional geometry. In comparison to ingot-cast products, the chemical composition and mechanical properties are more uniform throughout the cross-sections for this type of castings. Moreover, continuous casting is highly automated and more efficient.

Formability: The ability of metals to be formed into different shapes is termed as formability. Formability is based on the flowability and ductility of the metal, which, in turn, depends on its crystal structure. The formability of metals is affected by grain size, hot and cold working alloying elements, and heat treatment processes, e.g. annealing and normalising.

Low-carbon steel can be easily cold-worked as it has good metal-forming qualities. For metal forming processes, e.g. rolling, spining, drawing, extrusion, forging, etc., high carbon steel are unsuitable. In order to prevent the distortion of crystals during cold working, hot working processes are suitable for increasing the ductility as well as formability of crystals.

We may note that the metal to be deformed during forming process must not crack under the action of high deforming forces. Such cracks may arise if metal has reached its strain hardening capacity to the full extent. One can obtain the reliable information on the flow stress from tension or compression test but unfortunatively only at low strain rates. However, the testing of metals at high strain rate and high temperature is difficult and lacks testing facilities. For quite some time the *single blow test* under impact load has helped to provide indications whether a metal can be hot worked without developing cracks. To assess the materiability for hot forging and hot piercing, *hot twist test* has been used. In hot piercing process, solid cylinders are converted into stainless tubes. In hot twist test the metal cylinder is twisted at high rate and the torque required to cause given amount of twist is used to calculate the flow stress while number of twists prior to fracture provides the measure of ductility.

Example 1 *The Young's modulus of a certain material is* 180×10^3 *mega Newton/m*^{2} *and its true surface energy is 1.8 J/m*^{2}*. The crack length is 5* μ*m. Show that its fracture strength is* 273×10^6 *N/m*^{2}*. Interpret your result.* [BE, 1999]

Solution The fracture strength is

$$\sigma_f = \sqrt{\frac{2\gamma E}{\pi c}}$$

$\gamma = 1.8$ J/m^2

$E = 180 \times 10^9$ N/m^2

$c = \frac{5}{2} \times 10^{-6}$

$$= \sqrt{\frac{2 \times 1.8 \times 180 \times 10^9}{3.14 \times 5/2 \times 10^{-6}}}$$

$$= 273 \times 10^6 \text{ Newton/m}^2 = 273 \text{ MN/m}^2$$

We note the above calculated strength is approximately ¼ of the actually observed value. The reason for the difference between the two values is as follows. When the material is deformed plastically and the crack propagate, then the plastic work is done by the crack. When the crack is plastically deformed, energy is also consumed. Since we have not taken into account the plastic work required to propagate the crack and hence the calculated value is lower. One can calculate the amount of plastic work required to propagate the crack by the equation

$$\sigma_g = \left[\frac{2E(\gamma + p)}{\pi c}\right]^{1/2}$$

where σ_g is actual fracture strength and p is the plastic work required to propagate the crack.

$$\therefore \qquad 1000 = \left[\frac{2 \times 180 \times 10^9 (1.8 + p)}{3.14 \times 2.5 \times 10^{-6}}\right]$$

$$p = 25.32 \text{ J/m}^2 = 12.4\ \gamma$$

which is 12.4 times the surface energy. This shows that the effective surface energy that includes plastic work is 12.4 times higher than its true surface energy.

Example 2 *A tensile test specimen having a diameter of 12.7 mm was loaded upto a load of 76 kN and its diameter was measured as 12 mm. Compare true stress and strain with engineering stress and strain.* [BE, 2000]

Solution Initial area of cross-section $(A_0) = \frac{\pi}{4} d_0^2 = \frac{\pi}{4} (12.7)^2 = 126.6$ mm^2

The area of cross-section after load of 76 kN (A)

$$= \frac{\pi}{4} d^2 = \frac{\pi}{4} (12)^2 = 113.04 \text{ mm}^2$$

$$\text{Engineering stress } (\sigma) = \frac{P}{A_0} = \frac{76000}{126.6} = 600.3 \text{ N/mm}^2$$

$$\text{True stress } (\sigma') = \frac{P}{A} = \frac{76000}{113.04} = 672.33 \text{ N/mm}^2$$

$$\text{True strain } \varepsilon' = \log \frac{A_0}{A} = \log \frac{126.6}{113.04} = 0.113$$

$$\text{Also } \varepsilon' = \log(1 + \varepsilon)$$

$$= 0.113$$

$$\therefore \qquad 1 + \varepsilon = e^{0.113} = 1.12 \quad \therefore \; \varepsilon = 0.l2$$

Example 3 *A cast iron bed plate for a pump has a crack length of 100 μm. The Young's modulus of C.I. is 210 GN/m^2 and the specific surface energy is 10 J/m^2. Show that the fracture strength required is 1.6 $\times$ 10^8 N/m^2.* [BE, 2002]

Solution

$$\sigma = \left(\frac{2\gamma E}{\pi c}\right)^{1/2}$$

$$= \left(\frac{2 \times 10 \times 210 \times 10^9}{3.14 \times 50 \times 10^{-6}}\right)^{1/2}$$

$$= (2.67 \times 10^{16})^{1/2} \text{ N/m}^2$$

$$= 1.6 \times 10^8 \text{ N/m}^2$$

Example 4 *A piece of copper originally 305 mm long is pulled in tension with a stress of 276 MPa. If the deformation is entirely elastic, find the resultant elongation?* [BE, 2002]

Solution Since the deformation is elastic, strain is dependent on stress,

$$\sigma = \varepsilon E \tag{i}$$

Moreover, the elongation Δl is related to original length l_0 as,

$$\varepsilon = \frac{\Delta l}{l_0} \tag{ii}$$

Combining (i) and (ii) and solving for Δl, we have

$$\Delta l = \frac{\sigma l_0}{E}$$

$$= \frac{(276 \text{ MPa}) \times (305 \text{ mm})}{110 \times 10^3 \text{ MPa}}$$

σ = 276 Mpa
l_0 = 305 mm
ε = 110 GPa
= 110 $\times$ 10^3 Mpa
= 0.77 mm

Example 5 *Compute the strain-hardening exponent n in $\sigma_T = K\varepsilon_T^n$, for an alloy in which a true stress of 415 MPa produces a true strain of 0.10. Given K = 1035 MPa.* [BE, 1998]

Solution $\sigma_T = K\varepsilon_T^n$

Taking log of both sides and solving for n, we have

$$n = \frac{\log \sigma_T - \log K}{\log \varepsilon_t}$$

$$= \frac{\log(415 \text{ MPa}) - \log(1035 \text{ MPa})}{\log(0.1)}$$

$$= 0.40$$

SUGGESTED READINGS

1. T.H. Courtney, *Mechanical Behaviour of Materials*, 2nd ed., McGraw-Hill (2000).
2. K.G. Budinski and M.K. Budinski, *Engineering Materials, Properties and Selection*, 7th ed., Prentice Hall, NJ (2002).
3. N.E. Dowling, *Mechanical Behaviour of Materials*, 2nd Ed., Prentice Hall, Upper Saddle River, NJ (1998).

REVIEW QUESTIONS

1. Explain the term machinability. Describe the method of improving machinability. [AMIE]

2. Explain weldability and machinability of metal. [AMIE]

3. Differentiate between brittle fracture and ductile fracture. How they are caused. [B. Tech.]

4. Draw a stress-strain diagram for a low carbon steel specimen, indicating the proportionality limit, elastic limit, yield point and the point of maximum loading and rupture. Explain the above data. [AMIE, B.E.]

5. Compare and contrast elastic and plastic deformation of materials. [B. Tech.]

6. Define hardness. Discuss its role in characterising mechanical behaviour of materials. [B.E.]

7. What is the effect of stress and temperature on a creep curve? [B.E.]

8. Describe in detail the purpose and procedure for testing of metals for (a) tensile strength and (b) impact strength. [AMIE]

9. Explain how dislocations produces plastic deformation in materials. [AMIE, B.Tech.]

10. Sketch the conventional and actual stress-strain diagrams and state the significance of each. [AMIE]

11. Explain the different types of important technological properties of metal. [BE]

12. Explain the phenomenon of creep in metals. How is it controlled? [BE]

13. Sketch a typical creep curve and indicate important stages. Discuss the effect of grain size on creep strength. [AMIE]

14. What is creep? Draw a typical creep curve and explain the different stages on creep. Elaborate your answer with reference to steel. [AMIE]

15. Mention the technological properties that control the weldability of metals. [AMIE]

16. Draw S-N curve. Explain different stages occurring in the curve. [BE]

17. When does Hooke’s, law appears to fail? [AMIE]

18. Explain the ‘notch’ sensitivity of metals. [AMIE]

19. What is meant by creep? Explain the different types of creep with the help of creep curve. [BE]

20. With the help of creep curve, discuss three stages of creep. [BE]

21. What is the critical resolved shear-stress [AMIE]

22. Do all metals have endurance limits? Explain. [BE]

23. State the various factors which affect the mechanical properties of materials and state their effects. [BE]

24. Write factors which contribute to the onset of fatigue failure and those which tend to resist fatigue. [BE]

25. What are the different methods of determining the hardness of a metal? State their advantages and disadvantages. [BE]

26. Under what conditions may a ductile material fail in brittle structure? [AMIE]

27. Discuss the examples of ductile and brittle fracture. [AMIE]

28. What is the creep in metallic materials? Why is the knowledge of creep resistance important industrially? [AMIE]

29. Explain the role of fatigue failure behaviour of metals. [AMIE]

30. Which element is responsible for the high machinability of free cutting steels? [AMIE]

31. What is meant by the term 'hardenability'? Discuss, how hardenability of a steel can be estimated. [AMIE]

32. Describe the methods for minimizing creep of materials at low and high temperatures. [AMIE]

33. Describe tensile testing of metals. What are the data one can get from such a test? [AMIE]

34. Write factors which contribute to the onset of fatigue failure and those which tend to resist fatigue. AMIE]

35. Why are brittle materials used more often in compression than in tension in structural design? [BE]

36. Describe in detail the purpose and procedure for testing of metals for (i) tensile strength and impact strength. [AMIE]

37. Write short notes on the following

(a) Creep of materials *(b)* Fatigue failure
(c) Ultimate tensile strength *(d)* Griffith's crack theory
(e) Ductility *(f)* Ultimate stress
(g) Fracture *(h)* Weldability
(i) Elastic strain *(j)* Endurance limit
(k) Transition temperature *(q)* Grain-boundary Creep
(m) High temperature creep [AMIE, BE]

38. Differentiate between the following:

(i) Creep and fatigue *(ii)* Ductility and rigidity
(iii) Ductility and malleability *(iv)* Ductile and brittle fractures
(v) Kc and Kic of material, and its importance
(vi) Hardness and hardenability
(vii) True stress and Engineering stress
(viii) Elastic limit and proportional limit
(ix) Yield stress and proof stress
(x) Endurance limit and fatigue limit
(xi) Toughness and resilience
(xii) Upper and lower yield points
(xiii) Primary creep, secondary creep and tertiary creep. [AMIE, BE]

39. Discuss the principle of non-destructive testing by the following methods: *(a)* magnetic dust method *(b)* penetrating liquid test and *(c)* ultrasonic test method. [AMIE, BE]

40. Discuss the effect of dislocation climb and movements of dislocation jogs on creep. [AMIE]

41. What do you understand by creep resisting materials? Give few examples of such materials. [AMIE]

42. Describe the effect of the following on fatigue

(i) stress concentration on fatigue *(ii)* size
(iii) change of size of specimen and *(iv)* surface [AMIE]

4**3.** Discuss briefly the various hardness tests and state the advantages and disadvantages. [AMIE]

44. Discuss the principle of non-destructive testing by the following methods: *(i)* magnetic dust method *(ii)* penetrating liquid test and *(iii)* ultrasonic test. [AMIE, BE]

45. Name five important mechanical tests which give valuable information about metals and alloys. [AMIE]

46. What does the impact test signify? Explain with necessary formulations the procedures to be adopted in an impact test conducted on a pendulum-type impact testing machine. [AMIE, BE]

47. What do you understand by brittle to ductile fracture transition? What is transition temperature? Describe the factors on which this temperature depends. [AMIE]

48. *(a)* Explain the Charpy impact test.

(b) Explain the Brinell hardness test for mild steel specimen.
(c) Differentiate between the fracture of mild steel and cast iron specimens in tensile testing.
(d) What factors determine the mechanical properties of materials? Give examples.
(e) How is impact strength measured? Write its SI unit. [AMIE]

49. What are the different non-destructive tests? Explain briefly their fields of application. [AMIE]

50. Explain the Griffith's theory of metals. Derive the formula for average applied stress at which crack spreads. [BE]

51. How does corrosion have a deterimental effect on fatigue? How is corrosion fatigue propagated? What steps would you recommend to minimize it? [AMIE]

PROBLEMS

1. An aluminium rod of 10 mm dia and 300 mm length is subjected to a tensile load of 500 kg. Calculate the tensile strain, the increase in length and decrease in diameter. The Youngs's modulus of metal is 0.8×10^8 kg/cm^2 and Poisson's ratio is 0.33. [BE]

2. For a brass alloy, the stress at which plastic deformation begins is 345 MPa and the modulus of elasticity is 103 GPa. *(a)* What is the maximum load that may be applied to a specimen with a cross-sectional area of 130 mm^2 without plastic deformation? *(b)* If the original specimen length is 76 mm, what is the maximum length to which it may be stretched without causing plastic deformation?
[BE], [*(a)* F = 44,350 N *(b)* l = 76.25 mm]

3. Consider a cylindrical specimen of some hypothetical metal alloy that has a diameter of 10 mm. A tensile force of 1500 N produces an elastic reduction in diameter of 6.7×10^{-4} mm. Find the elastic modulus of this alloy, given Poisson's ratio is 0.35. [BE], [E = 100 GPa]

4. The true strain is ε_T is defined as

$$\varepsilon_T = \log \frac{l_i}{l_o}$$

Show that this may also be represented by

$$\varepsilon_T = \log \left(\frac{A_o}{A_i}\right)$$

When specimen volume remain constant during deformation. Which of the two expressions is more valid during necking? Why?

5. Find the toughness for a metal that experiences both elastic and plastic deformation. Assuming $\sigma = \varepsilon E$ for elastic deformation, that the modulus of elasticity is 172 GPa, and the elastic deformation terminates at a strain of 0.01. For plastic deformation assume the relationship $\sigma_T = K\varepsilon_T^n$, values for K and n are 6900 MPa and 0.30, respectively. Note that plastic deformation occurs between values of 0.01 and 0.75, at which point fracture occurs. [B.E.], [Toughness = 3.65×10^9 J/m^3]

OBJECTIVE QUESTIONS

1. True or False and Justify:
(a) Ductility is measured by the percentage reduction in area before rupture of a test piece.
(b) True stress is lower than engineering stress.
(c) The level of stress at which strain ceases to be wholly elastic is known as elastic limit.
(d) Hardenability is the property of material which enables it to resist abrasion. [BE, AMIE]

2. The measure of ductility is
(1) Percentage elongation (2) modulus of resilience

(3) modulus of toughness (4) ultimate tensile strength (1)

3. High elastic modulus in materials arises from

(1) weak bonds (2) high strength of bonds

(3) shall potential well (4) none of the above (2)

4. Fine grain sizes are obtained by

(1) fast cooling (2) slow cooling

(3) increasing nucleation rate (4) decreasing growth rate (1, 3, 4)

5. For a tensile test, it can be demonstrated that necking begins at

(1) $\frac{d\sigma_T}{d\varepsilon_T} = \sigma_T$ (2) $\frac{d\sigma_T}{d\varepsilon_T} = \sigma_T^2 - 1$

(3) $\frac{d\sigma_T}{d\varepsilon_T} = \sqrt{\sigma_T} - 1$ (4) $\frac{d\sigma_T}{d\varepsilon_T} = \frac{1}{\sigma_T}$ (1)

6. If l_0 is the original length before any load is applied, and l_i is the instantaneous length, then the engineering strain ε is given by ($\Delta l = l_i - l_0$):

(1) $\varepsilon = \frac{\Delta l}{l_o}$ (2) $\varepsilon = \frac{l_i + l_o}{l_o}$

(3) $\varepsilon = \frac{(l_o + l_i)^2}{l_o}$ (4) $1 - \frac{l_i}{l_o}$

7. Proof stress corresponds to

(1) elastic limit (2) lower yield point

(3) higher yield point (4) a specified strain (4)

8. In a tensile test, the engineering stress corresponding to the maximum load is termed as

(1) Upper yield stress (2) Yield strength

(3) Tensile strength as well as UTS (4) None of the above (3)

9. In a tension specimen the elongation at the time of fracture is

(1) localized near the ends

(2) localized in the region of necking

(3) delocalized in the centre of the length

(4) none of the above (2)

10. In a tensile test, necking starts at

(1) Ultimate tensile stress (2) Lower yield stress

(3) Upper yield stress (4) None of the above (1)

11. Increasing strain rate in tension test

(1) decreases yield and ultimate tensile strength

(2) causes yield point to disapper

(3) causes yield point to appear in materials which did not show yield point at low strain rate

(4) none of the above (3)

12. Which phenomenon is not used in the measurment of hardness?

(1) Wear (2) Fracture

(3) Scratch (4) Indentation (2)

13. Which combination of alloying increases ductility transition temperature?

(1) Si and Mn (2) C and Fe

(3) C and O_2 (3) C and Mn (3)

14. For copper, the yield stress (σ_y) and the brittle fracture stress (σ_f) are related as
(1) $\sigma_y < \sigma_y$ (2) $\sigma_y > \sigma_f$
(3) $\sigma_y = \sigma_f$ (4) $\sigma_y >> \sigma_f$ (1)

15. The method to increase the yield strength of a crystalline material are
(1) Annealing (2) Grain refinement
(3) Solute additions (4) None of the above (2, 3)

16. Which hardness method can measure hardness of a grain?
(1) Knoop (2) Shore (3) Rockwell (4) Vickers (1)

17. If K_F is fatigue strength reduction factor due to stress concentration and K is stress concentration factor, then fully notch sensitive materials have
(1) $K_f = K$ (2) $K_f = 1$ (3) $K_f < K$ (4) $K_f > K$ (1)

18. If q is notch sensitive index, then materials insensitive to notch are characterized by
(1) $K_f = 1$ (2) $K_f = K$ (3) $K_f = q$ (4) $K_f > q$ (1)

19. The relation between K_F, K and q is
(1) $q = \dfrac{K_f - 1}{K_{f+1}}$ (2) $q = \dfrac{K_f + 1}{K_{f-1}}$
(3) $q = K_f + K$ (4) $q = \dfrac{K_f - 1}{K - 1}$ (4)

20. In the Griffith equation, the fracture stress is proportional to
(1) C (2) $2C$ (3) $\sqrt{C}$ (4) $1/\sqrt{C}$ (4)

21. The fatigue strength of mild steel is
(1) equal to its yield strength
(2) lower than its yield strength
(3) equal to its tensile strength
(4) lower than its tensile strength (2)

SHORT QUESTION-ANSWERS

1. What is stress?

Ans. Stress is a measure of an applied mechanical load or force, normalised to take into account cross-sectional area. Two different stress parameters are: *(i)* engineering stress and *(ii)* true stress.

2. What is strain?

Ans. It represents the amount of deformation induced by a stress. Both engineering and true strains are used.

3. How you will account for ordinary elastic behaviour of materials?

Ans. This can be attributed to the instantaneous stretching of atomic bonds. The stress in this case is proportional to strain, as given by Hooke's law. Strongly bonded materials have high modulus while weakly bonded materials have low modulus.

4. What are the methods available to increase the elastic modulus of a material?

Ans. There are several methods available, e.g. making composites and aligning chain and sheet molecules in directions of high bond strength.

5. What are simple-strain tests which can be used to ascertain of the mechanical characteristics of metals?

Ans. There are four test types: tension, compression, torsion, and shear. Tensile are the most common.

6. What is Poisson's ratio?

Ans. This represents the negative ratio of transverse and longitudinal strains.

7. When rubber like elasticity arises?

Ans. This arises from entropy changes during stretching, when the coiled long chain molecules uncoil.

8. When the phenomenon of yielding occurs?

Ans. This occurs at the onset of plastic or permanent deformation; yield strength is determined by a strain offset method from the stress-strain behaviour, which is indicative of the stress at which plastic deformation begins.

9. What is tensile strength and ductility?

Ans. Tensile strength corresponds to the maximum tensile stress that may be sustained by a specimen, whereas percents elongation and reduction in area are measure of ductility—the amount of plastic deformation that has occurred at fracture. Ductile materials are normally tougher than brittle ones.

10. What is resilience?

Ans. It is the capacity of a material to absorb energy during elastic deformation; modulus of resilience is the area beneath the engineering stress-strain curve upto the yield point.

11. What do you understand by static toughness?

Ans. It represents the energy absorbed during the fracture of a material, and is taken as the area under the entire engineering stress-strain curve.

12. What is hardness? Is there any relation between hardness and tensile strength?

Ans. Hardness is a measure of the resistance to localized plastic deformation. In several popular hardness-testing techniques (Rockwell, Knoop, Brinell, and Vickers) a small indenter is forced into the surface of the material, and an index number is determined on the basis of the size or depth of the resulting indentation. Hardness and tensile strengths for many metals are approximately proportional to each other.

13. What is a safe stress?

Ans. Due to uncertainties in both measured mechanical properties and inservice applied stresses, design or safe stresses are normally utilized for design purposes. Safe stress for ductile materials is the ratio of the yield strength and the factor of safety.

14. What is a plastic deformation?

Ans. On a microscopic level, plastic deformation corresponds to the motion of dislocations in response to an externally applied shear stress, a process termed ‘slip’.

15. What is a slip?

Ans. Slip is a common mode of plastic deformation at ambient and elevated temperatures. Slip occurs on specific crystallographic planes and within these planes only in certain directions. A slip system represents a slip plane-slip direction combination, and operable slip system depend on the crystal structure of the material.

16. How you will account for the difference in observed and theoretical shear strengths?

Ans. The observed shear strengths of crystals is about three or four orders of magnitude smaller than the theoretical strengths. The motion of dislocations at low stresses accounts for this discrepancy. We may note that only in perfect crystals such as whiskers, the theoretical shear strength is attainable.

17. What is critical resolved shear stress?

Ans. This is the minimum shear stress required to initiate dislocation motion; the yield strength of a single crystal depends on both the magnitude of the critical shear stress and the orientation of slip components relative to the direction of the applied stress. For polycrystalline materials, slip occurs within each grain along the slip systems that are most favourably oriented with the applied stress; furthermore, during deformation, grains change shape in such a manner that coherency at the grain boundaries is maintained.

18. What do you understand by work hardening?

Ans. This is due to the formation of sessile (immobile) dislocations by the interaction of dislocations moving on non-parallel slip planes. The plastic flow strength increases during work hardening as the square root of the dislocation density.

19. How you can restore the microstructural and mechanical characteristics of a plastically deformed metal specimen to their performed states?

Ans. This can be achieved by an appropriate heat treatment, during which recovery, recrystallization, and grain growth processes are allowed to occur. During recovery there is a reduction in dislocation density and alternations in dislocation configurations.

20. What is grain growth?

Ans. It is the increase in average grain size of polycrystalline materials which proceeds by grain boundary motion.

21. What is the function of grain boundaries?

Ans. These provide effective obstables to dislocation motion. The yield stress of a polycrystalline material increases as the reciprocal of the square root of the grain diameter.

22. What is creep?

Ans. This is the thermally activated deformation that occurs as a function of time at temperatures above 0.4 T_m. Creep resistance can be improved effectively by dispersion hardening method.

23. What is fatigue?

Ans. This is a common type of catastrophic failure where-in the applied stress level fluctuates with time.

24. Mention characteristic fatigue surface features.

Ans. Beachmark and striations. Beachmarks form on components that experience applied stress interruptions; they normally may be observed with the naked eye. Fatigue striations are of microscopic dimensions and each is thought to represent the crack tip advance distance over a single load cycle.

9

Alloy Systems, Phase Diagrams and Phase Transformations

1. INTRODUCTION

Metals in actual commercial use are almost exclusively alloys, and not pure metals, since it is possible for the design engineer to realize an infinite variety of physical properties in the product by varying the metallic composition of the alloy. A homogeneous mixture of two or more metals or a metal and a non-metal when fused together at a certain temperature forms a new metal after solidification, termed as an alloy. Alloys are normally harder than their components, less ductile and may have a much lower conductivity, whereas the highly purified single crystal of a metal is very soft and malleable, with high electrical conductivity. This is why pure metals are used only for specific applications. The alloy is usually more corrosion resistant and less affected by atmospheric conditions. The conductivity of an alloy varies with the degree of order of the alloy and the hardness varies with the particular heat treatment.

The basic knowledge of structural properties of alloys is still in large part empirical, and indeed, it will probably never be possible to derive formulas which will predict which metals to mix in a certain proportion and with a certain heat treatment to yield a specified property or set of properties. However, a set of rules exists which describes the qualitative behaviour of certain group of alloys. These rules are statements concerning the relative sizes of constituent atoms for alloy formation, and concerning what kinds of phases to expect in terms of the valence of the constituent atoms. The rules were discovered in a strictly empirical way, and for the most part, the present theoretical understanding of alloys consists of rudiomentary theories which describe how the rules arise from the basic principles of physics.

In the solid state an alloy may be present in one or more of the following forms:

(i) As a solid solution
(ii) As an intermediate phase or intermediate chemical compound
(iii) As a finely divided mechanical mixture of solid solution
(iv) As a finely divided mechanical mixture of the metals
(v) As a finely divided mechanical mixture of chemical compounds of metals, the individual metals and solid solutions.

Alloys are classified as *binary alloys*, composed of two components; as *ternary alloys*, composed of three components; or as *multicomponent* alloys. Most commercial alloys are multicomponent. The composition of an alloy is described by giving the percentage (either by weight or by atoms) of each element in it.

Metal alloys by virtue of composition, are often grouped into two classes: *(i) ferrous* and *non-ferrous*. Ferrous alloys are those in which iron is the principal constituent, include steels and cast irons. The non-ferrous alloys are all alloys that are not iron based.

Alloys are widely used in industry because their physical and chemical properties can be easily varied to suit the exact individual requirement. One can achieve this by preparing alloys of different metals.

Metallic objects are invariably manufactured from alloys instead of pure metals. The alloying elements are added to improve one or more of the following properties: *(a)* tensile strength, hardness and toughness *(b)* corrosive and oxidation resistance, *(c)* machinability, *(d)* elasticity *(e)* hardenability *(f)* creep strength and *(g)* fatigue resistance, etc.

The characteristic of any series of alloys can be easily studied by using alloy *phase diagrams*. A phase diagram is a graphical description of the kinds and amounts of the phases that can be expected in an alloy as a function of its composition, temperature, and pressure, when it has reached thermodynamic equilibrium. The phases may be liquid, vapour or solid with various ordered and disordered crystal structures. A phase diagram does not provide information about how rapidly equilibrium can be reached. When a phase diagram is determined experimentally, it is necessary not only to find out what phases are present but also to assure that the alloy is in the stable equilibrium state. Equilibrium is reached when the Gibbs free energy of the system has reached its lowest possible value. The thermodynamic principle of minimum Gibbs free energy imposes restrictions on the possible forms a phase diagram can take, and even a complicated diagram with many phases is composed of only a few geometrical features.

The following terms are frequently used in the study of solid phases and phase diagrams:

(i) System: A system may be composed of solids, liquids, gases or their combinations and may have metals and non-metals separately or in any combination. It is explained as the whole complex of phases of one or several components at different pressures and compositions. A system is so isolated from its surroundings that it is unaffected by solids, liquids, gases or their combinations and is subjected to the change in the overall composition, temperature, pressure or total volume, only to the extent permitted by the investigator.

(ii) Components: These are the substances, either chemical elements or chemical compounds, whose presence is essential and sufficient to make a system. A pure metal is called a one-component system, an alloy of two metals is called a binary or two component system, etc.

(iii) Phase: It is a homogeneous portion of a system that has uniform physical and chemical characteristics. The number of phases in a system is the number of different substances that exist in it in a homogeneous system.

(iv) Phase diagram: A graphical representation of the relationships between environmental constraints (e.g. temperature and sometimes pressure), composition, and regions of phase stability, ordinarily under conditions of equilibrium.

Most phase diagrams are prepared by using slow cooling conditions whereby phases are in equilibrium. One can get following important informations from the phase diagrams

(a) Phases at different composition and temperature
(b) Equilibrium solubility of one element or compound in another element.
(c) Melting points of different phases in an alloy.
(d) Temperature of solidification or range of solidification of an alloy.

(v) Phase Equilibrium: The state of a system where the phase characteristics remain constant over indefinite time periods. At equilibrium the free energy is a minimum.

In an equilibrium diagram, liquid is one phase and solid solution is another phase.

(vi) Phase Transformation: A change in the number and/or character of the phases that constitute the microstructure of an alloy.

2. ALLOY SYSTEMS

The improvement in the properties of an alloy system depends upon the following factors:

(i) Manner in which the two or more metals are mixed with each other.
(ii) The percentage of different alloying metals/or elements.
(iii) Temperature at which these are cooled, etc.

It is possible that two or more metals may be soluble in each other in liquid state but may or may not be soluble in each other in solid state. It is possible that they may retain their identity even if they are soluble in liquid state, e.g. cadmium and bismuth are soluble in each other in liquid state but insoluble in each other in the solid state. It is also possible that the two or more metals may be soluble in each other in liquid as well as solid state, e.g. copper and nickel are soluble in each other in the liquid as well as in solid state. Obviously, one cannot distinguish copper from nickel.

When an alloy is in a liquid state the atoms of the constituent are distributed randomly through the liquid. When solidification takes place, there appears number of possibilities. A number of different types of solutions may form, as follows:

(a) Simple Eutectic Type In this case the two components of a alloy system (binary) are soluble in liquid state but separate out in the solid state, each maintaining its own separate identity. In the solid state, the two components are said to be insoluble in each other. The solution is called *eutectic type*.

(b) Solid Solution Type When the two components of a binary alloy remain completely mixed in each other both in liquid and solid state, the two components are said to be soluble in each other and a different type of solution may be formed. It is called a solid solution.

(c) Combination Type On solidifying, the two components of a binary alloy may show limited solubility in each other. This type of solution combines characteristics of both components.

(d) Inter-metallic Compounds It is observed that the elements may combine to form inter-metallic compounds on solidification, when their affinity is great. These types of compounds may find place in between the solid solution and chemical compound. A familiar example is of copper-zinc system. When the solutility, of copper in zinc is exceeded, a zinc rich β-phase appears with the Cu-rich α-phase.

In general, inter-metallic compounds are hard and brittle and can be used as bearing metals. These compounds wear resistant pads with a low co-efficient of friction, set in a matrix of a tough solid solution.

3. SOLID SOLUTIONS

A solid solution forms when, as the solute atoms are added to the host material, the crystal structure is maintained, and no new structures are formed. In other words, when elements completely dissolve in each other in liquid and or solid state the resulting phase is called *solid solution.* Solid solutions form readily when solvent and solute atoms have similar sizes and electron structures, so that it is compositionally or chemically homogeneous and the component atoms of the elements cannot be distinguished physically or separated mechanically. There is a homogeneous distribution of the constituents in the solid state so as to form a single phase or solid solution. Basically, solid solutions are of two types:

(a) Substitutional Solid Solution When the two metals in solid solution form a single face centred cubic lattice, i.e., in general solute or impurity atoms replace or substitute for the host atoms, is called as substitutional solid solution. One may also define solid solutions as crystalline phases of a variable composition. Atoms of a solute *B* can be arranged in the crystal lattice of a solvent *A* either by substituting for some atoms of the latter in the crystal lattice or by occuping the interatomic spaces between atoms, or interstices. We may note that the atoms of the two metals be of almost same size, i.e., the difference in atomic radii between the two atom types be less than about ±15%. Otherwise the solute atoms will create substantial lattice distortion and a new phase will form. For appreciable solid solubility the crystal structures for metals of both atoms must be same. If one element has more electropositivity and the other more electronegativity, then there is greater likelihood that they will form an intermetallic compound instead of a substitutional solid solution. It is also observed that other factors being equal, i.e. atomic size, crystal structure and electronegativity, a metal will have more of a tendency to dissolve another metal of higher valency than one of a lower valency. A familiar example of substitutional solid solution is found for copper and nickel to form monel. These two elements are completely soluble in another at all proportions. Brass, an alloy of copper and zinc, is another example, which forms solid solutions most readily as the atoms of these two elements have similar sizes and electronic structure unlimited solubility in the solid state can be observed in copper-gold, and germanium-silicon alloys. Polymorphous metals may possess unlimited solubility within a single modification of the space lattice. For example, Fe_α can form a continuous series of solid solutions with Cr (BCC lattices) and Fe_γ, a continuous series of solid solutions with Ni (FCC lattices). The formation of solid solutions is always associated with an increase of electric resistance and decrease of the temperature coefficient of electric resistance. Solid solutions are usually less plastic (except for copper-based solid solutions) and always harder and stronger than pure metals. Substitutional solid solutions are of two types:

(i) Random substitutional solid solutions and *(ii)* Ordered substitutional solid solutions.

When there is no order in the substitution of the two metal elements (Fig. 9.1(a)), the chance of one metal element occupying any particular atomic site in the crystal is equal to the atomic percent of the element in the alloy. Obviously, in such a situation the concentration of solute atoms can vary considerably throughout the lattice structure. Such type of the resulting solid solution is called a *random* or *disordered* substitutional solid solution.

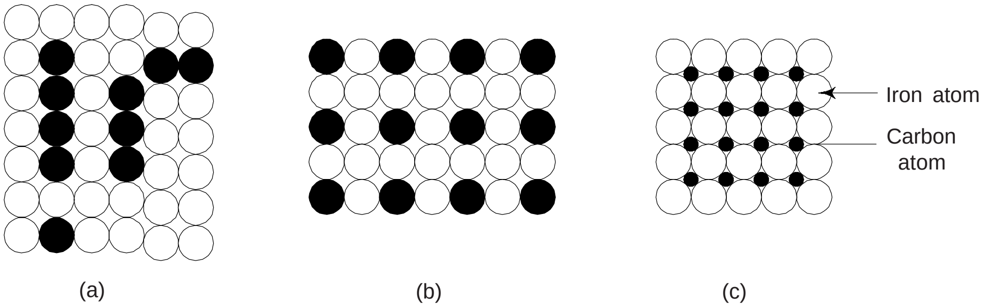


Fig. 9.1 (a) Solid solutions (b) substitutional solid solutions and (c) interstitial solid solution

Many substitutional solid solutions can be in an ordered state at relatively low temperatures, i.e. unlike atoms of two elements are distributed between the sites of the space lattice not statistically, but in a perfectly definite order. Such solid solutions are called *ordered*; the term '*superstructure*' is also in use. Such ordering is common at low temperatures because greater thermal agitation tends to destroy the orderly arrangement. [Fig. 9.1(b)].

The change from a disordered to ordered state takes place at a definite temperature or in a definite temperature range. The temperature at which a solid solution becomes completely disordered is called the *Kurnakov point* (θ_K). Ordering usually takes place under the conditions when a solid solution is cooled slowly from the temperature region above θ_K.

Ordered solid solutions can be found in systems with either substantial or unlimited solubility in the solid state; in that case complete ordering can occur with concentrations of the solid solution elements corresponding to simple atomic ratios of the components of the type, say AB or AB_3. Partial ordering may be observed in systems with concentration ratios close to the indicated ones. The arrangement of atoms in ordered solid solutions of two Cu and Au alloys of the corresponding to $AuCu_3$ and AuCu are shown in Fig. 9.2.

The appearance and disappearance of an order in the arrangement of atoms in solid solutions is associated with changes in the properties of an alloy.

Ordering increases the electrical conductivity, temperature coefficient of electric resistance, hardness, and strength, but decreases the ductility of an alloy. In ferromagnetic alloys, it changes magnetic properties; for instance, the magnetic permeability of perm alloys (magnetic iron-nickel alloys) can drop on ordering to a small fraction of the original value. Some alloys are paramagnetic in disordered state, but become ferromagnetic after ordering, such as Heusler (Mn-Cu-Al) alloys.

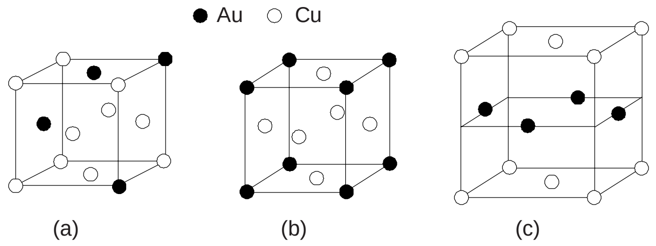


Fig. 9.2 Crystal lattice of Cu-Au alloys: (a) disordered solid solution; (b) ordered solid solution ($AuCu_3$ alloy); (c) ordered solid solution (AuCu alloy)

Usually, substitutional solid solutions have a *random arrangement* of the constituent atoms on the atomic sites, especially at elevated temperatures. This is so, as configurational entropy makes a greater contribution in lowering the free energy with increasing temperature, as we have $G = H – TS$. This random arrangement of the constituent atoms in a solid solution may change over to an ordered arrangement on cooling to lower temperatures, if ordering lowers the enthalpy of the crystal sufficiently.

(b) Interstitial Solid Solutions These can form, for instance, on melting together transition metals and non-metals with a small atomic radius (*H*, *N*, *C* or *B*).

The possibility of obtaining an interstitial solution is mainly determined by the size factor; i.e., the size of a solute atom must be equal to or slightly smaller than the size of an interstitial void.

Interstitial solid solutions always have a limited solubility and form preferably in solvents having an HCP or FCC lattice with interstices of a radius of 0.41 *R*, where *R* is the radius of a solvent atom. In BCC lattices, the solubility is low, since the size of interstices does not exceed 0.29 *R*.

Examples of interstitial solid solutions of commercial application are solid solution of carbon in Fe_{γ}, and Fe_{α}, Fe_r, which has an FCC lattice, can dissolve upto 2.14% (by mass) of carbon; Fe_{α} (BCC lattice) does not practically dissolve carbon, the maximum solubility being around 0.02 (by mass). *Other elements like:* nitrogen for maintaining stainless steel in austenitic condition for nitriding condition and hydrogen when introduced into steels during welding operations, acid cleaning and plating results in hydrogen embrittlement and causes a sharp decrease in ductility.

Lattice distortions appearing on the formation of interstitial solid solutions are stronger than those in the substitutional solid solutions, because of which the properties of alloys can be changed more drastically. As the concentration of a solute in a solid solution increases, this increases noticeably the electric resistance, coercive force, hardness and strength, but decreases the ductility and toughness.

Interstitial solid solutions usually have very limited solubility and are normally considered of secondary importance.

Alloys containing more than two elements can form either substitutional or interstitial solution. For example, when iron is melted together with manganese and carbon, manganese and iron form substitutional solid solution while carbon and iron, interstitial solid solution.

In conclusion we can say that solid solutions are crystals whose properties are close to those of the solvent, since they retain the same crystal lattice and type of bond. In particular, solid solutions of metals are distinguished by high plasticity: they are readily deformable in the hot state and many of them, in the cold state.

Solid solutions are the bases of most commercial structural and special alloys.

4. THE FAMILIES OF ENGINEERING ALLOYS

Engineering alloys are generally placed under following three groups:

(i) The light alloys, based on aluminium, magnesium and titanium. Aluminium Alloys: About 20% of the world production of aluminium is used for cast products. Aluminium alloys have a relatively low melting

temperature, but exhibit a high shrinkage during solidification. Shrinkage of between 3.5 and 8.5% may occur, and allowance has to be made for this in mould design in order to achieve dimensional accuracy. This group includes cast aluminium alloys, aluminium-silicon alloys, aluminium-copper alloys, aluminium-magnesium alloys, aluminium-zinc-magnesium alloys, wrought aluminium alloys. There are two important aluminium-alloy families under the heading *non-heat-treatable alloys*: aluminium-manganese alloys (3 XXX series) and aluminium-magnesium alloys (5 XXX series). There are three important aluminium families under the head *heat-treatable alloys*: aluminium copper alloys (2 XXX series), aluminium-magnesium-silicon alloys (6 XXX series) and aluminum-zinc-magniesium alloys (7 XXX series).

Magnesium Alloys: This includes cast magnesium alloys, wrought magnesium alloys.

Titanium Alloys: There are three types of Ti alloy microstructure, namely those with α, β and mixed α/β structures.

Copper Alloys: Copper and its alloys are widely used because of their high electrical and thermal conductivity, corrosion resistance, and ease of fabrication. Copper alloys mainly includes: Copper alloy of low solute content, copper alloys of high solute content, copper-nickel alloys, copper-tin alloys or bronzes, copper-rich aluminium alloys known as aluminium bronzes, etc.

Lead Alloys: The most significant applications of lead alloys are lead-acid storage batteries, and it is also widely used for building construction materials such as sheet because of its relative inertness to atmospheric attack. Lead is also finding increased applications as a material for controlling sound and mechanical vibration, due to its high damping capacity.

Because of its low melting point (327.5°C), pure lead undergoes creep at room temperature, so it usually strengthened by small additions of solute. Thus 1.5 to 3% antimony or traces of calcium are often present in the plates of car batteries, and different manufactures may use different solutes.

The important families of more concentrated alloys are the lead-tin solders and lead-based bearing alloys, which may contain antimony, tin and arsenic, for use in internal combustion engines.

Zinc Alloys: The main uses of Zn are as an alloying element (e.g. in copper) and as a protective coating for steel (galvanising, etc.). However, Zn-based alloys are also used for the production of gravity castings, and pressure die-castings of high dimensional tolerance.

The casting alloys are based on the Zn – Al system, and are close to eutectic composition of 5% Al. The hyperutectic alloys solidify with Zn-rich dendrites, whereas the hyperutectic contain Al-rich dendrites. The commercial alloys contain strengthening additions of copper and magnesium, but the greatest care has to be taken to prevent excessive pick-up of harmful impurity elements such as Pb, Cd, Sn, and Fe. The alloys themselves are prepared from high-purity components, otherwise intergranular embrittlement of the castings develops over a period of time.

Nickel Alloys: Commercially pure nickel offers excellent corrosion resistance to reducing environments (in contrast to those metals and alloys which owe their resistance to the presence of a tenacious oxide film), and are found in the chemical processing industry, as well as in food processing applications. As an electroplated coating, nickel is widely used in the electronics industry.

Ni-Cu Alloys: These alloys possess excellent corrosion resistance, notable in sea water. The monel (~30% Cu) series of alloys is used for turbine blading, valve parts and for marine propeller shafts, because of their high fatigue strength in sea water.

Ni-Cr Alloys: These alloys form the basic alloys for jet engine development – the Nimonic alloys, or Ni-based superalloys. The limitation of the conventionally cast supperalloys was their lack of creep ductility due to cavitation at the grain boundaries lying perpendicular to the maximum tensile stress, so directionally solidified and eventually single crystal alloys were developed.

Steels: This forms a large group of engineering alloys. Usually these alloys are divided into three groups: *(i)* low carbon steels, *(ii)* engineering steels, and *(iii)* stainless steels. Low carbon steel group includes —strip steels and structural steels. Engineering steel group includes heat-treated steels, pearlitic steels and marageing steels. Stainless steel alloys owe their passivity to the presence of a thin protective film of Cr_2O_3. This group of alloys include ferrite stainless steels, martensitic steels and austenitic stainless steels.

Cast Irons: These are shaped by casting into a mould rather than by forging in the solid-state. They are alloys of iron that usually contain between 2.5 and 4% carbon (and from 1 to 3% silicon, which tends to promote the appearance of the carbon as graphite, rather than as carbide, Fe_3C). A binary Fe – 4% C alloy is close to eutectic composition and have a low melting temperature. Cast irons are very fluid when molten and have good casting characteristics, but the casting usually have relatively low impact resistance and ductility, which may limit their use.

The mechanical properties of cast irons depend strongly on their microstructure, and there are three basic types—white iron, grey iron and ductile iron.

Nodular, or Spherodal Graphite (SG) cast iron: This contains graphite spheroids in the as-cast state, by the addition of cerium and/or magnesium to the iron. SG iron can be made with a pearlite matrix, or ferrite with appropriate heat-treatment, or with accicular or austenitic matrix when suitably alloyed, and they behave as more or less normal ductile ferrous materials.

In view of the large number of compositions of steels and cast commercially available, and the fact that their properties may be varied over a wide range by appropriate heat treatment, it is beyond the scope of this book to present the details over here.

Whenever higher strengths are required, then *brazed* joints are employed. *Soldering and brazing* are the processes whereby the basic metals are wetted by the filler metals, with subsequent filling of the joint gaps by capillary action. If the process is carried out in air, oxidation of the metals will occur and therefore fluxes are commonly used in both techniques to dissolve the oxide films and to ensure wetting by the filler metal of the metals to be joined. For higher strengths, normally *Al-based brazing alloys*, *Cu-based brazing alloys*, *Ni-based brazing alloys* are used.

Brazed joints of all composition are significantly stronger than those in soft solder, with copper-based joint strengths being in the range 250-400 MPa and nickel-based joints in the range 300-600 MPa, thus often equalling that of the metals being joined. It is however to welded joints that the highest strengths are normally encountered.

5. HUME-ROTHERY'S RULES

While developing an alloy, it is frequently desirable to increase the strength of the alloy by adding a metal that will form a solid solution. Hume-Rothery has framed *empirical rules* that govern the choice of alloying elements in the formation of substitutional solutions. We may note that if an alloying element is chosen at random, it is likely to form an objectionable intermediate phase instead of a solid solution. Extensive solid solubility by substitution occurs, when

(i) the solute and solvent atoms do not differ by more than 15% in size, i.e. diameter. Within this limit of size factor, each of the metals will be able to dissolve appreciably (to the order of 10%) in the other metal. However, if the atomic size factor is greater than 15%, solid solution formation tends to be severely limited and is usually only a fraction of one percent.

(ii) the electronegativity difference between the elements is small. If the chemical affinity of two metals is greater, then the solid solubility will be more restricted. When the chemical affinity of two metals is great, they tend to form an intermediate phase rather than a solid solution.

(iii) the valency and the crystal structures of the elements are the same. If the alloying element has a different valence from that of the base metal, the number of valence electrons per atom (called the electron ratio), will be changed by alloying. We may note that the crystal structures are more sensitive to a decrease in the electron ratio than to an increase. Obviously, a metal of high valence can dissolve only a small amount of lower valence metal; whereas the lower valence metal may have good solubility for a higher valence metal. Moreover, only metals that have the same type of lattice (for example FCC) can form a complete series of solid solutions. For complete solubility, it is found that the size factor must usually be less than 8%. Examples of binary and ternary systons exhibiting complete solid solubility are Cu-Ni and Ag-Au-Pt respectively.

Ag-Au, *Cu-Ni* and *Ge-Si* are the systems which satisfy Hume Rothery conditions very well (Table 9.1). Obviously, these systems form *complete solid solutions*, i.e. the elements mix in each other in all proportions.

Table 9.1 Parameters for few systems relevant to Hume Rothery Rules

System		*Crystal Structure*	*Atomic radius (Å)*	*Valency*	*Electro-negativity*
(i) Ag-Au	Ag	FCC	1.44	1	1.9
	Au	FCC	1.44	1	2.4
(ii) Cu-Ni	Cu	FCC	1.28	1	1.9
	Ni	FCC	1.25	2	1.8
(iii) Ge-Si	Ge	DC	1.22	4	1.8
	Si	DC	1.13	4	1.8

6. INTERMEDIATE PHASES OR INTERMEDIATE COMPOUNDS (OR INTERMEDIATE SOLID SOLUTIONS)

Crystals formed by various elements and having their own type of crystal lattice which differs from the crystal lattices of the component elements are called intermediate phases.

Fe_3C (iron carbide), a common constituent of steels, is an example of intermediate phase. It has a complex crystal structure referred to an orthorhombic lattice and is hard and brittle.

Intermediate phases may have any type of bond which depends on their components and determines largely the properties of the crystals, in particular, electric properties.

The arrangement of atoms (or ions) in the lattice may be disordered or partially or completely ordered. As in solid solutions, ordering sharply changes the properties, in particular, bring about the effect of super conductivity.

Like solid solutions, intermediate phases are crystals whose composition varies within a certain, sometimes very narrow, range of concentrations. Their variable composition can be explained by either the presence of interstitial 'extra' atoms (or ions) in the crystal lattice of intermediate phases or deficiency of atoms in the lattice site (Fig. 9.3).

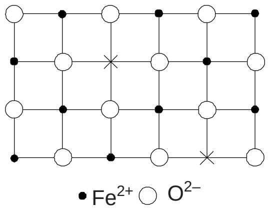


Fig. 9.3 Crystal lattice of FeO with a deficiency of metal ions

As solid solutions, intermediate phases are designated by Greek letters. It is also permissible to designate them by chemical formulae which reflect their stoichiometric composition, i.e., that of crystals having no defects: interstitial atoms or vacancies

Numerous and diverse intermediate phases have not been classified properly yet. The structure of an intermediate phase is noticed to depend on three factors: *(i)* relative size of atoms, *(ii)* valencies, and *(iii)* positions of elements in the Periodic Table of elements which determine their electronic structure. The intermediate phases of variable composition which do not obey the valency law are called *electron phases* or *electron compounds*. Hume Rothery has shown that electron phases occur at certain definite value of free electron to atom ratio in the alloy such as 3 : 2, 21 : 13 and 7 : 4. Few typical examples of electron phases are CuZn (3 : 2), Cu_5Zn_8 (12 : 13) and $CuZn_3$ (7 : 4).

7. PHASE DIAGRAMS

The study of phase relationships plays an important and vital role in the better understanding of the properties of materials. Much of the information about the control of microstructure or phase structure of a particular alloy system is properly displayed in what is called a *phase diagram*, also called as an *equilibrium or constitutional diagram*. Phase diagrams are clear maps that give the relationships between phases in thermodynamic equilibrium in a system as a function of temperature, pressure and composition.

The phases may be liquid, vapour, or solid with various ordered and disordered crystal structures. A phase diagram does not provide information about how rapidly equilibrium can be reached; when a phase diagram is determined experimentally it is necessary not only to find out what phases are present but also to assure that the alloy is in the stable equilibrium state. Equilibrium is reached when the Gibbs free energy of the system has reached to its lowest possible value. The thermodynamic principle of minimum Gibbs free energy imposes restrictions on the possible forms a phase diagram can take, and even a complicated diagram with many phases is composed of only a few geometrical features.

Many microstructures develop from phase transformations, the changes that takes place between phases when the temperature is altered (ordinarily upon cooling). This may involve the transition from one phase to another, or the appearance or disappearance of a phase. One can predict the phase transformations and the resulting microstructures with the help of phase diagrams, which may have equilibrium or non-equilibrium character.

We may note that the phases as determined from the microstructure of a material may or may not correspond to those included in the phase diagram. Interestingly, the thermal treatment given to a material often results in phases other than those in the diagram. However, this does not diminish the importance of the study of phase diagrams as a initial step in getting an insight into the control of microstructure.

Phase diagrams are of following three types:

(i) Unitary or single-component phase diagram
(ii) Binary or two-component phase diagram
(iii) Ternary or three-component phase diagram

Binary phase diagrams are extensively used.

8. THE PHASE RULE OR GIBB'S PHASE RULE OR CONDENSED PHASE RULE

This expresses mathematically the general relationships for the existence of stable phases corresponding to the equilibrium conditions. It enables us to predict and check the processes that occur in alloys during heating or cooling. Using this rule, it is possible to determine whether the solidification process takes place at a constant temperature or within a certain temperature interval; it can also indicate the number of phases that can exist simultaneously in a system.

The phase rule enunciated by J.W. Gibbs relating number of phases P, number of components C, and number of degrees of freedom F has a simple form:

$$P + F = C + n \tag{1}$$

$$P + F = C + 2$$

$\because$ n = number of external factors = 2 (temperature and pressure)

In applying the phase rule to metal systems the effect of pressure is neglected, leaving only one variable factor,—temperature. Equation (1) reduces to

$$F = C + 1 - P \qquad 1(a)$$

The number of degrees of freedom is essentially the number of independent variables, both internal (composition and phases) and external ones (temperature, pressure, concentration etc.), which can be changed without changing the number of phases in equilibrium. The number of indepenent variables cannot be more than the number of variables, i.e.

$$F = C - P + 2 \le P(C - 1) + 2 \tag{2}$$

where $P(C-1)$ denotes the total number of compositional variables when P phases are there in the system. Including the two external variables (pressure and temperature), the total number of variables is $P(C-1) + 2$.

In equilibrium all factors have definite values, hence the degrees of freedom cannot be less than zero,

$$C - P + 1 \ge 0$$

$$\text{then} \quad P \le C + 1$$

Obviously, the number of phases in a system cannot exceed the number of components plus one.

The components of a system may be elements, ions or compounds. The components refer to the independent chemical species that comprise the system. In the ice-steam system, the component is H_2O, in the Cu-Ni system the components are the elements Cu and Ni, whereas in the Al_2O_3-Cr_2O_3 system, one can take the two oxides to be components. In the Fe-C system, although Fe and graphite can be taken as components, but it may be convenient to choose Fe and Fe_3C (iron carbide) as the components.

As follows from the phase rule, the number of phases existing simultaneously in a binary system cannot be more than three. These three phases can only exist at a definite phase composition and a definite temperature. If the number of phases in a binary system turns out to be more than three, this means that either the alloy is not in the equilibrium state or the number of phases and constituents has been determined incorrectly. In a ternary system no more than four phases may be in equilibrium. When only one phase is present in a system, the degrees of freedom are equal to the total variables, with the increase in number of phases, the degrees of freedom decrease. The degree of freedom cannot be less than zero. Obviously, we have an upper limit to the number of phases that can exist in equilibrium in a given system.

At solidification temperature, a pure metal is a one-component system consisting of two phases of identical composition

$$F = 1 + 2 - 2 = 0$$

Clearly, the number of phases and number of degrees of freedom equals zero, i.e. $F = 0$. This is known as *non-variant equilibrium*. When the number of phases is less than the maximum possible number by one, the number of degrees of freedom will also increase by one ($F = 1$). Such type of system is called as *monovariant*.

An alloy of two metals is a two-phase and two-component system at solidification, $F = 1$. When $F = 2$, the system is said to be a *divariant*. Obviously, a system may be in equilibrium at different temperatures and concentrations.

One can represent all transformations occurring in alloys and depending on temperature and concentration ($C = 2$) by equilibrium diagrams. The equilibrium diagrams are plotted with concentration as the abscissa and temperature as ordinate.

9. COOLING CURVES (TIME-TEMPERATURE CURVES)

It is interesting to study the manner in which temperature changes with time as the liquid metal solidifies. Figure 9.4(a) shows a cooling curve which is distinctly divided into two portions while exhibiting the fall of temperature of time, the cooling curve exhibit that the temperature remains practically constant over a period of time.

This constant temperature is called as the point of arrest. The solidification occurs during temperature arrest. During this period, heat is still lost from the mass of metal but release of kinetic energy compensates the heat loss whereby temperature remains constant. The released heat at constant temperature is called the *latent heat*. Due to this, at certain stages the fall in temperature of the metal or alloy is totally arrested for a specific time. Such points are termed *critical points*. One can determine the critical points by the abrupt inflection of curves due to thermal effects in transformations. Following three types of cooling curves are commonly used in plotting the phase diagrams.

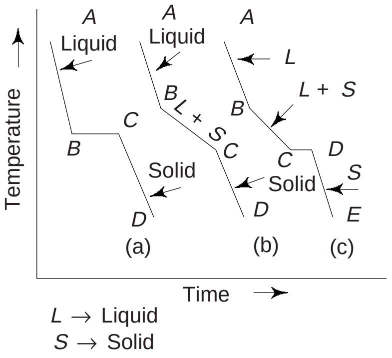


Fig. 9.4

(i) Curve (a): Applying Eq. 1 under constant pressure, for region *AB* when $P = 1$, $C = 1$, $F = 1$, i.e. system has single degree of freedom (called univariant). Obviously, only variant that changes is temperature. Between *B* and *C* both liquid and solid phases are present, $P = 2$, $C = 1$ so that $F = 0$, i.e. system has no degree of freedom (called as *non-variant*). Obviously, temperature remains constant (pressure is already constant) and the mass between *B* and *C* is *marshy* (partly liquid and partly solid). On further cooling from *C* to *D* the system

reaches room temperature. The slips of lines *AB* and *CD* indicate specific heat of liquid and solid metals, respectively (Fig. 9.4a). There is also another way to plot the cooling curve. One can plot the temperature against time taken to cool through a specific temperature interval (say 5°C) as shown in Fig. 9.5. Such cooling curves are called derived or inverse rate curve.

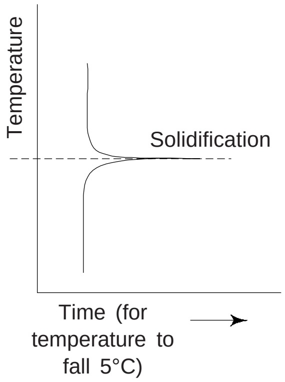


Fig. 9.5 Inverse rate or derived cooling curve.

We may note that the single temperature arrest during latent heat loss in cooling curve (Fig. 9.4) is the property of the pure metal. The temperature of solidification is the characteristic of that metal.

When two or more metals are mixed in liquid state to form an alloy and allowed to cool, the solidification occurs over a range of temperature. Figure 9.4(b) shows a cooling curve for an alloy of metals *A* and *B*. We can see that curve *AB* is the same as for pure metals. The freezing line *BC* drops until the whole mass is solid at point *C*. Applying Eq. (1) to the system between *B* and *C* (Fig. 9.4b) with $P = 2$, $C = 2$, $F = 1$, i.e. there is one degree of freedom. The temperature will change (the pressure is constant). Obviously *BC* portion of the curve is due to the fall in temperature recorded during freezing in a binary alloy. From point *C*, the solid further cools along the line *CD* to reach room temperature.

Figure 9.4(c) is the freezing curve for another binary system whose two components are completely soluble in liquid state but not at all soluble in solid state. They are liquid along *AB* upto point *B* of the cooling curve. At point *B* the component with larger content starts solidifying and temperature falls along *BC*. At point *C* the components solidify simultaneously at constant temperature, the lowest for a given alloy system, and are termed as eutectic alloys. At *D* the only phase that is present is solid and cools along *DE* as usual. Point *D* on the curve is called as *eutectic point*.

We may now consider the cooling curve for pure metal for evaluating the effect of slow and fast cooling. Figure. 9.6 shows two cooling curves for the same pure metal. We may note that constant temperature of crystallization is available only if cooling maintains equilibrium with surrounding. From Fig. 9.6, we can see that if cooling is rapid enough than temperature does not remain constant during solidification. Obviously, the more rapid the cooling lower is the temperature at which crystal formation begins to occur. This is due to rearrangement of molecules or atoms during freezing will need some time and rapid cooling may not be very conductive for such rearrangement. From Fig. 9.3 it is evident that the solidified portion of the metal will release heat, which is latent heat of fusion and will thus increase the temperature of surrounding molten metal. The phenomenon of temperature rise of molten metal is called *recalescence* which may sometimes may even cause the growing of molten metal.

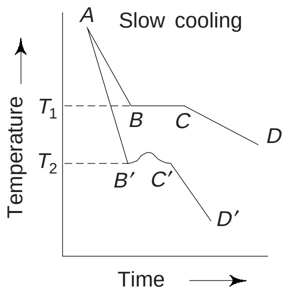


Fig. 9.6 Slow and rapid cooling curves for a pure metal

10. CONSTRUCTION OF A PHASE DIAGRAM OR CONSTITUTIONAL DIAGRAM

Depending upon the number of components and solubility characteristic, the phase diagrams are usually categorised as follows:

(i) Solid Solution Type: In this case two metals are completely soluble in solid as well as in liquid state. They have the same type of lattice and similar atomic size. Copper and Nickel form an isomorphous system.

(ii) Eutectic Type: When two metals are completely soluble in the liquid state but partly or completely insoluble in the solid state, is termed as *eutectic type*. Fe-C, Al-Mn, Pb-Sn form an eutectic system.

(iii) Peritectic Type: In this case liquid and solid combine to form a new solid. The melting points of two metals differ considerably. Ag and Pt form such a system.

In the eutectic system, the crystals of β-solid solution precipitated at the beginning of solidification react with the liquid alloy having definite composition to form new crystals of α-solid solution. In the peritectic reaction, two phases are used to produce one different phase with reaction just the opposite of eutectic reaction.

The *peritectic reaction* similar to other systems of solidification of different metals, but is comparatively less common. This reaction also occurs at constant temperature. One can write this reaction as

$$\text{Liquid} + \text{Solid } 1 \xrightleftharpoons[\text{heating}]{\text{Cooling}} \text{New solid } 2$$

i.e.,

$$L + \beta \xrightarrow[\text{heating}]{\text{Cooling}} \alpha$$

The temperature at which peritectic reaction occurs is denoted by T_d. Below T_d, the liquid phase disappears upon completion of the transformation. Pt-Ag is a good example of peritectic reaction.

There is also *peritectoid reaction.* This is the reaction of two solids into a third solid. One can represent this reaction as

$$\alpha + \beta \xrightleftharpoons[\text{heating}]{\text{Cooling}} \gamma$$

(iv) Monotectic Type: In this case the two liquid solutions are not soluble in each other over a certain composition range, i.e., there is a miscibility gap in liquid state between the two metals. In this type one liquid decomposes into another liquid solid. Cu and Pb form monotectic system.

(v) Eutectoid Type: In this one solid decomposes into two different solids. Obviously, solid to solid transformation takes place. Fe-C, Cu-Zn, Al-Cu, Cu-Sn, etc form eutectoid system. Eutectoid transformation takes place at a constant temperature and the products of transformation are present as intimate mixture having appearance under the microscope. The eutectoid reaction is of the form: $y \xrightleftharpoons[\text{heating}]{\text{Cooling}} \alpha + \beta$

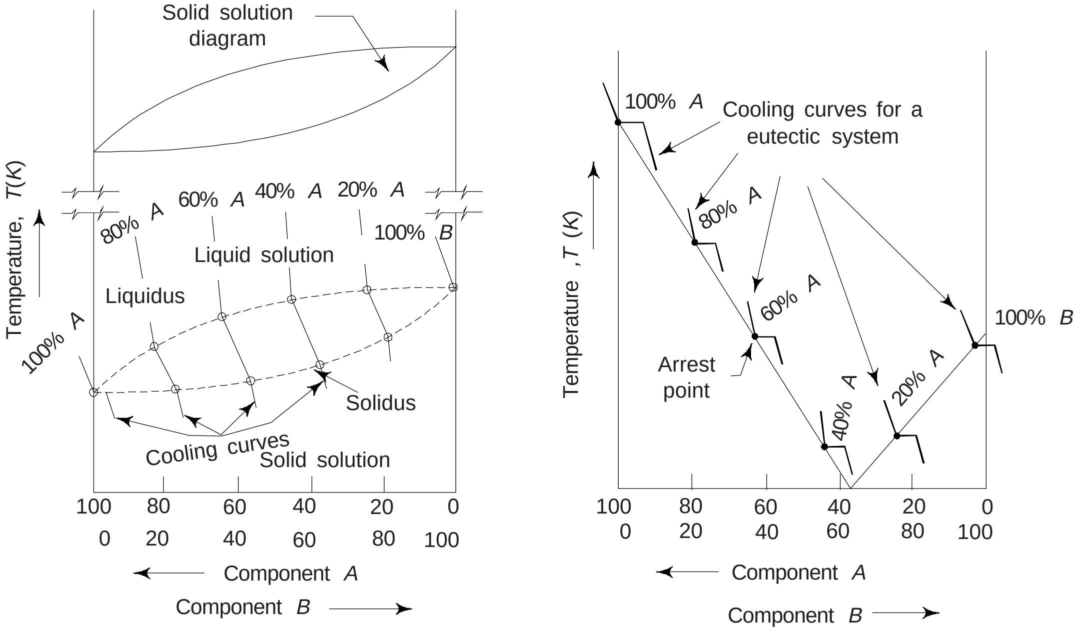
The eutectoid structure, frequently lamellar, is produced by precipitation from the solid solution. The crystal structure of new phase is known as the *Widmanstatten structure*. The eutectoid reaction occurs in Fe-C diagram in which austenite (a solid solution C in γ-Fe) is decomposed into pearlite. The most important example of the use of this reaction is heat treatment of steel.

Construction of a phase diagram for alloys of two metals *A* and *B* in their various compositions, using the data from the cooling curves is shown in Fig. 9.7. In order to get the actual phase diagram, the arrestment points have been joined. There is complete intersolubility in the liquid and solid phases in the diagram. The upper line corresponds to the temperature at which the alloys starts to solidify the liquids. The lower line (Fig. 9.7a) shows the complete solidification and is called the *solidus*. A mixture of solid and liquid exists between two lines. i.e. between *liquidus* and solidus and represents the alloys in a *semisolid* state. The figure is known as *equilibrium diagram*. To designate solid solutions in phase diagrams. Greek letters α, β, γ etc. are commonly used.

Phase diagrams are classified on the basis of the number of components in the system. Single component systems have unitary diagrams, two component systems have binary diagrams, three component systems give rise to ternary diagrams and so on.

11. THE LEVER RULE

This rule helps to calculate the relative proportions of solid and liquid material present in the mixture at any given temperature. The number and composition of phases can be obtained from the phase diagram. If the composition and temperature position is located within a two-phase region, things are more complex.



(a) (b)

Fig. 9.7 A phase diagram resulting from a series of cooling curves

In a two-phase region, one can determine the relative amount of each phase that is present from the phase diagram, using a relationship known as *lever rule* (or the inverse lever rule), which is applied as follows:

(i) Construct the tie line across the two-phase region at the temperature of the alloy.
(ii) The overall alloy composition is located on the tie line.
(iii) The fraction of one phase is computed by taking the length of the line from the overall alloy composition to the phase boundary for the other phase, and dividing the total tie line-length.
(iv) One can determine the fraction of the other phase in the same manner.
(v) In case if phase percentages are desired, each phase fraction is multiplied by 100. When the composition axis is scaled in weight percent; the phase fractions computed using the lever rule are *mass fractions* – the mass (or weight) of a specific phase divided by the total alloy mass (or weight). The mass of each phase is computed from the product of each phase fraction and the total alloy mass.

In order to determine the relative amounts of two phases, errect an ordinate or vertical line at a point on composition scale which gives the total composition of the alloy. The intersection of this ordinate with the given isothermal line denotes the fulcrum of a simple lever system. From Fig. 9.8(a) it is clear that the ordinate *KL* intersects the temperature line at a point *M*. However, the relative lengths of lever arm *OM* and *MP* (Fig. 9.8(b)) multiplied by the amount of phases present must balance. From Fig. 9.8(a) it is clear that the length *MP* represents the amount of liquid and the length *OM* represents the amount of solid. Therefore,

$$\text{The percentage of solid present} = \frac{OM}{OP} \times 100 = \frac{OP - MP}{OP} \times 100$$

$$\text{The percentage of liquid present} = \frac{MP}{OP} \times 100 = \frac{OP - OM}{OP} \times 100$$

From Fig. 9.8 $OM + MP = OP$, which represents the total composition of alloy between liquids and solidus, say at temperature t_p. The *OMP* (isothermal) can be considered a tie line since this line joins the composition of two phases in equilibrium at a specific temperature t_p.

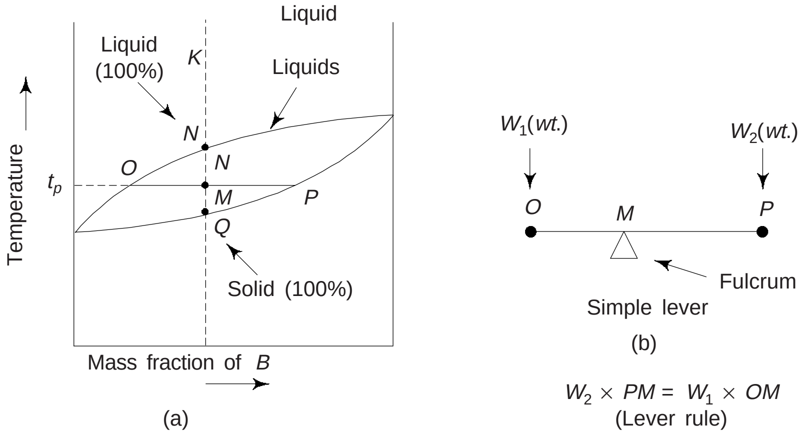


Fig. 9.8 Lever rule derivation using the phase diagram

We may note that in Fig. 9.8 *bivariant* regions occur above solidus and below liquids. Obviously, both temperature and composition can thus vary without causing phases appearing or disappearing. The regions between the solidus and liquids, however, is universal and consists of two phases. This means either temperature or composition can be varied independently without the disappearance of a phase.

Lever rules are valid for any two-phase region of the constitutional diagram and have no sense for single-phase regions. Using these rules together with the phase rule, one can 'read' any intricate constitutional diagram consisting of many branches and regions.

12. EQUILIBRIUM DIAGRAMS FOR BINARY ALLOYS FORMING EUTECTIC

The solubility of one metal into the other in the solid state is important criterion for the formation of alloy because the alloy is finally used in solid state. Any two given metals may have varying solubilities. Two cases of interest are : *(a)* alloys of two metals which are mutually soluble in liquid but completely insoluble in solid state and *(b)* alloys of two metals which are mutually soluble in liquid state but partially soluble in solid. We shall now consider these cases:

(a) Alloys with no solid solubility The phase diagram for all the alloys belonging to this group is as shown in Fig. 9.9.

Figures shows the cooling curves of liquid metal mixtures of *A* and *B*. The line *FEG* (liquids) is plotted as temperature of first arrest for a specific composition. *F* and *G* are denoting the melting points of metals

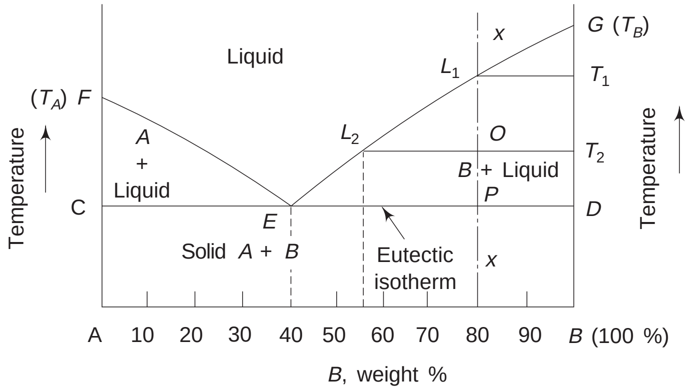


Fig. 9.9 Binary equilibrium diagram for an alloy system exhibiting no solid solubility

A and *B* respectively. *E* shows the position of *eutectic point*. *CED* line is termed *eutectic isotherm* while it is also *solidus*. Depending upon the composition, the solidification of alloy of any composition will begin anywhere on the liquids. The temperature corresponding to eutectic point is called *eutectic temperature*. Below the eutectic temperature the material is fully solid for all compositions. The composition which remains fully liquid upto the eutectic temperature during cooling is called *eutectic composition*.

Solidification of an alloy of eutectic type takes place at a constant temperature, called *eutectic temperature*, at which liquid of eutectic composition decomposes into two solid solutions. Temperatures remains constant till the transformation is completed. The eutectic reaction can be expressed as

$$\text{Liquid} \underset{\text{heating}}{\overset{\text{Cooling}}{\rightleftharpoons}} \text{Solid}_A + \text{Solid}_B$$

Fe-C, Al-Si, Al-Ni, Mg-Al are few examples which form such a type of solution.

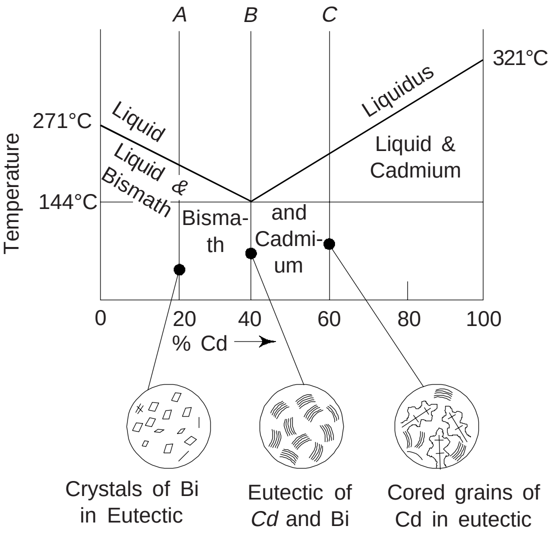


Fig. 9.10 Equilibrium diagram of the Cd-Bi system

Figure 9.10 exhibit the equilibrium diagram of Cd-Bi system which have no solid solubility with each other. The eutectic is the finely-divided intimate mixture of Cd and Bi. Three different compositions of alloys have been shown on the diagram (as typical examples).

For an alloy of 20% Cd and 80% *Bi* (represented by line *A* in Fig. 9.10), the mixture becomes saturated with Bi as the liquids is reached. Dendrites of pure Bi begin to form at this stage. More Bi solidifies as cooling continues further, making remaining liquid richer in Cd. When the eutectic point is attained (at temperature of 144°C) the remaining liquid is solidified as Cd-Bi eutectic which surrounds the grains of pure Bi in the microstructure.

When a mixture of 60% Cd and 40% Bi is allowed to cool, dendrites of pure *Cd* begin to form at liquids and remaining liquid gradually becomes richer in Bi. At the eutectic the alloy solidifies with Cd embedded in the matrix of the eutectic mixture.

If a mixture of eutectic proportions is allowed to cool the whole of the mixture solidifies at temperature of 144°C. However, no coring occurs in eutectic alloys.

(b) Binary alloy systems exhibiting partial solid solubility There are certain solvent metals, which can accommodate only a certain percentage of solute metal. This results in limited solid solution. It is observed that the partial solid solubility varies tremendously with temperature and almost in every case decreases with decreasing temperatures. A eutectic is formed in some such solid solutions.

It is observed that the solubility of a constituent (solute) in the other (the solvent) mainly depends upon relative size of two atoms. With the rise in temperature the distance between the atoms increases whereby the lattice is capable of absorbing greater distortions and hence the solubility increases.

The alloy of Pb and Sn, commonly know as solders is an typical example of this type. The equilibrium diagram for this alloy is shown in Fig. 9.11.

We may note that no pure metal exists in the solid alloy of any composition. It is the solid solution of one metal in the other that exists in the solid state. The phase which solidifies first is termed as α-phase while the one which separates later is termed as β-phase. For example, in the solution of Sn in Pb is referred as α while that of Pb in Sn is β. *OE* line in Fig. 9.11 shows the amount of Sn which can be dissolved in Pb at temperature below the solidus. This amount is referred as *solvus*. 20% of *Sn* can be accommodated

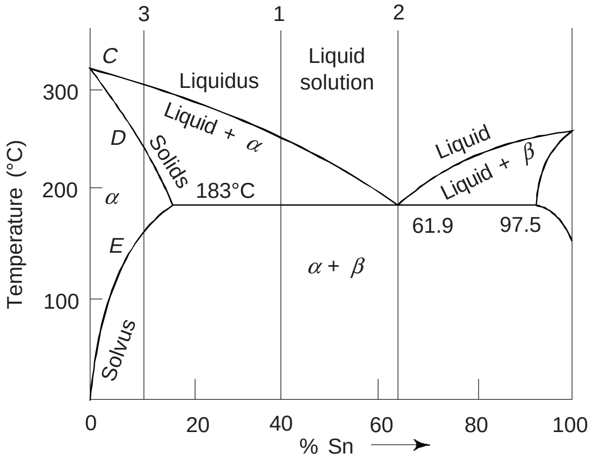


Fig. 9.11 Pb-Sn equilibrium diagrams

in Pb at 183°C. At room temperature this falls to 3%. The rejected tin from the solid solution sets up its own lattice into which some Pb is dissolved. Obviously, it so forms β-phase which adds to β-phase already existing in the solution. The solubility of Pb in Sn at room temperature is 1%. The structure of alloys with composition between these solubility limits, i.e. between 97% Pb–3% Sn and 1% Pb–99% *Sn* is a mixture of α and β phases. The eutectic composition of this system is 62% *Sn*. In alloys below this percentage of Sn, the structure is made up of α and eutectic while in alloys above this composition the structure is β and eutectic. The eutectic itself is made up of layers of α and β phases which have solidified simultaneously from the molten metal at temperature 183°C.

One can use the information gained from the equilibrium diagram in selection of a particular alloys for some purposes. In the present example, if the solder is required to solidify fast and be melted a lowest temperature, the eutectic composition will be most suitable. Electrical soldering possess such requirements where quickly solidifying solders at lower temperatures provide strength and safety. However, in plumbing work one may have to manipulate the joining material and hence require a solder that retains its mushy state over a range of temperature. This is why the plumber’s solder contains 65% of lead.

In order to retain a particular structure that exists at higher temperatures, the alloy may be suddenly cooled to room temperature. For such decisions, one finds the equilibrium diagrams come quite handy and mostly used for deciding heat treatments.

13. CERAMIC AND TERNARY PHASE DIAGRAMS

Ceramic and ternary and more complex alloys are widely used in engineering. Here, it is very important to mention that we should not assume that phase diagrams exist only for metal-metal systems. Phase diagrams that are very useful in the design and processing of ceramic systems have been experimentally determined for quit number of these materials.

Phase diagrams have also been determined for metallic (as well as ceramic) systems containing more than two components. However, their representation and interpretation may be exceedingly complex.

Nickel is frequently added to steel to improve its toughness and lead to brass to improve its machinability. Obviously, addition of alloys (components) possessing certain properties to the pure metals or binary alloys is very significant for development of alloys in some industrial applications, e.g. high speed steels, stainless steels and ceramics, etc.

In a ternary system such as Ag-Au-Cu, the composition is represented by two variables, for e.g. the percentage of Ag and the percentage of Au. The temperature is also still a variable. The phase diagram must be drawn in three dimensions. The Gibbs energy functions are surfaces, and the common tangent is a tangent plane rather than a line. It is possible to find a plane tangent to three surfaces over a range of temperatures, so that for ternary systems the maximum number of phases in equilibrium is four, and increases by one for each element added. The same principles of construction of three-dimensional phase diagram work for ternary systems, with one to four phases in equilibrium in each region, but the representation of the diagrams can become complicated.

14. APPLICATIONS OF PHASE DIAGRAMS

One of the most important application of phase diagram is in the *Zone refining* of materials. It is also termed as *fractional solidification*. A separation is brought about by crystallization of a melt without adding solvent. A massive solid is formed slowly with a sizeable temperature gradient imposed at the solid-liquid interface.

Consider the schematic phase diagram shown in Fig. 9.12. We need to purify or refine *A*, which has *B* as the impurity content. Let the starting composition be C_1, which starts to solidify at temperature T_1. Just below this temperature, a small quantity of solid of composition C_2 purer in *A* than C_1, separates from the liquid. Let us imagine that we stop cooling further, throw away the liquid part and remelt the left-over solid. This small quantity of liquid is now purer, as its component is C_2. This overall composition will now start to solidify at a higher temperature T_2. Again the first solid to separate will have a composition C_3 purer in *A* than C_2. By repeating this sequence of operations a few times, one can obtain very pure *A*, even though the quantity of purified material will be extremely small compared to the starting material.

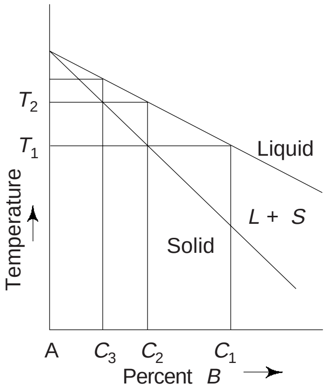


Fig. 9.12 Zone refining process of materials. This is based on the fact that the solid to crystallize first from a melt is usually purer in *A* than the liquid

This principle of phase separation is used in zone refining. Zone refining can be applied to the purification of almost every type of substance that can be melted and solidified. One can achieve the purity level of less than 1 part per million (PPM), i.e. better than 99.9999%.

The increase in carbon content increases the cementite component and finally results in an increase of hardness and decrease of ductility. This helps in the selection of material for several items, e.g. rivets, rails, chain links, file, knife, etc. Rivets and chain links have a low carbon content ~ 0.1% because of good ductility requirement. Rails have 0.5 to 0.6% carbon content (medium carbon steel) as it should have toughness with hardness and at the same time wear resistance. Knife and file having carbon content 1.3 to 1.4% have high hardness and wear resistance.

The decrease in melting point of a pure solid when one adds another component has also some applications. In western cold countries the depression of the freezing point of water by the addition of salt is used to melt ice on roads. Low melting eutectic alloys are used as safety devices in fire fighting equipment and petroleum storage tanks. Au-Si eutectic finds uses in the manufacture of semiconductor devices. A tea spoon made out of a quartenary (four components) eutectic of Pb, Sn, Bi and Cd melts inside a hot cup at tea at 70°C.

15. CORING

In thermal equilibrium diagram, it is assumed that cooling will be slow enough for equilibrium to be maintained. However, during actual operating condition where rate of cooling is more rapid, e.g. the production of Cu-Ni alloy, there is insufficient time for complete diffusion to take place. This leads to lack of uniformity in the structure of the metal. This is termed a *cored* structure, which give rise to less than the optimal properties. As a casting having a cored structure is reheated, grain boundary regions will melt first in as much as they are richer in low-melting component. This produces a sudden loss in mechanical integrity due to the thin-liquid film that separates the grains. Moreover, this melting may begin at a temperature below the equilibrium solidus temperature of the alloy. Coring may be eliminated by a homogenization heat treatment carried out at a temperature below the solidus point for the particular alloy composition. During this process, atomic diffusion occurs, which produces compositionally homogeneous grains.

16. PHASE TRANSFORMATIONS

The development of microstructure in both single-and two-phase alloys ordinarily involves some type of phase transformation. Phase transformation means a change in the number and/or character of the phases that constitute the microstructure of an alloy. Metallurgical processes often involve the transformation of a metal from one phase to another. One can consider that every transformation depends on two factors: *(i)* a thermodynamic factor that determines whether the transformation rate is possible and *(ii)* a kinetic factor that determines whether transformation is possible at a practical rate.

There are a variety of phase transformations are playing important role in the processing of materials, and usually they involve some alteration of microstructure. These transformations are usually divided into three groups: *(i) Simple diffusion-dependent transformations* in which there is no change in either the number or composition of the phase present. These include solidification of pure metals, allotropic transformations and recrystallization and grain growth; *(ii)* another type of diffusion dependent transformations in which there is some alteration in phase compositions and often in the number of phases present. The final microstructure ordinarily consists of two phases. The eutectoid reaction

$$\gamma\,(0.76 \text{ Wt \% C}) \underset{\text{heating}}{\overset{\text{Cooling}}{\rightleftharpoons}} \alpha\,(0.022 \text{ Wt\% C}) + \text{Fe}_3\text{C}\ (6.7 \text{ Wt \% C})$$

is of this type; *(iii) diffusionless transformation*, wherein a metastable phase is produced.

The time taken for a transformation to go to completion is quite important in the control of the structure of a material. On the basis of this knowledge, one can suppress or induce a phase transformation. The time taken mainly depends on the nature and the mechanism by which a phase transformation is brought about. The change in free energy during the transformation also determines its rate. As the transformation occurs on cooling a material from an elevated temperature, so the cooling rates may be extremely fast, very fast, normal, slow or extremely slow. A qualitative time scale for these phase transformation is shown in Table 9.2.

Table 9.2 Time scale for phase transformations

Qualitative nature of phase transformation	*Time for phase transformation*	*Suppressibility*
• Extremely fast phase transformation	microseconds	insuppressible
• Very fast phase transformation	milliseconds to seconds	suppressed by very fast cooling
• Normal phase transformation	seconds to hours	suppressible
• Slow phase transformation	hours to days	suppressed easily
• Extremely slow	years	transformation virtually impossible

17. THE KINETICS OF SOLID STATE REACTIONS

It is observed that most solid-state transformations do not occur instantaneously because obstacles impede the course of reaction and make it dependent on time. For example, since most transformations involve the formation of at least one new phase that has a composition and/or crystal structure different from that of parent one, some atomic arrangements via diffusion are desired. We know that diffusion is a time-dependent phenomenon. A second impediment to the formation of a new phase is the increase in energy associated with the phase boundaries that are created between parent and product phases.

From the microstructure standpoint of view, the first process to accompany a phase transformation is *nucleation*. Nucleation refers to the formation of very small (often submicroscopic) particles, or nuclei, of the new phase, which are capable of growing. It is found that the favourable positions for the formation of these nuclei are imperfection sites especially grain boundaries. The second stage is *growth*, in which the nuclei increase in size. However, during this process, of course, some volume of the parent phase disappears. The transformation reaches completion if growth of these new phase particles is allowed to proceed until the equilibrium fraction is attained.

(a) Nucleation (Mechanism of Phase Transformation) Let us consider a simple transformation where a crystal β that is in contact with the liquid of the same will grow as liquid freezes. We may consider that in the beginning, a container is holding both the liquid and solid phases of a pure substance at its melting point. When heat is removed from the container, the liquid solidifies. We may note that the change from the liquid to the solid state reflects the tendency of the substance to maintain itself in its most stable, i.e., equilibrium state. Since the liquid and the solid can coexist in the equilibrium state at the melting point, and hence both states are equally stable at this temperature. However, at lower temperature the solid state is the stable one, while equilibrium at higher temperature demands the liquid state.

At the equilibrium melting point T_m, the free energies of two phases are equal. The Gibbs free energy ΔF is considered to be a useful measure of the tendency for a transformation to occur. However, at the equilibrium point (T_m), there is no tendency for the transformation to occur (enthalpy change), i.e.

$$\Delta F_v = \Delta H_v - T_m \, \Delta S_v = 0 \qquad [3]$$

or

$$\Delta S_v = \frac{\Delta H_v}{T_m} = \text{Change in entropy}$$

Here, we have assumed that ΔF_v and ΔS_v do not vary with temperature. Thus the free energy change accompanying the solidification at any temperature can be determined as

$$\Delta F_s = \Delta H_v - T\Delta S_v$$

$$= \Delta H_v \left(\frac{T_m - T}{T_m} \right) = \frac{\Delta H \Delta T}{Tm} \qquad (4)$$

where ΔT is the degree of supercooling.

A plot of the variation of free energy during nucleation as a function of particle radius r for different temperatures is shown in Fig. 9.13.

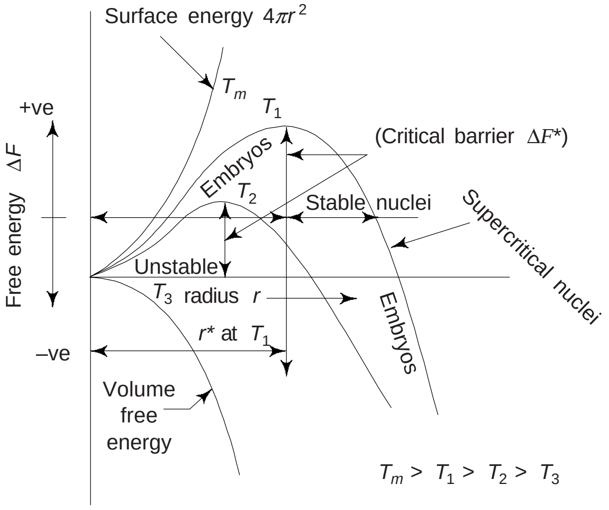


Fig. 9.13 **Variation of free energy change ΔF during nucleation (solidification), as a function of particle radius r at different temperatures**

Now, if ΔF is the free energy change accompanying the formation of a spherical new phase particle, we have

$$\Delta F = 4\pi r^2 \; \Delta F_v + \frac{4}{3}\pi r^2 \gamma \tag{5}$$

where r is the radius of the particle and γ is the surface energy per unit area of the interface separating the initial and product phases. First term in Eq. (5) corresponding to the surface energy is always positive. If ΔF_v is negative, the function ΔF passes through a maximum. Initially as the new particle starts to form, the energy of the system increases, as the surface energy term is dominant. The variations with r of the surface energy and the volume (Gibbs) free energy exactly balance each other at the maximum. Thereafter, the variation in the volume term becomes dominant and, as this term is negative, one finds that there is a continuous decrease in the energy of the system.

Setting $(d\Delta F/dr) = 0$, one obtains the values corresponding to the maximum called the critical values and denoted by r^* as

$$r^* = -\frac{2\gamma}{\Delta F_v} \tag{6}$$

and

$$\Delta F^* = \frac{16\pi\gamma^3}{3(\Delta F_v)^2} \tag{7}$$

The drops of liquid described by the curve ΔF versus r arise due to successive random collision of atoms taking place in pairs. Obviously, there will be drops of the whole range of sizes per unit volume of the parent phase n. Particles which are smaller than the critical size are called *embryos*. Those which are larger than the critical size are termed *nuclei* (this should not be confused with atomic nuclei). The critical sized particle is called a *critical nucleus.*

The number of critical sized particles per unit volume can be obtained from Boltzmann's equation, as

$$n^* = n \exp(-\Delta F^*/kT) \tag{8}$$

where k is Boltzmann's constant.

When the radius of the drop or sphere is less than the critical value (r^*), the sphere will revert back to the liquid phase. This sphere is termed as an embryo. When the sphere or drop has a radius greater than the critical radius, it is called a nucleus and serves as a nucleation site. We may note that the size of the critical radius depends on the degree of super cooling of the liquid. Combining Eqs. (4) and (7), one obtains

$$\Delta F^* = \frac{16\pi\gamma^3 T_m^2}{3(\Delta H)^2 (\Delta T)^2} \tag{9}$$

The rate of formation of the new phase particles or the rate of *nucleation*, dN/dT (=1) can be determined from the rate at which these nuclei are struck by additional atoms (z). This is given by

$$I_v = zn^* \, 4\pi (r^*)^2 \tag{10}$$

We may note that the rate of nucleation increases with decreasing temperature as ΔF^* is decreasing.

(b) The Kinetics of Solid State Reactions Figure 9.14 show a plot for solid state transformations displaying the kinetic behaviour, the fraction of transformation y is a function of time t as follows:

$$y = 1 - \exp(-kt^n) \tag{11}$$

where k and n are time-independent constants for the particular reaction. Equation (11) is usually called as the *Avrami equation*.

By convention, we can take the rate of a transformation r as the reciprocal of time required for the transformation to proceed halfway to completion, $t_{0.5}$, or

$$r = 1/t_{0.5} \tag{12}$$

We may note that the temperature is one variable in heat treatment process that is subject to control, and it may have a profound influence on the kinetics and thus on the rate of a transformation. For most reactions and over specific temperature ranges, rate increases with temperature according to

$$r = A\ \exp(-Q/RT) \qquad (13)$$

where R is the gas constant, T is absolute temperature, A is the temperature independent constant and Q is an activation energy for the particular reaction.

We may note that the diffusion coefficient also has the same temperature dependence. Processes with rates that exhibit relationship (13) with temperature are sometimes referred as *thermally activated.*

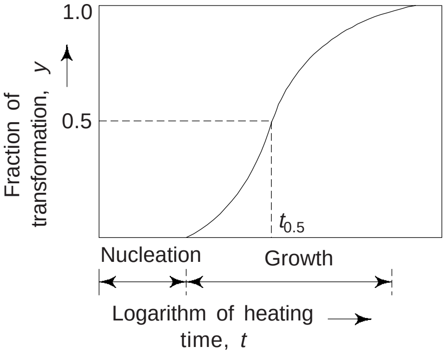


Fig. 9.14 Fraction reacted versus the logarithm of time typical of many solid-state transformations in which temperature is kept constant

(c) Homogeneous and Heterogeneous Nucleation In a *homogeneous nucleation*, the probability of nucleation occurring at any given site is identical to that at any other site within the assembly or the volume of the parent phase. We have already discussed homogeneous nucleation in section 17(a) in terms of free energy distribution among the liquid drops. We have noted that for this type of nucleation, extremely large amounts of super cooling was required, so that one may obtain a practical rate of nucleation of drops in the volume of liquid.

In *heterogeneous nucleation*, the probability of nucleation occurring at certain preferred sites in the assembly is much greater than that at other sites. Water boiling in a pot is common example of this type of nucleation. The steam bubbles usually originate at a particular location on the bottom of the pot are called heterogeneous nucleation sites. During solidification of a liquid, inclusions of foreign particles in the liquid and the walls of the container holding the liquid provide preferred nucleation sites. In a solid-solid transformation, foreign inclusions, grain boundaries, interfaces, stacking faults and dislocations can provide sites for preferred nucleation.

We may note that the mechanism of heterogeneous nucleation is similar to that the mechanism for homogeneous nucleation, already discussed. However, in addition to the difference in the ΔF^* term (discussed above), the number of sites will be preferred only.

(d) Nucleation and Growth Figure 9.15 shows the effect of a liquid solidifying on several nuclei. The black dots in Fig. 9.15(a) represent nuclei or very tiny particles identical in structure to the β crystal. These form in the liquid phase and grow to large size till the transformation is complete, i.e. over. The liquid solidifies on the nuclei and crystals grow as the temperature of the melt diminishes. The parallel lines within the boundaries shown in Fig. 9.15(b) represent the same direction in each crystal. However, the crystals are not in alignment with each other. We note that their orientation is random. The crystals join each other and form boundaries as solidification progresses (Fig. 9.15c). We note that each boundary marks a discontinuity

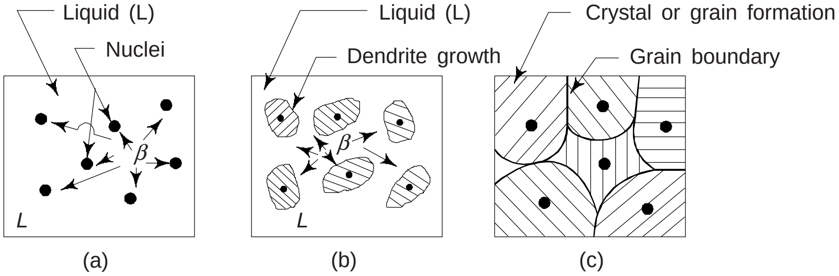


Fig. 9.15 Illustration of solidification process of the liquid to β-crystals by nucleation and growth

in the crystal structure. The individual crystals are named as *grains* and the boundaries are termed as *grain boundaries*. When solidification begins at a large number of nuclei, the resulting solid will have many grains and is termed to be *polycrystalline*. We may note that most engineering materials are *polycrystalline*.

18. MULTIPHASE TRANSFORMATIONS

Phase transformations may be wrought in metal alloy systems by varying temperature, composition, and the external pressure. However, temperature changes by means of heat treatments are most conveniently utilized to induce phase transformations. This happens to correspond to crossing a phase boundary on the composition temperature phase diagram as an alloy of given composition is heated or cooled.

An alloy, during a phase transformation proceeds toward an equilibrium state that is characterized by the phase diagram in terms of the product phases, their compositions, and relative amounts. We may note that most phase transformations require some finite time to go to completion, and the speed or rate is often important in the relationship between the heat treatment and the development of microstructure. There is one limitation of phase diagrams, i.e. they are unable to indicate the time period required for the attainment of equilibrium.

We may note that the rate of approach to equilibrium for solid systems is so slow that true equilibrium structures are rarely achieved. Equilibrium conditions can only be maintained if we carry heating or cooling at extremely slow and unpractical rates. For other than equilibrium cooling, transformations are shifted to lower temperatures than indicated by the phase diagram; for heating, the shift is to higher temperatures. These phenomena are known as *super cooling* and *super heating*, respectively. The degree of each phenomena depends on the rate of temperature change and the more rapid the cooling or heating, the greater is the super cooling or super heating. For example, for normal cooling rates the Fe-C eutectoid reaction is typically displaced 10 to 20°C below the equilibrium transformation temperature.

For several technologically important alloys, the preferred state or microstructure is a *metastable* one, intermediate between the initial and equilibrium states. Sometimes, a structure far removed from the equilibrium one is desired. This is why, it becomes imperative to investigate the influence of time on phase transformations. In many cases, this kinetic information is of greater value than a knowledge of the final equilibrium state.

19. APPLICATIONS OF PHASE TRANSFORMATIONS

Phase transformations are usually observed in microstructural changes in cooling or freezing (dendrite formation), in castings, in amorphous structures (solidification phenomenon in glassy structures), in heat treatment or in the binary phase diagram of Fe-Fe_3C system and TTT diagrams for eutectoid steels. Phase transformations are also observed in recrystallization and grain growth during mechanical working. Obviously, phase transformations are of great technological importance.

(a) Dendrite Formation and Structure of Ingots The shape and size of crystallized grains depend on the conditions of their growth during solidification, mainly on the rate and direction of heat removal, the temperature of molten metal and concentration of impurities.

A freely growing crystal acquires a dendritric (tree-like) shape (Fig. 9.16). It has been established that crystals grow with the highest rate along the planes and directions where atoms are packed most closely. Thus, long branches grow first, which are called the *first-order* dendritic axes. Then second-order axes branch off from them and third-order axes from the second-order ones, and so on. Finally, the metal remains between dendritic axes solidifies.

Dendrites grow until they interfere with one another. After that the interaxial spaces are filled in and the dendrites turn into continuous crystals of an irregular external shape. Such crystals are called *grains* or *crystallites*. If there is enough liquid metal to fill in interaxial spaces (for instance, at the opened end of a mould where a shrinkage cavity forms), some crystals may retain the dendritic shape. On solidification, impurities concentrate between dendritic axes at grain boundaries; further these are the most probable places for the formation of voids, because of shrinkage and impaired access of liquid metal to the solidification front.

The structure of *ingots* depends on many factors, the principal ones being as follows: the concentrations and properties of impurities in a pure metal or of alloying elements in an alloy; the temperature of casting the rate of cooling during solidification; and the shape, temperature, thermal conductivity and the state of internal surface of the casting mould. Macrostructures of ingots produced in a simple vertical metallic mould are shown in Fig. 9.17.

Gravity segregation is minimized by cooling ingots and castings at high rates. This defect can be eliminated fully in space technology, since the gravitational forces in the outer space are extremely small.

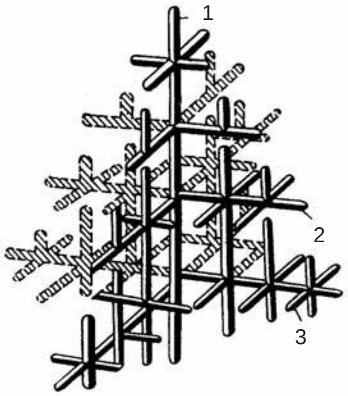
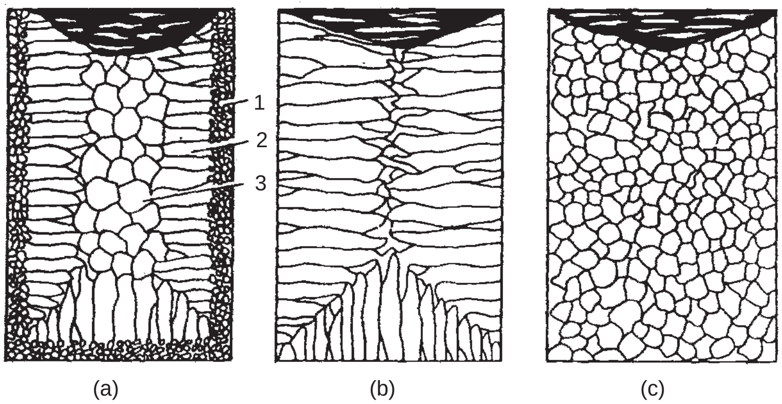


Fig. 9.16 Structure of a dendrite: 1, 2 and 3 - first- second-, and third order dendritic axes

(b) Casting This is very important process from practical point of view. In this process, liquid or viscous material is poured into performed cavity wherein it sets or solidifies to a well defined shape. Thermal gradients and temperature changes can become complex during the process. These may affect the microstructure of the cast metal or casting.



(a) (b) (c)

Fig. 9.17 Macrostructure of ingots

The practice of casting is also called *founding*. The foundary is the collection of all equipments in a place for casting mainly metals. Mould is the cavity in which molten material is poured and practice. The art of preparing the mould is termed *moulding*.

The best casting properties, i.e., high fluidity, small shrinkage, low (if any) liability to hot cracking, and high gas tightness are shown by alloys of Al-Si system (silumins), which can be explained by a high content of eutectic in their structure.

A mould for casting molten metal is prepared from sand, plaster or metal, the means by which it is achieved is called *pattern.* The mould is the heart of the foundry practice. The mould provides the cavity for depositing the molten metal gives the final dimensions and shape to the casting and holes and supports it during the process of pouring and solidification. The provisions are made for opening of the mould and withdrawal of the pattern. In most casting processes, the mould is considerably cooler than the melt pouring. We may note that the mould is often at room temperature.

The most common material in which pattern is made is *wood*. The other materials usually used are *plaster* or *plastic*. *Foamed polystyrene* and *wax* are also preferred in evaporative casing processes.

A flat pattern is perhaps simplest of all and has a flat side or bottom. There are some essential features of pattern for its removal from mould without causing and damage, e.g. such characteristics as fillet, draft and round (Fig. 9.18). *Split* pattern is made in two halves in order to facilitate the moulding. The pattern is so split that each part can be withdrawn from the sand mould.

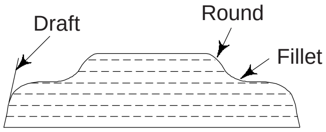


Fig. 9.18 A flat pattern

Flat and split patterns together are called as *loose* patterns. A pattern permanently fixed to a plane passing through its mid section is called a *mounted* pattern.

Casting techniques are used when *(i)* the finished shape is so large or complicated that any other method would be impractical, *(ii)* a particular alloy is so low in ductility that forming by either hot or cold working would be difficult, and *(iii)* in comparison to other fabrication processes, casting is most economical. Furthermore, the final step in the refining of even ductile metals may involve a casting process. A number of different casting processes are usually employed, including *sand*, *die investment* and *continuous* casting.

(c) Glass Transition : Amorphous (or glassy) structures The glass transition occurs in amorphous and semicrystalline polymers, and is due to a reduction in motion of large segments of molecular chains with decreasing temperature. Upon cooling, the glass transition corresponds to the gradual transformation from a liquid to a rubbery material, and finally to a rigid solid. The temperature at which the polymer experiences the transition from rubbery to rigid states is termed the *glass transition temperature* (T_g). However, this sequence of events occurs in the reverse order when a rigid glass at a temperature below T_g is heated. Moreover, abrupt changes in other physical properties accompany this glass transition, e.g. stiffness, heat capacity, and coefficient of thermal expansion.

Metallic materials with an amorphous structure can also be produced by superfast cooling from the gaseous (vapour or ionized) state, by electrolysis, and by cathode sputtering with a high rate of deposition.

(d) Phase Transformation in Fe-C Systems Iron and carbon make a series of alloys which include a number of steels and cast iron and form a most important subject in the study of steels and cast irons. Iron and carbon are polymorphous elements. Steels are the alloys which contain upto 1.2% of carbon with cast irons contain carbon within range of 2.3% to 4.2%. Alloys with carbon greater than 4.6% have poor properties and are not preferred. Fe-C alloys are very important to modern industry due to the wide ranging applications of ferrous metallurgy in fabrication, casting and production processes. Theoretically speaking, it is based on phase transformations in metals, e.g. in iron manufacture there is a transformation on cooling liquid iron to the solid state, as in casting on metals; heat treatment processes where different mechanical properties can be attained by changing the crystal structure.

We may note that Fe-C system provides the most important knowledge on heat treatment, based on polymorphic transformations and eutectoid decomposition. The phases than can exist in Fe-C alloys are mostly the liquid solution, ferrite, austenite, cementite, and free carbon (as graphite).

20. MICRO-CONSTITUENTS OF Fe-C SYSTEM

We can observe the various constituents of Fe and Steel produced due to decomposition of austenite under a microscope. The following are the important micro-constituents of Fe-C system:

(i) Austenite: This is the solid interstitial solution of carbon in gamma iron (Fe_{γ}). It has a FCC lattice in which the interstices are larger than in the BCC lattice, because of which the solubility of carbon in Fe_{γ} is much higher and attains 2.14%. Austenite is ductile and has a higher strength than ferrite (*HB* 160-200) at a temperature 20-25°C. On cooling below 723°C it starts transforming into pearlite and ferrite. We may note that the austenite in a eutectoid steel is unstable at all temperatures.

(ii) Ferrite: This is a soft and ductile phase. Basically, this is a BCC iron phase with very limited solubility of carbon. Like austenite, ferrite may have other alloying elements in substitutional solid solution. Ferrite is the name given to pure iron crystals. The solubility of carbon in ferrite is 0.025 at 723°C. Ferrite has

the following mechanical properties: σ_t = 300 MPa, δ = 40%, ψ = 70%, KCU = 2.5 MJ/m^2, and HB = 80-100. Below the critical temperature, the slow cooling of low carbon steel produces ferrite structure. Ferrite is very soft and highly magnetic and does not harden when cooled rapidly.

(iii) Cementite (Cem): This is essentially an iron carbide Fe_3C (of almost constant composition). It contains 6.69% C and has a complex rhombic lattice and under normal conditions it is extremely hard in nature and brittle. It is weakly ferromagnetic, but loses this property on heating to 210°C. The melting temperature of cementite is difficult to determine, since cementite decomposes on heating. In experiments with laser-beam heating, its melting point has been measured to be 1260°C. The brittleness and hardness of cast iron is mainly controlled by the presence of cementite in it.

(iv) Pearlite: Pearlite is a mechanical mixture of about 87% ferrite and 13% cementite having a two-phase microstructure and found in some steels and cast irons. Pearlite results from the transformation of austenite of eutectoid composition and consists of alternating layers (or lamellae) of α-ferrite and cementite. A steel with 0.8% carbon is wholly pearlite, with less than 0.8% carbon is hypoeutectoid and with more than 0.8% carbon is hypereutectoid steel. The former contains ferrite and pearlite and is soft while the latter contains pearlite and cementite which are hard and brittle.

(v) Bainite: This is a ferrite cementite aggregate, i.e., an austenitic transformation product found in some steels and cast irons. This is formed by the growth of a ferrite nucleus. It forms at temperatures between those at which pearlite and martensite transformations occur. It is the product of isothermal decomposition of austenite. The microstructure consists of α-ferrite and a fine dispersion of cementite. Bainite is present in two forms: the feathery bainite obtained in the upper part of the temperature range and needle-like or accicular bainite produced by lower reaction temperature.

(vi) Martensite: This is a body-centered, tetragonal metastable iron phase produced by entrapping carbon on decomposition of austenite when cooled rapidly. Obviously, this is supersaturated in carbon that is the product of a diffusionless (athermal) transformation from austenite. It is essentially a supersatured interstitial solid solution of carbon in Feα. Martensite crystals have a lamellar shape and grow at an enormous rate equal to the speed of sound in steel (roughly 500 m/s). The growth of martensite crystals can be hampered by boundaries of austenitic grains or martensite lamellae that have formed earliers (Fig. 9.19). Martensite is magnetic and is made of a needle-like fibrous mass and contains carbon upto 2% and extremely hard and brittle. We may note that the decomposition of austenite below 320°C starts the formation of martensite.

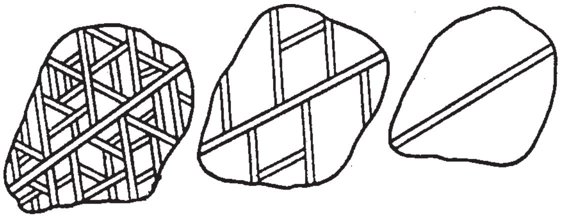


Fig. 9.19 Scheme of formation of martensite platelets in a grain of austenite

(vii) Troostite: This is obtained by quenching tempering martensite and composed of cementite phase in a ferrite matrix that can be resolved by a light microscope. In comparison to martensite, this is less hard and brittle and also weaker than martensite. It is also produced by cooling the metal slowly until transformation begins and then rapidly to prevent its completion at about 580-550°C. At the transformation temperature, the interlamellar spacing decreases to $1 \times 10^{-7} - 2 \times 10^{-7}$ m. It has dark appearance on etching.

(viii) Sorbite: This is also produced by the transformation of tempered martensite at a temperature of 640-590°C. It is produced when steel is heated at a fairly rapid rate from the temperature of the solid solution to normal room temperature. Its properties are intermediate between those of pearlite and troostite and has a good strength. It is practically pearlite. We may note that troostite and sorbite division of pearlitic

structures is quite conventional, since the mixture dispersity increases monotonously with decreasing temperature of transformation.

21. ALLOTROPIC FORMS OF IRON

Pure substances may exist in more than one crystalline form and each such crystalline form is stable over more or less well defined limits of temperature and pressure. This is termed as allotropy or polymorphism. Pure iron is relatively soft and ductile and its melting point is 1539°C. The pure iron exists in three important allotropic forms, i.e. alpha (α), gamma (γ) and delta (δ) iron. The existence of phases depends upon the temperature to which the iron is heated. An ideal curve for pure iron, showing the temperature ranges over which each of these crystallographic forms are stable at atmospheric pressure is shown in Fig. 9.20.

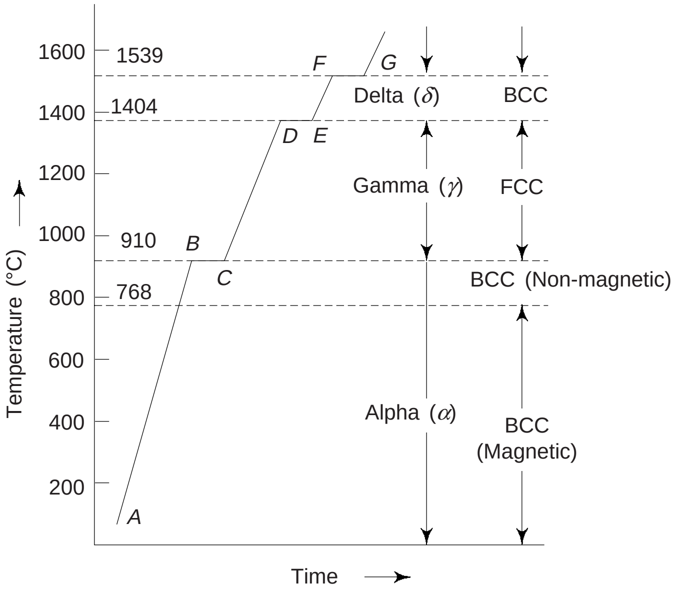


Fig. 9.20 Ideal heating curve for pure iron.

From Fig. 9.20, it is evident that from room temperature to 910°C pure iron has a BCC structure and is called alpha (α) iron (Fe_α). It is highly ferromagnetic and remains so upto 768°C (Curie point). On heating it becomes non-magnetic, i.e., ferromagnetism disappears. However, the crystal structure still remains BCC. Non-magnetic α-iron is stable upto 910°C. The non-magnetic α-iron was earlier known as β-iron. However, the X-ray crystallography revealed no change of crystal structure at 768°C. In order to avoid confusion, the original naming of the sequence retained with the β-phase deleted. Hence it is known as α-iron (Fe_α).

Above 910°C, it is transformed into FCC and allotropic change takes place. It is transformed from α to γ-iron structure.

Upon heating to 1404°C, again allotropic change takes place and γ-iron is transformed back into the BCC structure called δ-iron. It is stable upto the melting point, 1539°C of pure iron. The BCC structure δ-iron has a longer cube edge than BCC structure of α-iron.

We may note that the above transformations are reversible. The δ-iron is changed to γ-iron at 1404°C. Evolution of heat takes place on cooling at 910°C and γ-iron changes to α-iron

Critical Points: The temperature at which the above structural changes takes place are termed as *critical points* or *arrest points*. These points are designated by the symbol A_r for cooling and A_c for heating. The symbol *A* stands for arrest and '*r*' for refroidissment (a French word used for cooling) and '*c*' for chauflage (a French word used for heating). The critical points correspond to temperatures, 910°C and 1404°C and are referred to as A_{r_3} and A_{r_4} respectively. We may note that one can distinguish between the points A_{r_4} and A_{c_4} as well as A_{r_3} and A_{c_3} respectively. It is found that there exists approximately 30°C difference between a pair of similar points.

21. IRON-CARBON SYSTEM

Iron and carbon are polymorphous elements.

Iron-carbon alloys exist in different phases in steels and cast irons. In steels, the iron and carbon exists as two separate phases, ferrite and cementite. There are two following phase diagrams of iron-carbon system:

(i) Iron-iron carbide (Fe-Fe_3C) phase diagram

(ii) Iron-carbon (Fe-C) phase diagram

22. IRON-CARBON EQUILIBRIUM OR PHASE DIAGRAM

Figure 9.21 shows the iron-carbon equilibrium diagram representing the entire range of iron-carbon alloys. Diagram indicates transformations that take place in an alloy of iron-carbon from pure iron to cementite

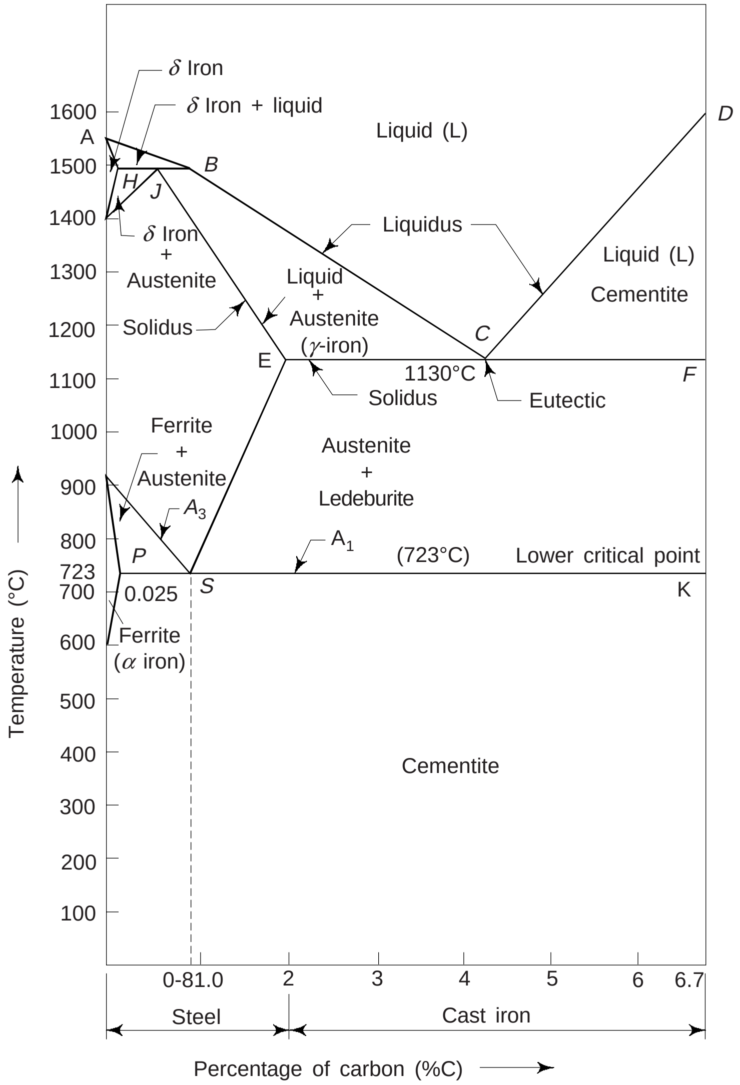


Fig. 9.21 Iron-carbon equilibrium diagram representing the entire range of iron-carbon alloys

(carbon content 6.67%). The carbon percentages is represented on the horizontal axis and the temperature on the vertical axis. We have also shown the names of the phases existing at temperatures and concentrations determined by the lines (boundaries) of these areas on the equilibrium diagram.

The curve *ABCD* is called the liquidus. All alloys represented by compositions and temperatures in the region above *ABCD* are completely liquids. Point *A* in the diagram represent the melting point of pure iron (1539°C). Point *D* represent the melting point (1539°C) of iron carbide or cementite. With the fall in the temperature of the liquid along the line *ABC*, crystals of austenite separate from the liquid. Similarly, in the same way, crystals of iron carbide (Fe_3C) separate from the liquid along the line *CD*. High temperature transformation ($\gamma -$ Fe $\rightleftharpoons$ $\delta -$ Fe or $Fe_\gamma \rightleftharpoons Fe_\delta$) take place at upper left hand portion of phase diagram (Fig. 9.21). The peritectic reaction *HJB* represents the formation of austenite, i.e. solid solution of carbon in gamma iron (γ-Fe or Fe_γ). crystals of δ-iron (Fe_δ) begin to separate from the liquid along the line *AB*. We know that a solid solution of carbon in α-iron (Fe_α) at high temperatures is known as δ-iron (Fe_δ).

The complete solidification of iron-carbon alloys proceeds along the line *HJECF* (Fig. 9.21) called the solidus. The alloys containing 0.18 to 1.7% carbon represent the solidus *HJE* and those with carbon contents ranging from 1.7 to 6.67% become solid along the line *ECF* at temperature 1130°C. (Fig. 9.21). At point *C* in Fig. 9.21 (4.3% carbon), austenite and cementite are simultaneously precipitated from the alloy to form the eutectic, also called as Lede-burite. The iron-carbon phase diagram (Fig. 9.21) indicates a peritectic point *J*, a eutectic point *C*, and a eutectoid point *S*. The following reactions takes place at these points:

(i) During the *peritectic reaction* (horizontal line *HJB* at 1401°C)

$$\delta\text{-Iron } (Fe_\delta) + \text{liquid} \underset{\text{Heating}}{\overset{\text{Cooling}}{\rightleftharpoons}} \text{Austenite}$$

or $$F_H + L_B \rightarrow A_J$$

(ii) At point *C* the *eutectic reaction* takes place (4.3%, 1130°C). This can be represented as

$$\text{Liquid} \xrightarrow[\text{Heating}]{\text{Cooling}} \text{Austenite + Cementite} \qquad \text{(Lede-burite eutectic mixture)}$$

or $$L_C \rightarrow [A_E + C_{em}]$$

(c) At point *S* (723°C), the eutectoid transformation can be expressed as

$$\text{Solid} \xrightarrow[\text{Heating}]{\text{Cooling}} \text{Ferrite + Cementite or Pearlite} \qquad \text{(eutectoid mixture)}$$

or $$A_S \rightarrow [F_P + C_{em}]$$

The eutectoid mixture of ferrite and cementite is called *pearlite*.

The eutectoid (pearlite with 0.8% carbon) and eutectic (ledeburite with 4.3% carbon) are considered as independent structural constituents which can influence markedly the properties of alloys. Most often pearlite has a lamellar structure and rather high strength properties.

23. MODIFIED IRON-CARBON PHASE DIAGRAM

In phase diagram of Fig. 9.21, the upper left hand portion, the region *ABJN*, in which δ-iron may exist is not very important from practical point of view because temperature near 1401°C are neither used for heat treatment and not in mechanical working. This is why, it is essential to discuss all the practical aspects of primary solidification with the modified iron-carbon phase diagram shown in Fig. 9.22.

The study of following two important transformations will clarify the interpretation of phase diagram, Fig. 9.22:

(i) *Primary solidification*: i.e, transformation from liquid to solid state.
(ii) *Secondary solidification or crystallization*: i.e, solid state transformation.

23.1 Primary Solidification

To study these transformations, let us consider the sequence of events when liquid alloys of various carbon

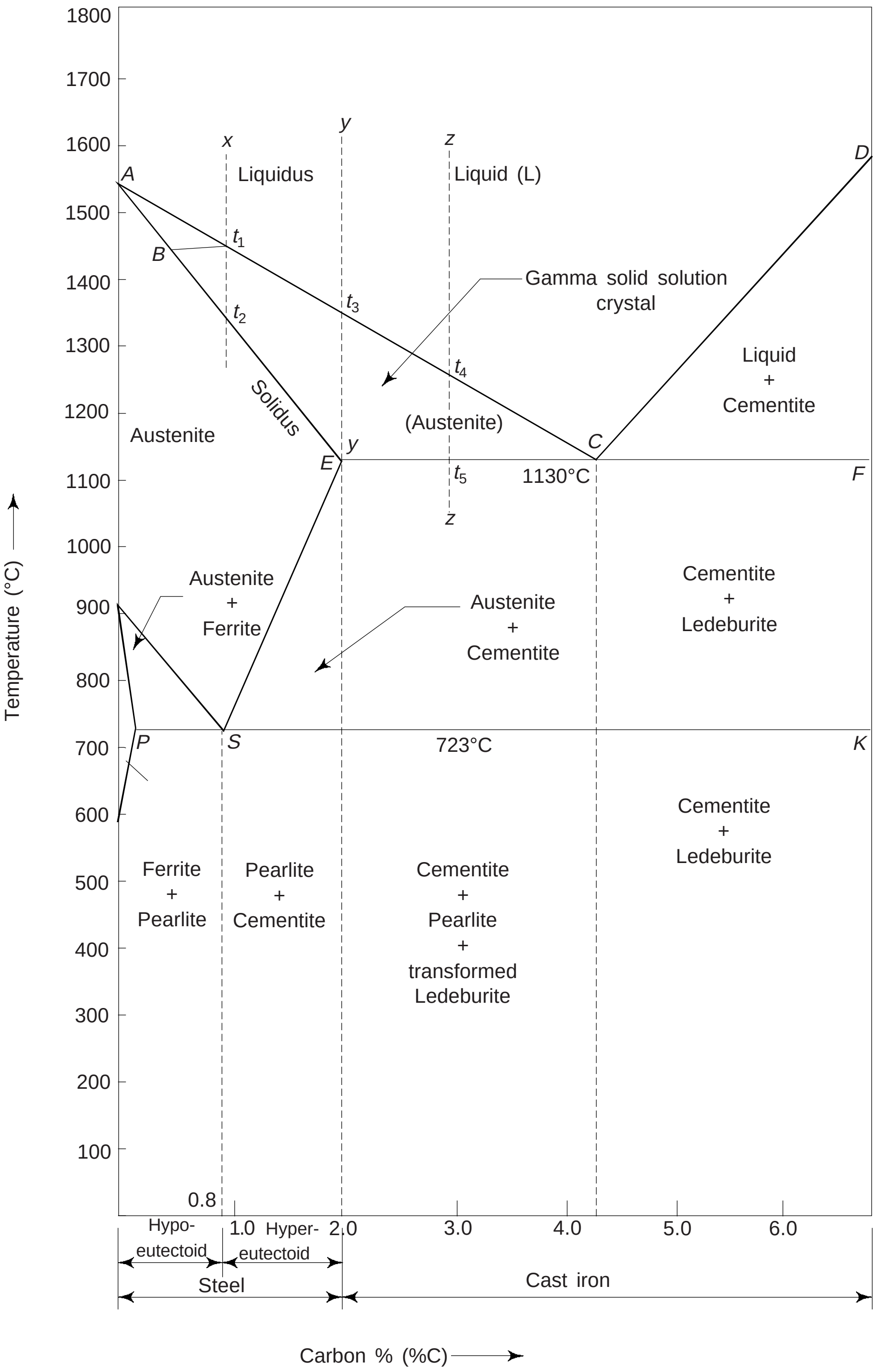


Fig. 9.22 Modified iron-carbon equilibrium or phase diagram.

contents are cooled to a temperature just below the eutectic temperature, 1130°C, as shown by line ECF Fig. 9.22). If an alloy containing 0.8% carbon is cooled from a point '*x*' lying above the liquids line *AC*, it will remain in its liquid state up to a temperature t_1 on the liquidus line *AC* (Fig. 9.22). At this temperature, the crystals of austenite (or γ-phase) begin to precipitate from the liquid alloy. The composition of austenite crystals at t_1 may be determined by drawing a horizontal line (called as tie-line) at t_1. The intersection of the tie-line with the solidus line *AE* (point *B*) gives the composition of austenite crystal at t_1. The amount of precipitation increases with the fall in temperature, and the liquid completely solidifies at temperature t_2. The composition of austenite, formed during cooling, varies along the solidus line *AB* and that of the liquid phase along the liquidus line *AC*. However, alloys of any composition upto 1.7% carbon will solidify in a similar manner and will consist of only austenite.

When a liquid alloy containing 1.7% of carbon is cooled from point '*y*', a similar process will take place and the austenite crystals will begin to form at temperature t_3 on the liquidus line *AC* shown in Fig. 9.22. As the crystals separate, the liquid becomes richer in carbon and will solidify when the eutectic temperature at 4.3% carbon is achieved. We may note that at this temperature the completely solidified alloy will consist entirely of austenite crystals containing 1.7% carbon dissolved in solid solution in γ-iron.

Now, we consider a liquid containing 3% carbon (point '*z*') being cooled. It begins to precipitate austenite crystals from the liquid alloy at the temperature t_4 lying on the liquids line *AC*. In this case, also austenite crystal precipitate from the liquid and this increases continuously as the temperature falls, till it attains t_5 (1130°C) when the alloy becomes completely solid. At this temperature, the liquid phase enriched in carbon upto 4.3% gets finally solidified and also the crystals of austentite and cementite separate simultaneously to form a eutectic mixture (i.e., ledeburite). All alloys between 1.7% and 4.3% carbon will solidify in this manner and will produce austenite and ledeburite.

Now, we consider an alloy containing 4.3% of carbon. The solidification of this alloy takes place at a constant temperature (point *C*) of 1130°C, forming ledeburite (eutectic at saturated austenite and cementite).

The solidification of alloys with 4.3% to 6.67% carbon starts along the line *CD* (Fig. 9.22) by the precipitation of cementite (of composition 6.67% carbon) from the liquid alloys. The cementite phase is richer in carbon and hence its separation will change the composition of the liquid alloy along the liquidu line *CD* (Fig. 9.22). This reaction occurs in the (hyper-eutectic cast iron) solidification of alloy which will be consisting of primary cementite crystals and ledeburite. Concluding, we can say that all alloys containing upto 1.7% carbon consist of austenite after the primary crystallization. The alloys having carbon percentage greater than 1.7% will consist of ledeburite plus primary crystals of excess austenite or cementite.

23.2 Secondary Solidification or Transformation

The transformation which occur in solid state, on cooling iron-carbon alloys below the eutectoid temperature, i.e., 723°C, are called as secondary transformations. Since these transformations occur in the steel section of the main diagram and hence are very important for all practical considerations. We know that the temperatures at which structural changes in steel on heating and cooling occur are called arrest points or critical points designated by $\mathrm{A_c}$ (for heating) and A_r (for cooling) where *A* stands for arrest.

Secondary transformations in iron-carbon alloys are accompanied by *(i)* the decomposition of austenite, and *(ii)* the transformation of γ-iron to α-iron. In Fig. 9.22, the point *G* at 910°C corresponds to the transformation of α-iron into γ-iron and vice-versa. Along the line *GS*, the decomposition of austenite takes place and the crystal of ferrite separate from the austenite during this process. The excess carbon from the austenite decomposition also forms the cementite. A solid phase reaction takes place at point *S* (0.8% carbon) at 723°C, in which simultaneous separation of ferrite and cementite forms a new structure called the pearlite. The point *S* is called the *eutectoid point* and the line *ESK* as eutectoid line.

The critical points along the line *GS* are denoted as A_{c_3} (heating) and A_{r_3} (in cooling). Cementite is separated during austenite decomposition along the line *SE* and all temperatures along this line are marked as A_{cm} points. The temperature at which pearlite is formed in cooling is denoted as A_{r_1} and the transformation of pearlite into austenite in heating is denoted by A_{c_1}. As stated earlier, the pearlite structure consists of alternate thin plates of ferrite and cementite. The formation of pearlite for all carbon alloys takes place along the line *PSK*.

GPQ region indicates the area of ferrite formation and the point *P* represents the solubility of carbon in α-iron at 723°C, which is 0.025% carbon. As shown by the line *PQ*, this solubility is reduced on cooling. For a given carbon content between the points *P* and *Q*, the composition of ferrite can be determined with the help of Lever's rule. The excess cementite which is separated is always present in this region along the boundaries of the ferrite grain. Similarly, one can observe the reduction of carbon solubility in γ-iron, during cooling along the line *SE*.

Now, we consider the changes that occur in steels containing 0.025 to when 0.8% carbon, when cooled from the solid solution or austenite region. However, the changes occur only when the temperature drops to the line *GS*. The alloys in this case have a single-phase austenite structure. Below this line *GS*, a two phase state, i.e. γ-iron autenite and α-iron ferrite, exists at temperatures because ferrite precipitates from the austenite. Along the line *GS*, the carbon concentration of the austenite will increase continuously on cooling and it will reach 0.8% at a temperature of 723°C as indicated by the point A_{c_1} (eutectoid point). At the constant temperature of 723°C, austenite of the eutectoid composition will decompose and pearlite will be formed (ferrite + cementite mixture)

The steels containing less than 0.8% carbon are known as *hypoeutectoid* steels and steels containing more than 0.8% carbon are known as *hypereutectoid* steels. When completely cooled, hypoeutectoid steels consists of ferrite and pearlite. The higher the carbon content, they will contain more pearlite and less ferrite. When temperature falls below the line *ES*, for hypereutectoid steels, the austenite is saturated with carbon and precipitates as cementite upon further cooling. Obviously, below line *ES*, a two-phase state exists consisting of austenite and secondary cementite for all completely cooled hypereutectoid steels (0.8–1.7% carbon).

Now, we consider the secondary solidification characteristics of cast iron. When completely cooled, hypoeutectic cast irons will have a structure consisting of pearlite, ledeburite (pearlite + cementite) and secondary cementite. In hypereutectic cast irons, the structure consists of ledeburite (pearlite + cementite) and primary cementite. We may note that eutectic (4.3%C) consists of ledeburite alone.

24. Formation and Decomposition of Austenite

Pure iron, upon heating, experiences two changes in crystal structure before it melts. At room temperature the stable form, called *ferrite* (or α-iron), has a BCC structure. Ferrite experiences a polymorphic transformation to FCC *austenite* (or γ-iron), at 812°C. This austenite persists to 1394°C, at which temperature the FCC austenite reverts back to a BCC phase known as δ-ferrite, which finally melts at 1538°C.

Formation of austenite from pearlite is the transformation that takes place on heating steel slightly above the equilibrium temperature (A_{c_1}), the free energy of pearlite is more than that of austenite. We may note that it is based on the nucleation and subsequent austenite crystal growth process. Obviously, this transformation is of the diffusion type and also depends upon the movement of carbon atoms over considerable distances. Regions of austenite are formed due to the dissolution of cementite and disappearance of ferrite on heating above the equilibrium temperature A_{c_1}. The rate of carbon diffusion will be greater with higher temperatures and thus pearlite to austenite transformation will be higher. We can say that this process completes in three stages: *(i)* formation of the austenite nuclei and their growth by taking up cementite and ferrite, *(ii)* dissolution of the cementite and homogenisation of austenite or *(iii)* equalising the austenite composition throughout the crystals.

When an austenite with 0.8% carbon (called eutectoid composition) is cooled below the eutectoid temperature (723°C), it decomposes into a mixture of ferrite and cementite (or iron cabide). This eutectoid mixture is called *pearlite*. Its microstructure contains alternate layers of ferrite and cementite. The fraction of ferrite and cementite present in the pearlite are:

Fraction of ferrite in pearlite, $f_\alpha = \dfrac{6.7 - 0.8}{6.7 - 0.0} = 0.88,$

and the fraction of cementite in pearlite

$$f_{Fe_3C} = \frac{0.8 - 0.02}{6.7 - 0.02} = 0.12$$

Obviously, microstructure of pearlite, is composed of 88% ferrite and 12% cementite. We may note that the structure of any alloy with carbon content below 0.8%, at room temperature will consist of partly proeutectoid ferrite and partly pearlite. Similarly, the structure of any alloy with carbon content above 0.8% at room temperature will consist of partly pearlite and partly proeutectoid cementite, e.g. steel with 0.2% carbon consists of about 75% of proeutectoid, ferrite and about 25% of pearlite. On the other hand, a steel with 0.6% carbon consists of about 38% proeutectoid ferrite and 62% pearlite. Concluding, we can say that with the increase in carbon content in steel, the amount of pearlite increases, until we obtain a fully pearlite structure at 0.8% carbon. However, if we increase the carbon content in the steel beyond 0.8%, the amount of pearlite decreases due to the presence of proeutectoid cementite.

24.1 Austenite Grain Size

We have read that grain size is of prime importance in the behaviour of metals under different loads. Grain size is also a very important factor in relation to the various physical properties of steel. Usually, grain size refers to austenite grain size. While deciding the rate at which steel is to be heated, one must take into consideration the effect of austenite grain growth or increase in temperature. Increase in austenite grain size may have the following effects:

(i) helps to improve the strength of steel after heat treatment, in fine grains.
(ii) help to improve machining finishes and machinability.
(iii) mechanical properties, e.g. creep strength, tensile strength and hardenability increases.
(iv) quenching cracks and distortion in fine grains reduces.
(v) improves the fabrication properties of a fine-grained steel. One may heat it to higher temperature without the fear of over heating, i.e. appreciable coarsening of grains.

Grain size is governed by the following factors:

(i) Composition of steel
(ii) Nature and amount of deoxidizers
(iii) Heat treatment processes, environment and working temperatures.
(iv) Metallic and non-metallic inclusions
(v) Mechanical working processes, e.g. rolling, forging, etc.
(vi) Grain size, heating and cooling time and tendency to grain growth.

24.2 Measurement of Grain Size

The following relationship may be used for determining the austenite grain size in steels.

$$n = 2^N - 1 \tag{14}$$

where N is Americation society for testing and materials (ASTM) grain size number and n is the number of grains per square mm as seen in a specimen viewed at a magnification of 100.

For the measurement of grain size, ASTM has recommended the following methods: *(i)* Comparison method *(ii)* intercept method and *(iii)* planimetric method.

The grain size is measured by comparison in the comparison method, under a microscope with a magnification of × 100 (after etching), with standard grain size ASTM charts. Usually grain sizes of steel are graded into eight classifications. Steels with grain size numbers from 1 to 5 are called coarse grained steels, whereas grain size numbers 6 to 8 are fine grained steels. By trial and error absest match is determined and one can note the grain size of the steel from the index number of the matching chart.

A photo-micrograph is used in the intercept method. On any straight line, one can divide the length of the line drawn in mm divided by the average number of grains intercepted by it gives the grain size. This method is found suitable for non-equiaxed grains. The last method, i.e. the plantimetric method, when used on a photomicrograph, indicates the number of grains per square mm in a drawn area of a circle or triangle.

For most steels, first of all, the specimen is carburised at 930°C for 8 hours. On normalisation or slow cooling, the austenite grain size will reveal a cementite network due to the transformation of austenite into pearlite. The grain size prior and after the test remains unchanged, as the heating temperature of the standard test does not exceed 930°C.

25. TYPES AND PROPERTIES OF CARBON-STEELS

The properties of carbon-steels depend upon the structure, i.e., iron-carbon system. Ferrite is relatively a soft and ductile material, whereas pearlite is harder and less ductile. Obviously, carbon steels having these substances will have therefore different properties.

The hardness and tensile strength increases with an increase in carbon content. However, toughness, machinability and ductility decreases with an increase in carbon content (Fig. 9.23). However, we note that toughness, machinability, and ductility decreases with an increase in carbon content.

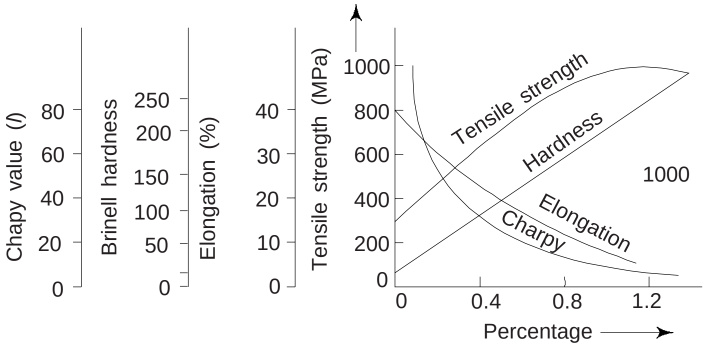


Fig. 9.23 Effect of carbon percentage (% C) on the structure of carbon-steel

If it contains other alloying elements, its properties will change further. Therefore carbon-steels are generally classified into following three groups:

(i) Mild steel or Low Carbon Steel: It contains carbon upto 0.25%. It is tough, soft, ductile and low tensile strength steel. It is a general purpose steel and can be easily worked and welded. It is used where hardness and tensile strength are not the most important requirements. Typical applications of mild steel are body work for cars and ships, screws, wires, nails, structural steels, etc.

(ii) Medium Carbon Steel: Carbon content in this steel is 0.3 to 0.55%. Strength and hardness are improved while ductility is reduced. This type of steel can be forged, rolled and machined. It finds uses for agricultural tools, fasteners, dynamo and motor shafts, crank shaft, connecting rods, gears, etc.

(iii) Tool Steel or High Carbon Steel: It is hardest of plain carbon steels and contains 0.6 to 1.5% carbon. It possess good tensile strength. Tool steel is for withstanding wear, where hardness is more necessary requirement than ductility. Tool steel is generally used for machine tools, saws, hammers, cold chisels, punches, axes, dies, taps, drills, razors. These are always used in hardened and tempered condition.

Mechanical properties of carbon steels are summarized in Table 9.3.

Table 9.3 Mechanical properties of carbon steels

Properties	*Unit*	*Low C.S.*	*Medium C.S*	*High C.S.*
U.T.S	Kg-f/cm^2	4000-6000	7500	9000
Young's Modulus (E)	Kg-f/cm^2	2.1×10^6	2.1×10^6	2.1×10^6
Shear Modulus	Kg-f/cm^2	0.8×10^6	0.8×10^6	0.8×10^6
Elongation	%	31	21	10
Hardness	HB	150	200	240

Typical applications of carbon steels with varying carbon content are given in Table 9.4.

Table 9.4 Typical applications of carbon-steels

Name	*% Carbon*	*Application*
(i) Extra low C.S.	upto 0.1	Car bodies, wire, rod, tubing, nails, sheet and strip for press work.
(ii) Low C.S.	0.1 to 0.25	Structural work chain links, wire and rod for nails, screws.
(iii) Medium C.S.	0.3 to 0.45	Shafts, high tensile tubing, anchor bolts.
	0.4 to 0.55	Shafts, gears, railway tyres, crank shafts.
(iv) High C.S.	0.55 to 0.65	Forging dies, rails, rails valve springs.
	0.65 to 0.75	Hammers, saws, cylinder liners, wrenches.
	0.75 to 0.90	Forging die blocks, punches, leaf springs.
(v) Tool steel	0.50 to 1.0	Milling cutters, drills, taps, knives, reamers.
	1.0 to 1.1	Helical springs, shaper, planner tools.
	1.1 to 1.2	Ball bearings, drills, lathe tools.
	1.2 to 1.5	Files, mandrels, wire drawing dies, razors.
(vi) Grey C.I.	3.0 to 3.5	Machine castings, machine beds.

26. ISOTHERMAL TRANSFORMATIONS-TTT DIAGRAM

Time-Temperature-Transformation (TTT) diagram or *S*-curve refers to only one steel of a particular composition at a time, which is applicable to all carbon steels. This diagram is also called as *C*-curve isothermal (decomposition of austenite) diagram and Bain's curve. The effect of time-temperature on the microstructure changes of a steel can be shown by TTT diagram. These diagrams are extensively used in the assessment of the decomposition of austenite in heat-treatable steels. We have seen that the iron-carbon phase diagram does not show time as a variable and hence the effects of different cooling rates on the structures of steels are not revealed. Moreover, equilibrium conditions are not maintained in heat treatment. Although, the iron-carbon equilibrium diagram reveals on the phases and corresponding microstructures under equilibrium conditions but several useful properties of the steels can be obtained under non-equilibrium conditions, e.g. variable rates of cooling as produced during quenching and better transformation of austenite into pearlite and martensite.

The steels with different percentage of carbon, give different TTT diagrams. The diagram shows the rate at which austenite is transformed, at a given temperature, from all austenite to coarse pearlite; to fine pearlite, to upper bainite to lower bainite, martensite plus residual austenite, depending upon the carbon content. The transformation of the austenite takes place at a constant temperature, i.e., the liquid bath temperature in which the component is cooled for the required time. Since the changes takes place at a constant temperature, it is known as *Isothermal Transformation.* One can determine the amount of microstructural changes by the microscopic examination of the sample.

In order to construct a TTT diagram, a number of small specimens of steel are heated to a temperature at which austenite are stable and then rapidly cooled to temperatures, e.g. 650°C, 600°C, 500°C, 250°C, etc. The specimens are held isothermally at these temperatures for different periods of time until the austenite is completely decomposed. Experimentally, it is observed that at the start of the cooling shown by the points B_1, B_2, B_3 and B_4, there is no decomposition of austenite. This time period is referred as *incubation period.* After this, austenite starts to decompose into the ferrite-cementite mixtures. After the lapse of a certain period of time, the process of decomposition of austenite is stopped, as shown by points E_1, E_2, E_3 and E_4 (Fig. 9.24a). Experimentally, it is observed that the rate of decomposition of austenite is not constant and initially it is rapid and slows down gradually.

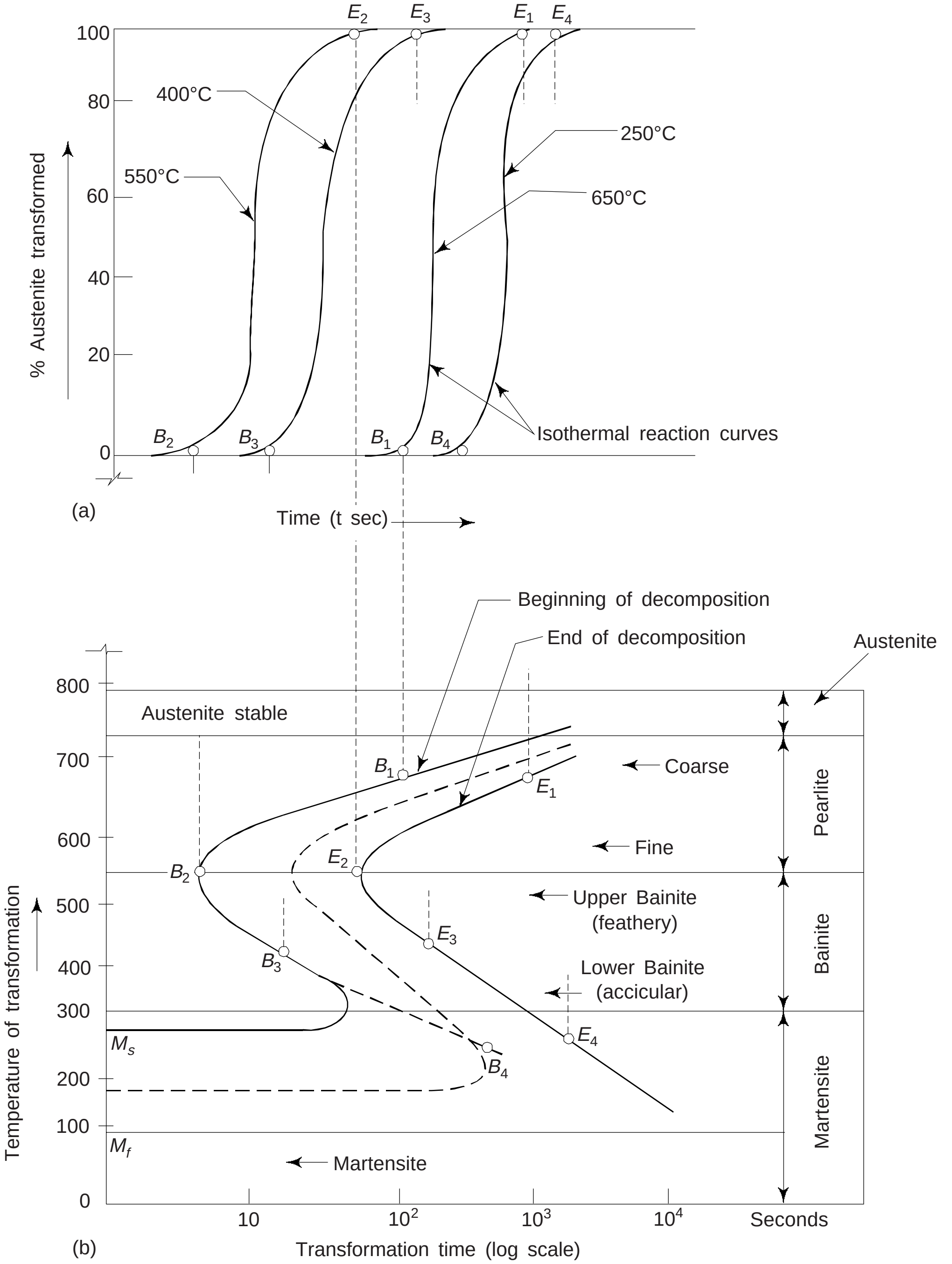


Fig. 9.24 **TTT diagram for plane carbon steel (a) isothermal transformation curves (b) construction of TTT diagram**

The isothermal transformation of the austenite in steel containing 0.8% carbon is shown in Fig. 9.24. For the sake of convenience, the time scale in such diagrams is logarithmic, since the decomposition of austenite takes from a fraction of a second to hours. One obtains two different curves by plotting the starting and end points of the decomposition of austenite. The transformation of austenite takes place in the area between these two curves. We can see from Figure 9.24, the important products of austenite decomposition.

Above 723°C, for the steel represented in Fig. 9.24, austenite is the only stable phase. Below this temperature, the stability first rapidly decreases with temperature in cooling rate. Experimentally it is found to be least stable at about 500-550°C and below this temperature range starts to increase. A diffusionless transformation of austenite into a hardened steel structure called martensite is observed at temperatures between 250 and 50°C. The temperatures, at which the martensite starts and finishes are marked by M_s and M_f. M_s stands for the temperature at which the formation of martensite begins and M_f stands for the temperature at which the formation of martensite ends. We may note that this is a super-saturated solid solution of carbon in α-iron.

A ferrite-cementite mixture with coarse pearlite is obtained at temperatures near A_{r_1} (700°C). The product of austenite decomposition (a mixture of ferrite-cementite), lamellar in structure, and finer pearlite is obtained between 700°C and 550°C, which is known as *sorbite.* An evenly dispersed mixture of ferrite-cementite (troostite) is obtained, when the temperature is lowered too between 550°C to 500°C. A needle-like structure known as accicular troostite (or bainite) is obtained when the temperature is lowered from 500°C to 300°C. We may note that Bainite is harder, stronger and tougher than pearlite. Bainite steel is more ductile than pearlitic steel for same level of hardness. We note that with fall in temperature, the ferrite-cementite mixture will be harder due to the increased rate of dispersion.

The TTT diagram is changed by alloying other elements added to carbon-steel. The effect of Cr, Ni, molybdenum is to shift the TTT diagram to the right, i.e. more time is required, at any particular temperature, for transformation from pearlite to bainite.

Normally, the TTT diagrams are derived from data obtained for transformation at constant temperature. If the continuous cooling is considered (as stated above); the TTT diagram is shifted to the right and downward. Such a diagram is referred to as a *continuous cooling transformation curve*.

27. TRANSFORMATION OF AUSTENITE UPON CONTINUOUS COOLING

Let us consider a number of specimens of eutectoid steel heated to a temperature *t*, above the critical points (Fig. 9.25). At this temperature, the steel is present in the form of stable austenite. Let the specimen of steel is cooled continuously below the lower critical point, i.e., 723°C at various cooling rates. Let the inclined curves V_1, V_2, V_3, . . . on temperature-time graph (Fig. 9.25) represent these cooling processes.

The slowest cooling rate is represented by the curve V_1. Slightly higher cooling rate is represented by the curve V_2. Still more rapid cooling rates are represented by the V_3, V_4 and V_5. We may note that these curves are straight lines.

Now, we superimpose these cooling curves (V_1, V_2, V_3 . . .) on the time-temperature transformation diagrams as shown in Fig. 9.25. From Fig. 9.25, we note that the curve V_1 crosses transformation curves 1 and 2 at points a_1 and b_1 respectively. Obviously, on slow cooling, the austenite completely transforms. into a ferrite cementite mixture. Since the transformation takes place at the highest temperature therefore the ferrite cementite mixture is pearlite. The curve V_2 also intersects both the transformation curves at points a_2 and b_2 respectively. Austenite, at this cooling rate completely transforms into ferrite-cementite mixture. The transformation occurs at lower temperature (as compared to V_1), the resulting ferrite-cementite is sorbite. Similarly, the curve V_3 also intersects both the curves at points a_3 and b_3 respectively. The resulting ferrite-cementite mixture is troostite.

We note that the curve V_4 does not cross both the transformation curves. It intersects only the curve 1 at point a_4, and does not reach the stage of completion. Obviously, a part of austentic grains transform into ferrite-cementite mixture, while the other does not transform due to insufficient time. It is observed that the remaining part of austenite which has not been transformed, undergoes transformation into martensite on

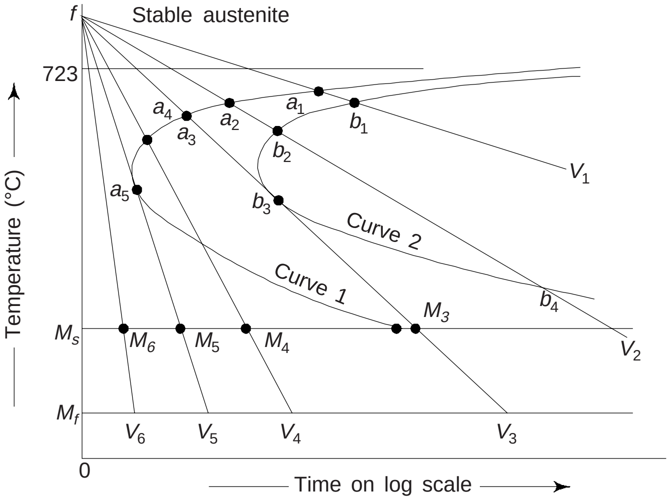


Fig. 9.25 Transformation of austenite upon continuous cooling

reaching the temperature M_s. It is shown by the intersection of the curve V_4 and temperature M_s at the point M_4. Obviously, the structure of steel, cooled at the rate of V_4, consists partly of troostite and partly of martensite. This type of structure is common to all steels, which are cooled at a rate faster than those represented by V_3 and slower than by V_5. This cooling rate for carbon steels is achieved by quenching in oil.

We may note that at any cooling rate, higher than V_5, e.g. curve V_6, of austenite does not transform into ferrite-cementite mixture. However, the austenite is transformed into martensite. Point M_5 and M_6 in Fig. 9.25 represent this transformation. This cooling rate for carbon steel corresponds to the quenching in water. We may note that the austenite is never transformed into martensite. This untransformed austenite is known as retained austenite. The minimum cooling rate, at which all the austenite is rapidly cooled to temperature M_s and is transformed into martensite is called as critical cooling rate. It is represented by straight line V_5, i.e. the tangent line drawn to the curve 1.

We may see that curves V_2 and V_3 and others between them have more slope also intersect the line M_s. Obviously, the martensite is formed at the end of transformation. However, it has reported that martensite is never formed at such cooling rates. Perhaps, this may be due to the fact that the curves V_2 and V_3 and others intersect both the transformation curves. Thus the complete transformation of the austenite takes place at points b_2 and b_3 respectively. No austenite is left in the steel beyond these points. This means that nothing is to be transformed into martensite. This is why that point M_3 in Fig. 9.25 has no physical sense.

28. TRANSFORMATION OF AUSTENITE TO MARTENSITE

Martensite is a metastable phase of steel. In the formation of martensite, the FCC structure of austenite changes to body centered in a special way which does not involve diffusion, but results from a shearing action. The crystal structure of martensite is body centered tetragonal (Fig. 9.26). Since the carbon is present in the martensite, it is hard, strong and brittle. Martensite is formed by the transformation of austenite below M_s temperature.

The hardness of martensite is much greater than that of pearlite and this enhanced hardness is of great importance. This high hardness provides an extremely high resistance to abrasion and deformation. However, martensite is a brittle. In order to reduce its brittleness, martensite is reheated without loss of its hardness. The process is called *tempering*. During tempering, metastable martensite decomposes into more stable phases of ferrite and cementite.

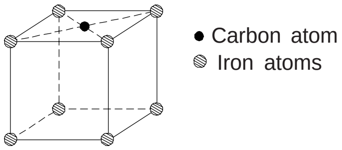


Fig. 9.26 Space lattice of martensite

Martensite is an interstitial supersaturated solid solution in iron with a body centered tetragonal (BCT) lattice and accicular of needle like structure (Fig. 9.27). The needles, in fact, are the plates of martensite, which appear in the microsection as long and thin needle. The rate of formation of martensite needles is very high (~ 50 milliseconds). However, the formed martensite needles do not grow in length or breadth. However, as the transformation proceeds further the subsequent needles become shorter and shorter. We may note the martensite transformation cannot proceed isothermally like the pearlite, sorbite transformation. Obviously, when the steel is held at a constant temperature (below the M_s point) the martensite transformation ceases very rapidly. However, with the further lowering of temperature, the martensite transformation begins again. The temperature at which the transformation is finally ceased, is denoted by M_f. We may note that M_f is different for each type of steel. This shows that the martensite transformation occurs over a wide range of temperatures from M_s to M_f.

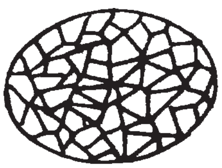


Fig. 9.27 Martensite structure

The martensite transformation range is determined by the percentage of carbon in steel. It is reported that higher the percentage of carbon in steel, lower is the temperature of beginning and end of the martensite transformation. Figure 9.28 shows the variation of the martensite temperature range with the percentage of carbon in steel.

We note from figure that above 0.7% C, the M_f temperature is below 0°C. Obviously 100% martensite cannot be produced even if the high carbon steels are quenched in ice cold water. This means that some austenite is always left untransformed. This austenite is known as residual or retained austenite. It is observed that higher the percentage of carbon in steel, greater will be the amount of retained austenite. Interestingly some of the retained austenite still remains in the structure of steel, even if this is cooled below the M_f temperature. The characteristic properties of martensite are its high hardness and extremely low impact strength.

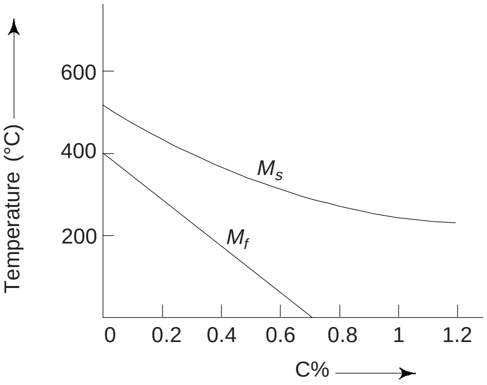


Fig. 9.28 Temperature versus percentage of carbon (C%) curves for martensite

29. METALS FOR NUCLEAR ENERGY

The aim of nuclear energy (fission and fusion) and nuclear engineering and technology is to produce electric power at a cost comparable to that of the same produced from conventional sources, e.g. coal, hydro power, etc. The raw materials for producing nuclear energy are known as fuels. The various metals for producing nuclear energy are used as materials for moderators, reflectors, fuel elements, control elements, and pressure vessel metals. Materials used to prevent corrosion by coolant media and escape of radioactive fission products are termed as canning materials. The important metals used for nuclear energy are uranium, thorium, plutonium, zirconium, beryllium, niobium and their alloys. These are used as raw material, moderators, reflectors and construction materials for pressure vessels.

Good number of materials are used for nuclear energy and engineering, e.g. structural heat resistant, stainless steels, alloys, non-ferrous metals, etc. Materials used for nuclear energy have special properties. During the nuclear fission reactions, neutrons are released.

$$_{92}U^{235} + {}_0n^1 \rightarrow {}_{92}U^{236} \rightarrow {}_{36}Kr^{90} + {}_{56}Ba^{144} + {}_{2 0}n^1 + 200 \text{ MeV} \quad (15)$$

We may note that there are also other ways in which U^{236} may undergo fission giving rise to $_{57}La^{146}$ and $_{35}Br^{87}$ in place of fission fragments given in Eq. (15). The release of two or three neutrons per nuclear fission is important for what is called the chain reaction. It is probable that one or two nearby U^{235} atoms will capture these neutrons and undergo fission and thus making the fission reaction self-sustaining. The absorbing capacity of biological protection and fission reaction controlling materials should be high and low for shell materials. The neutron absorbing capacities and melting temperatures of few materials are given in Table 9.5.

Table 9.5

Property	Cd	B	Hf	Ni	Fe	Al	Zr	Mg	Be
(i) Effective neutron capaturing cross-section area in 10^{-28} m^2	2400	729	115	4.5	2.4	0.23	0.18	0.063	0.01
(ii) Melting point temperature (°C)	321	2000	2130	1455	1539	660	1845	651	1280

For the operation of a nuclear power plant is necessary to be able to control the rate of production of power and also it is necessary to shut the plant whenever required. For this purpose, control rods are used in nuclear power plants. The fundamental parts of a nuclear power plant or nuclear reactor are: *(i)* fuel, *(ii)* moderator *(iii)* coolant *(iv)* control rods *(v)* pressure vessel *(vi)* reflector *(vii)* radiation shield.

Nuclear fuel is made from a fissionable materials like U^{237}, U^{235}, Pu^{239} or Pu^{241}. Of these U^{235} is the only one which is available relatively in plenty. Modern nuclear power plants utilise the uranium fuel as an oxide (UO_2) or a carbide (UC) or a nitride (UN). The main parameters which decide the use of nuclear fuels in power plants are melting point, specific heat, thermal conductivity, density of metal, and irradiation stability.

The heat produced in nuclear power plants is intense. Water and gases can be used as *heat carriers*. Liquid metals with higher heat conductivity and latent heat and low melting point are employed as heat carriers. Bismuth and its alloys and sodium are also used. Bismuth is most suitable as it is light and has low melting point, 217°C. Pb-Bi eutectic alloys (43.5 Pb, 56.5% Bi) is more suitable with lower melting point of 215°C, less expensive and more fluid although has slightly lower uranium stability and high absorption capacity for neutrons. Sometimes Lithium is also used.

In order to slow down 2 MeV neutrons produced in the fission process to 0.025 V energies, moderators are used. Hydrogen, deuterium, beryllium, carbon and oxygen having low atomic mass and small neutron capture cross-section are used. Boron, cadmium, samarium, hafnium are found suitable for use as a control material.

Example 1 *Find the percentage of proeutectoid ferrite and pearlite in a slowly cooled 0.5% carbon steel.* [Diploma]

Solution Let fulcrum is at 0.5% carbon. We have from lever rule

$$\% \text{ Proeutectoid ferrite} = \frac{0.80 - 0.5}{0.80 - 0.0} \times 100 = 37.5\%$$

$$\% \text{ Pearlite ferrite} = 100 - 37.5 = 62.5\%$$

Example 2 *Determine the degrees of freedom of a system of two components, when the number of phases is one, two, three, etc.* [B.E., M.Sc. (M.S.)]

Solution We have $F = C - P + N$

For $C = 2$,

one can list the total variables and the degrees of freedom in the following table

No. of phases	*Total variables* $P(C - 1) + 2$	*Degrees of freedom* $C - P + 2$
1	3	3
2	4	2
3	5	1
4	6	0

In equilibrium, the system cannot have more than four phases.

Example 3 *A material of unknown composition at atmospheric pressure (arbitrarily chosen) exhibit four phases at 987 K. Find the minimum number of components in the system?* [BE]

Solution At arbitrarily chosen pressure, one can used the modified form of the phase rule, i.e.

$$F = C - P + 1$$

The minimum number of compounds would correspond to minimum degrees of freedom, which is zero. Thus, taking $F = 0$, we have

$$0 = C - 4 + 1$$

or $$C = 3$$

SUGGESTED READINGS

1. P. Gordon, *Principles of Phase Diagrams in Material Systems*, McGraw Hill, New York (1968). Reprinted by Krieger Publishing Company, Melbourne, FL (1983)
2. *ASM Handbook*, Vol. 3, *Alloy Phase Diagrams*, ASM International, Materials Park, OH (1985).
3. H. Okamoto, *Desk Hand book* : *Phase Diagrams for Binary Alloys*, ASM International, Materials Park, OH (2000).

REVIEW QUESTIONS

1. What is a solid solution? Explain *(i)* substitutional solid solution and *(ii)* interstitial solid solution. Write two examples of each in alloy system. [BE]

2. State Hume Rothery's rules for the formation of substitutional type of solid solutions. [BE]

3. Carbon when dissolved in iron, forms interstitial solid solution. Explain. [BE]

4. What are three variables that determine the microstructure of an alloy. [Diploma]

5. What thermodynamic condition must be met for a state of equilibrium to exist? [Diploma]

6. What is a solid solution? Discuss the similarities and differences between substitutional and interstitial solutions.

7. What is solid solution? Give types of solid solution. [AMIE]

8. Draw a schematic TTT curve for a 0.8% plain carbon steel. Label all phase regions. [BE]

9. Draw Fe-C phase diagram. Label all the phases and temperatures properly. Describe the phase changes during solidification of Fe – 0.45% C alloy. [BE]

10. State and explain 'Gibb's phase rule', defining all terms with examples. [BE]

11. Draw the cooling curves for a binary system forming eutective solution. Explain, how cooling curves are useful. [BE]

12. Write short notes on the following

(i) Hume-Rothery's rules
(ii) Interstitial compounds
(iii) Gibb's phase rule
(iv) Zone refining
(v) Cooling curves
(vi) Various types of solid solutions
(vii) Zone refining
(viii) Eutectoid and Peritectic reaction
(ix) Coring
(x) Precipitation
(xi) T.T.T. curve [AMIE, BE]

13. Describe the allotropic transformations in iron and explain their importance in practical applications. [BE]

14. Draw Fe-C phase diagram. Label all the phases and temperatures properly. Describe the phase changes during solidification of Fe – 0.45% alloy. [BE]

15. Define critical cooling rate. Give critical cooling rates for different types of carbon steels. [BE]

16. What do you understand by critical points and critical range and how are they related to heating and cooling of steel. Explain their importance in phase transformations. [BE]

17. Why there are tie lines for three phase equilibria but not for two phase equilibrium in a two component system? [Diploma]

18. Describe the working of Fe-C diagrams and list the advantages and limitations of these diagrams when applied to heat treatment. [BE]

19. Explain the working and importance of TTT diagrams. What information is supplied by these diagrams? [BE]

20. Describe the process of austenite decomposition of alloyed steels with TTT diagrams. [Diploma]

21. Write short notes on

(i) Martensite *(ii)* Cementite *(iii)* Pearlite *(iv)* Ferrite *(v)* Pearlite
(vi) Ferrite *(vii)* Cementite *(viii)* Bainite *(ix)* Retained Austenite [Diploma]

PROBLEMS

1. Determine the raction of β-phase in an alloy of 80% Sn in the Pb-Sn system at 184°C and 182°C [BE], [*Ans.* 0.51, 0.78]

2. Calculate the volume percentage of phase present in an alloy of 16% by weight silicon and 84% by weight aluminium. Given density of Si = 2.35 gm/cc and density of aluminium = 2.7 gm/cc. [BE]

3. A Pb-Sn alloy containing 10% Sn is cooled from 200 to 100°C. How much β (Wt%) precipitates from the alloy as a result of the cooling? [Diploma]

4. What is the fraction of proeutectoid cementite in a *(i)* 1.4% C, *(ii)* 1.0% C, and *(iii)* 0.7% C steels? [*Ans.* *(i)* 0.10, *(ii)* 0.034 and *(iii)* 0] [Diploma, BE]

5. Derive the degrees of freedom for a system, which has equal number of components and phases. [Diploma] [*Ans.* 2]

6. In the *Pb-Sn* system, calculate the alloy composition at which the fraction of total α is 5/2 times the fraction of the β phase at 182°C. [BE] [*Ans.* 40.6% Sn]

SHORT QUESTION ANSWERS

1. Why intermetallic compounds are used as bearing materials?

2. Why salt is added into the ice on road in cold countries?

3. Why equilibrium phase diagrams are preferred?

Ans. These are a convenient and concise way of representing the most stable relationships between phases in alloy systems.

4. What is the use of tie lines and lever rule within two phase regions?

Ans. Tie lines and lever rule are used for computations of phase composition and mass fraction within two phase regions.

5. What are isomorphous diagrams?

Ans. These are the diagrams for which there is complete solubility in the solid phase. The copper-nickel system displays this behaviour.

6. What is the significance of phase rule?

Ans. This gives the degrees of freedom that are permissible for a system of a given number of components and phases in equilibrium.

7. What are the binary phase diagrams?

Ans. These are temperature composition plots, where the pressure variable is omitted.

8. What are the solid state analogue of the eutectic and peritectic reactions?

Ans. Eutectoid and peritectoid reactions are the solid state analogues of eutectic and peritectic reactions respectively.

9. What are the advantages of brasses over capper?

Ans. Brasses have better mechanical properties and are more economical than copper.

10. What is the significance of a 'lever rule'?

Ans. This is a simple rule of mass balance, giving the relative amounts of two phases (or microconstituents) for a given overall composition.

11. Which phase diagram do you consider technologically most important?

Ans. Iron-iron carbide.

OBJECTIVE QUESTIONS

1. The degrees of freedom for a system having equal number of components and phases will be
(1) 1 (2) 2 (3) 3 (4) 5 (2)

2. In a single component system, the maximum number of phases that can coexist in equilibrium is
(1) 5 (2) 7 (3) 3 (4) 2 (3)

3. Pearlite phase in steel is made up of
(1) alternate layers of martensite and cementite
(2) alternate layers of ferrite and cementite
(3) alternate layers of ferrite and martensite
(4) alternate layers of bainite and cementite (2)

4. In which of the following phases of steel cementite is in particle form?
(1) Martensite (2) Ferrite (3) Pearlite (4) Bainite (4)

5. In which of the following phases of steel cementite is in lamellar form?
(1) Ferrite (2) Bainite (3) Martensite (4) Pearlite (4)

6. The reaction that yields two solid phases on cooling a single solid phase is called
(1) eutectic (2) peritectoid (3) congruent (4) eutectoid (4)

7. When FCC iron and BCC iron coexist in equilibrium, the degrees of freedom are
(1) 1 (2) 2 (3) 3 (4) 0 (1)

8. The maximum number of coexisting phases in a C-system is
(1) C–1 (2) $C + 2$ (3) $P(C–1)$ (4) $C–P + 2$ (2)

9. When ice, water and water vapour coexist in equilibrium, the degrees of freedom are
(1) 0 (2) 1 (3) 2 (4) 3 (1)

10. The pearlite content in plain carbon steel
(1) increases with carbon content upto 8% and then decreases
(2) increases with increasing carbon content upto 1.2%
(3) decreases as carbon content increases
(4) none of the above (1)

11. During cooling, the complete transformation of austenite takes place from liquid state
(1) at 723°C (2) just above 723°C
(3) just below 723°C (4) none of the above (3)

12. The reaction that on heating one solid phase yields another solid phase together with one liquid phase is termed
(1) peritectic (2) peritectoid (3) eutectic (4) eutectoid (1)

13. On heating, if one solid phase splits into two solid phases, the reaction is
(1) eutectoid (2) eutectic (3) peritectic (4) peritectoid (4)

14. The fraction of pearlite in a 0.55% C-steel is
(1) 0.33 (2) 0.69 (3) 0.63 (4) None of the above (2)

15. If the rate of cooling of a liquid metal is rapid, the temperature of freezing/crystallization will
(1) decrease (2) increase
(3) remain constant (4) none of the above (1)

10

Heat Treatment

1. INTRODUCTION

Heat treatment refers to the heating and cooling operations required to alter the properties of metals, alloys plastic and ceramic materials. Changes in material's properties result from changes made in microstructure of the material. Heat treatment can be applied to ingots, castings, semi-finished products, welded joints and various elements of machines and instruments.

During heat-treatment of a metal piece, when it is heated to a definite temperature followed by cooling at a suitable rate, there occur changes in the micro-constituents of the metal. These changes in the micro-constituents of the metal may be in their nature, form, size and distribution in the metal piece. Obviously, temperature of heating and rate of cooling are the main controlling factors of changes in micro-constituents. These changes in micro-constituents then control the changes in physical and mechanical properties of heat treated metal specimen. For various fabrication and manufacturing operations, heat treatment is a very important process. The purpose of heat treatment is to achieve any one or more objectives cited as follows:

(i) To remove strain hardening of a cold worked metal and to improve its ductility.

(ii) To relieve internal stresses set up during cold-working, casting, welding and hot-working treatments.

(iii) To remove gases from castings, to soften a metal to improve its machinability, and to increase the resistance to wear, heat and corrosion.

(iv) To improve the cutting ability, i.e., hardness of a steel tool, to improve grain structure after hot working a metal and to remove effects of previously performed heat-treatment operations.

(v) To improve magnetization property, especially of steels, for producing permanent magnets.

(vi) To refine grain structure after hot working a metal.

(vii) To soften and toughen a high carbon steel piece.

(viii) To produce a single phase alloy in stainless steel, and to produce a hard, wear resistant case on a tough core of a steel part.

(ix) To harden non-ferrous metals and alloys, especially aluminium alloys and to produce a single phase alloy in stainless steel.

(x) To produce a hard, wear resistant case on a tough core of a steel part and to toughen a hardened steel piece at the cost of its hardness.

The principle of the theory of heat treatment is that when an alloy has been heated above a certain temperature, it undergoes a structural adjustment or stabilization when cooled to room temperature. The cooling rate plays an important role in this operation. The structural modification is mainly based on the cooling rate. The heat treatments are normally applied to hypo-eutectoid carbon steels. For steel the eutectoid reaction in the iron-carbon diagram involves the transformation and decomposition of austenite into pearlite, cementite or martensite. Figure 10.1 shows the temperature ranges and heat treatment processes. Common microstructure of steel obtained during heat treatment is shown in Fig. 10.2. Figure shows the iron-iron carbide phase diagram in the vicinity of the eutectoid. The horizontal line at the eutectoid tem-

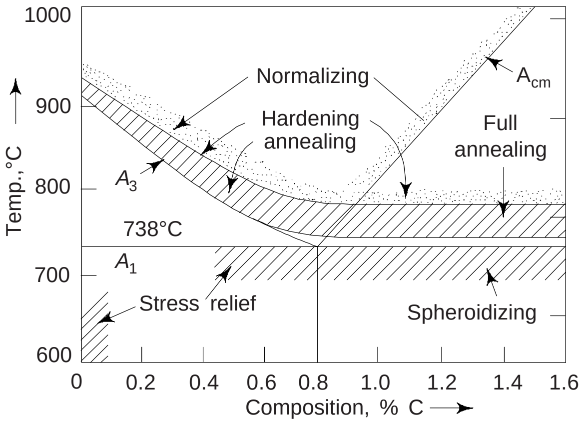


Fig. 10.1 Heat treatment range for carbon steels

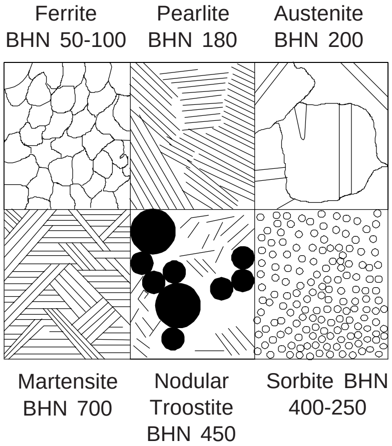


Fig. 10.2 Microstructures of steel

perature, conventionally labelled A_1, is termed the *lower critical temperature*, below which, under equilibrium conditions, all austenite will have transformed into ferrite and cementite phases. The phase boundaries denoted as A_3 and A_{cm} represent the *upper critical temperature* lines, for hypoeutectoid and hyper-eutectoid steels respectively.

For temperatures and compositions above these boundaries, only the austenite phase will prevail. Other alloying elements will shift the eutectoid and the positions of these phase boundary lines.

2. HEAT-TREATMENT PROCESSES

The principal kinds of heat treatment are:

(i) Annealing *(ii)* Normalising *(iii)* Hardening *(iv)* Tempering *(v)* Case hardening *(vi)* Surface hardening and *(vii)* Ageing. Each of them has a number of varieties.

3. ANNEALING

This is a kind of heat treatment after which a metal or alloy acquires a structure close to the equilibrium one. A material is exposed to an elevated temperature for an extended time period and then slowly cooled. Normally, annealing is carried out to *(i)* relative stresses *(ii)* increase softness, ductility and toughness; and/ or *(iii)* produce a specific microstructure.

The temperature of heating in annealing depends on the composition of an alloy and the particular kind of the process; the rate of cooling from the annealing temperature is usually not high *(*within 30-200°C/h*)*.

There are variety of annealing heat treatments possible. These are characterized by the changes that are induced, which many times are microstructural and are responsible for the alteration of the mechanical properties.

An annealing process consists of three stages: *(i)* heating to the desired temperature, *(ii)* holding or 'soaking' at that temperature, and *(iii)* cooling, usually to room temperature. In these annealing procedures, time is an important parameter. There exist temperature gradients between the outside and interior portions of the piece during heating and cooling. The magnitudes of temperature gradients depend on the size and geometry of the piece. If the rate of temperature, change is too great, temperature gradients and internal stresses may be induced that may lead to wraping or even cracking. Moreover, the actual annealing time must be long enough to allow for any necessary transformation reactions. An annealing temperature is also an important consideration. Since diffusional processes are normally involved and therefore annealing may be accelerated by increasing the temperature. The various types of annealing operations are: *(i)* Full annealing, *(ii)* Process annealing, *(iii)* Spheroidise annealing and *(iv)* Diffusion annealing.

4. ANNEALING OPERATIONS

4.1 Full Annealing

This operation removes all structural imperfections by complete recrystallization. This operation is often utilized in low and medium carbon steels that will be machined or will experience extensive plastic deformation during a forming operation. This operation consist of:

(i) Heating the hypoeutectoid steel to about 50-70°C above the upper critical temperature *(*for hypoeutectoid steels*)* and by the same temperature above the lower critical temperature for hypereutectoid steels until equilibrium is achieved. This ensures that the metal is heated thoroughly and phase transformation has taken place throughout the whole volume.

(ii) The alloy is then furnace cooled; i.e., the heat-treating furnace is turned off and both furnace and steel cool to room temperature at the same rate, which takes several hours.

The microstructural product of full anneal is coarse pearlite (in addition to any proeutectoid phase) that is relatively soft and ductile. The full-annealing cooling procedure is time consuming; however, a microstructure having small grains and a uniform grain structure results.

4.2 Process Annealing or Partial Annealing

This is a heat treatment that is used to negotiate the effects of cold work, i.e. to soften and increase the ductility of a previously strain hardened metal. Process annealing is commonly utilized during fabrication procedures that require extensive plastic deformation, to allow a continuation of deformation without fracture or excessive energy consumption. It is the recrystallization of cold work, i.e., recovery and recrystallization processes are allowed to occur. Ordinarily a fine grained microstructure is desired, and therefore, the heat treatment is terminated before appreciable grain growth has occurred. In other words, the exact temperature depends upon the extent of cold working, grain size, composition and time held at heat. Surface oxidation or scaling may be prevented or minimized by annealing at relatively low temperature (but above the recrystallization temperature) or in a non oxidizing atmosphere. Process annealing is very useful in mild steels and low carbon steels (Fig. 10.3).

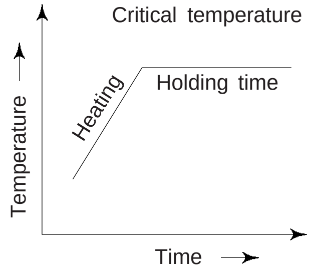


Fig. 10.3 Process annealing

Process annealing or sub-critical annealing which is done on cold-worked low carbon steel sheet, wire or tubing to relieve internal stresses and to soften the material. The process is as follows:

(i) The steel is heated to 550-650°C, which is just below the lower critical temperature on iron-carbon diagram for steel.

(ii) Stresses throughout the metal are relieved and recrystallization causes new grains to form and grow. Changes taking place during process annealing are shown in Fig. 10.4.

Heating period is followed by slow cooling. Prolonged annealing causes the cementite in the pearlite to “ball up” or spheroidise. Ferrite grain growth also occurs. Obviously, annealing time and temperature control is very essential for proper process annealing.

Patening: It is mainly applied to medium to high carbon steels prior to drawing of wire or between drafts. This mainly increases ductility for wire drawing. The process is as follows:

(a) Heating to a temperature above the transformation range and

(b) then cooling to a temperature below that range in air or in a bath of molten lead or salt maintained at a temperature appropriate to the carbon content of the steel and the properties required of the finished product.

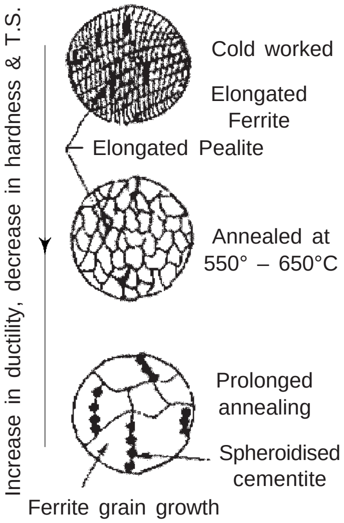


Fig. 10.4 Changes for low carbon steel in process annealing

Stress-Relief Annealing

Internal residual stresses may develop in metal pieces due to: *(i)* plastic deformation processes such as matching and grinding; *(ii)* non uniform cooling of a piece that was processed or fabricated at an elevated temperature, such as a weld or a casting; and *(iii)* a phase transformation that is induced upon cooling where in parent and product phases have different densities. Distortion and warpage may result in case if residual stresses are not removed. They can be eliminated by this process-in which the piece is heated to the recommended temperature, held there long enough to attain a uniform temperature, and finally cooled to room temperature in air. We may note that the annealing temperature is ordinarily a relatively low one such that effects resulting from cold working and other heat treatments are not affected, castings, forgings, weldment and other work pieces may have residual stresses.

In this process, the work pieces are first heated to about their recrystallization temperature and then these are cooled slowly. Like process annealing this stress-relief may also be performed on any metal.

Double Annealing: It is quite useful for steel castings. It removes the strains. It coalesces the sulphide films (which embrittle the steel) in the ferrite and produces homogeneity by rapid diffusion. The steel is ultimately obtained with refined grains and in soft condition.

In this process, steel is heated to a temperature considerably over the A_3 point, i.e. above the critical range for a time period. Then rapid cooling is performed to a temperature below the lower critical temperature. Now, immediately reheating is done to a point just over the upper critical temperature for a time period. Finally, slow cooling allowed to room temperature.

Normalizing

This is used as a finishing treatment for carbon steels giving higher strength than annealing. There is no serious loss of ductility too. Heating and soaking in this process is same as in the full annealing but part is allowed to cool in air so that cooling rate is much faster. An annealing heat treatment called normalizing is used to refine the grains (i.e., to decrease the average grain size) and produce a more uniform and desirable size distribution. Fine grained pearlite steels are tougher than coarse-grained ones. The fine grain structure increases the yield and ultimate strengths, hardness and impact strength. Normalizing is accomplished by heating at approximately 55 to 85°C above the upper critical temperature, which is, of course, dependent on composition.

Normalizing often applied to castings and forgings is stress relieving process. To some extent, it increases strength of medium carbon steel. It improves machinability, when applied to low carbon steel.

Alloy steels in which the austenite a procedure termed austenizing is very stable can be normalized to produce hard martensitic structure. Cooling in air produces high rate of cooling which can decompose the austenitic structures in such steels and martensite is produced. This increases the hardness to great extent.

The advantages of this method are:

(i) In comparison to fully annealed material, normalizing produces stronger material.
(ii) Normalizing refines the grains.
(iii) Normalizing produces homogenised structure.
(iv) Normalizing is used to improve properties of steel castings instead of hardening and tempering.
(v) Strength and hardness are increased.
(vi) Better surface finish is obtained in machining.
(vii) Resistance to brittle fracture is increased in hot-rolled steel.
(viii) Crack propagation is checked.

Spherodising

Tool steels for cutting tools, measuring instruments and cold forming dies have a high concentration of carbon (07-2%), which provides high hardness and makes these steels poorly machinable. The hardness of tool steels can be reduced by annealing. Besides, spheroidizing annealing of hypereutectoid steels prepares their structure to hardening. This type of heat treatment produces carbide in the form of round or globular (spheroids) instead of plates as in pearlite. This structure gives: good machinability, high ductility and improvement in formability.

The hardness of a steel is the lowest when the steel structure consists of granular pearlite with inclusions of rounded-off (spherical) cementitie grains. From this the name 'spheroidizing' annealing transpires.

The spheroidizing heat treatment consists of heating the alloy at a temperature just below the eutectoid phase diagram (line A_1 in Fig. 10.1, or at about 700°C] in the $\alpha + Fe_3C$ region of the phase diagram. If the precusor microstructure contains pearlite, spheroidizing times will ordinarily range between 15 and 25 hours. During this heat annealing there is coalescence of the Fe_3C to form the spheroid particles.

4.3 Hardening

It is a kind of heat treatment which forms a non-equilibrium structure in an alloy. Non-equilibrium structures can be produced by heat treatment only in cases when the alloy being treated undergoes certain transformations in solid solutions, decomposition of a high-temperature solid solution by eutectoid reaction, etc. To form a non-equilibrium structure in an alloy, it is heated above the temperature of the phase transformation in the solid state and then cooled (chilled) quickly; fast cooling is essential for preventing the equilibrium transformation during cooling.

Structural and tool-making alloys are hardened in order to increase their strength. Alloys undergoing a eutectoid transformation under the equilibrium conditions can be strengthened by hardening quite substantially. Their strength increases either due to the martensitic phase change or due to a reduction of the temperature of eutectoid reaction; in both cases there forms a fine grained eutectoid mixture. If the hardening procedure has resulted in that the metal at room temperature (20-25°C) has the fixed state of high-temperature solid solution, the strengthening effect immediately after hardening is insignificant; it will be pronounced mainly on a repeated low-temperature heating or after holding at 20-25°C.

In alloys possessing special properties, hardening makes it possible to change the structure-sensitive physical and chemical properties, for instance increases the electric resistance, coercive force or corrosion resistance.

Hardening capacity and *hardenability* are important characteristics of steels. Hardening capacity is defined as the surface hardness of a hardened article and depends mainly on the carbon content of the steel. We may note that the influence of alloy composition on the ability of a steel alloy to transform to martensite for a particular quenching treatment is related to a parameter *hardenability*. Moreover, for every different steel alloy there is a specific relationship between the mechanical properties and the cooling rate. 'Hardenability' is a term that is used to describe the ability of an alloy to be hardened by the formation of martensite as a result of a given heat treatment. We may note that hardenability is not "hardness", which is the resistance to indentation; rather, hardenability is a qualitative measure of the rate at which hardness drops off with distance into the interior of a specimen as a result of diminished martensite content.

There are two testing methods to measure the hardenability of steels: *(i) cylinder series test* and *(ii) Jominy end-quench test*. The first method gives a single value of hardenability. The value is stated in terms of percentage of martensite at the centre when quenched in a certain manner. Severity of quench is an index that quantitatively defines the quenching condition. This index is denoted by H and defined as

$$H = \frac{\text{Heat transfer coefficient between steel and fluid}}{\text{Thermal conductivity of steel}}$$

When $H \to \infty$, it represents the severest condition of quench, i.e., surface of steel immediately reaches the temperature of the quenching medium. Table 10.1 summarizes relative values of H that can be obtained in various quench media under different condition with value of one for still water as base.

Table 10.1 Relative quench severities

Agitation of quenching medium	*Movement of piece*	*Severity of quench*			
		Air	*Oil*	*Water*	*Brine*
None	None	0.02	0.3	1.0	2.2
None	Moderate	—	0.4-0.6	1.5-3.0	—
None	Violent	—	0.6-0.8	3.0-6.0	7.5
Violentor spray	—	—	1.0-1.7	6.0-12.00	—

The Jominy End-Quench Test

This is most convenient and widely used laboratory test for hardenability. With this test, except for alloy composition, all factors which may influence the depth to which a piece hardens (i.e., specimen size and shape, and quenching treatment) are maintained constant. A standard specimen (25.4 mm in diameter and 100 mm long) is austenitized at a prescribed temperature for a prescribed time. After removal from the furnace, it is quickly mounted in a fixture (Fig. 10.5a). The lower end of the specimen is then quenched by a standard jet of water, resulting into a varying rate of cooling.

The cooling rate at the jet end is about 300°C/s while that at the other end is about 3°C/s. After the piece has cooled to room temperature, shallow flats 0.4 mm deep are ground along the specimen length and Rockwell hardness measurements are made for the first 50 mm along each flat (Fig. 10.5b); for the first 12.8 mm, hardness observations are noted at 1.6 mm intervals, and for the remaining 38.4 mm, every 3.2 mm. This varying cooling rate produces a wide range of hardness along the length of Jominy specimen. A hardenability curve is obtained when hardness is plotted as a function of position from the quenched end.

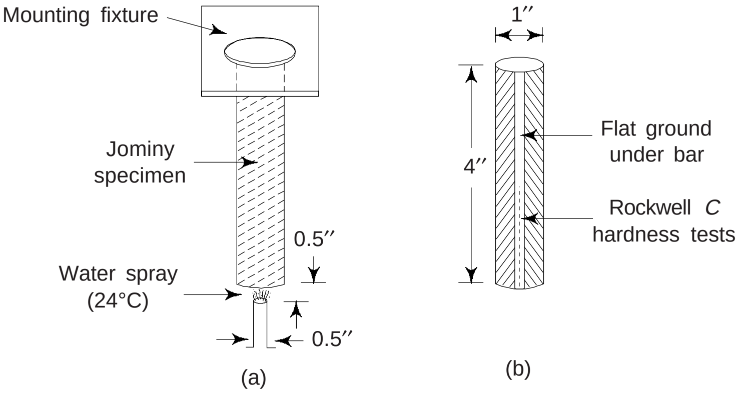


Fig. 10.5 Jominy end-quench specimen (a) mounted during quenching, and (b) hardness is measured along the length of the specimen

Hardenability Curves

Figure 10.6 shows a typical hardenability curve. The quenched end is cooled quite rapidly and exhibits the maximum hardness. For most steels, 100% martensite is the product at this position. Cooling rate decreases with distance from the quenched end, and the hardness also decreases. With diminishing cooling rate proper time is allowed for carbon diffusion and the formation of a greater proportion of the softer pearlite, which may be mixed with martensite and bainite. Obviously, a steel that is highly hardenable will retain large hardness values for relatively long distances and a low hardenable will not. We may note that each steel alloy has its own unique hardenability curve.

Sometimes, one may find it convenient to relate hardness to a cooling rate rather than to the location from the quenched end of a standard Jominy specimen. The correlation between position and cooling rate is the same for plain carbon and many alloy steels as the rate of heat transfer is nearly independent of composition. Usually, cooling rate or position from the quenched end is specified in terms of Jominy distance (1 Jominy distance unit = 1.6 mm).

The hardenability curves for five different steel alloys all having 0.40 Wt% C, yet differing amounts of other alloying elements are shown in Fig. 10.7. From figure, we note that all five alloys have identical hardnesses at the quenched end (57 HRC). However, this hardness is a function of carbon content only, which is the same for all these five alloys.

Perhaps, the most significant feature of these hardenability curves is shape, which relates to hardenability. The hardenability of the plane carbon 1040 steel is low as the hardness drops off precipitously (to about 30 HRC) after a relatively short Jominy distance of 2 in (50 mm). The hardnesses of the 4340 and 8640 steel alloys are about 50 and 32 HRC, respectively. Obviously, out of these two alloys, the 4340 alloy is more hardenable. Importantly, a water-quenched specimen of the 1040 plain carbon steel would harden only to a shallow depth below the surface, whereas for the other four alloys the high quenched hardness would persist to a much greater depth.

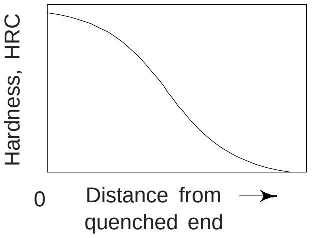


Fig. 10.6 Hardenability curve of Rockwell *C* hardness as a function of distance from the quenched end.

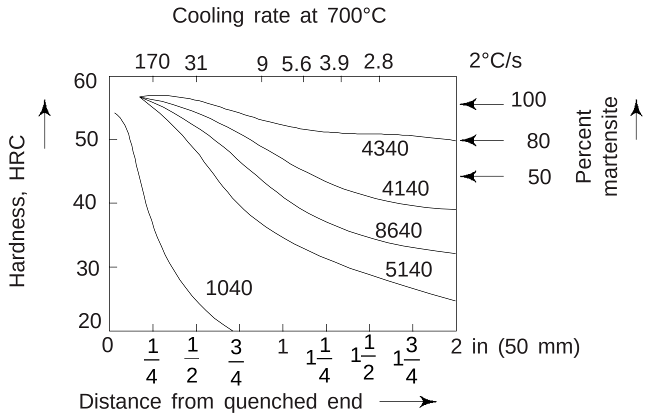


Fig. 10.7 Hardenability curves for five different steel alloys. Each alloy contains 0.4 Wt% C. Alloy composition (Wt%) are approximately as: (i) 4340 – 1.85 Ni, 0.80 Cr, and 0.25 Mo; (ii) 4140 – 1.0 Cr, and 0.25 Mo; (iii) 8640 – 0.55 Ni, 0.50 Cr, and 0.20 Mo; (iv) 5140 – 0.85 Cr, and (v) 1040 an unalloyed steel

Hardenability curves shown in Fig. 10.6 also shows the influence of cooling rate on the microstructure. We may see that at the quenched end, where the quenching rate is approximately 600°C/s, 100% martensite is present for all five steel alloys. When the cooling rates are less than about 70°C/s or Jominy distances greater than about 6.4 mm, the microstructure of the 1040 steel alloy is predominantly pearlitic, with some proeutectoid ferrite. However, the microstructure of the other four steel alloys consist primarily of a mixture of martensite and bainite. The bainite content increases with decreasing cooling rate.

One can explain the disparity in hardenability behaviour of the said five steel alloys (Fig. 10.7) by the presence of nickel, chromium, and molybdenum in alloys. These alloying elements delay the austenite-to pearlite and/or bainite reactions. This permits more martensite to form for a particular rate, yielding a greater hardness.

We may note that the hardenability curves also depend on carbon content. Figure 10.8 demonstrates this effect for a series of steel alloys in which only the concentration of carbon is varied. We note that the hardness at any Jominy position increases with the concentration of carbon.

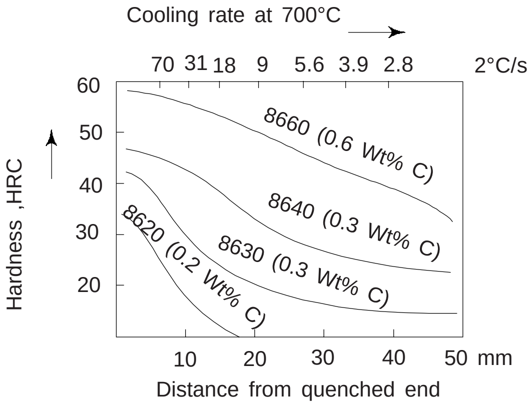


Fig. 10.8 Hardenability curves for four 8600 series alloys of steel exhibiting the effect of variation of carbon concentration

Influence of Quenching Medium, Specimen, Size, and Geometry

So far we have discussed the influence of both alloy composition and cooling or quenching rate on the hardness. The cooling rate of a specimen depends on the rate of heat energy extraction, which is a function of the characteristics of the quenching medium in contact with the specimen surface and also on the size and geometry of the specimen.

"Severity of quench" is widely used to express the rate of cooling. Interestingly, the more rapid the quench, the more severe the quench. Water, oil and air are the three most common quenching media. Out of these three, water produces the most severe quench followed by oil, which is most effective than air. The degree of agitation of each medium also influences the rate of heat removal. The quenching effectiveness can be enhanced by increasing the velocity of the quenching medium across the specimen. Oil quenches are found suitable for the heat treating of many alloy steels. A water quench is too severe for higher. Carbon steels, because cracking and wraping may be produced. Air cooling of austenitized plain carbon steels ordinarily produces an almost totally pearlitic structure.

During the quenching of a steel specimen, heat energy must be transported to the surface before it can be dissipated into the quenching medium. As a consequence, the cooling rate within and throughout the interior of a steel structure varies with position and depends on the geometry and size.

The heat energy is dissipated to the quenching medium at the specimen surface and the rate of cooling for a particular quenching treatment depends on the ratio of surface to the mass of the specimen. The larger this ratio, the more rapid will be the cooling rate and, consequently, the deeper the hardening effect. Interestingly, irregular shapes with edges and corners have larger surface-to-mass ratios than regular and rounded shapes (e.g. spheres and cylinders) and are thus more amenable to hardening by quenching.

We may note that there are a multitude of steels that are responsive to martensitic heat treatment, and one of most important criteria in the selection process is hardenability.

Hardening may cause few defects due to prolonged or uneven heating, temperature higher than required, improper cooling. Few of these defects are: *(i)* hard and soft spots due to uneven or prolonged heating, *(ii)* oxidation and decarburisation due to high temperature, *(iii)* mechanical properties achieved different from anticipated, *(iv)* wrappage due to high temperature, *(v)* change in dimension due to heating or cooling, and *(vi)* cracks either of circular or vertical nature due to uneven heating.

4.4 Tempering

Tempering and ageing are the kinds of heat treatment which are applied to hardened alloys; they involve certain phase transformations which make the metal structure approach the equilibrium.

A combination of hardening and tempering or hardening and ageing is almost always aimed at obtaining higher properties (hardness, strength properties, coercive force, electric resistance, etc.) as compared to those resultant from annealing.

In most alloys, hardening forms a supersaturated solid solution (or a mixture of solid solutions); in that case the main process occurring in the subsequent tempering or ageing is the decomposition of the supersaturated solid solution.

The temperature and holding time are chosen so as to avoid the formation of the equilibrium state in the alloy (as in the case of annealing). The rate of cooling from the temperature of tempering or ageing has no effect on the structure and properties of alloys (for only a few exceptions).

The term 'tempering' is usually applied to steels and other alloys undergoing a polymorphic transformation in hardening (two-phase aluminium, bronzes, some titanium alloys) and the term 'ageing' to alloys which undergo no polymorphic transformations in hardening (aluminium alloys, austenitic steels, nickel alloys, etc.).

Tempering releases the stresses and reduces the brittleness. Tempering causes the transformation of the martensite into less brittle structure, i.e., a fine pearlistic structure termed as *troostite*. Troostitie is much tougher, although somewhat softer than martensite. Most c.s. cutting tools have this type of structure. Once the tempering temperature has been reached, it is normal to quench the steel. All structures resulting from tempering are termed tempered martensite. The changes taking place during various temperature ranges are as follows:

(a) 100°C-220°C Below 200°C tempering temperature only relieves the hardening stresses and very little change occurs in the micro-structure. However, the stress relieving treatment is given when maximum hardness is desirable and brittleness is a problem. The strain is relieved due to removal of carbon atoms from their trapped positions.

(b) 240°C-400°C Above 220°C the martensite starts to change into a fine pearlistic structure termed as *troostite*. In the range of temperature 240°-400°C, martensite decomposes rapidly into emulsified form of pearlite called as secondary troostite. This type of material is very fine in nature and hence provides good shock resistance. The fine edge tools are usually tempered within the range 270°C-300°C.

(c) 400°C-550°C Tempering above 400°C causes the cementite particles to ball up giving a coarse structure called *sorbite*, which is more ductile and tougher than troostite. Within this range, the precipitate troostite begins to coalesce forming a coarser form of globular pearlite called as sorbite. We may note that both troostite and sorbite are now preferably called tempered martensite. This treated is usually preferred in such components as beams, springs and axles.

(d) 600°C-700°C Within this range, heating hardened steel causes *spheroidisation*, the structure being known as *spherodite*. This structure is formed due to further coalescence of the carbide within the alloy. Spheroidised steels exhibit fairly good machinability as the hard carbide particles are embedded in the soft ferrite matrix and consequently do not have to be cut by the cutting tool. When the shperoidized steel is heated to just above its lower critical temperature the pearlite present will alter to austenite and cooling to room temperature will yield a structure of lamellar pearlite plus pro-eutectoid ferrite or cementite depending upon carbon content.

Usually the temperature of tempering is judged by colour appearance on shop floors. However, for accuracy, the exact temperature measurement are to be made. Table 10.2 gives the tempering temperatures and colour appearance for various tools.

Effect of Tempering on Mechanical Properties

Tempering improves the ductility and toughness of quenched steel while decreases hardness. Figure 10.9 shows how these properties are affected by tempering. The tempering temperature is so chosen that it results

Table 10.2 Tempering temperatures and colours for various tools

Tool or component	*Temperature (°C)*	*Colour*
Planning tools	230	Pale straw
Drills, Milling cutters	240	Dark straw
Taps and screw cutting dies	250	Brown
Punches, drill bits	260	Purplish-brown
Press tools	270	Purple
Cold chiesels	280	Dark purple
Wood saws, springs	300	Blue
Turning tools	225	
Light lathe tools	230	
Razors, steel blades	235	
Boring cutters, reamers	245	
Twist drills	270	

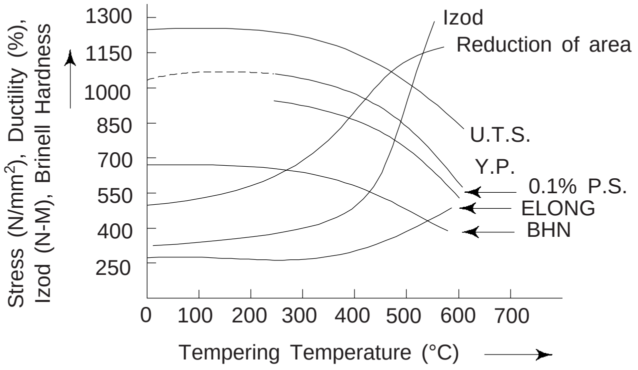


Fig. 10.9 Property chart or tempering diagram for water quenched 28 mm diameter bars of En-12 steel. (Scale for stress only)

in the desired combination of the properties. There are some steels which exhibit drop in impact values in the certain tempering temperature range. This drop is an indication of brittleness and one should avoid such range. For mild steel this brittleness (termed as blue brittleness) occurs at about 300°C tempering temperature.

The applicability of a particular kind of heat treatment can be determined in principle by means of the *phase diagram*. In that connection it is possible to distinguish the following main groups of alloys:

(a) alloys which undergo no phase transformations in the solid state,

(b) alloys with a variable solubility of the constituents in the solid state and

(c) alloys undergoing a eutectoid transformations

4.5 Special Treatments

In case of large sections where the water quenching may produce cracks, special treatments are preferred for hardening. The cracks on the surface are caused due to fast skin cooling and changes into martensite while the inner core cools slowly and transforms later accompanied by dilation. This dilation causes outer skin to crack. In order to avoid this type of cracking special treatments have been developed. Prior to their discussion, it would be proper to have some idea about isothermal transformation.

(i) Isothermal Transformation: Usually austenite is not converted into martensite instantaneously but the process continues for some time. It is observed that different steels take different time for full transformation and the time depends upon the temperature from where cooling begun. Conditions of

constant temperature are termed *isothermal*. The related phase diagrams are referred to as *isothermal transformation diagrams*, or sometimes as **time-temperature transformation** (or T-T-T) diagrams.

The selected specimen is austenized and then quenched in liquid bath held at temperature to be investigated. The specimen is held for a different length of time in the bath and then quenched in the water. One can study the resulting structure under microscope or any other associated property, e.g. hardness may be studied. It is observed that definite times are required for the initiation and completion of the transformation and these times vary with the temperature. The progress of transformation, say for 10%, 50%, or 90% may also be determined.

Figure 10.10 illustrates complete transformation diagram for eutectoid steel. The transformation temperature is lowered from A_1 temperature to about 550°C, and the nucleation and completion time decreases and pearlite lamellae becomes finer. The nucleation and completion time increases from about 550°C to 250°C. The transformation product is *bainite* (composed of two equilibrium phases that are ferrite and cementite). Minimum nucleation time is identified as the *nose* or *knee*. The transformation product is martensite below about 250°C. Although, this forms almost instaneously, but the amount formed depends upon the temperature. The upper and lower limits of martensite transformation temperature are represented by M_s and M_F respectively. These diagrams are also referred to as *S*-curves because of their shape.

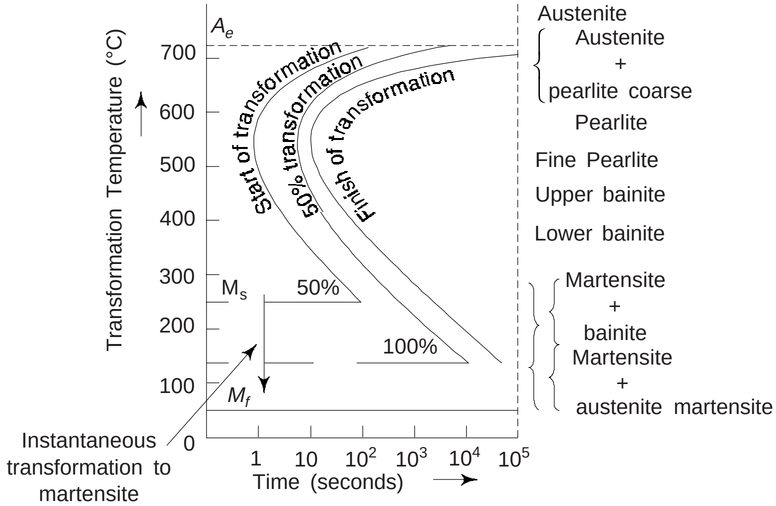


Fig. 10.10 **Isothermal transformation diagram for a eutectoid steel. Structures present after 10^5 seconds are shown on the right hand side**

(ii) Austempering or Isothermal Quenching: The component or specimen to be hardened is first austenized and then quenched into a lead or salt bath held at just above the martensite transformation temperature. The component is kept in the bath until the bainite transformation is completed. Now, the component is removed from the bath and cooled in air till the room temperature is reached. The bainite so produced is somewhat softer than martensite of the same carbon content and distortion is also minimum. Moreover, the austempered steel has improved shock resistance and low notch sensitivity. Figure 10.11 shows the process of austempering.

Molten salts, 45% $NaNO_3$ and 55% KNO_3 or molten alkalies 20% NaOH and 80% KOH are used as quenching medium. Their temperature is maintained between 150-450°C.

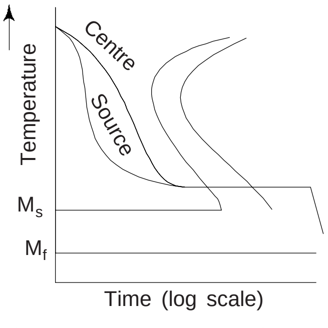


Fig. 10.11 **Austempering on the TTT curve**

Interestingly, austempering is a heat treatment and no reheating is required as in tempering.

Austempering is often limited to section thickness of 20 mm. This is the only limitation in austempering, i.e. only small sections are suitable for austempering as big sections cannot be cooled rapidly to avoid the formation of pearlite. Austempering is applicable to a few plain carbon steels and requires facility of molten salt bath. One may regard this as a disadvantage over quenching and tempering.

(iii) Martempering or Steeped Quenching: The sample or the specimen to be hardened is fully austenized and then quenched into a lead or salt bath held at a temperature just above that at which martensite starts to form. It is maintained at this temperature until its temperature becomes uniform throughout, i.e. there remains no difference in outside and inside temperatures. Then it is water quenched to form complete martensitic structure and bainite formation is prevented. This process separates successfully the cooling contraction from the austenite-martensite expansions and thus prevents quenching in large size articles. The steel can be tempered to low temperature to further refine the structure.

Effect of such treatment is to minimise cracking, distortion, and the thermal shock of the quenching. The hardness and ductility are usually similar to those obtained by direct quenching to the martensitic state followed by tempering. However, the impact toughness may be better.

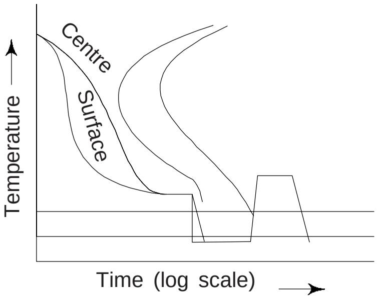


Fig. 10.12 Martempering on the TTT curve

4.8 Surface Hardening

Surface hardening of articles, like most methods of surface strengthening (chemical heat treatment, strain hardening, knurling, etc.) offers an additional advantage that large compressive stresses appear in the surface layers of hardened articles.

A number of components require only a hard surface to resist wear and tear and a tough core to resist shock loads instead of complete component being made hard. These two properties do not exist in one steel. For toughness, one finds that the core should not exceed 0.3% carbon content, also sometimes it is not desirable to harden complete component. For example, it is undesirable to have case harden screw threads. The threads would be brittle and distortion during hardening would need expensive thread grinding operations to correct the distortions. Surface hardening is classified into two types:

(i) Selective heating of the surface layers: In this case, the time available to change the micro-structure of steel is short. Obviously, hardened and tempered steels respond well but annealed steels do not. Moreover, composition of the steel must be such that the quenching will produce martensite and so harden the steel. Clearly, carbon content to be 0.4% or more.

Flame hardening can be one of the method of selective heating. The surface of a component is heated to 850°C with an oxyacetylene flame and quenched with cold water immediately. Flame heating transforms the structure of the surface layers to austenite and quenching changes the austenite surface layers to martensite resulting into a hard surface. Depth of hardening depends on the heat supplied per unit time. Flame hardening is mostly used in case of carbon and alloy steels having carbon content 0.4% to 0.6, e.g. shafts, gears, cylinder linears, crank shafts bearing journals, etc.

Another method of selective heating is *induction hardening*. Steel components are placed within a coil through which a high frequency is allowed to pass. Surface layers of the component are heated between 850-1000°C. Subsequently cooling transforms the austenite to martensite. The heating coil is often made of tube perforated with fine spray holes so that it can be used both for heating and quenching. We may note that the depth of heating produced by this method is related to the frequency of the alternating current. Higher the frequency of a.c. current, less will be the hardened depth, but there will be more rapid rise in temperature.

(ii) Case Hardening: In this method C and/or N_2 are introduced in the surface layer. In *carburizing* the part is surrounded by material or atmosphere rich in carbon and on heating this carbon is first released and then absorbed in steel. More recently case carburizing is more effectively performed by heating steel part in the atmosphere of natural gas, coke even gas, butane or propane or the volatized form of liquid hydrocarbons like terpenes and benzene or the volatilized form of liquid hydrocarbons like terpenes and benzene. Volatilized form of alcohol and glycols or ketones have also been used.

In these cases the thickness of hardened layer is proportional to root of the time of treatment in hour. The depth of hardness depends upon the number of times the process is repeated. The component is finally raised to a red heat and quenched in water. However, one will have to take the precaution that the part to be hardened should be covered with the compound.

Carburising

This is another method of surface hardening. The composition of surface layers are changed. This process is usually carried out on a steel containing less than 0.2% carbon. Carburizing is usually employed for treating certain types of machine elements which have to have a wear-resistance working surface and tough core, gear wheels, shafts, pins, camshafts, cams, worms, etc.

The initial medium for carburizing (carbon saturation) is usually called a *carburizer*. Two methods of carburizing are in use: in a solid and in a gaseous carburizer. In both cases, however, the carburizing process passes through a gaseous phase. The most popular solid carburizer consists of charcoal with an addition of 20-25% barium carbonate to intensify the process and of 3-5% of $CaCO_3$ to prevent the carburizer particles from caking. Articles to be carburized and the carburizer are placed into a container (steel box) and heated in a furnace to 910-930°C. During heating, charcoal reacts with the oxygen of remaining air

$$2C + O_2 \rightarrow 2CO$$

$$BaCO_3 + C \rightarrow BaO + 2CO$$

The leading reaction of carburizing that takes place in the surface of metal is the disproportioning reaction:

$$2CO \rightleftharpoons CO_2 + C$$

The active carbon that forms by the reaction is absorbed in the surface of metal, which is in the austenitic state, while CO_2 reacts with charcoal to form new portions of CO.

Thus, the carbon formed in the reversible disproportioning reaction is transferred into the surface being saturated.

In gas carburizing, the carburizer is diluted natural gas (consisting almost fully of methane), a controlled atmosphere produced in special gas producer and a liquid hydrocarbon (kerosene, benzene, etc.) supplied dropwise into an airtight furnace where it forms an active gaseous medium. The main leading reaction for methane is

$$CH_4 \rightarrow 2H_2 + C$$

In some cases when the gas mixture contains CO (an endothermic-base controlled atmosphere), the reaction $2CO \rightarrow CO_2 + C$ is also possible. Depending on the composition of gaseous mixture and the concentration of carbon in steel, the atmosphere in the furnace may be carburizing, decarburizing or neutral. We may note that carburizing will take place when the carbon concentration in the surface of steel is smaller than the carbon potential of gaseous mixture at the given temperature. Normally, this process is carried out on a steel containing less than 0.2% carbon. To achieve adequate hardness the surface of the component should have a carbon content of 1% approximately. In this process the carbon is usually added to the surface layers upto 0.7 to 0.8% by a carefully regulated depth, which is followed by a quenching process to convert the surface layers into martensite.

The structure obtained on carburizing is coarse-grained, since the metal is held for an appreciably long time at the carburizing temperature. The time of isothermal holding in carburizing depends on the grade of steel and the desired depth of layer.

The process of gas carburizing occurs more quickly than that of solid carburizing, since no time is lost for heating the carburizer box. Besides, gas carburizing can be more easily controlled and automatized.

By carburizing process hardness can be increased from 200 HB (0.8 carbon content in the core) to 750 HB (0.7% carbon content in the surface).

Other methods used for carburising are:

Pack Carburizing: The component to be carburised is heated above A_3 temperature between 850°C to 1060°C for about five hours in a sealed metal box containing charcoal and barium carbonate as mentioned earlier. Oxygen present in the box reacts with the carbon to produce carbon monoxide. Carbon present in the atmosphere gets diffused into the surface austenitic layers of the component. When carburising is complete, metal boxes are allowed to cool down so that they can opened and unpacked. The components are then cleaned and made available for subsequent heat treatment. This method is mainly used for large components or where a thick surface layer has to be hardened.

Salt Bath Carburizing: This method involves the component in a bath of suitable carbon rich salts between 750-850°C. Usually, sodium cyanide is mainly used as a medium. Carbon from the molten salt diffuses into the component. The advantages and disadvantages of cyanide salts are:

(i) Cyanide salt is poisonous in nature. Obviously, one will have to take utmost care while working with such a salt. A few grains of this salt under the fingernails could prove fatal if transformed to mouth via smoking or food.

(ii) After the treatment, removal of the salt from the hardened surface is more problematic in case of blind holes or threaded parts.

(iii) In this process heating and carburisation are more uniform with less chance of distortion.

(iv) The components can be hardened by quenching straight out of cyanide salt without need for further heat treatment.

Nitriding: This is the process of diffusion saturation of steel surface with nitrogen. The component is heated in a mixture of ammonia and hydrogen so that nitrogen diffuses into the surface layers and hard nitride compounds are formed. The process is employed for increasing the wear resistance and endurance limit of machine elements (crankshafts, cylinder sleeves, worm gears, shafts, etc.).

Prior to nitriding, steel articles are subjected to hardening and high temperature tempering (improving heat treatment) and finishing. After nitriding, they are ground and polished.

In a nitriding process, the component is heated in a mixture of ammonia and hard nitride compounds are formed. Common nitriding is carried out at a temperature of 500-600°C in a muffle or container through which dissociating ammonia is being passed. This process is used with those alloys of steel which contain elements that form stable nitrides, e.g. chromium, molybdenum, tungsten, vanadium, aluminium, etc. It is probable that the reaction of ammonia dissociation takes place at the steel surface; nitrogen ions are absorbed by the surface and then diffuse into the depth of the metal. The time taken for nitrogen to react with the steel surface is about 100 hours.

If ammonia is heated in an isolated volume, only the reaction $2NH_3 \rightarrow N_2 + 3H_2$ is possible and the molecular nitrogen formed by this reaction cannot diffuse into the metal without being ionized.

The depth to which the nitrides are formed depend on the temperature and time allowed for the reaction, and in normal conditions, it is unlikely to exceed 0.7 mm. The components are cooled slowly in the ammonia hydrogen atmosphere. The process of nitriding is usually completed with the formation in the metal surface of the ε-phase with an HCP lattice and ordered arrangements of atoms in a wide range of concentrations.

When nitriding carbon steels, the rate of nitrogen diffusion decreases with an increase of carbon content and the formation of carbonitride phases becomes possible.

A thin white layer of nitrides is formed on the surface of the component, which should be removed by mechanical or chemical means as this layer adversely affects the mechanical properties.

Nitriding is a rather time-consuming process. Nitrided layers of alloyed steels have higher values of hardness and wear resistance compared with carburized layers. Since no quenching treatment is required with nitriding, cracking and distortions are less likely to occur. With this process very high surface hardness can be obtained with special alloys. The hardness is retained upto about 500°C compared to carburising in which case the surface becomes softer at above 200°C. The nitriding process is however used less frequently than carburizing, mainly in view of its longer duration and smaller depth of the strengthened layer, which limits the allowable contact loads onto the metal surface.

In recent time, sources of highly concentrated energy (electron and laser beams) have come into use for heat treatment of some articles. Pulses of electron and laser beams can be used efficiently for local heating of the surface of articles which makes it possible to carry out surface hardening of working of cutting tools and certain areas of the articles to be subjected to wear. In some cases, a thin surface layer is heated to melting and then cooled quickly so that a fine-grained or amorphous structure forms in it.

In surface hardening by means of sources of highly concentrated energy, water quenching becomes needless, since a thin locally heated surface layer is cooled quite rapidly by the underlying metal. Electron accelerators and continuous wave gas lasers and pulsed lasers can be employed for the purpose.

Age Hardening (Precipitation Hardening): One can enhance the strength and hardness of some metal alloys by the formation of extremely small uniformly dispersed particles of a second phase within the original phase matrix. However, this must be accomplished by phase transformations that are induced by appropriate heat treatments. The process is termed as *precipitation hardening*, as the small particles of the new phase are termed “precipitates”. This process is also called as ‘age hardening’, as the strength develops with time, or as the alloy ages. The alloys that are hardened by this process are aluminium-copper, copper, berryllium, copper-tin, and magnesium-aluminium. Few ferrous alloys are also precipitation hardenable.

Age hardening is a characteristic of alloys with a falling solid solubility, e.g. the solubility of copper in aluminium to form the solid solution is only 9.5% at room temperature. It increases with rise in temperature and reaching a maximum of 5.7% at 548°C. Obviously, the solubility of copper in aluminium depends on the temperature. When an alloy containing 4% copper is cooled slowly from the alpha region, coarse particles of $CuAl_2$ get precipitated. In this condition, the alloy is comparatively weak and brittle. When carried out the age hardening treatment as follows, this improves mechanical properties:

(i) The alloy is heated to a prescribed temperature of 500°C to dissolve and other alloying element.

(ii) The alloy is quenched from the above temperature to super-saturate the alpha solid solution at room temperature keeping precipitation rate very low. A super-saturated solid solution of 4% copper in aluminium is obtained. Precipitation of copper as $CuCl_2$ is prevented by rapid cooling.

(iii) Finally the alloy is kept at room temperature and hardening (age) occurs spontaneously.

We may note that when the solute atoms form a solid solution with solvent atoms it produces an alloy which is harder and also has more strength than the pure solvent metal. This is termed as *solute hardening*.

Interestingly, precipitation hardening and the treating of steel to form tempered martensite are totally different phenomena, even though the heat treatment procedures are similar. The principal difference between the two lies in the mechanisms by which hardening and strengthening are achieved.

A comparison between various surface hardening processes is being made in Table 10.3

Table 10.3 A comparative study of main surface hardening processes

Process	*Temperature (°C)*	*Case depth (mm)*	*Hardness (HRC)*	*Applications*
Pack carburizing	850-1100	0.25-3.0	450-750	Low carbon and carburising alloy steels. Large components
Gas carburising	850-1000	0.07-3.0	450-750	Low carbon and carburising alloy steels. Large number of components

(*Contd.*)

Table 10.3 (*Contd.*)

Process	*Temperature (°C)*	*Case depth (mm)*	*Hardness (HRC)*	*Applications*
Cyaniding	750-850	0.02-0.7	550-700	Low carbon and light alloy steels
Nitriding	500-530	0.07-0.7	550-850	Alloy steels
Flame hardening	850-1000	upto 0.8	600-750	0.4 to 0.7% carbon steels
Induction hardening	850-1000	0.5-5.0	600-750	0.5 to 0.7% carbon steels

5. MASS EFFECT

Variation in hardness across a section of the components having higher thickness, through heat treatment is termed as '*mass effect*'.

Hardness of plain carbon steel depends upon its carbon content and the rate of cooling from the hardening temperature. A part having less thickness will cool more quickly in comparison to a part having higher thickness, when both are cooled in the same quenching bath. In a thicker component, outer layer will cool faster than the core and the heat will get trapped at the centre. This causes to a variation in hardness across a section of the component. The result may be another layer of martensite and an inner core of pearlite. This variation in hardness is generally referred to as mass effect and shown in Fig. 10.13.

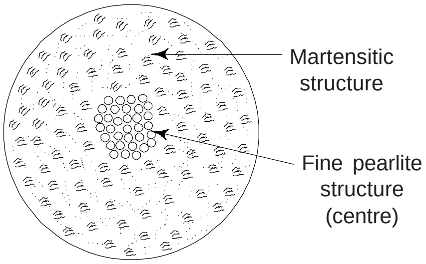


Fig. 10.13 Mass effect

Due to mass effect, plain carbon steels having large sections are said to have a poor hardenability, as they cannot be fully hardened throughout. A 3 percent nickel steel containing 3% carbon will harden uniformly through its section, as it has a relatively low critical cooling rate. However, due to its higher carbon content the plain carbon steel will exhibit a very much higher surface hardness.

Obviously, the rate of cooling that takes place for a component when quenched in say water, depends on the size of the component. Slower will be its rate of cooling for bigger the size. Therefore, the quenching medium has to be chosen keeping in view the particular size of the component, e.g., for a 12 mm diameter bar, air cooling may result in a micro-structure involving ferrite, pearlite. With oil quenching the outcome is bainite plus martensite. However, with water quenching the microstructure in only martensite. If martensite structure is desired, then for this particular diameter, bar water quenching must be used.

6. PRINCIPAL EQUIPMENT FOR HEAT TREATMENT

The principal equipment for heat treatment of metals includes *heat furnaces* and installations and *cooling arrangements*. There are two major types of furnaces: *batch furnace* and *continuous furnace*. Also depending on the source of heat, heating furnaces are divided into *electric* and *fuel fired* (the fuel being gas or, less frequently, fuel oil).

The selection of furnace requires consideration of several factors particularly because furnaces consume a good amount of energy. To make efficient use of energy in heat furnaces, their design and use of proper insulation requires full consideration. In selecting a heat furnace, the man power requirement for operation, the initial cost and convenience of maintenance and repair are some other important considerations.

To avoid oxidation and decarburization of steel articles during heating, the working space of modern heating furnaces is vacuumized or filled with an inert gas forming a protective gaseous atmosphere. For increasing the productivity of heat treatment of small articles (elements of machines and instruments), high speed heating is employed, i.e., articles are placed in a preheated furnace. Temporary thermal stresses appearing during heating in small articles cannot cause their cracking or buckling. High speed heating may however be dangerous for large articles (rolling rolls, shafts, large housings, etc.) because of which these kinds of articles are heated slowly (together with the furnace) or by a stepwise schedule. In some cases, high-speed heating is carried out in molten-salt baths (heating of drills, screw taps and other small tools). Since temperature control and temperature cycle control are very important in heat treatment and therefore *electronic* and *computer controlled* furnaces are taking precedence over older type furnaces.

Oil or gas furnaces have distinct advantage over electrical furnaces due to the former often introduce products of combustion in the heating space and thus affecting the part to be treated. On the other hand, the electrical heating has slower start up and also not easy to control.

6.1 Batch Furnaces

In order to place the job in place, the basic requirements of these furnaces are an insulated chamber, the heating system and a door or several doors. The parts to be heat treated are loaded and unloaded in individual batches. A furnace which is easy to use, simple to construct in various sizes and also having versatility to accommodate various sizes in a *box furnace* which could be horizontal rectangular. Several times a flat platform on wheels is used to carry the parts in the furnace.

Pitch Furnace: This is made in form of a vertical pit below the ground level. The parts to be treated are lowered in the pit. One can suspend long parts like rods, bars, tubing, shafts, etc. in the space of the pit furnace. If placed in horizontal position in the box furnace, these parts are susceptible to distortion.

Bell Furnace: This furnace does not have bottom and is lowered on the stack of parts to be treated. The furnace chamber could be round or rectangular.

Elevator Furnace: In this furnace, the parts to be treated are placed on the rolling platform which rolled in the proper position and lifted to the heating chamber of the furnace. By placing a quenching tank directly below the surface the savings on quenching time and space are made.

6.2 Continuous Furnaces

These are the furnaces in which parts to be treated are placed on some sort of conveyer which move into the furnaces according to programmed heating cooling cycle. This helps in saving the time for loading-unloading and also handling of job is reduced. These furnaces are most suitable for high production rate and better control of heat treating cycle.

6.3 Salt Bath Furnaces

To ensure good control of uniform temperature and high heating rates as compared to air or gas, salt bath furnaces are most suitable. Metals or molten salts have higher conductivity than air or gas. The salt may be heated from outside of it is non-conducting or by passing a low voltage alternating current between electrodes placed in the salt. In order to avoid electrolysis of salt; direct current is not used. Among metals, lead is commonly used. One can obtain wide range of temperature from such furnaces.

6.4 Induction Heating

The most perfect method of surface hardening is induction hardening in a special installation where articles are heated by high frequency currents. This method is highly productive, can be completely automatized, ensures high rise production, and minimizes their buckling and oxidation.

As is known, with an increase of the frequency of a.c. electric current, the *skin effect* is enhanced, i.e. the current density in the external layers of a conductor turns out to be many times greater than in the core. As a result, the heat is liberated almost entirely in the surface layer and the latter is heated much more substantially than the core.

High frequency a.c. current heating of articles is effected by an inductor coil. If an article has a relatively small length (or height), its entire surface can be heated simultaneously to the hardening temperature. With longer articles, heating is carried out successively by moving the article relative to the inductor coil at a proper speed. Figure 10.14 shows induction coils, which normally are made in copper or copper based alloy which are water cooled. The induction coil is also designed for quenching the part after heating.

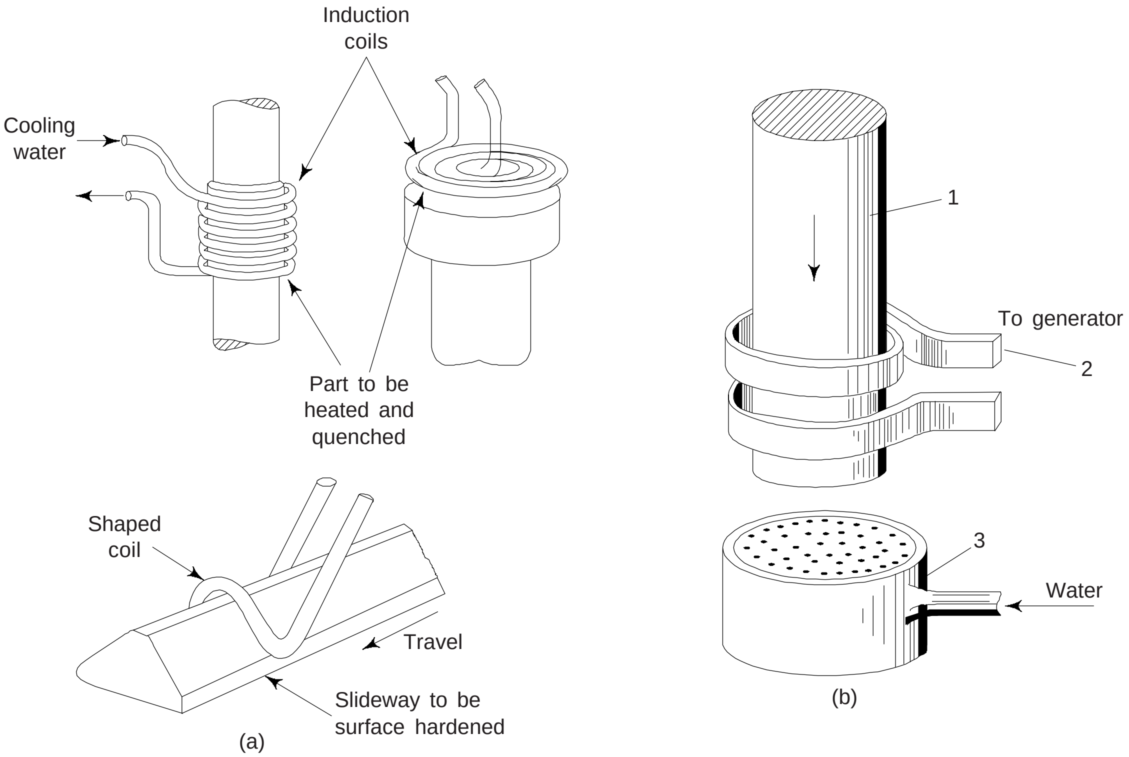


Fig. 10.14 **(a) Coils for induction heating (b) Arrangement of a workpiece, inductor coil and sprayer for hardening by induction heating: 1 - article; 2 - inductor coil; 3 - sprayer**

Cooling of articles in induction hardening is usually effected by water which is supplied through a sprayer, a perforated tabular ring embracing the article.

6.5 Furnace Atmospheres

When the heating is not through salt bath, then the job to be treated is subject to varying atmosphere, e.g. atmospheric air or any one of several gases present in the atmosphere. We have read that surface oxidation, tarnishing and decarburization are the main problems which the metals face. Oxygen may cause oxidation, rusting and scaling. Depending upon the concentration, CO_2 may cause decarburization in furnace atmosphere. In the presence of water vapour, thin blue film is formed. For improving surface appearance, *bluing* of surface is done. Nitrogen provides a neutral atmosphere. Often vacuum furnaces are used for small and accurately finished parts provide complete safety from effects of atmosphere.

7. MAJOR DEFECTS IN METALS OR ALLOYS DUE TO FAULTY HEAT TREATMENT

The faulty heat treatment may cause following major defects in a metal or a alloy:

(i) Overheating: Prolonged heating of a metal or a alloy above a temperature marked by A_3 line (Fig. 10.1) leads to the formation of very large actual grains. It is usually termed as overheating. On cooling a metal, this yields a structure containing crystalline martensite (usually called as Widman tatten structure). Such a structure has reduced ductility and toughness. By means of usual annealing, it is possible to retrieve an overheated metal or alloy. One may use double annealing for considerable heating.

(ii) Burning: Heating a metal or an alloy near melting point for a longer duration may lead to burning. This leads to the formation of iron oxide inclusions along the grain boundaries. Burnt metal or alloy has a *stoney fracture* and such a metal or alloy is irremediable and such metals are rejected.

(iii) Oxidation: Sometimes a metal or alloy gets oxidized due to oxidizing atmosphere inside the furnace. It is usually characterized by a thick layer of scale on the surface of a metal or alloy. One can prevent it by making use of controlled atmosphere in the furnace or using molten salt baths.

(iv) Decarburization: Decarburization is the loss of carbon in the surface layers of metals or alloys. It is caused by the oxidizing furnace atmosphere. In order to prevent it, the metal or alloy should be heated in a natural or reducing atmosphere or inside boxes with cast iron chips or in molten salts baths.

(v) Cracks: These occurs in quenching when the internal tensile stress exceed the resistance of metal or alloy to separation. The tendency of a metal or alloy to crack formation increases with carbon content, hardening temperature and cooling rate in the temperature interval of martensite transformation. Cracks also increases with the hardenability of metal or alloy. Concentration of local stresses is the another reason for the formation of cracks.

We may note that the cracks are irremediable defects. In order to prevent the formation of cracks, it is advisable to *(i)* avoid stress concentrations, e.g. sharp corners on projections, acute angles, abrupt changes from thicker to thinner cross-sections, etc. in component design, *(ii)* conduct quenching from the lowest possible temperature, *(iii)* cool the metal or alloy slowly in the martensite interval of temperatures by quenching in two media and stepped quenching (i.e., martempering), and *(iv)* apply isothermal quenching.

(vi) Distortion and Wraping: This is consisting in changes in the size and shape of heat-treated work and is mainly due to thermal and structural stresses. Asymmetrical distortion of work is often called wraping in heat-treating practice. Usually, it is observed in case of overheating or non-uniform heating for hardening, when the work is quenched in the wrong position and when the cooling rate is too high in the temperature interval of martensite transformation. By eliminating these causes, one may substantially reduce warping.

8. SURFACE FINISH AFTER HEAT TREATMENT

To remove the defects mentioned in section 7, certain finishing operations are carried out prior to use of assembly of the same. Few of these are:

(i) Acid Pickling: If the heat treatment is not performed in a controlled gas atmosphere, oxide or scale layer is formed on the surface of heat treated component. Descaling can be done by acid pickling either by *mechanical circulation* or by *direct dissolution of a component in an acid bath*. The reaction is

$$FeO + H_2SO_4 \rightarrow FeSO_4 + H_2O$$

For the above purpose, 8-12% H_2SO_4 or 20% HCl acid is taken. The acid bath is usually used at 50-70°C. The components removed from the pickling bath are thoroughly cleaned with steam and then neutralized in a week alkaline bath for about 5-10 minutes. Finally, they are rinsed by warm water.

(ii) Sand Blasting: A jet of compressed air carrying sand, composed of dry, sharp, quartz grains of 1-12 mm is directed against the surface to be cleaned after heat treatment. This process provides fine surface and quite useful for large areas to be cleaned. However, in this process, a good amount of dust is generated which requires precautions for operators. To avoid inhalation of dust, protection measures are essential. To avoid inhalation of dust, operators are provided with respirators and sand blast hoods.

There is another method of surface cleaning called as *shot blast*. Surface cleaning is performed by centrifugal action of abrasive material like iron shorts instead of sand blowing with pneumatic pressure. We may note that it also produces cold case hardening of the surface, increasing the fatigue resistance of the articles.

(iii) Degreasing and Alkaline Detergent Cleaning: When treated in salt baths, a number of components have salt deposits. These can be removed by soaking the component in a boiling water. Tools tempered and quenched in oil bath have a greasy material deposited on the surface. These are usually cleaned by 10% (by volume) soda ash solution at 80-90°C or 3% solution of caustic soda. Both salts and oil deposits can be cleaned by immersing the components in a tank containing boiling soda solution and then washing in water. The operation is known as *vat* or *sak cleaning*.

(iv) Straightening: During heat treatment, certain tools and machine parts get distorted or warped. To bring them back to their original shape, straightening is carried out. Straightening of rolled stock is usually done by Roll Straighteners or Hydraulic straightening presses.

9. MEASUREMENT OF HIGH TEMPERATURES AND PYROMETERS

Radiation Pyrometers

The physics of measurement of very high temperatures above 1600°C is called *pyrometry*. The devices measuring such high temperatures are called *pyrometers.* Some of these pyrometers use thermal radiations from hot bodies which depend for black bodies, only on temperatures ($E \propto T^4$ by Stefan's law) of the body and hence they are known as radiation pyrometers. Two different types of radiation pyrometers are used for the measurement of high temperatures. They are:

(i) Total Radiation Pyrometers: In these types of instruments the total energy of the radiation emitted by the body under test is measured and the temperature is obtained by applying Stefan's law ($E = \sigma T^4$) assuming that the body is black.

(ii) Optical or Spectral Pyrometers: These are based on the Wien's distribution law. The intensity of radiation of a certain wavelength emitted by the body is compared with that of the radiation of same wavelength emitted by a standard body at a known temperature. The temperature of the body is deduced from Wien's distribution law, or more accurately from Planck's law.

(a) Fery's Total Radiation Pyrometer Fery in 1902 devised the first total radiation pyrometer shown in Fig. 10.15. It is virtually a reflecting telescope, with a concave mirror M, of nickel-plated copper or gold-coated glass, and with an eye piece E, fitted coaxially into a hole in its centre. Mirror can be moved by a rack and pinion arrangement. A metal disc, with its front face blackened, is placed at the focus of the mirror and receives the radiation from the body whose temperature is to be measured, after reflection at the mirror. Two plane semi-circular mirrors m, m inclined to each other at about 5° with an opening of about 1.5 mm at the centre are placed in front hole in the mirror. Any direct radiation is prevented from falling upon the disc by placing a suitable enclosure around it. At the back of the disc is attached one junction of a thermo couple (Th), whose colder junction is shielded from direct radiation by means of a toungue T. A millivolmeter (mV) is connected from the leads of the thermo couple.

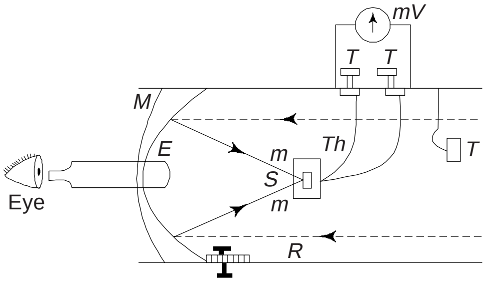


Fig. 10.15 Fery's total radiation pyrometer

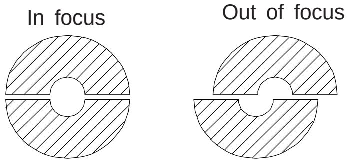


Fig. 10.16 (a) (b)

Working of the Pyrometer: The radiations from the hot body under test falls on the concave mirror M which focusses it on the opening between m, m. The opening is seen with the help of an eyepiece E through the hole in M. The position of mirror E through M is so adjusted that the appearance through the eyepiece is as shown in Fig. 10.16(a). Clearly, the radiation which is properly focussed on the opening between m, m falls on S which is therefore heated and produces a thermo *e.m.f.* Reading in the millivoltmeter is recorded.

Calculation of Temperature: If V is the reading of millivoltmeter, then

$$V = a\left(T^4 - T_0^4\right)$$

Here T is the absolute temperature of the source, T_0 is the absolute temperature of the strip and a is the constant which is not the same as Stefan's constant. One can determine the value of a by obtaining the value of V corresponding to a known temperature T. If T is very much higher than T_0, then $V = a\,T^4$

The instrument is calibrated by receiving radiations from a black body, maintained at known different temperatures. The temperatures are measured by a standard thermocouple and the corresponding readings of the millivoltmeter are recorded. A calibration curve is plotted between Log V vs. Log T which is a straight line,

$$\text{Log } V = K \text{ Log } T$$

The millivoltmeter may also be directly calibrated in degrees.

Modification for Measurement of Very High Temperatures

To measure temperature above 1400°C, a rapidly rotating sector as shown in Fig. 10.17 is interposed between the hot body and the pyrometer. If θ is the angle of the sector, then it will allow only a fraction $\theta/360$ of the incident radiation to enter the pyrometer. The pyrometer indicates a temperature which is less than the true temperature of the body. Since the radiation is proportional to the fourth power of the absolute temperature therefore if T be the true temperature of the body and T_1 that indicated by pyrometer, then one can write

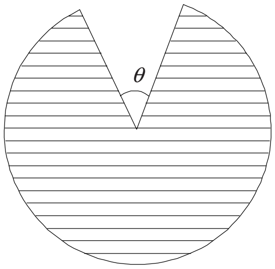


Fig. 10.17

$$T_1^4 = \frac{\theta}{360}\, T^4 \tag{1}$$

or

$$T = T_1\left(\frac{360}{\theta}\right)^{1/4} \tag{2}$$

Thus T, the black body temperature of the body can be easily obtained.

Advantage and Drawbacks of the Instrument

Advantage: The most important advantage of the total radiation pyrometer is that it can be easily adapted for a continuous recording of temperature, *e.g.*, the roof of an open-hearth.

Drawbacks: (i) It is not suitable for the measurement of temperature of smaller bodies, whose heat image is likely to be too small to cover the whole the disc. *(ii)* Calibrated in the manner as described above, its utility is confined to the limited range, 800°C to 1600°C. *(iii)* The readings of the apparatus, while theoretically quite independent of the distance from the hot body, are not really so in actual practice, *(iv)* Due to conduction along the wires of thermocouple, the Stefan's fourth power law is not strictly applicable. The power of temperature T varies from 3.8 to 4.2.

(b) Optical Pyrometers Two types of optical pyrometers are commonly used *(i)* the disappearing filament pyrometer and *(ii)* polarising pyrometer.

(i) Disappearing Filament Pyrometer: A modified form of the disappearing filament due to Holborn and Forsy is shown in Fig. 10.18. It consists of a telescopic objective L and an eyepiece E. The cross wires of telescope are replaced by the tungsten filament of a highly evacuated electric lamp, flanked by two limiting diaphragms D_1 and D_2. The filament is heated by the current from a two volt battery and its temperature controlled by adjusting the current with the help of rheostat R.

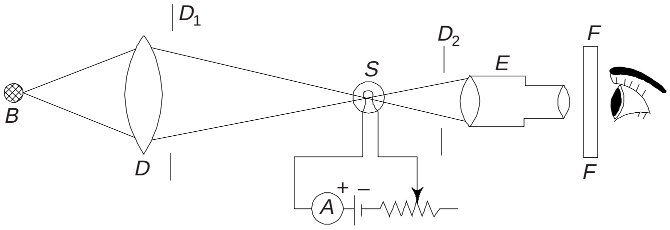


Fig. 10.18 Disappearing filament pyrometer

The telescope is directed towards the hot body whose temperature is to be determined. By adjusting the position of the objective of telescope L, an image of black body is focussed on the filament F. The temperature of the filament is adjusted by means of the rheostat R, until the filament F disappears against the image of the black body. When the brightness of filament F and that of the black body have been exactly matched, the reading of the ammeter A is noted.

Calculation of the Temperature: When the brightness of the black body and that of the filament F are same, then their emissive powers will be equal. Evidently, the temperature of the black body will be equal to that of the filament. The temperature of the filament, *i.e.,* of black body, is then calculated from the relation

$$i = a + bT = CT^2 \tag{3}$$

Here i is the current flowing through the filament, and a, b and c are constants. These constants can be determined by calibrating the instrument against known temperature.

This type of pyrometer can normally measure temperatures from 700°C to 1500°C. The range can be extended to 2700°C by using a rotating sector.

Advantages: (i) It is portable *(ii)* It can be used to determine the temperature of even small bodies, since the tip of the filament can be matched with any part of the image seen through the telescope *(iii)* It can be easily constructed and this makes it highly precise and accurate *(iv)* The theory underlying it is relatively less affected by a departure of the body from a perfectly black body; so that the corrections involved are comparatively much smaller.

Drawback: It requires some wavelengths in the visible part of the spectrum for its working and cannot, therefore, be used for measuring temperatures of bodies which are not actually glowing. Further, there is some uncertainty as to the actual wavelength of the radiation used.

(ii) Polarising Optical Pyrometer: Wanner devised in 1901, an optical pyrometer in which a ray of a particular colour from the hot body is compared with a ray of the same colour from a standard electric lamp, by means of a polarising device. The instrument is more of a polarising pyrometer. Its essential parts are shown in Fig. 10.19.

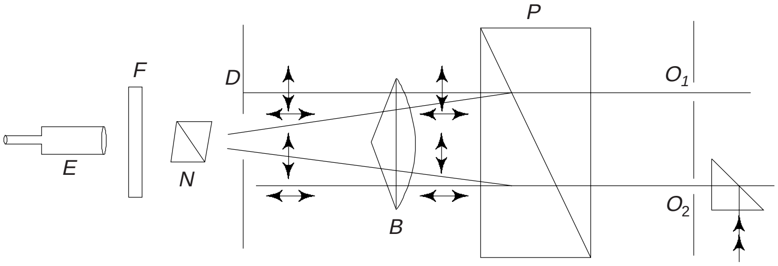


Fig. 10.19 Polarising optical pyrometer

Radiations starting from the hot body after passing through a hole O_1 fall on a calcite prism P. Calcite being doubly refracting crystal, breaks the beam into two components polarised at right angles to each other. Now, these beams pass through a biprism B which bends them in such a way that the horizontal component passes through a limiting diaphragm D, where the vertical component (↕) is cut off. Similarly, a radiation beam from a standard body at a steady temperature is allowed to pass through the system, but in this case the vertical component passes and horizontal component (↔) is cut off. These two mutually perpendicular beams emerging from D pass through a Nicol prism (N) and then from a red filter (F). Two circular patches of light are observed when we look through an eyepiece E.

The Nicol prism N is rotated. In one particular position of N, the light from hot body transmits fully through N, whereas the light from the standard body is completely cut off. When the Nicol is rotated by 90° from this position then vice versa is happened. From this position the Nicol prism N is gradually rotated till the two patches of light appear equally bright, *i.e.*, light from both the bodies is transmitted equally through N. The angle through which the Nicol is turned for the purpose is read on a circular scale attached to it. Let it be θ.

Let E_λ be intensity of radiation after filtration through F from the hot body and S_λ be the intensity of radiation of same colour from the standard body, then

$$\frac{E_\lambda}{S_\lambda} = \tan^2 \theta \tag{4}$$

Let $E_{\lambda'}$ be the intensity of radiation from the second hot body (in place of first body) and θ', be the angle of rotation of N, then

$$\frac{E_{\lambda'}}{S_\lambda} = \tan^2 \theta' \tag{5}$$

From (4) and (5), one obtains

$$\frac{E_\lambda}{E_{\lambda'}} = \frac{\tan^2 \theta}{\tan^2 \theta'} \tag{6}$$

From Wien's distribution law, we have

$$E_\lambda = A\ \lambda^{-5}\ e^{-B/\lambda T} \quad \text{and} \quad E_{\lambda'} = A\ \lambda^{-5}\ e^{-B/\lambda T'} \tag{7}$$

Here A and B are constants and T and T' are temperatures of hot body under test and second body respectively. Thus, we have

$$\frac{E_\lambda}{E_{\lambda'}} = e^{\frac{B}{\lambda}\left[\frac{1}{T'} - \frac{1}{T}\right]} \tag{8}$$

From (6) and (8), one obtains

$$\frac{\tan^2 \theta}{\tan^2 \theta'} = e^{\frac{B}{\lambda}\left[\frac{1}{T'} - \frac{1}{T}\right]}$$

Taking log of both sides, one obtains

$$2\,(\log \tan\ \theta - \log \tan\ \theta') = \frac{B}{\lambda}\left[\frac{1}{T'} - \frac{1}{T}\right] \tag{9}$$

whence, we have a linear relation between log tan θ and $1/T$, of the form

$$\log \tan\ \theta = a + b/T \tag{10}$$

Here a and b are constants and one can obtain them by calibrating the Pyrometer against known temperatures. Knowing a, b and θ, one can easily evaluate the temperature of the given hot body.

The normal range of this pyrometer is from 700°C to 1400°C but the range can be raised to 4000°C by putting absorption glasses in the path of the incident radiation.

Example 1 *Determine the grain diameter of an ASTM No. 8.* [Diploma]

Solution The grain size specified by the American Society for Testing and Materials (ASTM) is given by

$$n = 2^{N-1}$$

where n is number of grains per inch square at a magnification of 100X, and N is ASTM grain size number. Grain size number 1 to 5 indicates the coarse grain size steel. ASTM number one, is the coarsest grain which represents one grain per inch square at 100X. ASTM No. 3 represents 4 grains per inch square at 100X and so on. Grain size number 6 and above indicates fine grained steel.

Here $N = 8,\ n = 2^{8-1} = 128$

∴ Number of grains/inch^2 without magnification = 128 × 100 × 100

∴ Number of grains/mm^2 = $\frac{128 \times 100 \times 100}{(25.4)^2}$

∴ Average area of each grain = $\frac{(25.4)^2}{128 \times 100 \times 100}$ mm^2

∴ Average grain diameter = $\frac{25.4}{100 \times \sqrt{128}}$ = 0.022 mm

The relationship between grain size number and diameter of the grain is given in Table 10.4.

Table 10.4 Relationship between grain size number and actual diameter of the grain

ASTM grain size number, N	*Grain per inch square at 100 X*	*Actual grain diameter*	
		(inch)	*(mm)*
1	1	0.0100	0.25
2	2	0.0071	0.18
3	4	0.0050	0.125
4	8	0.0035	0.091
5	16	0.0025	0.062
6	32	0.0018	0.044
7	64	0.0012	0.032
8	128	0.0009	0.022
9	256	0.0006	0.016

SUGGESTED READINGS

1. C.R. Brooks, *Principles of the Heat Treatment of Plain Carbon and Low Alloy Steels*, ASM International, Materials Park, OH (1995).
2. D.P. Henkel and A.W. Pense, *Structures and Properties of Engineering Materials*, 5th Ed., McGraw Hill, New York (2001).
3. ASM Handbook, Vol. 4, *Heat Treating*, ASM International, (1991).

REVIEW QUESTIONS

1. What do you understand by heat treatment? Mention the various stages of heat treatment procedure.

2. How is a piece of steel austenitized? Why are time and temperature of austenitizing very important in heat treatment of steels? [BE]

3. Explain why: (a) Carburising heat treatment is generally on low carbon steels? (b) Alloying elements in general increase hardenability of steels and (c) Metals undergoing cyclic loading are case hardened. [Roorkee]

4. What are three steps in an age hardening (heat) treatment? Describe what happens in each step? [Roorkee]

5. *(i)* Explain the various purposes of heat treatment. *(ii)* Describe the flame hardening process and its application, *(iii)* Explain the process of annealing. [AMIE]

6. Why must water quenching be used to harden plain carbon steels, while some alloy steels can be hardened by air cooling? [BE, M.Sc. (M.S.)]

7. Distinguish between hardness and hardenability of a steel. State the factors which affect hardenability. [AMIE]

8. What factors have to be considered in determining the type of quenching to be used to harden a steel? [AMIE]

9. Distinguish between full annealing and process annealing. [AMIE]

10. What is meant by the term hardenability? Describe how hardenability of a steel can be estimated. [AMIE]

11. Explain different types of annealing treatments and their objects. [AMIE]

12. What is the minimum carbon content percentage required in steel so that it may respond to hardening by heat treatment? [AMIE]

13. Explain the theory of '*tempering*'. What are the effects of tempering on the mechanical properties of steel. [BHU]

14. *(i)* Define the term 'heat treatment'. Why are the steels heat treated? *(ii)* Discuss the major defects in steel due to faulty heat treatment. [AMIE]

15. Differentiate between hardness and hardenability. Draw hardness and hardenability curves for steel. [AMIE]

16. Discuss the use of hardenability curves. [AMIE]

17. *(a)* Describe the flame hardening process with the aid of a sketch. *(b)* Describe the main features of martensite transformation *(c)* State the objectives of heat treatment of metals. State the process of tempering. [AMIE]

18. What type of heat treatment is given for die steels? [AMIE]

19. Distinguish between
(a) Hardness and Hardenability
(b) Full annealing and Process Annealing
(c) Martempering and Tempering
(d) Annealing and Tempering
(e) Hardness and Hardening
(f) Solute and Precipitation Hardening [AMIE, BE]

20. Outline the significance of hardenability. How is it measured? Describe some hardening methods. [AMIE]

21. Define carburising. Describe the various carburising processes. Why are they preferred over each other? [BE]

22. Outline the principle of high frequency induction hardening. How is induction hardening carried out. Mention its advantages, disadvantages and applications. [BE]

23. Describe flame hardening and compare it with induction hardening. Describe its merits and demerits and mention its applications. [BE, B.Sc. (H)]

24. Write short notes on the following:
(i) Age hardening heat treatment *(ii)* Normalising *(iii)* Full annealing and process annealing *(iv)* Case hardening *(v)* Annealing *(vi)* Nitriding *(vii)* Cyaniding *(viii)* Spheroidising [AMIE]

PROBLEMS

1. Two specimens of an Fe-3% Silicon-alloy, having grain sizes of ASTM 1 and 8 respectively, yield at 118 and 207 MNm^{-2}. Calculate the yield strength of a single crystal of the alloy. [IIT]

2. A cylindrical piece of 8660 steel is to be austenitised and quenched in moderately agitated oil. If the hardness at the surface of the piece must be at least 50 HRC. Determine the maximum allowable diameter. [*Ans.* 95 mm] [BE]

SHORT QUESTION-ANSWERS

1. What is heat treatment?

Ans. It is the controlled heating and cooling of materials in order to deliberately alter their mechanical properties. Changes in material's properties result from changes made in the microstructure of the material.

2. What is the use of time-temperature transformation (T-T-T) curves?

Ans. These curves provide feedback on the time and temperature requirements for heat treatment. Using these curves, the heat-treatment parameters can be adjusted to achieve the best results.

3. What are the different methods through which the purposes of heat treatment may be served?

Ans. Annealing, normalising, hardening, tempering and a number of case hardening and surface hardening methods.

4. What is the purpose of annealing and normalizing a metal?

Ans. It is often important for a metal to be more ductile than hard and strong. Annealing and normalizing are used to obtain a softer, more ductile, and less distorted metal that is easy to machine and form and less likely to crack or distort. Both processes involve slowly cooling the heated metal.

5. What is a quenching process?

Ans. Quenching is the controlled cooling process used to freeze particular grain structures in metals. Before quenching, the metal is heated to high temperatures.

6. Mention few quenching media.

Ans. Water, brine (salt water), oil and air.

7. Why metals are often tempered after hardening?

Ans. After hardening, metals are often tempered to remove stresses and improve their machinability.

8. What is the purpose of case-hardening the softer materials?

Ans. To improve their toughness and wearability.

9. What is martensite?

Ans. This is a transformation product in steels, results when austenite is cooled very rapidly. It is a metastable and single-phase structure that may be produced in steels by diffusionless and almost instantaneous transformation of austenite. We may note that two different morphologies (lath and lenticular) are possible.

10. Which method is used primarily for hardening non-ferrous alloys?

Ans. Age-hardening or precipitation hardening.

11. Is hardening a two stage process?

Ans. Yes, solution heat treatment and aging.

12. What is hardenability?

Ans. It is that property which determines the depth and distribution of hardness obtained by quenching. It depends on cooling methods, grain size, chemical compositions and alloying elements. Hardenability is normally determined by Jominy or end quench test.

OBJECTIVE QUESTIONS

1. Fill in the blanks:

(a) The steel that has been ____________ annealed, will harden more uniformly, will lesser chance of cracking.

(b) Brine gives ____________ cooling rate than water.

(c) Finer grain size steels have ____________ strength than the coarse grained steel.

(d) ____________ is mainly used as a medium of heating the component in case of salt bath carburising.

(*e*) In case of selective heating, time available to change the ___________ of steel is short.

(*f*) Jominy test also gives the hardness at different ___________ rates. [Diploma]

2. State whether the following statements are True/False

(*a*) Oil quenching is required with nitriding.

(*b*) The size of austenite grains increases with increase in temperature and time of heating.

(*c*) Tempering temperature below 200°C relieves the hardening stresses.

(*d*) Age hardening is mainly employed to improve the physical properties of ferrous alloys.

(*e*) When the steel is heated to its hardening temperature, it becomes pearlitic. [Diploma]

3. Which one of the following is not equilibrium heat treatment

(1) Austenetising (2) Annealing

(3) Normalizing (4) Precipitation (4)

4. Chemical heat treatment involves use of some chemical in conjunction with ___________ (heating)

5. Austenitizing is a function of two variables ___________ and ___________. (temperature, time)

6. Spheroidising is principally used for ___________ steels which are more difficult to form (shape) (high carbon)

7. The hardenability is not affected by

(1) air (2) chemical composition of steel

(3) critical cooling rate (4) quenching medium and method of quenching (1)

8. The slowest cooling rate is obtained when steel is quenched in

(1) air (2) brine (3) fused salt (4) mixture of oil and water (1)

9. The fastest cooling rate is achieved when steel is quenched in

(1) air (2) oil (3) water (4) brine (4)

10. Which one of the following is not correct

(1) Martensite has a BCC structure

(2) Austenite has FCC structure

(3) Martensite is a solid solution of carbon in BCC iron

(4) The martensite which is formed during quenching is too brittle (1)

11. Hardenability of steel is assessed by

(1) impact test (2) Jominy end-quench test

(3) hardness test (4) non-destructive test (2)

12. Annealing temperature is

(1) same as normalizing temperature

(2) greater than normalizing temperature

(3) less than normalizing temperature

(4) sometimes greater and sometimes lesser than normalizing temperature (3)

13. Heat treatment that requires heating a part below A_1 temperature, i.e. between 550°C and 650° is called as

(1) hardening (2) normalizing

(3) process annealing (4) full annealing (3)

11

Deformation of Materials

1. INTRODUCTION

Deformation of materials is the change in their dimensions or forms under the action of applied forces. Deformation in a material takes place either by the mechanical action of the external forces or by the various physical and physico-chemical processes, e.g. changes in volume of separate crystallites in phase transformations of as a result of temperature gradient. Normally all materials undergo some deformation in their daily use by the application of external load. We are mainly concerned with this type of deformation. A deformation can be classified as either: *(i) recoverable* or *elastic* after the removal of external force or applied load; such a deformation is termed as *temporary* deformation, *(ii) non-recoverable* or *plastic* or *visco-elastic deformation*, this deformation remains even after the removal of the deforming force or applied load; this type of deformation is termed as *permanent deformation*. Both types of these deformations may be either independent or dependent on time. Metals exhibit less elastic but more plastic deformation even at room temperature while steady flow is predominant at higher temperature ranges.

The difference between the geometrical coordinates of individual metallic body in the loaded and unloaded position represent the *displacements*. The existence of these displacements in a material body exhibit the deformation of the body. When a change in length, i.e., increase or decrease, takes only, the deformation is said *longitudional*. However, when the change in the angles between the various faces of the body takes place, then the deformation is said to be *angular*.

The deformation of a material body is made up of the change in its initial volume, i.e., the *change of density* and the change of its initial stage, i.e., the *distortion.* Obviously, the total deformation can be viewed to split up in *(i)* volumetric, and *(ii)* distortional components.

2. ELASTIC DEFORMATION

When subjected to a force or load, many solid materials behave elastically, i.e., the solid material deforms when loaded but returns to its original shape and size, i.e., position when unloaded. Such type of behaviour is common to metals, ceramics, rubbers and polymers. Elastic deformation in a solid can take place due to change in pressure, or by an application of force or load. We have read that elasticity has its origin in the forces between atoms of the solid, and therefore depends on both the chemical bonding and the structure of solid. However, the elastic properties of a material are independent of crystalline imperfections, e.g. vacancies and dislocations. Elastic deformation is illustrated in Fig. 11.1.

The deformation is said to be an *ideal deformation* which takes place instantaneously upon application of force or load and disappears completely on removal of the force or load. Such deformations in a solid materials obey *Hooke's law* (Stress & Strain). Strain related to this stress should not exceed more than one percent of the original length of the material. Elastic deformation in other non-crystalline materials such as rubber is a few hundred percent as compared to about one percent in metals. We may note that ideal deformation occurs with comparatively smaller deformation forces, i.e., such forces can keep working stresses within the elastic limit.

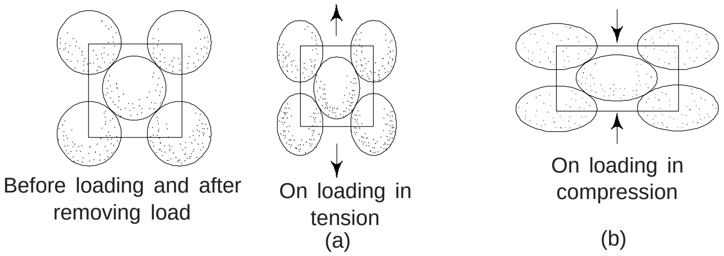


Fig. 11.1 Elastic deformation in a solid material

In engineering practice, material's crystal structures are also subjected to shearing loads. A shearing stress set up by the shearing load produces displacement of one plane of atoms relative to adjacent planes of atoms in a material (Fig. 11.2). Obviously, shear stress tends to slide each layer of atoms past the next. The modulus of elasticity does not apply directly to shear. However, shear stress bears a direct relationship to the elastic shear strain. Upto the limit of elastic displacement the shear strain is proportional to the shear stress.

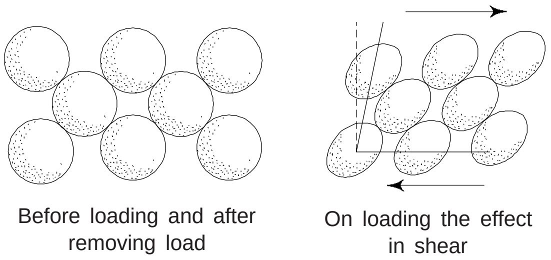


Fig. 11.2 Elastic shear strain (not to the scale)

There are many materials, where Hooke's law does not hold good for all engineering applications where deformation is large, i.e. for these materials, stress & strain, relationship is no longer applicable. Non-ferrous metals, mild steel, etc. are few examples which show *deviation from perfect elastic behaviour*.

3. PLASTIC DEFORMATION

For most metallic materials, elastic deformation persists only to strains of about 0.005. As the material is deformed beyond this point, Hooke's law ceases to be valid, i.e. stress is no longer proportional to strain, and *permanent, nonrecoverable*, or *plastic deformation* occurs. Figure 11.3 shows the tensile stress-strain behaviour into the plastic region for a typical metal. The proportional limit *P*, and the yield strength σ_y has been determined using the 0.002 strain offset method.

The transition from elastic to plastic is a gradual one for most metals; some curvature results at the onset of plastic deformation, which increases more rapidly with rising stress.

Plastic deformation can occur under tensile, compressive and torsional stresses. The rate of plastic deformation is controlled by applied stress, strain rate and temperature control.

The ability of metals to undergo plastic deformation is called *ductility*. In a plastic deformation of a metal, a change in shape occurs concurrently with changes in a number of metal properties; in particular, cold deforming increases the strength of a metal. The ductility of a metal ensures the structural strength of elements under a load and neutralizes the effect of stress concentrations. To

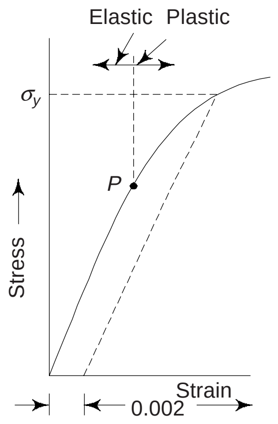


Fig. 11.3 Stress-strain behaviour for a metal showing elastic and plastic deformations

perform processes such as rolling, forging, extrusion, spining, pressing, drawing, stamping, etc., plastic deformation is intensionally carried out. We may note that machinability and wear resistance are entirely dependent on the plasticity of the metals.

From an atomic perspective, plastic deformations are based on irreversible displacement of atoms through substantial distances from their equilibrium. Obviously, plastic deformation corresponds to the breaking of bonds with original atom neighbours and then reforming bonds with new neighbours as large number of atoms or molecules move relative to one another, upon removal of the stress they do not return to their orignial positions. Fig. 11.4 shows plastic deformation of a long range material. The mechanism of this deformaton is different for crystalline and amorphous materials. For crystalline solids, deformation is accomplished by means of a process called *slip,* which involves the motion of dislocations. Plastic deformation in non-crystalline solids (as well as liquids) occurs by a *viscous flow mechanism.*

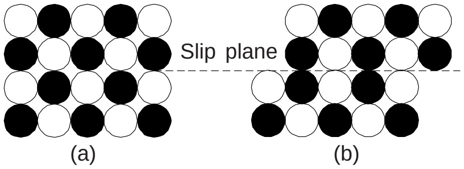


Fig. 11.4 Plastic deformation (a) Prior to deformation (b) After the passage of unit dislocation

3.1 Mechanism of Plastic Deformation

Let us consider the mechanism of plastic deformation taking a single crystal as an example. Plastic deformation in a single crystal is effected by the shear of one portion of the crystal relative to the another portion under the action of *tangential stresses* when they exceed the critical value (τ_{cr}).

Plastic deformation of a single crystal occurs by two principal kinds of shear: *(i)* by *slip* and *(ii)* by *twinning* (Fig. 11.5). In the former case, one portion of the crystal is displaced parallel to its other portion along a plane called the *slip plane* or shear plane (Fig. 11.5a). *Slip* is the main kind of shear in metals and alloys. Deformation by *twinning* is essentially the rearrangement of a portion of crystal into a new position which is mirror-symmetrical relative to undeformed portion (Fig. 11.5b). The plane of the mirror symmetry is called the *twinning plane*. In twinning plane, the atomic planes of a crystal are displaced parallel to the twinning plane through different distances. The portion of the crystal in which the reorientation of the crystal lattice has taken place due to twinning is called the *deformation twin*. Twinning plays a less important part than slip. The role of twinning increases in cases when slip is impeded. In *BCC* and *FCC* metals, twinning is observed only at low temperatures or high deformation rates. In *HCP* metals, deformation may occur both by slip and by twinning under normal conditions.

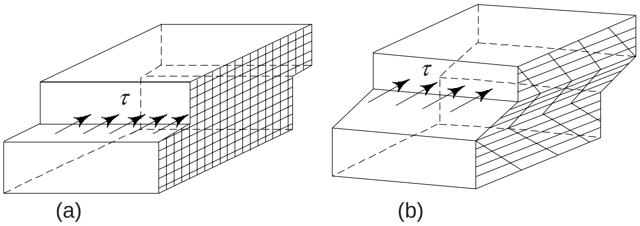
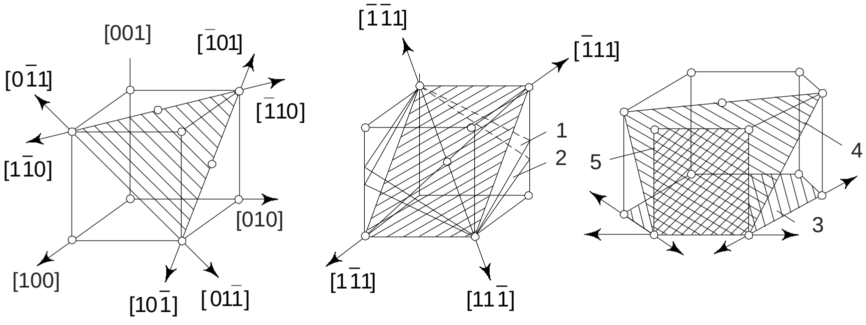


Fig. 11.5 Schemes of plastic deformation (a) by slip and (b) by twinning

3.2 Deformation by Slip

The process by which plastic deformation is produced by dislocation motion is termed *slip*. On the microscopic level, plastic deformation corresponds to the motion of dislocations in response to an externally applied shear stress. Macroscopic plastic deformation simply corresponds to permanent deformation that results from the movements of dislocations, or slip, in response to an applied shear stress.

Slip occurs on specific crystallographic planes and within these planes only in certain directions. A slip system represents a slip plane-slip direction combination, and operable slip systems depend on the crystal structure of the material.



(a) (b) (c)

Fig. 11.6 Slip systems in (a) FCC (b) BCC and (c) HCP lattice metals; planes of the easiest slip are shown hatched: 1 – (112) plane; 2 – (123) plane; 3 – basal plane; 4 – pyramidal slip plane 5 – prismatic slip plane.

Deformation by slip can proceed only along the planes and directions having the most dense packing of atoms (Fig. 11.6). A slip plane together with the direction of slip belonging to that plane forms a *slip system*. The slip system depends on the crystal structure of the metal and is such that the atomic distortion that accompanies the motion of a dislocation is a minimum. The number of slip systems in various metals is different and depends on the type of crystal lattice. In FCC metals (Cu, Al, Ni, etc.) slip occurs in {111} planes and <110> directions. The four slip planes in the FCC lattice and three directions of slip in each of them give 12 equivalent slip systems.

In metals having a BCC lattice, slip develops in {110}, {112} and {123} planes and <111> directions and the number of slip systems is 48.

In metals with an HCP lattice with $c/a \geq 1.63$ (Mg, Zn), slip develops in the basal plane which has three equivalent directions. These metals have a lower ductility than BCC or FCC metals. The number of slip systems can increase if the critical shear stresses in other planes with less densely packed atoms are decreased. With an increase of the number of slip systems, the capacity of a metal for plastic deformation increases. In particular, with $c/a < 1.63$ (as in zirconium and titanium), slip takes place in the basal plane and the pyramidal and prismatic planes, since the critical shear stresses in these planes have very close values. For that reason these metals are more ductile than magnesium or zinc in which slip can occur only in the basal plane. Slip planes, slip directions and number of slip systems observed in various crystal structures are given in Table 11.1.

Table 11.1 Slip planes, slip directions and slip systems in crystal structures

Crystal structure	*Slip planes*			*Slip direction*	*Number of slip systems*
FCC		{111}		<110>	4 × 3 = 12
BCC	More common	{110}		<111>	6 × 2 = 12
	Less common	{112}, {321}			12 × 1 = 12
					24 × 1 = 24
NaCl		{110}		<110>	1 × 6 = 6 or 6 × 1 = 6
HCP	More common	{0001}	Basal Plane	<11$\bar{2}$0> closed packed directions	3 × 1 = 3
	Less common		Prismatic and Pyramidal planes		
Al_3O_2		{0001}		<1120>	1 × 3 = 3

Slip systems for few metals are given in Table 11.2.

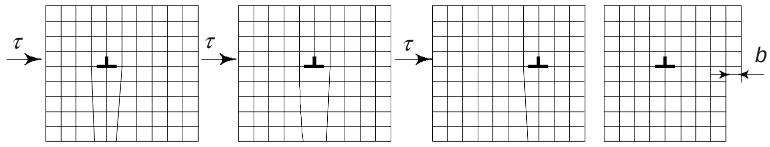
Table 11.2 Slip systems for FCC, BCC and HCP metals

Metals	*Slip plane*	*Slip direction*	*No. of slip systems*
Cu, Al, Ni, Ag, Au	FCC		
	{111}	$<1\bar{1}0>$	12
	BCC		
α – Fe, W, Mo	{110}	$<\bar{1}11>$	12
α – Fe, W	{211}	$<\bar{1}11>$	12
α – Fe, K	{321}	$<\bar{1}11>$	24
	HCP		
Cd, Zn, Mg, Ti, Be	{0001}	$<11\bar{2}0>$	3
Ti, Mg, Zr	$\{10\bar{1}0\}$	$<11\bar{2}0>$	3
Ti, Mg	$\{10\bar{1}1\}$	$<11\bar{2}0>$	6

An elementary act of shear is the displacement of one portion of a crystal relative to its other portion through one interatomic spacing (Fig. 11.7). In a perfect crystal which is free of structural defects, all atoms in the shear plane must simultaneously participate in slipping. For such a 'rigid' synchronous shear, it is required, as has been shown by calculations, that the critical shear stress be $\tau_{cr} = G/2\pi \simeq 0.16$ G (where G is the shear modulus). The stress τ_{cr} is called the theoretical strength of the crystal. In real crystals, shear through one interatomic distance requires a stress around 10^{-6} G, i.e., only one-thousandth of the theoretical value. The low strength of real crystals is attributed to their structural imprefection. Energy (E) for the dislocation line is given by

$$E \propto LGb^2 \tag{1}$$

where L is the length of the dislocation line, G is shear modulus and b is Burger's vector. Energy required for deformation will be minimum when the shear modulus (G) and Burger's vector (b) are minimum. Dislocations having the shortest slip vector (b) are easier to generate and develop for plastic deformation.



(a) (b) (c) (d)

Fig. 11.7 Movement of an edge dislocation by slip

Plastic deformation in real crystals is effected by successive movements of dislocations. A dislocation can move easily in the plane in which the dislocation line and its Burger's vector lie. Under the action of a shear stress, an excessive half-plane at the top of a crystal combines with the underlying portion of an adjacent atomic plane and forms a single atomic plane. The remaining portion becomes a new half-plane which is ended by an edge dislocation. A slight displacement of a small group of atoms turn out to be sufficient for movement of a dislocation through one interatomic distance (Fig. 11.7a and b). With a continuously acting stress, the dislocation is relay raced, as it were from one atomic plane to another and displaces each neighbouring 'regular' row of atoms in this process. The process goes on (Fig. 11.7c) until the dislocation comes onto the surface of the crystal and the top portion of the latter will be shifted relative to the bottom portion through one interatomic distance (Fig. 11.7d). Thus a step of the size of the Burgers vector forms on the surface of the crystal.

A slip plane usually contains tens of dislocations. Their successive displacements along the slip plane develop the process of plastic flow. The height of a step in this process increases proportionally to the number of dislocations that emerge at the surface of the crystal. A step observed with a microscope is called a *slip line*. Deformation develops non-uniformly and distances between the slip lines may be different. Groups of closely spaced slip lines form what is called slip bands (Fig. 11.8). Plastic deformation of single crystals develops preferably in the slip system where the tangential stresses attain the critical value sooner. The initial stage of deformation is called the stage of *easy slip*; the deformation of single crystals at this stage may reach a few tens of per cent. With an increase of deformation, slip is spread over other slip systems, resulting in what is called *multiple slip*. At this stage, dislocations move in intersecting planes, the resistance to their movement increases, and a *complex dislocation* structure forms.

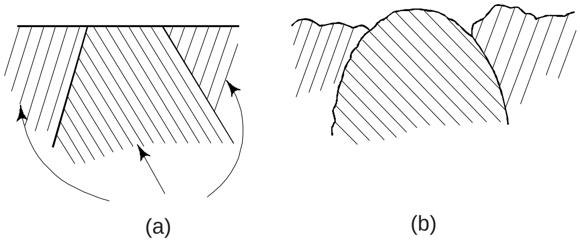


Fig. 11.8 Slip bands (a) Before deformation (b) After deformation

We may note that dislocation slip is not associated with *diffusion*, since it takes place without mass transfer. This explains why this kind of dislocation movement can occur easily even at sub-zero temperatures when the rate of diffusion is quite low.

3.3 Critical Resolved Shear Stress for Slip

The stress at which slip starts in a crystal depends on the relative orientation of the stress axis with respect to the slip plane and the slip direction. When a tensile stress σ is applied to a crystal (Fig. 11.9), the shear stress τ_R resolved on a slip plane whose normal makes an angle ϕ_1 with the stress axis, along a slip direction inclined at an angle ϕ_2 to this axis, is given by

$$\tau_R = \sigma \cos \phi_1 \cos \phi_2 \qquad (2)$$

This is termed *resolved shear stress* (τ_R). The equation (2) is known as *Schmid's law* and the term $\cos \phi_1 \cos \phi_2$ is called *Schmid's sector*. In general, $\phi_1 + \phi_2 \neq 90°$, since it need not be the case that the tensile axis, the slip plane normal, and the slip direction all lie in the same plane.

The resolved shear stress τ_R should reach a critical value called the *critical resolved shear stress* τ_{cr} (CRSS) for plastic deformation to start. If the slip plane and slip direction are inclined at an angle 45° to the tensile axis, critical shear stress will be given by

$$\tau_{cr} = \sigma \cos 45° \cos 45° = \frac{\sigma}{2}$$

Obviously, the critical resolved shear stress will always be less than half the tensile stress. The component of shear stress in the direction of slip is called "*critical shear stress*" and is a constant for each material, and depends upon the type of material. It is affected by the rate of deformation purity of the material, temperature etc. With the rise in temperature, the critical shear stress value decreases, and the thermal mobility of material increases.

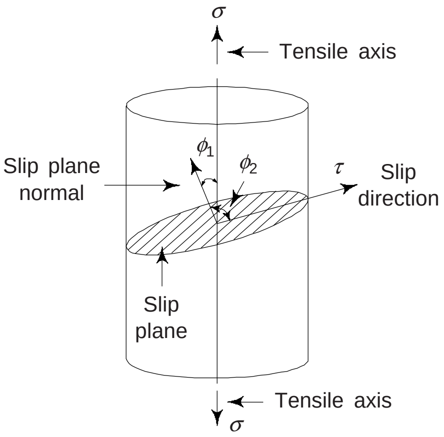


Fig. 11.9 The resolved shear stress (τ) on a slip plane along a slip direction

3.4 Factors Affecting Critical Shear Stresses

(i) The critical stress value is less for pure metals, i.e. purity of metals reduces the critical sresss.

(ii) Surface effects, e.g. surface films greatly enhance the critical shear stress.

(iii) The thermal mobility increases with rise in temperature but critical shear stress decreases.

(iv) Rate of deformation and the extent of initial deformation also help in raising the critical shear stress.

(v) Deformation and slip in *polycrystalline materials* is somewhat more complex. Because of the random crystallographic orientations of the numerous grains, the direction of the slip varies from one grain to another.

3.5 Characteristics of Dislocation

We have read that the process by which plastic deformation is produced by dislocation motion is termed slip. Dislocation motion is analogous to the mode of locomation employed by a caterpillar.

All metals and alloys contain some dislocations that were introduced during solidification, during plastic deformation, and as a consequence of thermal stresses that result from rapid cooling. The number of dislocations, or *dislocation density* in a material, is expressed as the total dislocation length per unit volume or, equivalently, the numbe of dislocations that intersect a unit area of a random section. The units of dislocations are per square millimeter. For heavily deformed metals the density may be as high as 10^9 to $10^{10}/\text{mm}^2$. Heat treating a deformed metal specimen can diminish the density to on the order of 10^5 to $10^6/\text{mm}^2$. A typical dislocation density for ceramic materials is between 10^2 and $10^4/\text{mm}^2$.

Strain fields that exist around dislocations, which are influential in determining the mobility of dislocations, as well as their ability to multiply. When metals are plastically deformed, some fraction of the deformation energy (~5%) is retained internally and the remainder is dissipated as heat. The major portion of this stored energy is as strained energy associated with dislocations. During plastic deformation, the number of dislocations increases dramatically.

3.6 Effects of Dislocations on Cold Working

It is found that if various types of dislocations interact, it can immobilize a dislocation and just reduce its mobility. With the increase in the density of dislocation, the dislocation mobility decreases because increase in density causes more interactions. Obviously, the yield stress of a material depends on its dislocation density. One can increase the dislocation density by deforming a crystalline material plastically.

Fig. 11.10 shows the effect of deformation on the tensile properties of the steel, i.e. tensile strength, yield point, hardness and elongation. However, the cold working (deforming it at room temperature or below) strengthens and hardens a material. Cold working diminishes the amount of elongation which a specimen can sustain before fracture. One can explain all these phenomena by the reduction in dislocating mobility caused by increasing the dislocation density. For hardness, it is found that penetration involves permanent deformation and depends on dislocation mobility.

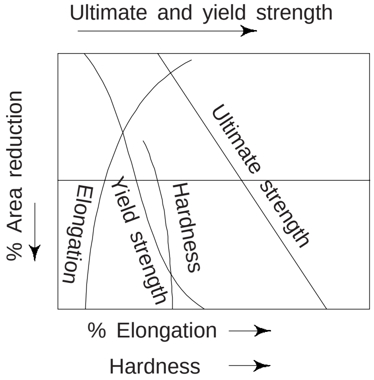


Fig. 11.10 The effect of cold working on the tensile properties (steel)

4. DEFORMATION BY TWINNING

Twinning is the movement of atoms in the lattice, resulting in a division of lattice into two parts which are symmetrical but differently oriented. The amount of movement of each plane in the twinned region is proportional to its distance from the twinning plane and therfore a mirror image is formed across the twin plane (Fig. 11.11). We may note that in twinning, each atom moves individually.

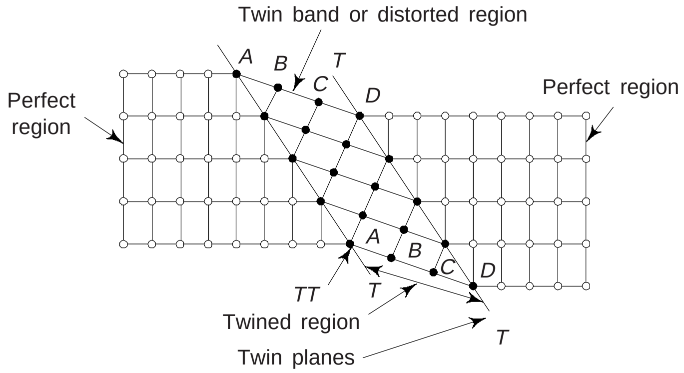


Fig. 11.11 Deformation in a crystal by twinning

Like slip, twinning almost takes place in special planes termed as *twinning planes*. Like that of slip planes, twinning planes are also mostly fixed. In FCC lattice (111) plane is the twinning plane; in BCC lattice (112) is the twinning plane and in HCP lattice (112) plane is the twinning plane. In Fig. 11.12, twinning plane is (111) plane and ML is twinning direction.

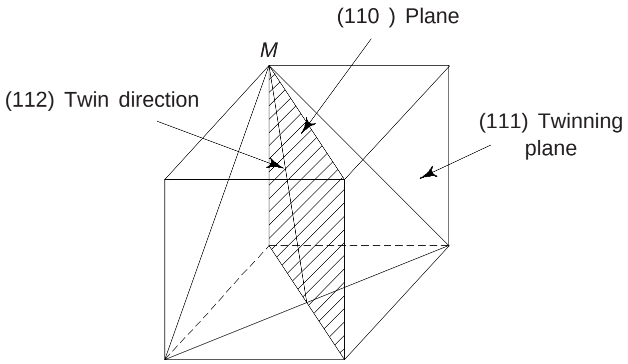


Fig. 11.12 Twin plane and twin direction

Generally, twinning is not a significant deformation mechanism for cubic metals, unless we are concerned with low temperature regions or fast forming processes. In the case of HCP metals twinning becomes more significant and is observed at room temperature. Twinning occurs on the $(1\bar{0}12)$ plane in the $\langle 1\bar{0}11\rangle$ direction in HCP crystals. Twinning reorients the crystal region between the twin planes and thereby facilitates further slip by placing the planes or potential slip in an orientation which is more favourable for dislocation slip. Twinning also plays the role of changing the shape of the surface. Since twinning is produced suddenly and it is accompanied by sound. Twinning is generally caused due to impact, thermal treatment and plastic deformation. In heat treatment, the movements of twin boundaries, as in annealing, is significant. However, the actual deformation produced by twinning is small, having a maximum of perhaps several percent for ductile HCP crystals.

Twins may be produced by mechanical deformation, or as the result of annealing following plastic deformation. The first types of twins are known as *mechanical twins*, the latter are termed as *annealing* *twins*. Mechanical twins are produced in BCC and HCP metals under conditions of rapid rate of loading, i.e. shock loading and decreased temperature. However, FCC metals are not ordinarily considered to deform by mechanical twinning, although gold-silver alloys twin fairly readily at low temperature. Mechanical twins have been produced in copper at 4 K.

Twins are produced in definite direction on a specific crystallographic plane for each type of crystal structures. Common twin planes and directions are given in Table 11.3.

Table 11.3 Twin planes and directions

Crystal structure	*Twin plane*	*Twin direction*
FCC	(111)	<112>
BCC	(112)	<111>
HCP	(1012)	<1011>

4.1 Neumann bands

We have read that twinning is generally caused due to impact, thermal treatment and plastic deformation. As a result of strain or low temperature, when the applied stress is higher than required, *thin lamellar twins* appears. These are called *Neumann bands* (Fig. 11.13). These bands appear when BCC iron and its alloys are rapidly loaded at lower temperatures.

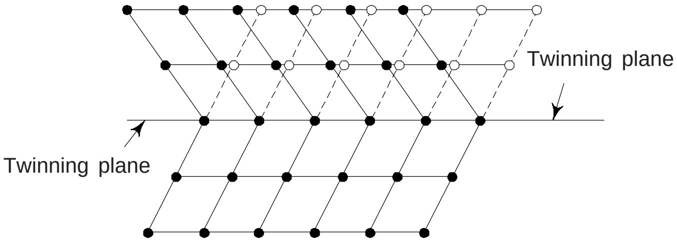


Fig. 11.13 Newmann bands—The arrangement of atoms on either side of a twin surface.

5. COMPARISON BETWEEN SLIP AND TWINNING

A comparison between slip and twinning is presented in Table 11.4.

Table 11.4 Comparison between Slip and Twinning

	Slip *(1)*	*Twinning* *(2)*
(i) Orientation	• The orientation of the crystal above and below the slip plane remains same after deformation.	• Twin causes orientation difference in the crystal across the twin plane.
(ii) Movement during the process	• Slip usually occurs in discrete multiples of atomic spacing.	• The movement of atoms are much smaller in comparison to atomic spacing.
(iii) Deformation planes and directions	• Slip occurs on relatively wide planes. Usually, the slip plane is the plane of greatest atomic density, and the slip direction is the closest packed direction within the slip plane.	• In twinning, every atomic plane is involved in the deformation in the twinned region of the crystal. For each crystal twin occurs in a definite direction on a specific crystallographic plane.

(*Contd.*)

Table 11.4 (*Contd.*)

	Slip *(1)*	*Twinning* *(2)*
(iv) Time	• Slip takes several million seconds to occur, i.e. there is a delay time of several million seconds prior to forming of one slip band.	• Twins can form in a few microseconds only.
(v) Stress	• Slip takes place when the shearing stress on the slip plane in the slip direction reaches a threshold value called the *critical resolved shear stress.*	• Resolved critical shear stress in twinning almost has no role.

6. PLASTIC DEFORMATION OF POLYCRYSTALLINE MATERIALS

Deformation and slip in polycrystalline materials, as compared to single crystals is somewhat more complex. Due to the random crystallographic orientations of the numerous grains and the effect of neighbouring atoms, the direction of slip varies from one grain to another, e.g., in brass. As regard to common metals, they are polycrystalline aggregates because they are made up of a number of small crystals or grains. We may note that the slip mechanism operates because of the geometric regularity of a crystal. For each, dislocation motion occurs along the slip system that has the most favourable orientation.

Gross plastic deformation of a polycrystalline specimen corresponds to the comparable distortion of the individual grains by means of a slip. During deformation, mechanical integrity and coherency are maintained along the grain boundaries, i.e., the grain boundaries usually do not come apart or open up. As a consequence, each individual grain is constrained to some degree in the shape it may assume by its neighbouring grains. Prior to deformation the grains are equiaxed, i.e. they have approximately the same dimension in all directions. For this particular deformation, the grains become elongated along the direction in which specimen was extended.

In case of polycrystalline metals, greater stresses are required to initiate slip and the attendant yielding since they are stronger than their single-crystal equivalents. This is to a large degree, also a result of geometrical constraints that are imposed on the grains during deformation. Even though a single grain may be favourably oriented with the applied stress for slip, it cannot deform untill the adjacent and less favourably oriented grains are capable of slip also, this requires a higher applied stress level. The characteristics of polycrystalline metals are as follows:

(i) When polycrystalline metals are loaded, various grains have random orientation for slip.

(ii) Greater stress is required due to random orientations and therefore, the resolved shear stress is not constant but varies with grain orientation.

(iii) Under load condition, all crystals of a polycrystalline metal are prefectly elastic. There are a number of slip planes within a crystal. Considerable structural changes result in deformation.

(iv) Contrary to single crystals, polycrystalline metals show elastic after effects. Application of this effect is in *work hardening*.

(v) A significant amount of permanent deformation and an appreciable amount of yielding can be obtained in the polycrystalline crystals of metals after elastic deformation.

7. WORK HARDENING OR STRAIN HARDENING

This is the phenomenon whereby a ductile metal becomes harder and stronger as it is plastically deformed. Ductile materials show increase in strength and hardness when plastically deformed at temperature lower than the crystallization temperature. Most metals stain harden at room temperature. We may note that the temperature at which deformation takes place is “cold” relative to the absolute temperature of the metal *cold working*. Cold working reduces ductility because part of the elongation that takes place is observed during testing. The cold-working destroys the randomness of grain orientations and the resulting properties of the material are quite different in various directions. During cold-working between 1 to 10% of the energy of plastic deformation is stored in the material as internal energy. Plastic working below 0.3-0.5 T_m is called work, where T_m is the melting point in Kelvin.

Sometimes it is convenient to express the degree of plastic deformation as *percent cold work* rather than as strain. Percent cold work (% CW) is defined as

$$\%\ \text{CW} = \left(\frac{A_0 - A_d}{A_0}\right) \times 100 \tag{3}$$

where A_0 is the original area of cross section that experiences deformation, and A_d is the area after deformation.

Figure 11.14 shows the strain hardening. Initially, the metal with yield strength σ_{y_o} is plastically deformed to point D. The stress is released and then reapplied with a resultant new yield strength, σ_{y_i}. The metal has thus become stronger during the process because σ_{y_i} is greater than σ_{y_o}.

One can explain the phenomenon of strain-hardening on the basis of dislocation-dislocation strain field interactions and the interactions of dislocations with the barriers which immobilize them when crossing through them. The dislocation density in a metal increases with deformation or cold work, due to dislocation multiplication or the formation of new dislocations. Consequently, the average distance of separation between dislocations decreases, i.e., the dislocations are positioned closer together. On the average, dislocation-dislocation strain interactions are repulsive. Obviously, the motion of a dislocation is hindered by the presence of other dislocations. As the dislocation density increases, this resistance to dislocation motion by other dislocations becomes more pronounced. Clearly, the imposed stress necessary to deform a metal increases with increasing cold work.

Work hardening phenomenon can also be explained from the intersection of various dislocations moving through the active slip planes. The resistance of the metal to deformation is increased to work or strain the metal in both the cases. This implies an increase in strength and hardness.

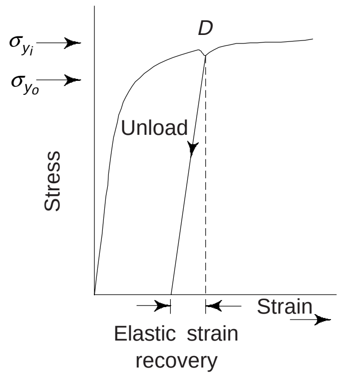


Fig. 11.14 **Tensile stress-strain diagram show in the phenomena of elastic strain-recovery and *strain hardening*. σ_{y_o} is the initial yield strength. The yield strength after releasing the load at point *D*, and then after upon reloading is σ_{y_i}**

We can show three stages of work hardening for work hardenable metal on the stress-strain curve (Fig. 11.15).

(i) Work hardening rate is practically constant—Stage I

(ii) Due to predominant of dislocation, the work hardening rate is higher than that of Stage I—Stage II

(iii) Due to the lower mobility of dislocation, the material continues to work harden but at a decreasing rate—Stage III.

The strain hardening process is of great significance in several metal-forming and fabrication operation in industries. This is often utilized commercially to enhance the mechanical properties of metals during fabrication procedures. As a result of work hardening, electrical conductivity can be decreased whereas the rate of chemical action increased. Work hardening might cause development of internal stresses, increase in corrosion and crack formation, elastic distortion and fragment of crystals. The effects of strain hardening may be removed by an annealing heat treatment.

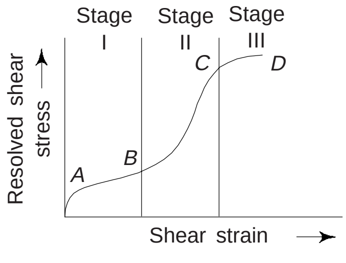


Fig. 11.15 **Three stages of workening for work hardeneable metal**

8. SEASON CRACKING

This causes disintegration of metal. The combined effects of internal stress and intercrystalline corrosion causes season cracking. Usually this term is applied to metals such as brass. After cold working, internal stresses of very high intensity may be left in a specimen. These stresses are susceptible to corrosion in parts of metals if these are stored for a long duration. Internal stresses make alpha brass quite appreciable susceptible to intergranular cracking in atmospheres containing ammonia. Corrosive agents like sea water industrial atmosphere etc., accelerates the development of cracks. One can overcome this defect by annealing the brass at 200–300°C.

It is interesting to note that the tendency of brass cracking also increases with increased zinc or alloying content which results in active response to inter-granular attack along the grain boundaries.

9. BAUSCHINGER EFFECT

This is the directionality of straining, i.e., if strain hardening in a material takes place due to application of stress in a particular direction it will be easier to deform in the opposite direction. We can also see that if a material has a yield point stress Y in tension and is continued to deform from tension to compression in the same cycle it will show a yield point in the compression at Y', where Y' is less than Y. The main reason of this effect is that lower stress is required to reverse the direction of slip on a certain slip plane than to continue slip in the origin direction.

To illustrate this effect, we consider Fig. 11.16. *OAC* is the stress-strain curve of a certain ductile material in tension. The initial yield stress of material in tension occurs at *A*. If the same material were tested in compression the yield would occur at *B*. Let another specimen of the same material be loaded in tension upto the point *C* which is beyond the yield point *A*. Now, the specimen is unloaded and it follows the path *CD*. A compressive stress is applied from the point *D* on the specimen. The plastic flow start begin at the point *E*. The stress corresponding to the point *E* is appreciably below that corresponding to point *B*. This decrease in yield stress is the *Bauschinger effect*. This effect is reversible in the sense that if compressive stress were applied first followed by tension then the material could have shown a lower yield strength in tension. The difference in strain between tension and compression curve at a given stress is a measure of the amount of Bauschinger effect. If, we now complete the loading cycle by further loading in compression to point *F*, then unloading and reloading in tension, the Bauschinger effect will result in closure of the stress-strain loop at *C*. This loop is known as *mechanical hysteresis loop*. The area enclosed within this loop is the loss of energy due to Bauschinger's effect.

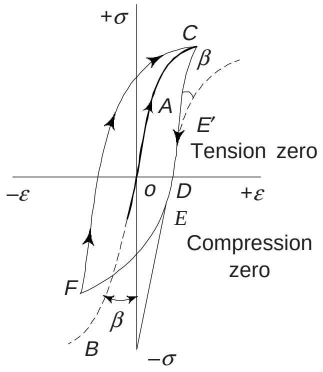


Fig. 9.16 **Bauschinger's effect and hysteresis loop**

The said phenomenon is observed in the cold working of metals. For example, cold rolled steel is strengthened by work hardening. Rollers compress the mild steel stock and cause elongation and thereby increasing the tensile yield stress. However, there is a decrease in tensile yield stress across the bar due to lateral compression.

10. ANELASTICITY

So far, we have assumed that elastic deformation is time independent, i.e., that an applied stress produces an instantaneous elastic strain that remains constant over the period of time the stress is maintained. We have also assumed that upon release of the load the strain is totally recovered, i.e., that the strain immediately returns to zero. However, in most engineering materials, there will also exist a time-dependent elastic strain component. Obviously, elastic deformation will continue after the stress application, and upon load release some finite time is required for complete recovery. This time-dependent elastic behaviour, i.e., strain lagging the stress is termed as *anelasticity*. This is due to time-dependent microscopic and atomistic processes that are attendant to the deformation. The anelastic component is normally small for metals and hence neglected. However, its magnitude for some polymeric materials is significant and in this case it is called *viscoelastic* behaviour. In an anelastic material if the load is repeating in a cyclic manner, there will be certain dissipation of energy due to anelasicity. For those materials which are required for absorbing the vibration, this dissipation of energy will enhance their capacity to damp out the vibrations. In other materials this dissipation of energy may lead to undesirable heating.

11. ADIABATIC AND ISOTHERMAL STRAINING

Stress and strain are related with temperature and entropy in the elastic range. This means that change in stress and strain will cause alteration of temperature and entropy. This relationship is termed as *thermo-elastic* effect. We can show that

$$\left.\frac{\partial T}{\partial \varepsilon}\right|_s = -V_m \ \alpha ET/C_v \tag{4}$$

where V_m is molar volume of the material, E is isothermal modulus of elasticity, α is coefficient of thermal expansion, T is absolute temperature, S is entropy and C_v is specific heat at constant volume.

If we want to stretch a crystalline material, then this can be achieved in two ways: *(i)* the material can be extended at a slow rate so that the sufficient time may be allowed for exchange of heat with the atmosphere resulting into its constant temperature, *(ii)* the stretching could be so fast that no time is allowed for exchange of heat with the atmosphere as a result the material will cool as is clear from Eq. (4). The former is known as *isothermal straining*, while the latter is termed as *adiabatic straining*. For a given stress the isothermal strain will be greater than the adiabatic strain and isothermal modulus of elasticity at any point will be smaller than adiabatic modulus of elasticity. For elastic deformation, the isothermal and adiabatic stress-strain relationship is shown in Fig. 11.17. In actual practice the process of loading a material is neither adiabatic and nor isothermal. There is, always, some exchange of heat take place with the surrounding during loading and unloading. In such a situation curves shown in Fig. 11.17 are not obtained. Instead the curves shown in Fig. 11.18 will result. The loop shown in Fig. 11.18 is essentially the elastic hysteresis energy lost and will be equal to area included in the loop.

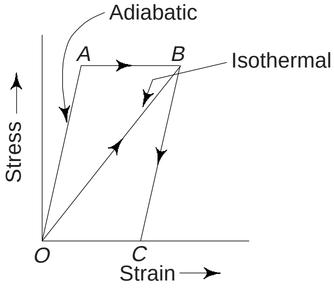


Fig. 11.17 Thermoelastic effect

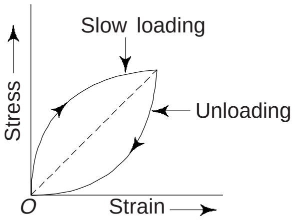


Fig. 11.18 Elastic hystresis

The shape and size of the hysteresis loop varies with the frequency of cyclic loading. The rate of loading and unloading at low frequencies is slow. The slow rate corresponds to nearly isothermal conditions and for such cases the width of hysteresis loop is negligible. This means, there will be no loss of energy. The rate of loading is substantially high at very high frequencies. The loading and unloading become nearly adiabatic. In this case also the width of the loop becomes very small and there will be no loss of energy. However, at frequencies between the two extremes (very low and very high) the hysteresis loop will have some width and hence the energy loss will occur. The variation of energy loss with respect to frequency is shown in Fig. 11.19.

This hysteresis loss of energy in the material in attributed to *internal friction* or *damping capacity*. The cyclic load which may cause vibration, this would put the material in a situation in which energy in form of hysteresis loop is lost continuously. However, if the material is capable of dissipating heat to the surrounding, its temperature will not rise. We may note that the internal friction or damping capacity is the direct function of degree of perfection of internal structure. A material free from defects tends to have

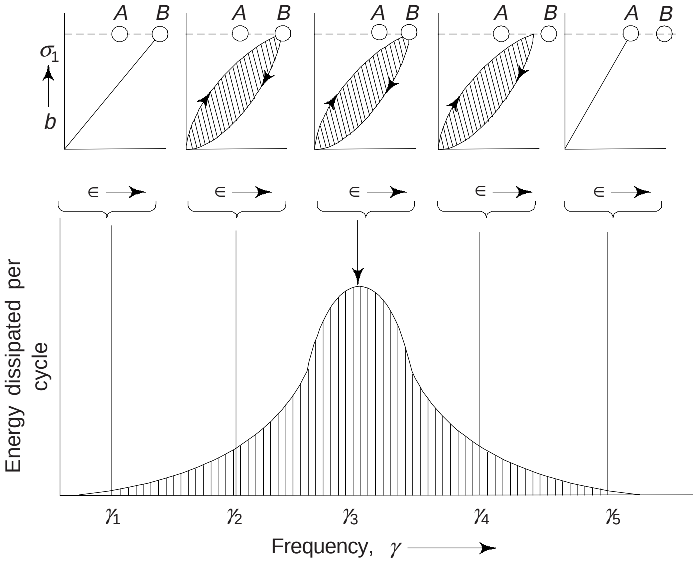


Fig. 11.19 Energy dissipation per cycle as a function of frequency

overall high strength but low damping capacity. However, the defects like dislocations and grain boundaries may have role to play. In polymers it is due to molecular structure of a specific material.

In practice, the vibration amplitudes have to be controlled and for that purpose external damping is provided. Internal damping is not dependable because it is small in high strength materials of which machine parts and structures are made. Heat has to be produced in hysteresis for better damping capacity. At the time of designing, a balance between vibration damping and heating of damping material will have to be achieved. The ratio of energy loss in a cycle to the energy corresponding to maximum strain of a body is termed as *specific damping capacity*. This varies from material to material and also with frequency for a given material. Mechanical treatments and temperature have effect on specific damping capacity. Metals have much lower damping capacity than polymer materials. However, impurities in metals increase damping.

12. YIELD POINT PHENOMENON AND RELATED EFFECTS

Engineers design most structures to ensure that only elastic deformation will result when a stress is applied. It is desirable to know the stress level at which plastic deformation begins, or where the phenomenon of *yielding* occurs. For metals that experience this gradual elastic-plastic transition, one can determine the point of yielding as the initial departure from linearity of the stress-strain curve, this is sometimes referred as the *proportional limit*, as shown by point *P* in Fig. 11.3. In such cases the position of the point *P* may not be determined precisely. For precise determination of point *P*, a straight line is constructed parallel to the elastic portion of the stress-strain curve at some specified strain off set, usually 0.002. The stress corresponding to the intersection of this line and the stress-strain curve as it bends over the plastic region is called as the *yield strength* σ_y (here, ‘strength’ is used in lieu of stress) Fig. 11.3. The units of yield strength are MPa.

For those materials having a non-linear elastic region, strain offset method is not useful. For such a situation, the usual practice is to define the yield strength as the stress required to produce some amount of strain, e.g. $\varepsilon = 0.005$.

Some steels and other materials shows the stress-strain diagram as obtained during tensile deformation when a sample of mild steel was loaded (Fig. 11.20). The elastic plastic transition is very well and occurs abruptly in what is termed a *yield point phenomenon*. At the upper yield point, plastic deformation is initiated with an actual decrease in stress. Continued deformation fluctuates slightly about some constant

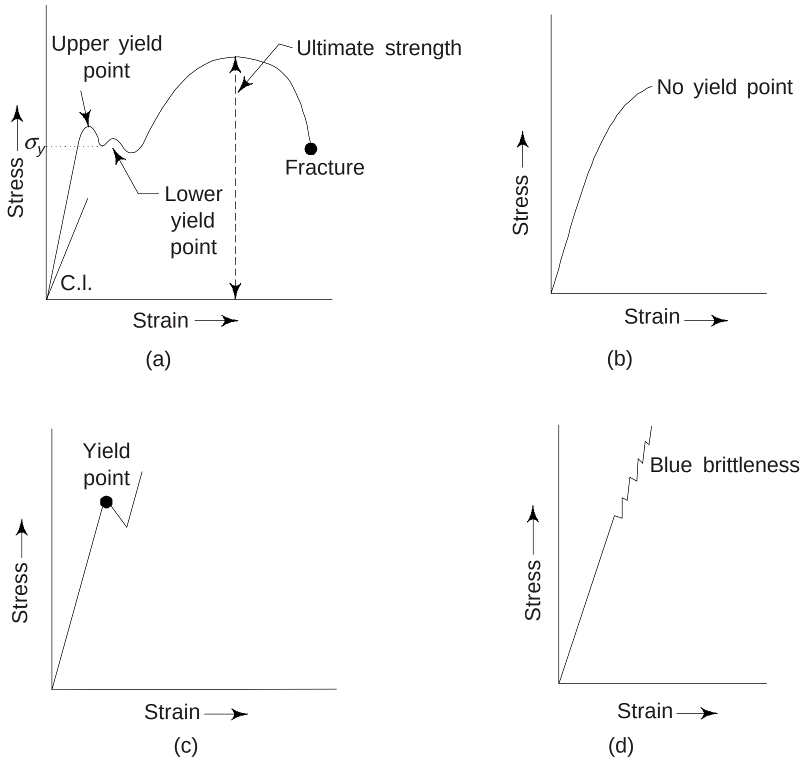


Fig. 11.20 Stress-strain behaviour founds for some steels demonstrating the yield point phenomenon and related effects. (a) Anchored atmosphere (first test) (b) Free atmosphere (unloaded) and (c) Strain aging (anchored atmosphere)

stress value, termed the lower yield point. However, stress subsequently rises with increasing strain. For metals that display this effect, the yield strength is taken as the average stress that is associated with the lower point. Obviously, it is not necessary to employ the strain offset method for these materials.

We may note that the magnitude of the yield strength for a metal is a measure of its resistance to plastic deformation. Yield strengths may range from 35 MPa for a low strength aluminium to over 1400 MPa for high-strength steels. In addition to iron and steel, yield points have also been recorded in polycrystalline metals like molybdenum and aluminium alloys. Yield point phenomenon is also important for several forming processes like stampings and drawing of thin sheets in the fabrication of automobile parts, etc.

The related effects of yield point are *(i)* Luders bands and stretcher strains, *(ii)* Strain ageing, *(iii)* Blue brittleness, and *(iv)* Orange peel effect.

(i) Luders Bands: These represents the markings on the surface of a tensile test sample, formed at the points of stress concentration like fillets. These bands are usually at 45° to the tensile axis when several bands are formed, the stress-strain curve is irregular during yield point elongation (Fig. 11.20a). Luders bands distinguish those portions of the sample that have yielded form those which have not. These bands are formed because of *jogs*. Arrival of the upper yield point is indicated by the formation of these. Luder bands can be observed at a number of positions on the sample, e.g. at both the ends on the gauge length. When Luder bands propagate and cover the whole length, the flow will increase with strain. The yield point elongation ends here. In drawing and stamping operations, Luders bands produce rough appearance on the surface of the metal sheet and have to be avoided on the finished products. In order to control this uneven yielding effect, one will have to overstrain the sheets prior to pressing operations, by means of temper roll pass. This helps to eliminate the yield point phenomenon.

(ii) Strain Ageing: This is associated with yield point phenomenon and refers to changes in the properties of an over strained alloy with time. In order to remove the yield point, if the test sample which has been overstrained, is allowed to rest or age after plastic deformation, the yield point returns with higher stress (Fig. 11.20c). However, if the specimen is unloaded and rested without ageing then the yield point will occur (Fig. 11.20b) due to weak dislocation effects from the atmosphere of carbon and nitrogen atoms. Carbon and nitrogen atoms form new interstitials and anchor the dislocations.

In strain ageing of iron, nitrogen plays a more important role than carbon because nitrogen has higher solubility and diffuse easily.

Strain ageing or strain-age-hardening is accompanied by hardening due to increase in stress value. When the sample, after test of Fig. (11.20a) is unloaded and allowed to rest or age for some time and then retested, when the yield point as in Fig. (11.20c) reappears, the said phenomenon is observed. Commercial iron and steel show this behaviour. This has an important effect on the study of fatigue and creep behaviour of some metals.

(iii) Blue Brittleness: When the temperature range of the test is raised; the yield point becomes less pronounced and blue brittleness effect is produced. For plain carbon steel the temperature range is 200 to 370°C. However, this effect appears at about 160 to 300°C (Fig. 11.20d). We can see that the section of the stress vs. strain diagram shows a steeper curve with rapid formation of successive yield points due to quick diffusion and dislocation effects.

(iv) Orange Peel Effects: During stretching, these markings appear on metals. This is observed in coarse-grained metals at low-temperature-forming operations and it is also a surface roughening phenomenon. This effect is eliminated by grain refinement.

Interstitial atoms (C, O, H, and N) can contribute largely to the strengthening effect if they accumulate on dislocations and lock these, thus forming segregations or *cottrell atmospheres*. Such atmospheres anchor the dislocations and higher stresses are required to free a dislocation. However, the value of this stress should be corresponding to upper yield points. We may note that free dislocations lower the yield point. Cottrell atmospheres gives yield point and blue brittleness as shown by stress-strain diagrams.

13. ATOMIC DIFFUSION–AN ELASTIC AFTER EFFECT

We have stated that diffusion is the process which causes the atoms to migrate to positions of lower energy. Obviously, the atomic structure is relaxed and further strain is allowed. When unloaded, the atoms may migrate back to their original positions. Steel is a very good example in which the interstitial carbon atoms occupy the positions at the centers of the edges of the cube in the BCC unit cell of iron. Figure 11.21 illustrates one such atom. The carbon atom is slightly larger to fit into these positions. If carbon atom does it, distorts the iron lattice. If the steel is unstressed these carbon atoms tend to distribute themselves uniformly among the cube edges. When we apply the stress, the edges of the unit cell parallel to the stress direction elongate. The edges of the unit cell perpendicular to the stress contract due to the Poisson's effect. The contraction becomes difficult, if one of these edges contains an interstitial carbon atom. The tendency of contraction will push the atom out and make it jump to a position in an edge parallel to the applied stress (Fig. 11.21). In case, if the stress is applied very slowly the carbon atom has sufficient time for such migration. Moreover, if the stress removal is also slow then the carbon atom has sufficient time to diffuse back to the original position. If the application of load is very rapid, the carbon may

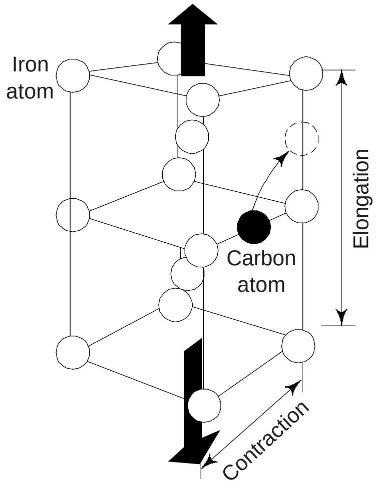


Fig. 11.21 Stress induced diffusion of carbon in iron

not have sufficient time to diffuse to an edge along the direction of the load. However, when sufficient time is allowed under stress the carbon atom will diffuse to the position which is more favourable to it and thereby causing the delayed elastic effect. Naturally, this elastic after effect will depend upon the rate of diffusion which in turn is governed by temperature. If the temperature is high the diffusion rate is high and obviously, the transient strain occurs rapidly. However, the transient strain exists until no more atoms can jump.

14. PREFERRED ORIENTATION

Usually, ordinary alloys show no effect of direction on their elastic properties. Such crystals are named as *isotropic*. Amorphous solids, liquids and gases are isotropic, i.e., their properties are same in every plane and every direction. The spacing of atoms in crystalline material, differ on different planes and different directions along these planes. We know that the physical and mechanical properties of a material are determined by the arrangement of atoms, these crystals are *anisotropic*, i.e., their properties change according to the plane or direction along which they are measured.

In single crystals, anisotropy is observed in elastic and plastic strength, ductility, hardness, diffusion rate, thermal and electrical conductivity, etc. For example, the elastic modulus of single iron crystals is 1.37×10^{10} kg-m^{-2} in [100] direction, 2.28×10^{10} kg-m^{-2} in [110] direction and 2.91×10^{10} kg-m^{-2} in [111] direction. The electrical resistivity of graphite crystal is 100 times greater in one direction than in the direction at right angle to it.

Alloys are made up of millions of very tiny crystals. If the crystals are randomly oriented, the elastic properties will be same in all directions. It is possible for crystals to assume identical orientation i.e., try to align themselves in a common axial direction by casting, rolling or heat treatment processes. Such a direction is known as *preferred orientation*. Obviously, preferred orientation occurs under severe deformation of metal in which certain crystallographic planes tend to orient themselves in a preferred number with respect to the direction of maximum strain (Fig. 11.22). The properties of the material will be different in different direction.

There are some individual crystals, which are strongly anisotropic. A poly-crystalline material is mostly isotropic because of random orientation character. One can develop the preferred orientation by mechanical working. The grain tends to rotate the slip plane along tensile axis (during plastic deformation).

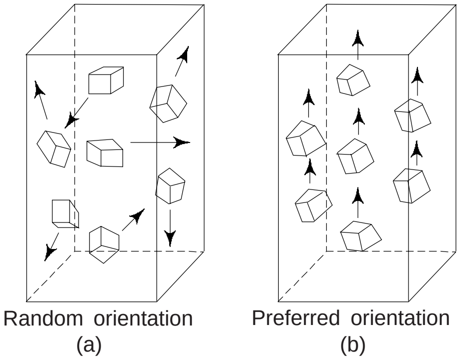


Fig. 11.22 Preferred orientation. Grains are represented by cubes: (a) random orientation and (b) preferred orientation

Preferred orientation resulting from plastic deformation is found to depend strongly on the slip and twinning systems. The slip planes of all crystals are caused to rotate into more favourable directions with respect to the direction of critical shear stress. X-ray methods are used to determine or detect them. We may note that heat treatment cannot eliminate preferred orientation, once it has been developed by mechanical working.

Preferred orientation finds following engineering applications:

(i) Directional properties, e.g. strength and magnetic permeability are quite useful in crystals of some metals, e.g. silicon steels, and quite undesirable in many others.

(ii) The formation of a strong preferred orientation can cause an anisotropy in mechanical properties. This means that during fabrication, different mechanical properties can be obtained in different directions.

(iii) The manufacture of transfer cores of iron sheets helps in orienting the core iron in the direction of the magnetic field. This helps in saving a large amount of electrical energy dissipated as heat in the iron core by eddy currents.

(iv) Preferred orientation causes the loss of ductility in semifinished products likes wires and sheets. In cold rolled sheets, the preferred orientation is developed which results in lower tensile strength, lower yield strength and lower ductility across the grain, i.e. transverse to the rolling direction, than parallel to it. If one wants to have more strength, ductility, etc., in transverse direction too, then he has to break up preferred orientation. One can achieve this by cross-rolling the sheets (for deep drawing purposes), i.e. rolling alternatively in two directions at right angles to each other.

(v) Preferred orientation has been found useful in cast metals having columnar growth.

15. RECOVERY, RECRYSTALLIZATION AND GRAIN GROWTH

Plastically deforming a metal specimen at temperatures that are low (~ 25°C) relative to its absolute melting temperature produces microstructural and property changes. These include: *(i)* a change in grain shape, *(ii)* strain hardening, and *(iii)* an increase in dislocation density. Some fraction of the energy spent in deformation of the material is stored in the metal as strain energy. This energy is associated with tensile compressive, and shear zones around the newly created dislocations. Moreover, other properties, e.g. electrical conductivity and corrosion resistance may also be modified as a consequence of plastic deformation.

One may revert back these properties and structures to the pre-cold worked states by appropriate heat treatment, i.e. annealing treatment. Such restoration of properties results from two different processes that occur at elevated temperatures: *recovery* and *recrystallization*, which may be followed by *grain growth*. Recovery takes place at relatively low temperatures (below 0.3 T_{mp}) and recrystallization at higher temperatures.

Recovery: This term implies all changes in the fine structure and properties of a metal which involve no changes in the microstructure of the deformed metal, i.e., recovery does not change the size and shape of grains. During recovery, some of the stored internal strain energy is relieved by virtue of dislocation motion (in the absence of an externally applied stress), as a result of enhanced atomic diffusion at the elevated temperature. There is some reduction in the number of dislocations, and dislocation configurations are produced having low strained energies. Also, physical properties such as electrical and thermal conductivities and the like are recovered to their precold-worked states.

Recovery is further subdivided into two stages: *strain-relief crystallization* and *polygonization*. The former always takes place during heating of deformed metals, whereas polygonization can develop only under appropriate conditions.

Strain-relief crystallization of a cold-deformed metal is the stage at which the number of point-defects, mainly vacancies, diminishes; in a number of metals, such as aluminium and iron, strain-relief crystallization also includes dislocation climb which is accompanied with interaction of unlike-sign dislocations and results in a noticeable decrease of dislocation density. Redistribution of dislocations also involves a release of residual stresses. Strain-relief crystallization decreases the electric resistance and increases the density of a metal.

In the general case, the hardness and strength decrease at most by 10-15% from their initial values and the ductility accordingly increases. After strain-relief crystallization, the metal has a higher resistance to corrosion cracking.

Polygonization is a kind of recovery involving the formation of new low angle boundaries within each grain. These boundaries can form by slip and climb of dislocations; as a result, a grain is divided into subgrains, or polygons which are free of dislocations (Fig. 11.23).

In commercial purity metals and solid-solution alloys, polygonization is observed only after slight deformation and not in all metals. For example, this process only rarely develops in copper and copper alloys, but is well pronounced in aluminium, iron, molybdenum, and alloys of these metals. Polygonization of a cold-deformed metal usually results in a decrease of hardness and strength. A block-like structure appearing on polygonization is quite stable and is retained in the metal almost up to the melting point. After the formation of a block-like structure, recrystallization cannot take place, i.e. polygonization and recrystallization turn out to be competing processes.

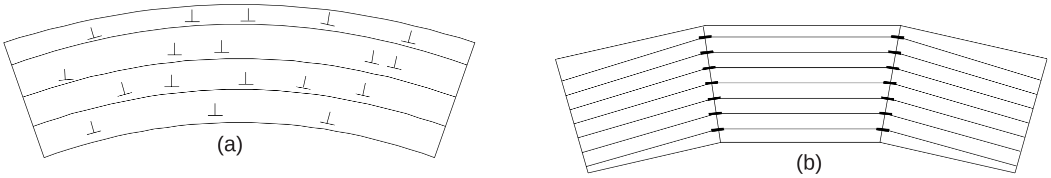


Fig. 11.23 **Scheme of polygonization; (a) strain hardened metal prior to polygonization; (b) after polygonization**

Recrystallization: This is understood as the nucleation and growth of new grains with a smaller number of structural defects; recrystallization results in the formation of entirely new, most often equiaxed crystals. Obviously, recrystallization is the formation of new set of strain-free and equiaxed grains (i.e., having approximately equal dimensions in all directions) that have low dislocation densities and characteristic of the precold-worked condition. The main driving force to produce this new grain structure is the difference in internal energy between the strained and unstrained material. The new grains form as very small nuclei and grow until they completely consume the parent material, processes that involve short-range diffusion. Recrystallization of a cold-worked metals may be used to refine the grain structure.

Moreover, during recrystallization, the mechanical properties that were changed due to cold working are restored to their precold worked values. This means that the metal becomes softer, weaker, yet more ductile.

Recrystallization in plastically deformed metals can take place only when the degree of deformation has exceeded a definite critical value called the *critical degree of deformation*. With the degree of deformation less than the critical value, nucleation of new grains during heating is impossible. The critical degree of deformation is not high (2–8%); it is close to 2% for aluminium and 5% for iron and copper.

Recrystallization is a process the extent of which depend on both time and temperature. The influence of temperature is shown in Fig. 11.24.

Recrystallization is also characterized by a definite *recrystallization temperature* (T_{rec}) which is the lowest temperature of heating at which nucleation of new grains is possible. The recrystallization temperature of a metal constitutes a certain fraction of its melting point temperature (T_{mp}).

$$T_{\text{rec}} = a\ T_{\text{mp}} \tag{5}$$

The coefficient a depends on the purity of metal and degree of plastic deformation. The recrystallization temperature for the brass alloy of Fig. 11.24 is about 450°C. Typically, it is between 1/3rd and 1/2 of the absolute T_{mp} of a metal or alloy. For commercially pure metals, a = 0.3-0.4 and decreases with increasing degree of deformation. Reducing the concentration of impurities can diminish the coefficient a to 0.1-0.2. For solid solutions, a = 0.5-0.6 and can reach 0.7-0.8 on dissolution of high-melting metals. For Al, Cu and commercially pure Fe, the lowest temperatures of recrystallization is equal to 100°C, 270°C and 450°C respectively.

Increasing the percentage of cold work enhances the rate of crystallization, with the result that T_{rec} is lowered, and approaches a constant or limiting value at high deformations (Fig. 11.25).

Grain-Growth

During recrystallization, the nucleation of new grains occurs in portions of the highest dislocation density, usually at boundaries of deformed grains. The greater the plastic deformation, the more recrystallization centres appear in a metal. These centres are essentially submicroscopic regions with the least quantity of point and linear defects in the structure. These regions appear due to redistribution and partial annihilation of dislocations. In this process, a low angle boundary appears between a recrystallization centre and deformed matrix.

After a certain time, centres of new grains increase in size due to the passage of atoms from the deformed surroundings to a more perfect lattice; the high-angle boundaries of new grains then move into the depth of strained-hardened metal.

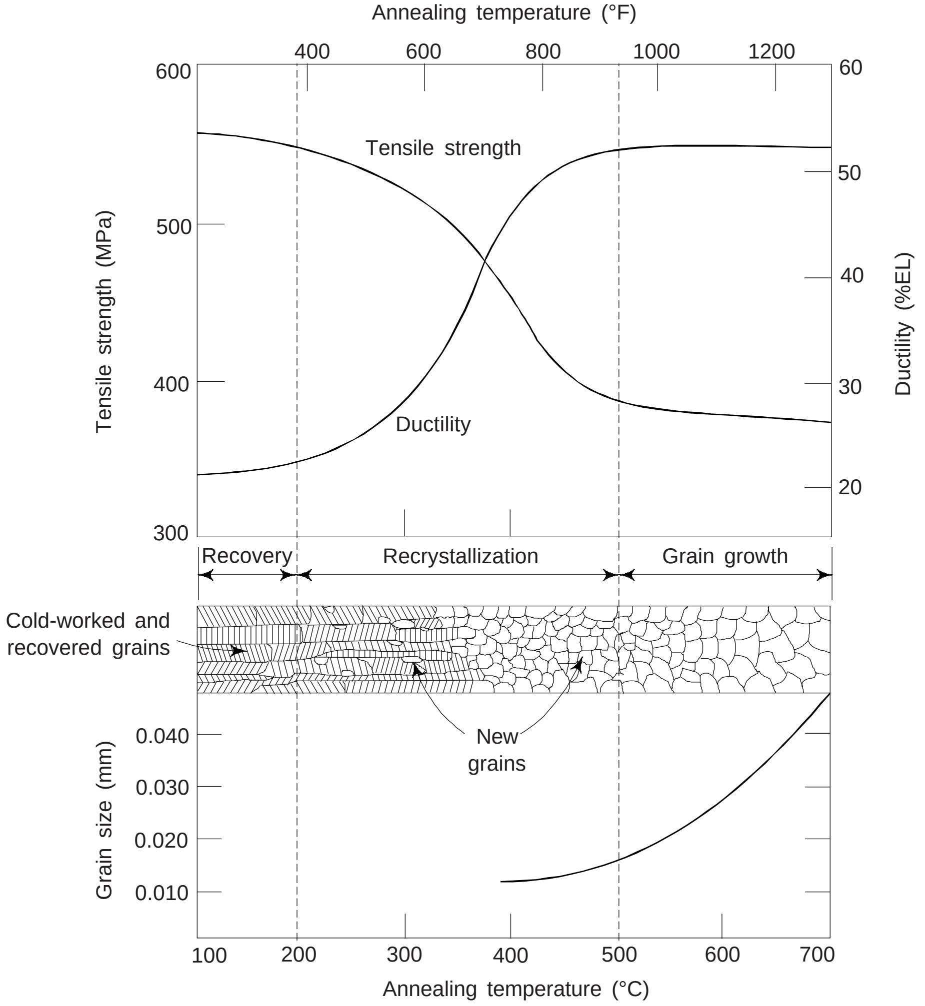


Fig. 11.24 The influence of annealing temperature on the tensile strength and ductility of a brass alloy. Grain size as a function of annealing temperature is also shown. Grain structures during recovery, recrystallization and grain growth stages are also shown schematically

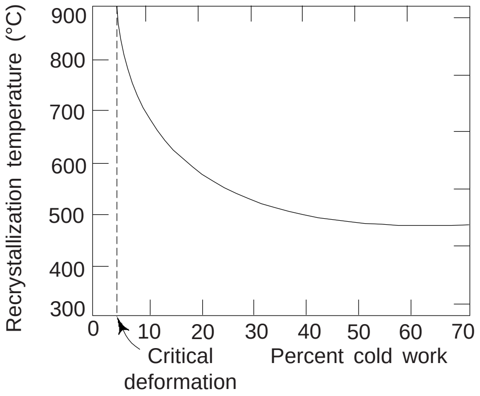


Fig. 11.25 Variation of recrystallization temperature (T_{rec}) with percent cold work for iron

Recrystallization of the kind just described is called *primary recrystallizatoin*. Primary recrystallization is over after complete substitution of old grains in the entire volume of deformed metal by new grains (Fig. 11.26).

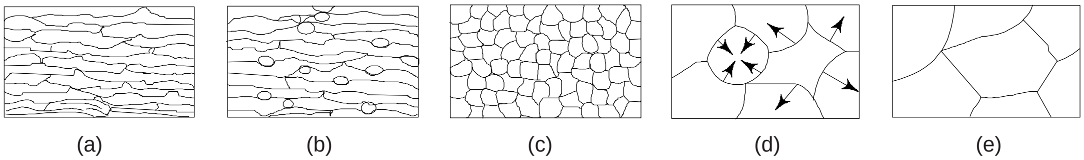


Fig. 11.26 Changes in microstructure of strain-hardened metal during heating: (a) strained hardened metal, (b) beginning of primary recrystallization, (c) end of primary recrystalliztion, (d) and (e) stages of grain growth

Subsequent heating and extension of the holding time, when the process of primary crystallization has been completed, lead to growth of certain crystallized grains at the expense of others. This stage of recrystallization is called *grain growth*. The process develops spontaneously, provided that the temperature is sufficiently high, since grain growth results in a decrease of free energy of metal owing to a decrease of the surface energy (with lager grains, the total surface area of grain boundaries is smaller). Grain growth does not need to be preceded by recovery and recrystallization; it may occur in all polycrystalline materials, metals and ceramics alike.

Grain growth occurs by the passage of atoms from one grain to a neighbouring grain through the boundary; in this process some grains diminish in size and even disappear, while others become large and devour adjacent grains. Grain growth is accelerated with an increase of temperature. The higher the temperature of heating, the larger the recrystallized grains.

For many polycrystalline materials, the grain diameter d varies with time t according to the relationship

$$d^n - d_0^n = Kt \tag{6}$$

Where d_0 is the initial grain diameter at time $t = 0$, and K and n are time-independent constants; the value of n is generally equal to or greater than 2.

At room temperature, the mechanical properties of a fine-grained metal are usually superior (i.e., higher strength and toughness) to those of coarse-grained ones. If the grain-structure of a single-phase alloy is coarser than the desired, refinement may be accomplished by plastically deforming the material and then subjecting it to a recrystallization heat treatment.

In a polycrystalline metal the yield stress (σ_y) increases with decrease in grain size; as per the following relation:

$$\sigma_y = \sigma_i + kd^{-1/2} \tag{7}$$

where σ_i is yield stress for a crystal having no grain boundaries, d is grain diameter and k is Hall-Petch constant (0.71 $\text{MNm}^{-3/2}$ for BCC Fe, 0.11 $\text{MNm}^{-3/2}$ for FCC Cu and 0.068 $\text{MNm}^{-3/2}$ for FCC Al).

Grain growth depends upon a number of factors, e.g. *(i)* annealing temperature and time, *(ii)* degree of previous cold work, *(iii)* effect of insoluble mixtures to the metal *(iv)* addition of some alloying elements, *(v)* rate of heating and cooling, etc.

Grain size is affected largely by the fine dispersion of second phase particles. They restrict the movement of grain boundary.

16. HOT-WORKING

When metals are worked above recrystallization temperature (or 0.7 T_{mp} approximately), they recrystallize while being deformed and they remain soft. Metal billets are stressed and given massive deformation. This process is called *hot working*. During this process the metal is in a plastic state and it readily takes the desired shape. We may note that the cold working process is carried out below the recrystallization temperature whereas hot working proceses are carried out at temperatures above the recrystallization range. Usually, hot working proceses are at a given compressive stress, e.g. as in rolling, forging extrusion, etc. The speed of hot working operation at a given temperature can be calculated by considering the amount of time required for recrystallization. Operation has to be performed rapidly at higher temperature.

The upper limit of hot working is the temperature at which metal begins to melt and usually it is about 40°C below the melting point of the material. In comparison to metals, alloys are more difficult to work. In hot working hardening due to deformation and softening due to annealing occur simultaneously. Most hot working operations are done in several *stages* or *passes*. The temperature in intermediate pass is kept well above the minimum hot working temperature so that energy in deforming metal may be saved. The working temperature in last pass is just above the minimum hot working temperature. It is just enough to ensure that grain growth during cooling is negligible. To have, finally the fine grain size structure, the working temperature in the last pass is lowered. This *finishing temperature* is usually above the minimum recrystallization temperature. The amount of deformation in the last pass has to be large for the recrystallization grain size. We may note that high rates of deformation are undesirable as they inherit the characteristics of the *cold working* process.

The common hot working processes are rolling, hot drawing, hot spinning, hot extruding, forging welding, hot piercing, etc.

Advantages of Hot-Working Processes

The important advantages of hot working processes are as follows:

(i) Power required is less and there is less danger of cracking the metal.
(ii) Due to the pressing together of cavities, cracks, and blow holes by the pressure used in the hot working, porosity in the metal is eliminated to a greater extent.
(iii) Grain refinement is possible, as coarse grains of casting are broken up and refined into smaller recrystallized grains.
(iv) Intermediate annealing is eliminated. Impurities are squeezed into fibres and distributed throughout the mass.
(v) Desirable directional properties are obtained from a fibre structure due to directional flow obtained in the hot working.
(vi) Relatively large deformation is possible without producing a fracture as compared to cold working.
(vii) Mechanical properties like strength, ductility, elongation percentage and impact resistance are improved in hot working due to grain refinement.
(viii) Hot working is economical and also rapid for forming all commercial metals.

The limitations of hot working are:

(i) At high temperatures, due to rapid oxidation and scaling of the surface of the metal, a poor surface finish of the final product is obtained.
(ii) Maintenance of temperature is usually difficult particularly for light sections as thin sheets and fine fibres cool very rapidly, i.e., they cannot be maintained at hot working range.
(iii) Due to poor surface texture, it is difficult to maintain close dimensional tolerances.
(iv) The metals which are deficient in plasticity cannot be hot worked. Such materials are termed as *Hot Short*.
(v) Hot working is usually performed in air, and hence lot of metal is lost in oxidation.
(vi) Hot working process is costly because the equipments used to withstand high-temperature operations, e.g., rolling, drop forging, etc. are very costly. Moreover, the life of the tools used in this process is reduced due to their use at high temperature.

Example 1 *An aluminium crystal slips on (111) plane and in the (110) direction, with a 3.5 MPa stress applied in the* $(1\bar{1}1)$ *direction. Determine the value of critical resolved shear stress.* [B.E., M.Sc.]

Solution We have, Miller indices of

Slip plane = (111), slip direction = (110)

and stress plane = $(1\bar{1}1)$

The angle ϕ between two planes of Miller indices ($h_1\,k_1\,l_1$) and ($h_2\,k_2\,l_2$) is given by:

$$\cos\ \phi = \frac{h_1h_2 + k_1k_2 + l_1l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2}\ \sqrt{h_2^2 + k_2^2 + l_2^2}}$$

∴ Angle α between stress plane and slip plane is given by

$$\cos\ \alpha = \frac{1\times1-1\times1+1\times1}{\sqrt{1^2+1^2+1^2}\ \sqrt{1^2+1^2+1^2}} = \frac{1}{3}$$

$$\therefore \qquad \sin\ \alpha = \sqrt{1-\frac{1}{9}} = \frac{2\sqrt{2}}{3}$$

Angle θ, between stress direction and slip direction is given by

$$\cos\ \theta = \frac{1\times1+1\times1+1\times0}{\sqrt{1^2+1^2+1^2}\ \sqrt{1^2+1^2+0^2}} = \frac{\sqrt{2}}{\sqrt{3}}$$

∴ Critical resolved shear stress

$$\tau_{cr} = \sigma \sin\ \alpha \cos\ \alpha \cos\ \theta \qquad \text{[Given } \sigma = 3.5 \text{ MPa]}$$

$$= 3.5 \times \frac{2\sqrt{2}}{3} \times \frac{1}{3} \times \sqrt{\frac{2}{3}} = 0.899 \text{ Mpa}$$

Example 2 *The axis of a cylindrical crystal of copper lies in the (111) plane. A tensile stress σ is applied along the axis of the crystal. Show that the resolved shear stress τ in any direction on this plane is zero. What inference will you draw from this?* [BE]

Solution We have

$$\tau_R = \sigma \cos\ \phi_1 \cos\ \phi_2 \qquad (1)$$

Here, the angle ϕ_1 = 90° for the plane (111) and hence cos ϕ_1 = 0. Obviously, from Eq. (1), the resolved shear stress τ_R for any value of ϕ_2 is zero. Since all planes in the family of {111} are not parallel to each other and hence we cannot conclude that the crystal will not deform plastically. There are planes other than (111) and definitely they will have a finite resolved shear. As soon as this will attain the critical value, plastic deformation will be initiated.

Example 3 *Find the yield stress for a polycrystalline alloy when the grain size is ASTM 8 (grain diameter = 0.002 mm). Assume σ_i = 80 MN m^{-2} and K = 0.63 MN $m^{-3/2}$.* [BE, AMIE]

Solution We have

$$\sigma_y = \sigma_i + Kd^{-1/2}$$
$$= 80 + 0.63 \times (0.22 \times 10^{-3})^{-1/2}$$
$$= 80 + 0.63 \times 213.2 \text{ MN m}^{-2}$$
$$= 214.3 \text{ MN m}^{-2}$$

Example 4 *The yield strength of a polycrystalline material increases from 120 MN m^{-2} to 220 MN m^{-2} on decreasing the grain diameter from 0.04 mm to 0.01 mm. Calculate the yield stress for a grain size ASTM 9.* [BE, AMIE]

Solution

$$\sigma_y = \sigma_i + Kd^{-1/2} \qquad (1)$$

Now, substituting the two yield stresses and the two grain diameters in Eq. (1), one obtains

$$120 = \sigma_i + K\,(0.04 \times 10^{-3})^{-1/2} \tag{2}$$

$$220 = \sigma_i + K\,(0.01 \times 10^{-3})^{-1/2} \tag{3}$$

Solving (2) and (3), we obtain

$$\sigma_i = 20 \text{ MN m}^{-2} \text{ and } K = 0.633 \text{ MN } m^{-3/2}$$

Grain diameter for ASTM 9 is

$$d = \frac{1}{\sqrt{10^4 \times 256/645}} = 0.0159 \text{ mm}$$

∴ Yield stress for this grain size is

$$\sigma_y = 20 + 0.633 \times (0.0159 \times 10^{-3})^{-½}$$
$$= 179 \text{ MN m}^{-2}$$

SUGGESTED READINGS

1. D. Hull, '*Introduction to Dislocations*', Butterworth-Heinemann, Woburn, UK (1984).
2. J.P. Hirth and J. Lothe, '*Theory of Dislocations*' 2nd ed., Wiley-Interscience, New York (1982).
3. J. Weertman and J.R. Weertman, '*Elementary Dislocation Theory*', Oxford University Press, New York (1992).

REVIEW QUESTIONS

1. Explain deformation of metals. How deformation take place? Mention its effect. [BE]

2. What are elastic and plastic deformations? Compare elastic and plastic materials when they are subjected to tensile loads. Mention difference between elastic and plastic deformation. [BE]

3. Explain the terms slip and twinning. How does slip occur? Explain slip directions and slip planes with diagrams. [AMIE, M.Sc.]

4. Explain the difference between slipping and twinning. How does twinning occur in metals? Name and explain two types of twins. [AMIE]

5. Describe the process of slip. How deformation take place by slip? What are slip systems and easy slip directions. [BE]

6. What is slip? On what crystallographic planes and in what directions it is likely to occur in BCC, FCC and HCP metals? [BE]

7. Explain different types of twins. How slip lines be distinguished from twin boundaries? [BE]

8. Explain the mechanism for dislocation in plastic deformation with neat sketches. [AMIE]

9. What do you understand by stress field around dislocation? Discuss the mechanism of parallel edge dislocations. [BE]

10. What is critical resolved shear stress? On what factors does it depend? [AMIE]

11. What is dislocation reaction and how it can be represented? Also discuss the decomposition of a perfect dislocation. How sessile dislocation block the movement of other dislocations?

12. *(a)* What is effect of cold working on the mechanical properties of metals and alloys?
(b) Discuss the change in properties when a severely cold-worked metal is annealed at successively higher temperatures. [B.E., AMIE]

13. What are the common changes that occur due to cold working of a metal? How can we restore the pre-cold properties? Specify the process of restoration. When does grain growth occur? [BHU, BE]

14. *(a)* What is dislocation and how it is formed *(b)* What is Bauchinger's effect? Explain, its effect in deformation of metals?

(c) Differentiate between the hot working and cold working. Explain the effect of each of them on material properties. [AMIE]

15. Distinguish between the term 'recovery' and the 'recrystallization' involved in the process of heating cold-worked metals. [AMIE]

16. How do grain boundaries influence recrystallization? Discuss the role of grain boundaries in imparting various properties to materials.

17. *(a)* Define Burger's vector and illustrate it on the sketch of an edge dislocation.

(b) What is the significance of dislocations in the plastic deformation of metals? [AMIE]

18. What are low angle grain boundaries? How they are produced?

19. What is hot and cold working? How do they differ? Explain their effects on the properties of materials. [AMIE]

20. What do you understand by preferred orientation? How does it affect the properties of a material? How will you detect that preferred orientation is present or not? [BE]

21. Describe the three stages of annealing of cold-worked metal. Discuss the mechanisms of recovery and recrystallization. [BE]

22. Differentiate primary and secondary recrystallization. Describe the factors which control recrystalliztion. [BE]

23. Explain the yield point phenomenon in materials in terms of dislocation. [AMIE]

24. Explain Bauschinger effect as applied to the deformation of metals. [AMIE]

25. Write short notes on the following:

(i) Season cracking of brass
(ii) Recrystallization
(iii) Dislocation theory
(iv) Point and line defects
(v) Grain growth
(vi) Grain boundary defects
(vii) Preferred orientation
(viii) Critical resolved shear stress
(ix) Burger vector
(x) Polygonisation
(xi) Strain ageing
(xii) Elastic and plastic deformation in steel
(xiii) Stacking faults in crystal
(xiv) Slip and twinning
(xv) Bauschinger effect. [AMIE, M.Sc.]

26. Distinguish between

(i) Slip and twinning
(ii) Elastic deformation and elastic after effect
(iii) Edge and screw dislocation
(iv) Elastic and plastic deformation
(v) Hot and cold working
(vi) Recovery and recrystallization
(viii) Shear stress and critical resolved shear stress [AMIE, BE]

PROBLEMS

1. Show that the resolved shear stress reaches a maximum value, when $\phi_1 = \phi_2 = 45°$ [BE]

2. In an FCC crystal, the tensile axis is along the [12 *T*] direction. Determine the slip system that is most favourably oriented for slip to start at the lowest applied tensile stress.

[*Ans.* (111) $[01\bar{1}]$ or $(\bar{1}1\bar{1})$ [110] [BE, M.Sc.]]

3. A grain size of ASTM 7 in a mild steel is refined to ASTM 14 by the addition of micro alloying elements. Estimate the yield strength of the steel. [*Ans.* 300 MN m^{-2}] [BE]

4. Compute the mean grain diameters in mm, corresponding to ASTM grain size numbers 3.

[*Ans.* 0.125 mm] [AMIE]

5. Show that the average grain diameter decreases by half, for every increase of 2 in the ASTM number for grain size. [M.Sc., BE]

OBJECTIVE QUESTIONS

1. Burger's vector of an edge dislocation is
(1) parallel to dislocation line
(2) perpendicular to dislocation line
(3) at any angle with dislocation line including 0 and 90°
(4) none of the above (2)

2. Yield point in mild steel occurs because
(1) carbon atom occupy the vacancy line just below the dislocation and anchor dislocation by reducing its energy.
(2) slip is promoted by carbon atoms to take place in several parallel plans simultaneously.
(3) carbon atoms form Fe_3C with iron
(4) none of the above (1)

3. If ASTM grain size number 7 corresponds to grain diameter of 0.03 mm, ASTM 4 corresponds to
(1) 0.36 mm (2) 0.085 mm (3) 0.055 mm (4) 0.008 mm (2)

4. The mean grain diameter corresponding to ASTM number = –0.5 is
(1) 0.87 mm (2) 0.43 mm (3) 0.24 mm (4) 0.13 mm (2)

5. Strain ageing in mild steel causes
(1) material to behave in brittle manner
(2) material to lose its ultimate tensile strength
(3) yield point to recur in alternate loading beyond yielding and unloading
(4) none of the above

6. The yield strength σ_y, internal frictional stress σ_i and diameters of grain d are related as
(1) $\sigma_y = \sigma_i + Kd^{-1/2}$ (2) $\sigma_y = \sigma_i + Kd^{-1}$
(3) $\sigma_y = \sigma_i - Kd^2$ (4) $\sigma_y = \sigma_i - Kd$ (1)

7. The minimum stress necessary to introduce yielding occurs when a single crystal is oriented such that

$\phi = \lambda = 45°$ in $\sigma_y = \dfrac{T_{crs}}{(\cos\phi\cos\lambda)_{max}}$, under these conditions

(1) $\sigma_y = T_{crs}$ (2) $\sigma_y = 2T_{crs}$
(3) $\sigma_y = T_{crs/2}$ (4) $\sigma_y = 4T_{crs}$ (2)

8. If ϕ represent the angle between the normal to the slip plane and the applied stress direction, and λ the angle between the slip and stress directions, then the resolved shear stress τ_R is
(1) $\tau_R = \sigma \sin\phi \sin\lambda$ (2) $\tau_R = \sigma \cos\phi \sin\lambda$
(3) $\tau_R = \sigma \cos\phi \cos\lambda$ (4) $\tau_R = \sigma \sin\phi \cos\lambda$ (3)

9. Recrystallization temperature is the temperature at which recrystalliztion just reaches completion in
(1) 3 hr (2) 2 hr (3) 1 hr (4) 1/2 hr (3)

10. The recrystallization temperature for brass alloy is about
(1) 1350°C (2) 770°C (3) 660°C (4) 450°C (4)

11. If T_m is absolute melting temperature, then the recrystallization temperature for pure metals in terms of T_m is normally
(1) 0.7 T_m (2) 0.9 T_m (3) 0.5 T_m (4) 0.3 T_m (4)

SHORT QUESTION ANSWERS

1. What is plastic deformation?

Ans. On a microscopic level, this corresponds to the motion of dislocations in response to an externally applied shear stress, a process termed slip.

2. What a slip system represents?

Ans. A slip system represents a slip-plane slip direction combination, and operable slip systems depend on the crystal structure of the material.

3. What is critical resolved shear stress?

Ans. It is the minimum shear stress required to initiate dislocation motion.

4. On what factors the yield strength of a single crystal depends?

(i) Magnitude of critical resolved shear stress.

and *(ii)* Orientation of slip components relative to the direction of the applied stress.

5. Can mechanical twinning cause limited plastic deformation under some circumstances in BCC and HCP metals?

Ans. Yes.

6. What is the role of grain boundaries?

Ans. Grain boundaries serve as barriers to dislocation motion; thus refining the grain size of a polycrystalline material renders it harder and stronger.

7. How the microstructural and mechanical characteristics of a plastically deformed metal specimen may be restored to their predeformed states?

Ans. By an appropriate heat treatment, during which recovery, recrystallization, and grain growth processes are allowed to occur.

8. What happens during recovery?

Ans. During recovery there is a reduction in dislocation density and alterations in dislocation configurations.

9. What do you understand by recrystallization?

Ans. This is the formation of a new set of grains that are strain free; in addition, the material becomes softer and more ductile.

10. What is grain growth?

Ans. This is the increase in average grain size of polycrystalline materials, which proceeds by grain boundary motion.

11. What is work hardening?

Ans. It is due to the formation of sessile (immobile) dislocations by the interaction of dislocations moving on non-parallel slip planes. The plastic flow strength increases during work hardening as the square root of the dislocation density.

12

Oxidation and Corrosion

1. INTRODUCTION

Most materials, to one degree or another, experience some type of interaction with a large number of diverse environments. Often, such interactions with environments impair a material's usefulness as a result of the deterioration of its mechanical properties, e.g. ductility and strength, other physical properties, or appearance.

Deteriorative mechanisms are found to be different for the three types of materials. In metals, there is actual material loss either by dissolution (*corrosion*) or by the formation of non metallic scale or film (*oxidation*). Ceramic materials are found to be relatively resistant to deterioration, which usually occurs at elevated temperatures or in rather extreme environments. The process is frequently also referred as *corrosion.* In case of polymers, mechanisms and consequences differ from those for metals and ceramics, and usually referred as *degradation.* Polymers may dissolve when exposed to a liquid solvent, or they may absorb the solvent and swell; also, electromagnetic radiation (primarily ultraviolet) and heat may cause alternations in their molecular structure.

The following types of corrosions are significant: *(i) Direct corrosion*, *(ii) Electrochemical corrosion*, *(iii) Galvanic corrosion*, and *(iv) High temperature oxidation corrosion*.

2. CORROSION-RESISTANT MATERIALS

Structural materials intended for operation in aggressive media must possess a high corrosion resistance, as well as the desired mechanical properties. The most corrosion-susceptible materials are metals and alloys, which can be explained by their high chemical activity and high electric conductivity.

Corrosion of metals is defined as the destructive and unintentional attack of a metal. It is electrochemical and ordinarily begins at the surface. The consequences of corrosion are all too common. Familiar examples of corrosion of metals include the rusting of automotive body panels and radiator and exhaust components.

Corrosion of metals is their spontaneous destruction due to chemical or electrochemical interaction with the surrounding medium. Corrosion-resistant metals and alloys are those which can withstand the corrosive action of a medium, i.e., corrosion processes proceed in them at a relatively low rate. Usually, it is distinguished between two principal kinds of corrosion: electrochemical and chemical. The following factors influence corrosion: *(i) internal structure*, *(ii) chemical nature*, *(iii) environment* or working conditions, e.g., stresses, temperature, or concentration, *(iv) surface film*, *(v)* presence of dirt, dust or any foreign matter, *(vi)* distribution of secondary phases, *(vii)* blow holes, inclusions and trapped gases, *(viii)* eddy electric currents, and (ix) nature of engineering application.

3. ELECTROCHEMICAL CORROSION

This kind of corrosion develops in liquid electrolytes: moist atmosphere and soil, salt and fresh water, aqueous solutions of salts, alkalies and acids. Electrochemical corrosion is characterized by the appearance of an electric current and dissolution of the metal due to its electrochemical interaction with the electrolyte. There is a transfer of electrons from one chemical species to another. Metal atoms characteristically lose or give up electrons in what is called an *oxidation* reaction. The surface of a metal is electrochemically inhomogeneous so that when the metal is placed into an electrolyte, this leads to the formation of *microgalvanic corrosion cells* (Fig. 12.1). In some portions of the surface, which are called *anodes*, there occurs reactions of the type (1) and other portions, called *cathodes*, of the type (2) and (3)

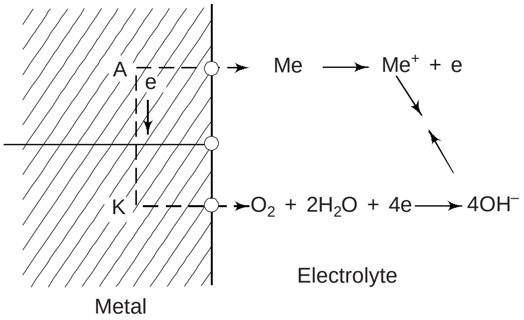


Fig. 12.1 Scheme of a corrosion cell

$$Me \rightleftharpoons Me^{n+} + ne^{-1} \tag{1}$$

$$O_2 + 2H_2O + 4e^- \rightleftharpoons 4(OH)^- \tag{2}$$

$$2H^+ + 2e^- \rightleftharpoons H_2 \tag{3}$$

Here Me represents the hypothetical metal. Examples in which metals oxidize are

$$Fe \rightarrow Fe^{2+} + 2e^- \tag{4}$$

$$Al \rightarrow Al^{3+} + 3e^- \tag{5}$$

The site at which oxidation takes place is called the *anode* and oxidation is sometimes called an *anodic reaction.*

The electrons generated from each metal atom that is oxidized must be transferred to and become a part of another chemical species in what is termed a *reduction* reaction (eq. 3). Other reduction reactions are possible, depending on the nature of the solution to which the metal is exposed. For an acid solution having dissolved oxygen, the reduction reaction

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O \tag{6}$$

may probably occur. For a neural or basic aqueous solution in which oxygen is also dissolved, reduction reaction given by (2) will take place. Any metal ions present in the solution may also be reduced. For ions that can exist in more than one valence state (multivalent ions), reduction may occur in accordance by

$$Me^{n+} + e^- \rightarrow M^{(n\ +\ 1)+} \tag{7}$$

in which the metal ion decreases its valence state by accepting an electron. It is also possible that a metal may be totally reduced from an ionic to a neutral metallic state as

$$Me^{n+1} + ne^- \rightarrow Me \tag{8}$$

The location at which reduction occurs is called the *cathode*. We may note that two or more of the reduction reactions above may occur simultaneously.

As long as corrosion cell remains open, the reactions at the anode and the cathode portions proceed at the same rate in both directions, i.e. they are *reversible*.

An overall electrochemical reaction must consist of at least one oxidation and one reduction, and will be the sum of them. We may note that often the individual oxidation and reduction reactions are termed as *half-reactions*. However, there can be no net electrical charge accumulation from the electrons and ions, i.e., the total rate of oxidation must be equal to the total rate of reduction. Obviously, all electrons generated through oxidation must be consumed by reduction.

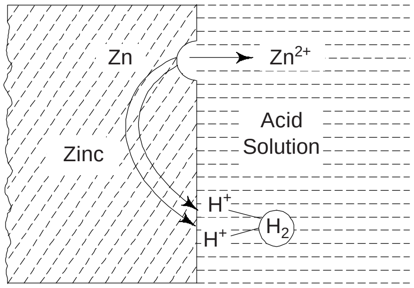


Fig. 12.2 Zinc in an acid solution. The electrochemical reactions with the corrosion of zinc

To make it clear, let us consider an example in which zinc metal is immersed in an acid solution containing H^+ ions. At some regions on the metal surface, zinc will experience oxidation or corrosion (Fig. 12.2),

$$Zn \rightarrow Zn^{2+} + 2e^- \tag{9}$$

Zinc, being metal, is a good electrical conductor. The electrons liberated in (9) may be transferred to an adjacent region at which the H^+ ions are reduced in accordance with

$$2H^+ + 2e^- \rightarrow H_2 \text{ (gas)} \tag{10}$$

when no other oxidation or reduction reactions occur, the total electrochemical reaction is just the sum of (9) and (10), i.e.

$$Zn + 2H^+ \rightarrow Zn^{2+} + H_2 \text{ (gas)} \tag{11}$$

As an another example, we can consider the oxidation or rusting of iron in water, which contains dissolved oxygen. This process completes in two steps. In the first step, Fe is oxided to Fe^{2+} [as $Fe(OH)_2$],

$$Fe + \frac{1}{2}O_2 + H_2O \rightarrow Fe^{2+} + 2OH^- \rightarrow Fe(OH)_2 \tag{12}$$

and in the second step, Fe^{3+} [as $Fe(OH)_3$] according to

$$2Fe\,(OH)_2 + \frac{1}{2}O_2 + H_2O \rightarrow 2Fe(OH)_3 \tag{13}$$

We may note that the compound $Fe\,(OH)_3$ is the all too familiar rust.

High-temperature oxidation of iron process is as follows:

$$2Fe + O_2 \rightarrow 2FeO$$

$$\underset{\text{(Metal)}}{M} + O_2 \rightarrow \text{Metal oxide}$$

or

$$Fe^{2+} + O^{2-} \rightarrow FeO \tag{14}$$

Silverware gets tarnished in the presence of the atmosphere by hydrogen sulphide. The reaction is as follows

$$Ag + H_2S \rightarrow AgS + H_2 \tag{15}$$

We may note that these reactions are common to all metals and the scales formed are uniform throughout the surface, i.e., direct attack proceeds uniformly over the entire surface of the metal.

As a consequence of oxidation, the metal ions may either go into the corroding solution as ions (reaction 9), or they may form an insoluble compound with non metallic elements (reaction 13).

Electrode Potentials

When the piece of iron rod (electrode) is immersed in polar solvent like water, some of the metal ions leave the crystal and go into solution. They get hydrated, i.e., they form bonds with the polar water molecules. As the metal continues to dissolve, more and more electrons are left back and a net negative charge builds up in metal. Similarly, the presence of excess positive ions builds up a positive charge in the adjacent solution. These opposite electrical layers discourage further dissolution of the metal in the solvent. (However, not all metallic materials oxidize to form ions with the same degree of ease). Obviously, a dynamic equilibrium is reached soon with no net flow of metal ions into the solution.

The potential developed by an electrode in equilibrium is a property of the metal forming the electrode. The difference in potential at the electrodes causes the reactions to occur. The electrode potential depends on the nature of the electrolyte and temperature. This potential is measured under standard conditions with a pure metal as the electrode and electrolyte containing unit concentration of ions of the same metal.

Let us consider the generalized reactions involving the oxidation of metal M_1 and the reduction of metal M_2 as

$$M_1 \rightarrow M_1^{n+} + ne^- - V_1^0 \tag{14}$$

$$M_2^{n+} + ne^- \rightarrow M_2 + V_2^0 \tag{15}$$

where $V^{0\prime}$ s are the standard potentials as taken from the standard emf series (Table 12.1).

Table 12.1 The standard emf series

	Electrode reaction	*Standard electrode Potential, V°(Volt)*
Noble	$Au^{3+} + 3e^- \rightarrow Au$	+1.420
↑	$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$	+1.229
	$Pt^{2+} + 2e^- \rightarrow Pt$	~+1.2
Increasing inert		
(Cathodic)	$Ag^+ + e^- \rightarrow Ag$	+0.800
	$Fe^{3+} + e^- \rightarrow Fe^{2+}$	+0.771
	$O_2 + 2H_2O + 4e^- \rightarrow 4(OH)^-$	+0.401
	$Cu^{2+} + 2e^- \rightarrow Cu$	+0.340
	$2H^+ + 2e^- \rightarrow H_2$	0.000
	$Pb^{2+} + 2e^- \rightarrow Pb$	–0.126
	$Sn^{2+} + 2e^- \rightarrow Sn$	–0.136
	$Ni^{2+} + 2e^- \rightarrow Ni$	–0.250
Increasingly active	$Co^{2+} + 2e^- \rightarrow Co$	–0.277
(anodic)	$Cd^{2+} + 2e^- \rightarrow Cd$	–0.403
↓	$Fe^{2+} + 2e^- \rightarrow Fe$	–0.440
	$Cr^{3+} + 3e^- \rightarrow Cr$	–0.744
	$Zn^{2+} + 2e^- \rightarrow Zn$	–0.763
	$Al^{3+} + 3e^- \rightarrow Al$	–1.662
	$Mg^{2+} + 2e^- \rightarrow Mg$	–2.363
	$Na^+ + e^- \rightarrow Na$	–2.714
	$K^+ + e^- \rightarrow K$	–2.924 Active

Since metal M_1 is oxidized, the sign of V_1^0 is opposite to that as it appears in Table 12.1. Addition of Eqs. (14) and (15) yields

$$M_1 + M_2^{n+} \rightarrow M_1^{n+} + M_2 \tag{16}$$

and the overall cell potential $\Delta V°$ is

$$\Delta V° = V_2^0 - V_1^0 \tag{17}$$

For this reaction to occur spontaneously, ΔV^0 must be positive. If it is negative, the spontaneous cell direction is just the reverse of Eq. (16). In a closed corrosion cell, the rates of the direct and reverse reactions become different. The reaction at the anode occurs preferably in the direction of metal ionization and that at the cathode, in the direction of reduction. Thus corrosion current appears as a result motions of electrons in the metal and of ions in the electrolyte. Under the effect of current, *irreversible electrode potentials*, (V_{ir}) are established at the anode and cathode. Compared with the reversible potentials (V_r), they are less negative for the anode and less positive for the cathode.

The difference between the V_r and V_{ir} is proportional to the current in a corrosion cell. The proportionality factors P_A and P_C are called the *polarizabilities*:

$$V_r^A - V_{ir}^A = P_A I \tag{18}$$

$$V_r^C - V_{ir}^C = P_C I \tag{19}$$

where P_A and P_C are the anode and cathode polarizabilities. Some metals have a high anode polarizability, P_A under particular conditions and are called *passivating metals*.

The corrosion current I that appears in a corrosion cell and determines the rate of corrosion is expressed by the relation

$$I = V_r^C - V_r^A /(R + P_A + P_C) \tag{20}$$

where R is the Ohmic resistance of the corrosion cell. This relation is used for qualitative estimation and finding of a particular factor that determines the corrosion resistance of a metal.

Influence of Concentration and Temperature on Cell Potential

We again consider the electrochemical reaction described by Eq. (16). If M_1 and M_2 electrodes are pure metals, the cell potential depends on the absolute temperature T and the molar ion concentrations $[M_1^{n+}]$ and $[M_2^{n+}]$ according to Nernst relation:

$$\Delta V = (V_2^0 - V_1^0) - \frac{RT}{nF}\ln\left[\frac{M_1^{n+}}{M_2^{n+}}\right] \tag{21}$$

where R is the gas constant, n is the number of electrons participating in either of the half cell reaction, and F (= 96,500 C/mol) is the Faraday constant—the magnitude of charge per mole (6.023×10^{23}) of electrons. At 25°C (~ room temperature),

$$\Delta V = (V_2^0 - V_1^0) - \frac{0.0592}{n}\log\left[\frac{M_1^{n+}}{M_2^{n+}}\right] \tag{22}$$

to give ΔV in volts. Again, for reaction, ΔV must be positive. As expected, for 1 M concentrations of ion types (i.e., $[M_1^{n+}] = [M_2^{n+}] = 1$), Eq. (21) simplifies to Eq. (17). If $M_2^{n+} < 1$, i.e., if the electrolyte is deficit in metal ions, the potential decreases and goes more towards the active end of Table 12.1. It the electrode is in the alloyed condition, $M_2^{n+} < 1$ and then the potential increases and goes towards the noble end.

With metals and alloys of commercial purity, their electrode potential V_r^A is determined by their composition. The corrosion stability is higher with a smaller difference between V_r^C of the probable cathode reaction and V_r^A of the process under the particular conditions of collision. This is why that the electropositive metals have a high corrosion resistance. The composition of the metal whose surface is being corroded determines the polarization of the anode and cathode, P_A and P_C. Both characteristics decrease with an increase of the surface area of the anode and cathode.

The phase composition and structure of alloys of the same chemical composition can influence all the electrochemical parameters mentioned. The possibility of changing the structure of a metal by heat treatment should be considered as an additional means for improving the resistance of metals and alloys to electrochemical corrosion. In chemical corrosion, this possibility is almost unavailable.

Alloys with a single phase-structure are usually more corrosion-resistant than those with multi-phase structures, since the anodes in them are various faults of the crystal structure: high-and low-angle boundaries and dislocations which have a rather low electrode potential. Only anode portions of the structure are destroyed by corrosion.

The corrosion resistance of metals can also be influenced by films of corrosion products formed on the exposed surface of a metal. The electrolyte can dissolve these films and thus impair the corrosion resistance of a metal. The solubilities of these films on various metals are different in various electrolytes.

Electrochemical corrosion is subdivided into several kinds: atmospheric, sea-water, soil, acid, and alkaline corrosion, depending on the conditions under which the corrosion process develops. By the nature of destruction, it is distinguished between uniform and local (isolated) corrosion. Besides, local corrosion is subdivided into the following kinds: galvanic (two-metal) corrosion, cavitation corrosion, intercrystalline (intergranular) corrosion, pitting (point) corrosion, corrosion fatigue, corrosion cracking.

4. GALVANIC (TWO-METAL) CORROSION

A reaction takes place when a metal comes in contact with an electrolyte, releasing the metal ions into the electrolyte. However, very soon an equilibrium is established and the reaction does not proceed further. Now, if a dissimilar metal forms a second electrode, there will be a similar equilibrium at this electrode also. Now, if the circuit is completed allowing for the flow of current, the equilibrium is disturbed. Let us consider Copper (Cu) as the first electrode, Zinc (Zn) as the second electrode (Fig. 12.3), and water as the electrolyte. This arrangement is called as *galvanic cell*.

If the current continues to flow, a less noble metal Zn (in comparison to Cu) will gradually dissolve in the electrolyte or in other words *corrode*. Zn is the anode of the galvanic cell. The more noble metal copper will act as cathode. The reactions that occur at the two electrodes are as follows:

$$\text{Zn} \rightarrow \text{Zn}^{2+} + 2e^- \text{ (anode reaction)} \qquad (23)$$

Copper will deposit on the cathode through the following reduction reaction

$$\text{Cu}^{2+} + 2e^- \rightarrow \text{Cu} \qquad (24)$$

If cupric ions are not available in the electrolyte, the reduction at the cathode may still proceed through other reactions. Hydrogen ion may get reduced, resulting in the evolution of hydrogen at the cathode.

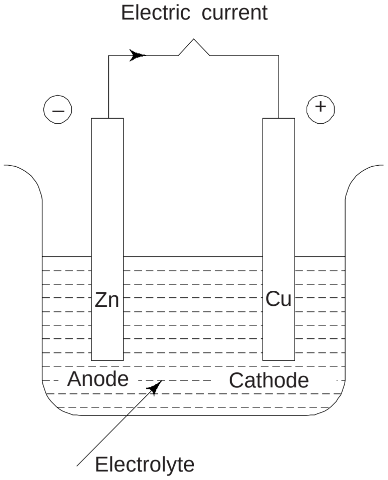


Fig. 12.3 Galavanic corrosion

$$2\text{H}^+ + 2e^- \rightarrow \text{H}_2 \text{ (cathode reaction)} \qquad (25)$$

Alternatively, the reduction may take place through oxygen and water combining with the excess electrons to produce hydroxyl ions:

$$\text{O}_2 + 2\text{H}_2\text{O} + 4e^- \rightarrow 4\text{OH}^- \qquad (26)$$

A galvanic cell may be set up due to the differences in concentration of the metal ion in the electrolyte (Fig. 12.3). The potential difference and the rate of the anodic metal will depend upon the relative positions of the two metals in the *galvanic series* (Table 12.2). A galvanic cell set up due to such differences in metal ion concentration is also called as *concentration cell.*

Table 12.2 Galvanic series

Increasingly inert (cathodic)	Platinum
	Gold
	Graphite
	Titanium
	Silver
	316 Stainless steel (passive) 304 Stainless steel (passive)
	Inconel (80 Ni-13 Cr-7Fe) (passive) Nickel (passive)
	Monel (70 Ni-30 Cu) Copper-Nickel alloys Bronzes (Cu-Sn alloys) Copper Brasses (Cu-Zn alloys)
	Inconel (active) Nickel (active)
	Tin
	Lead

(*Contd.*)

Table 12.2 (*Contd.*)

Increasingly active (anodic) ↓	316 Stainless steel (active)
	304 Stainless steel (active)
	Cast iron
	Iron and steel
	Aluminium alloys
	Cadmium
	Aluminium (commercially pure)
	Zinc
	Magnesium and alloys of magnesium

Corrosion of dissimilar metals in contact with each other like the galvanic cell is rapid. If we can avoid direct contact between dissimilar metals, we can avoid the corrosion of anodic metal. That is why, in fabrication work, one prefers the composition of electrodes in case of welding and of rivets in rivetting should be of the same chemical composition as the base metal. Galvanic corrosion also found to occur on the unprotected steel of ships due to the selective action of steel on the brass screws. An underground pipe line that passes through impervious clays in some regions and through porous sands in some other regions may corrode in clay region.

A galvanic cell may also be formed due to different residual stresses in the same metal. The stressed region is more active and is anodic with respect to a stress free region. Such types of galvanic cells are called *stress cells*. Such types of cells can be formed between regions of different dislocation density in a cold worked metal or in a polycrystalline metal, where the grain boundaries are anodic to the interior of the grains. Interestingly, a bent wire is likely to corrode at the bend, where it has been plastically deformed.

The Galvanic Series

Table 12.2 represents the relative reactivities of number of metals and commercial alloys in seawater (galvanic series). The alloys near the top of the table are cathodic and unreactive, where those at the bottom of the table are most anodic; no voltages are provided. Interestingly, a comparison of the standard emf and the galvanic series reveals a high degree of correspondence between the most relative positions of pure base metals.

In a wide variety of environments, most metals and alloys are subject to oxidation or corrosion to one degree or another, i.e., they are more stable in an ionic state than as metals. In going from metallic to oxidized states, thermodynamically there is a net decrease in free energy. This is why, essentially all metals occur in nature as compounds, e.g. oxides, hydroxides, carbonates, silicates, sulfides and sulphates. There are two notable exceptions; gold and platinum. For these two metals, oxidation in most environments is not favourable, and therefore, both of these two metals may exist in nature in the metallic state.

In order to reduce galvanic corrosion effects following measures may be taken:

(i) If coupling of dissimilar metals is necessary, one may choose two that are close together in the galvanic series.

(ii) One should use anode area as large as possible and avoid an unfavourable anode-to-cathode surface area ratio.

(iii) One should insulate electrically metals from each other.

(iv) Connect electrically third metal to the other two. This is known as cathodic protection.

5. CORROSION RATES

The rate of material removal as a consequence of the chemical action, or the corrosion rate, is an important parameter. One can express this as the *corrosion penetration rate* (CPR), or the thickness loss of material per unit time. It is expressed as

$$\text{CPR} = \frac{KW}{\rho A t} \tag{27}$$

Where W is the weight loss after exposure time t; ρ and A represent the density and exposed specimen area, respectively, and K is a constant, its magnitude depends on the system of units used. For convenience, CPR is expressed in terms of either mils per year (mpy) or millimeters per year (mm/yr). In the first case, $K = 534$ to give CPR in mpy (where 1 mil = 0.001 inch), and W, ρ, A and t are specified in units of milligrams, gms/cc, square inches and hours respectively. In the second case, $K = 87.6$ for mm/yr, and units for other parameters are the same as for mils per year, except that A is expressed in square centimeters. A corrosion penetration rate less than about 20 mpy (0.50 mm/yr) is acceptable for most applications.

As stated earlier, there is an electric current associated with electrochemical corrosion reactions. One can also express corrosion rate in terms of this current, or more specifically, current density i.e., the current per unit surface area of material corroding—which deginated by i. The rate r, in units of mol/m^2-s, can be determined with the help of the following expression

$$r = \frac{i}{nF} \tag{28}$$

where, again, n is the number of electrons associated with the ionization of each metal atom, and F is 96,500 C/mol.

6. HIGH TEMPERATURE OXIDATION OR DRY CORROSION

Ferrous alloys get rusted at high temperatures forms scales and oxides. One can prevent this corrosion by adding alloying elements which are resistant to oxidation in some metals and also have good properties of high temperature service.

When liquid metals flow past other metals, high temperature corrosion also occurs. Essentially, a corrosion reaction is a process of mass transfer and is not dependent upon local cell potentials for its driving force. Truly speaking, the corrosion is due to the tendency of the solid to dissolve in the liquid metal upto the solubility limit at a given temperature.

We may note that the liquid-metal attack may either form a simple solution of the solid metal, a chemical compound, or be the selective extracton of one of the component metals in a solid alloy. This occurs when there is a concentration gradient or temperature gradient within the solid-liquid system. Serious damage by liquid-metal attack has been reported in heat exchangers carrying (Bi and Na) liquid coolants. A portion of the solid goes into solution in the liquid as the solid container usually copper tubing, approaches equilibrium with the liquid-metal coolant in the hot zone of heat exchanger. The solids try to deposit on the walls of the heat exchanger tubes, as the liquid moves to the cooler part of the heat exchanger. Obviously, the hot zone is continually corroded and the cold zone becomes plugged with the deposited corrosion products.

One can prolong the life of heat exchangers by the addition of certain inhibitors to the liquid alloy to form protective films to prevent high temperature corrosion.

7. PASSIVITY

There are some normally active metals and alloys, which lose their chemical reactivity and become extremely inert under particular environmental conditions. This phenomenon is called as *passivity* and it is displayed by chromium, iron, nickel, titanium and several of their alloys. This passive behaviour results from the formation of a highly adherent and very thin oxide film on the metal surface. This film serves as a protective barrier to further corrosion. As a result of passivation, stainless steels are highly resistant to corrosion in a rather wide variety of atmospheres. They contain at least 11% chromium, which as a solid-solution alloying element in iron, minimizes the formation of rust; instead, a protective surface film forms in oxidizing atmospheres. We may note that stainless steels are susceptible to corrosion in some environments, and therefore are not always “stainless”.

On increasing the potential of a metal electrode, the current density increases at first. When the current density attains a critical value, it may abruptly fall to a much lower value and remain more or less constant for some further increase in potential. This phenomenon is usually known as *passivation*. For chromium, the critical current density just before passivation is about 200 A/m^2. The current after passivation is less than 0.1 A/m^2. Obviously, there is a big drop in current density. This big drop in current density is associated with the simultaneous formation of a thin oxide layer on the metal surface. This phenomenon of passivation affords an important means of corrosion prevention.

Aluminium is highly corrosion resistant in many environments because it also passivates. If damaged, the protective film usually reforms very rapidly. However, a change in the character of the environment, e.g. alteration in the concentration of the active corrosive species, may cause a passivated material to revert to an active state. Subsequent damage to a preexisting passive film could result.

8. ENVIRONMENTAL EFFECTS

The variables in the corrosion environment, e.g. fluid velocity, temperature and composition can affect the corrosion properties of the materials that are in contact with it. It is observed that in most instances, increasing fluid velocity enhances the rate of corrosion due to erosive effects. In majority of corrosion cases, the rates of most chemical reactions rise with increasing temperature. Increasing the concentration of the corrosive species, e.g. H^+ ions in acids, in many cases produces a more rapid rate of corrosion. However, for materials, which are capable of passivation, raising the corrosive content may result in an active-to-passive transition, with a considerable reduction in corrosion.

Cold working or plastically ductile metals is generally used to increase their strength. However, a cold-worked metal is more susceptible to corrosion than the same material in an annealed state, e.g. deformation proceses are used to shape the head and point of a nail. Obviously, these positions are anodic with respect to the shank region. We may note that differential cold working on a structure should be a consideration when a corrosive environment may be encountered during service.

9. SPECIFIC FORMS OF CORROSION

One can conveniently classify corrosion according to the manner in which it is manifest. Usually, metallic corrosion is classified into eight different forms: uniform, galvanic, crevice, pitting, intergranular, selective leaching, erosion-corrosion and stress corrosion. In addition, there is *hydrogen embrittlement*. Truly speaking, hydrogen embrittlement is a type of failure rather than a form of corrosion. However, hydrogen embrittlement is often produced by hydrogen that is generated from corrosion reactions.

(i) Uniform Corrosion: This is a form of electrochemical corrosion that occurs when the whole surface of the metal is corroded to the same degree and when the metal surface and environment are uniform. This type of corrosion often leaves behind a scale or deposit. Under these conditions one can easily estimate the useful life of a given material and there is no need to fear about unexpected failure. One can easily control the uniform corrosion attack by selecting suitable materials. In practice such a corrosion is rare. Uniform corrosion is generally observed in metals like Zn, Pb and Al.

(ii) Crevice Corrosion: Electrochemical corrosion can also occur as a consequence of concentration differences of ions or dissolved gases in the electrolyte solution, and between two regions of the same metal piece. For such a *concentration cell*, corrosion takes place in the local that has the lower concentration. Corrosion occurring in crevices and recesses or under deposits of dirt or corrosion products where the solid solution becomes stagnant and there is localized depletion of dissolved oxygen are few good examples of this type of corrosion. Corrosion preferentially occurring at these positions is known as *crevice corrosion*. In order to penetrate the crevice must be wide enough, yet narrow enough for stagnacy. We may note that usually the width is several thousandths of an inch.

After oxygen has been depleted within the crevice, oxidation of the metal occurs at this position (Eq. 8). Electrons from this reaction are conducted through the metal to adjacent external regions, where they are consumed by reduction (probably in accordance with reaction Eq. (2)). In several aqueous environments, the solution within the crevice has been found to develop high concentrations of H^+ and Cl^- ions, which are especially corrosive. Since protective films are often destroyed by the H^+ and Cl^- ions and therefore many alloys that passivate are susceptible to crevice corrosion.

One can prevent this type of corrosion by using welded instead of riveted or bolted joints, using non absorbing gaskets when possible, removing accumulated deposits frequently, and designing containment vessels to avoid stagnant areas, and ensure complete drainage.

(iii) Pitting Corrosion: This type of very localized corrosion attack occurs when the whole surface of the metal is corroded to the same degree and when the metal surface and environment are uniform. This type of corrosion attack produces fine pits or holes at various locations on the metal surface. They ordinarily penetrate from the top of a horizontal surface downward in a nearly vertical direction. Pitting corrosion results either of the following reasons: *(a)* localized difference in the composition of the metal surface, or its smoothness, *(b)* inhomogeneous external environment, *(c)* insoluble corrosion product *(d)* protective films on the metal surface having not been uniformly perfect *(e)* crystallographic directions having not been equal in reactivity, and *(f)* a small spot on the surface being continuously anodic with current flowing from it at a relatively high density.

Pitting corrosion results from an electrochemical reaction between the broken film as anode and the surrounding unbroken film as cathode which form a small galvanic cell and produce pits. The mechanism for pitting is probably the same as for crevice corrosion in that oxidation occurs within the pit itself. This type of corrosion is observed in aluminium, steel, copper and nickel alloys. Stainless steel alloys are somewhat susceptible to this form of corrosion. However, alloying with about 2% molybdenum enhances their resistance significantly. Figure 12.4 illustrates the pitting corrosion.

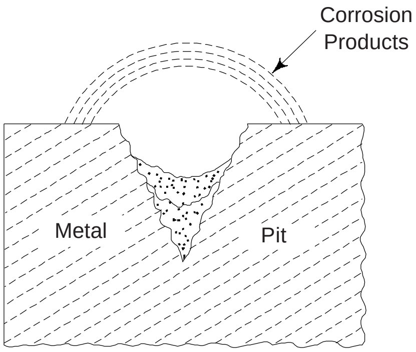


Fig. 12.4 Pitting Corrosion

A pit may be initiated by a localized surface defect such as a scratch or a slight variation in composition. However, pits are randomly formed on the surface. The shapes of pits vary widely, they are normally hemispherical with electropolished inner surfaces. It has been observed that specimens having polished surfaces display a greater resistance to pitting corrosion. Pits have considerable effect on mechanical properties of metals, for example fatigue.

(iv) Intergranular Corrosion: This type of corrosion occurs preferetially along grain boundaries for some alloys and in some specific environments. This occurs when a pronounced difference in reactivity exists between grain boundaries and the reminder of the alloy. This type of corrosion is especially prevalent in some stainless steels. When heated to temperatures between 500 and 800°C for sufficiently long time periods, these alloys become sensitized to intergranular attack. Probably, this type of heat treatment causes the formation of small precipitate particles of Chromium Carbide ($Cr_{23}C_6$) by reaction between the chromium and carbon in the stainless steel. To form the precipitates, both the chromium and the carbon must diffuse to the grain boundaries. This leaves a chromium-depleted zone adjacent to the grain boundary. This grain boundary region is highly susceptible to corrosion.

Selective corrosion attack between the grains of metals and alloys may occur due to potential differences set up there when: *(a)* when there is precipitation in the grain boundaries and/or *(b)* even because of difference in grain orientations.

Intergranular corrosion is normally observed in defective welding and heat treatment of stainless steels, copper and aluminium alloys. One can examine intergranular corrosion with the help of microscope.

Figure 12.5(a) shows the intergranular corrosion, where the grain boundaries have been etched, i.e., corroded, the atoms at the boundaries have a different electrode potential than the atoms in the grain proper. Obviously, anode and cathode are developed, resulting in galvanic stress cells causing corrosion of the anode, i.e. the grain boundary. We may note that a fine grained metal will have a higher corrosion rate than a coarse grained metal as there is no anode area.

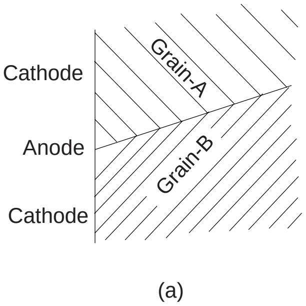


Fig. 12.5(a) **Intergranular corrosion, the grain boundaries act as the anode because the boundary atoms have a high potentials**

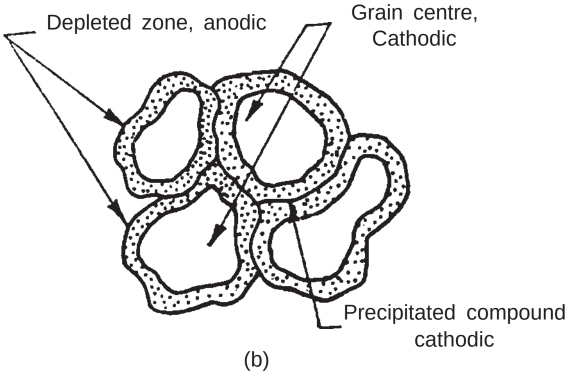


Fig. 12.5(b) **Intergranular corrosion caused by the precipitation of certain compound at grain boundaries**

The intergranular corrosion caused by the precipitation of certain compound at grain boundaries is shown in Fig. 12.5(b). We note that the solid solution adjacent to the grain boundary gets impoverished in one constituent. With respect to the centre of the grain and to the precipitated compound, the impoverished solid solution is anodic. This means that it will be selectively attacked by the corrosive medium. Specially, high strength aluminium alloys containing copper may be susceptible to this type of corrosion.

This type of corrosion is an especially severe problem in the welding of stainless steels, when it is often called *weld decay*.

One can protect the stainless steel from intergranular corrosion by the following measures: *(i)* subjecting the sensitized stainless material to a high-temperature heat treatment in which all the chromium carbide particles are redissolved, *(ii)* minimizing the carbide formation by lowering the carbon content below 0.03 Wt% C, and *(iii)* alloying the stainless steel with such a metal, which has a greater tendency to form carbides than does chromium so that the Cr remains in solid solution, e.g. niobium or titanium.

(v) Selective Leaching or corrosion and Dezincification: This type of corrosion is usually found in solid solution alloys and occurs when element or constituent is preferentially removed as a consequence of corrosion processes.

The most common example of selective corrosion is the dezincification of brass. In this type of corrosion, the zinc atoms are dissolved from the solid solution leaving behind copper as the spongy mass which has practically no strength (Fig. 12.6). The mechanical properties of the copper-brass zinc alloy are significantly impaired, since only a porous mass of copper remains in the region that has been dezincified. Moreover, the material changes from yellow to a red or copper colour. It is found that the selective leaching may also occur with other alloy systems in which Al, Fe, Co, Cr, and other elements are vulnerable to preferential removal.

In order to overcome from dezincification, one may add tin, antimony or phosphorus to brasses containing Zn more than 15% Zn.

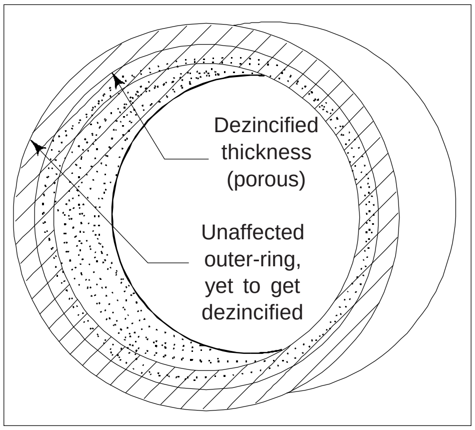


Fig. 12.6 **Illustration of uniform dezincification halfway through a brass water pipe**

(vi) Atmospheric Corrosion: This type of corrosion is very frequent on ferrous materials. Rain water and humidity, in air

are mainly responsible for this type of corrosion and act as electrolytes. This type of corrosion is primarily due to the formation and breakdown of films by the moisture and electrochemical attack on the exposed metal. Obviously, this follows the oxygen-absorption mechanism. It is observed that cracks and discontinuities in the film produced expose fresh areas to the formation of corrosion cells.

(vii) Erosion-Corrosion: This type of corrosion arises from the combined effect of chemical attack and mechanical abrasion or wear as a consequence of fluid motion. Mechanical abrasion is produced by the turbulent flow of liquids or the impingement of trapped gases. Normally, all metal alloys, to one degree or another, are susceptible to erosion-corrosion. Erosion-corrosion causes accelerated attack because it mechanically removes the protective layer that normally builds up on the corroding surface. Obviously, this type of corrosion is especially harmful to alloys that passivate by forming a protective surface film. When the coating is not capable of continuously and rapidly reforming as a protective barrier, erosion-corrosion may be severe. This type of corrosion action contributes to the formation of localised pitting cells. Usually, this type of corrosion is associated with the formation of cavities in the metal by fast-moving liquids. Relatively soft metals such as copper and lead are also sensitive to this form of corrosion attack.

This type of corrosion is generally found in piping, especially at bends, elbows, and abrupt changes in pipe diameter-positions where the fluid changes direction or flow suddenly becomes turbulent. Propellers, turbine blades, valves, and pumps are also susceptible to erosion-corrosion. Figure 12.7 illustrates the erosion-corrosion in a pipe-line.

In order to reduce erosion-corrosion, one can change the design to eliminate fluid turbulence and impingement effects. One can also utilize other materials that inherently resist erosion. Moreover, removal of particulates and bubbles from the solution may lessen its ability to erode.

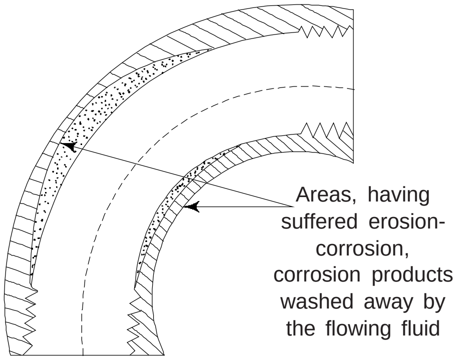


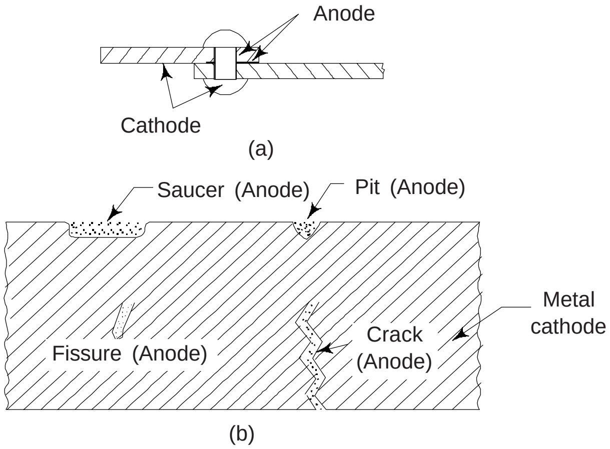
Fig. 12.7 Erosion corrosion in a pipe line

(viii) Stress Corrosion: This type of corrosion also termed as *stress corrosion cracking* is produced by the combined effect of mechanical stress and a corrosive environment on a metal. The stress that produces this type of corrosion need not be externally applied; it may be a residual one that results from rapid temperature changes and uneven contraction, or for two phase alloys in which each phase has a different coefficient of expansion.

This type of corrosion occurs predominantly in metal components having internal stresses (which may be due to cold working or otherwise) and being used in corrosive environment. In fact, some materials that are virtually inert in a particular corrosive medium become susceptible to this form of corrosion when a stress is applied. Pile up of dislocations at grain boundaries in cold-worked metals increases the energy in such regions and makes anodes in the structure in certain environments. Small cracks form and then propagate in a direction perpendicular to the stress, with the result that failure may eventually occur. Season cracking in brass, especially in the presence of moisture and traces of ammonia, and caustic embrittlement of steel exposed to solutions containing sodium hydroxide are few well known examples of stress corrosion. Most alloys are susceptible to stress corrosion in specific environments, especially at moderate stress levels. Pure metals are relatively immune to stress corrosion. The most effective control of stress corrosion is the elimination of tensile stress from the component part or lower the magnitude of the stress. This may be achieved by reducing the external load or increasing the cross-sectional area perpendicular to the applied stress. Furthermore, an appropriate heat treatment may be used to anneal out any residual thermal stresses.

In case of cold worked metal, the corrosion takes place at the strained portion. In the strained area, the atoms of metal are not in their lowest energy position. Consequently, strained area of the metal (having internal stresses) has higher electrode potential than the strain-free area. We may note that in the presence of corrosive environments, the strained area acting as anode while unstrained area as cathode give rise to a type of galvanic-cell, called *stress-cell*. Obviously, corrosion starts at the strained area due to electrochemical action of the cell.

It is found that the stress corrosion rate is dependent upon the corrosive agent, time and temperature of exposure, behaviour of protective films and stress magnitude. Under high tensile stress, the absorption of hydrogen in metal causes embrittleness. This phenomenon is usually known as hydrogen *embrittleness.* When mild steel exposed to alkaline solutions at high temperatures and stresses, one can observe hydrogen embrittleness (Fig. 12.8).



(a)

(b)

Fig. 12.8 Stress corrosion

(ix) Corrosion Fatigue: This is a type of failure that results from the simultaneous action of a cyclic stress and chemical attack. Now, it is well understood that corrosion becomes the cause of cracking while fatigue may cause the crack propagation. A corrosion crack may propagate even under static stresses. We may note that the combined action of corrosion and repeated stresses is far more serious than the sum of these two factors acting independently. Corrosion fatigue is also the reduction of fatigue strength of materials due to the presence of a corrosive medium. The alternates slip taking place at granular legal creates minute regions of high strain energy which tend to become anodic with respect to surrounding regions of low strain energy. The corrosion crack presents a mechanical stress concentration whereby the fatigue strength will reduce. The effect of corrosion fatigue is clearly observed in heat exchangers where variable stresses are produced because of thermal expansion and contraction. We may say that under corrosive conditions and fatigue loading the material is in most critical state.

One can express the effect of corrosion fatigue strength by the damage ratio as

$$\text{Damage ratio} = \frac{\text{Corrosion fatigue strength}}{\text{Normal fatigue strength}}$$

The damage ratio for salt as a corroding medium is ~0.2 for carbon steels, ~0.5 for stainless steels and bout 1.0 for copper. The treatment of the corroding medium and surface protection of the metal may be suitable protective measures against corrosion fatigue. It is found that nitriding of steels may be often useful for this process.

(x) Fretting: Fretting corrosion is a type of corrosion which occurs at interface between material surfaces under high pressure, if slight slip occurs. Usually, this type of corrosion takes place when two closely fitted metal surfaces, e.g. when bolted or rivetted joints, or press fitted hubs on shafts are subjected to vibrations. Molecular plucking of the surface particles takes place in the initial stage, which oxidise to form a debris. In case of steel this debris is like red rust.

Fretting corrosion depends on load, temperature, humidity, presence of oxygen, and metallurgical factors. One can minimise the effects of this type of corrosion by avoiding relative motion between the two surfaces by introducing compressive stresses and by heat treatment. Fretting corrosion reduces at high humidities and increases in the presence of oxygen. It decreases with the decrease in temperature. It is found to be more in soft materials than that in hard materials.

(xi) Hydrogen Embrittlement: A significant reduction in ductility and tensile strength is observed when atomic hydrogen (H) penetrates into various metal alloys, and specifically some steels. This phenomenon is usually referred to as *hydrogen embrittlement* and sometimes also called as *hydrogen-induced cracking* and *hydrogen stress cracking*. We can say that hydrogen embrittlement is a type of failure, in response to applied or residual tensile stresses. Brittle fracture occurs catastrophically as cracks grow and rapidly propagate. Interestingly, hydrogen in its atomic form (H as opposed to the molecular form, H_2) diffuses interstitially through the crystal lattice, and concentrations as low as several parts per million can lead to cracking. We may note that hydrogen-induced cracks are most often transgranular, although intergranular fracture is noticed in some alloy systems. In order to explain hydrogen embrittlement, number of mechanisms, most of them based on the interference of dislocation motion by the dissolved hydrogen have been proposed.

Hydrogen embrittlement is quite similar to stress corrosion. However, one may distinguish these two phenomena on the basis of their interactions with applied electric currents. The cathodic protection reduces or causes a cessation of stress corrosion, whereas, it may lead to the initiation or enhancement of hydrogen embrittlement.

In order that hydrogen embrittlement may occur, some source of hydrogen must be present, and, in addition, the possibilities for the formation of its atomic species. Pickling (a procedure used to remove surface oxide scale from steel pieces by dipping them in a vat of hot, dilute sulphuric or hydrochloric acid) of steels in H_2SO_4, electroplating, and the presence of hydrogen-bearing atmospheres (including water vapour) at elevated temperatures, e.g. during welding and heat treatments are such situations wherein these conditions are met. Moreover, the presence of poisons such as sulphur, i.e. H_2S and arsenic compounds accelerates hydrogen embrittlement. These substances retard the formation of molecular hydrogen and thereby increase the residence time of atomic hydrogen on the metal surface. Probably, hydrogen sulphide is the most aggressive poison and found in petroleum fluids, natural gas, oil-well brines, and geothermal fluids.

It is found that high-strength steels are susceptible to hydrogen embrittlement, and increasing strength tends to enhance the material's susceptibility. Martensitic steels are found to be especially vulnerable to this type of failure. Bainitic, ferritic and spheroiditic steels are found to be more resilient. FCC alloys (austenitic stainless steels, and alloys of copper aluminium and nickel) are found relatively resistant to hydrogen embrittlement mainly due to their inherently high ductilities. Strain hardening of these alloys may enhance their susceptibility to embrittlement.

In order to reduce the likelihood of hydrogen embrittlement some of the commonly used techniques include reducing the tensile strength of the alloy via a heat treatment, removal of the source of hydrogen, "baking" the alloy at an elevated temperature so that one may drive out any dissolved hydrogen, and substitution of a more embrittlement resistant alloy.

10. CORROSION PREVENTION AND CONTROL

Perhaps the most common and easiest way of preventing corrosion is the proper selection of the materials, once we have characterized the corrosion environment. In this respect, standard corrosion references are helpful. However, one has also to examine the economic feasibility prior to employing the material that provides the optimum corrosion resistance, for example, one can use noble metals in very limited applications such as ornaments and delicate scientific instruments. In order to control corrosion, the materials would have to be absolutely uniform without any heterogeneities in either composition or structure and at the same time the environment would also have to be entirely uniform.

The following methods are widely used for the control and preventation of corrosion:

(i) Using high-purity metals and alloy additions
(ii) Proper design against corrosion

(iii) Proper modification of corrosive environment
(iv) Making use of protective coatings
(v) Making use of inhibitors
(vi) Cathodic protection

(i) Use of High-Purity Metal and Alloy Additions: One can improve the corrosion resistance of a given metal by avoiding galvanic couples, i.e. limiting designs to only one metal. However, this is not always feasible. By selecting suitable electrically insulating metals of different composition or by alloy addition, one may avoid the cells in special circumstances. One can increase the corrosion resistance of most metals by alloying them with suitable alloying elements. Alloying may increase corrosion resistance by increasing the energy level of the solution, reducing the driving *emf* in the galvanic cell, and protective coatings and decreasing the mobility of the corrosive ions. Alloying may however help in the following ways:

(i) To avoid carbon precipitation in heat treatment.
(ii) To prepare steels with high chromium content. We may note that chromium corrodes less readily.
(iii) To make steels containing strong carbide formers. Titanium and tantalum are such elements. Interestingly, these elements do not permit carbon precipitation at the grain boundary.

These methods reduce intergranular corrosion in metals. This technique is mainly used in stainless steel which must be fabricated by welding.

(ii) Design Against Corrosion: One can design to avoid physical contact between dissimilar metals so that a galvanic couple does not form. In addition, the design should allow for complete drainage in the case of a shutdown, and easy washing. It is found that dissolved oxygen may enhance the corrosivity of many solutions, the design should, if possible, include provision for the exclusion of air. In order to prevent, one may ensure the following points:

(i) As far as possible avoid of dissimilar metal contracts. When contact between dissimilar metals is unavoidable, one may take care that the metal forming the anode does not have a small surface area as compared to the cathode.
(ii) There must be cathodic protection.
(iii) One must avoid recesses and sharp corners.
(iv) Avoidance of excessive concentration with proper fabrication.
(v) One must prevent the retention of liquid with air tight joints.
(vi) Surfaces must be clean, i.e., free from foreign matter like dust, dirt and soot.
(vii) Wherever possible, one must use corrosion-resistant materials.
(viii) We may take care that the two different metals should be as close as possible.

(iii) Modification of Corrosive Environment: It is found that changing the character of environment, if possible, may also significantly influence corrosion. One can achieve large saving in materials if these are used in less corrosive environments. Lowering the fluid temperature and/or velocity usually helps in the reduction in the rate at which corrosion occurs. Several times increasing or decreasing the concentration of some species in the solution can have a positive effect, e.g., the metal may experience passivation. One can modify the environment in the following ways:

(i) Lowering the temperature, pressure, concentration, velocity, etc.
(ii) By changing humidity, oxygen content, oxidizing agents and solid impurities, one may alter the chemical composition of environment. It is found that purified and dehumidified atmosphere around the structure helps in reducing corrosion.
(iii) By adding *inhibitors*, i.e. those substances in low concentrations to the environments, which decreases corrosiveness of the environment. For example, alkaline neutralizers are used to prevent the corrosive effect of the environment by neutralizing the acidic character of corrosive conditions. To prevent high-temperature corrosion of ferrous metals and alloys, vacuum and inert gases are also found useful. Of course, the specific inhibitor depends both on the alloy and on the corrosive environment.

(iv) Use of Protective Surface Coatings: Physical barriers to corrosion are widely applied on surfaces in the form of films and coatings. There are good number of diverse metallic and non-metallic coating materials available. It is essential that the coating should have a good corrosion resistance when in use, and maintain a high degree of surface adhesion, which undoubtedly requires some preapplication surface treatment. In several cases, the coating must be virtually non reactive in the corrosive environment and resistant to mechanical damage that exposes the bare metal to the corrosive environment. In general, all three material types—metals, ceramics and polymers are used as coatings for metals.

Metallic Coating

In the protection of the underlying metal by metallic coating, two factors are mainly involved: *(i)* mechanical isolation of the metal from the corroding environment and *(ii)* galvanic relation of coating metal and the base metal. If the metal coating is baser than the substrate (e.g., zinc or aluminium on steel), galvanic protection is offered to the underlying metal. The coating is anodic and corrodes first. However, Zn and Al become passive after the initial attack. Nickel and chromium plating give an decorative attractive appearance. Hard chromium plating produces a thicker, wear resistant surface. The tin coating on steel article is produced so that tin plate may have excellent corrosion and fabrication qualities. The most common use of tin plate is for making food containers. The following methods are widely used to apply metallic coatings:
(i) hot dipping, *(ii)* electroplating, *(iii)* metal cladding, *(iv)* high temperature diffusion, *(v)* metal spraying etc.

Zinc coating on steel (galvanized coating) and the tin coating on a steel article are usually produced by dipping clean sheet steel into molten zinc or a tin bath. Galvanized iron (G.I.) is produced by dipping a low carbon steel sheet in molton zinc bath at about 450°C. The zinc coating crystallizeds forming '*zinc flowers*' when cooled in air. Articles, e.g. buckets and drums prepared from galvanized iron are very suitable for aqueous environments. However, the lives of galvanized coatings depend on the thickness of the zinc layer and environmental conditions. In the presence of oxygen, zinchydroxide is precipitated as a protective layer.

One can obtain thinner, uniform coatings, of Zn and Sn by *electroplating*. Now a days, this process is popular for these metals. Chromium, cadmium and nickel platings are usually produced by electroplating (Fig. 12.9). In this electroplating process, the coating metal is deposited on the base metal by passing a direct current through the solution of electrolyte containing a salt of the coating metal. The base metal is made the cathode and the coating material is made the anode. We may note that the coating properties mainly depends on the current intensity, agitation, temperature of the solution and the chemical composition of the plating solution. Interestingly these coatings are relatively thick and are mostly applied to cast or machined parts.

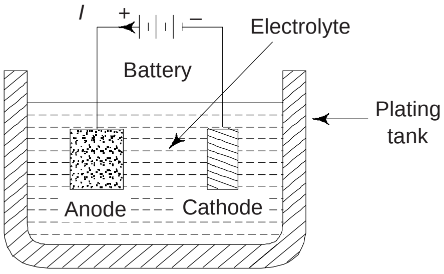


Fig. 12.9 Electroplating

A comparatively thick lining is placed upon the metal surface forming a strong alloy bond between plates of two metals in metal cladding. This strong alloy bond is formed either by casting or hot rolling. The principal objective for cladding is the production of a corrosion resistant surface. To make bimetallic strips for thermostats, metallic cladding is found quite useful.

Use of the powder metallurgy technique is made in the high-temperature diffusion process. The base metal is heated with the powder of the metal which forms the coating. When diffusion coatings are formed on steels with the help of high-temperature treatment, it is found that the innermost core is rich in the base metal followed by successive richer coating surfaces. In the presence of powered aluminium, this treatment is known as *calorizing*, with powered zinc as *shearadizing*, and with chromium as *chromising*. Interestingly, these coatings are also thick.

By metal spraying, metallizing or flame spraying, thick coatings of any metal can be produced. This involves the application of a spray of molten-metal droplets on the surface to be coated. We can use variety of metals. Deposits are porous, but one may offset this by applying a sealing coat of paint. If an overcoat is planned for, then this procedure may be beneficial. If the coating is anodic to the base metal as in the case of zinc or aluminium coated steel, then porosity is of less concern. Metal spraying is particularly very advantageous when a very large surface is to be treated, greater speed of working is desired and the coating is to be applied to hidden areas.

Normally, three types of metal-spraying processes are in use: wire, powder and plasma. In the wire and powder metal spraying techniques, the coating metal is fed into the spray gun, where it is melted. Compressed air then atomises and sprays the molten particles onto the prepared surfaces to be coated. The new plasma spray process also uses powdered metal. An inert gas, usually argon or nitrogen is dissociated in an electric area, and the energy produced by this raises the temperature of the gas stream. Powered metal is introduced into the gas stream where it immediately melts as it is driven toward the metal surface. The plasma spray process is also used for applying non-metals, e.g. ceramics, cermets, etc., although the reason for applying them is to improve abrasion resistance rather than for corrosion protection.

Non-Metallic Coatings

The most common non-metallic coatings used these days are enamels, oils, paints, tars, and chemical coatings. These act by simply excluding water and oxygen and by providing a layer of higher electrical resistance.

Paint is the most widely used protection against corrosion. For articles used indoors, it is often enough to coat them with one layer of paint. In order to provide a mechanical barrier against the corroding environment, a rust inhibiting action is provided by red lead. Paints used for the protection of outdoor structures such as bridges, elevator cranes, harbour equipment, railway carriages, automobiles and bicycles, which are normally exposed to the action of the atmosphere, are of more complex-composition. Moreover, their process of application is more intricate, involving several coatings. Bituminous paint, which consists of hard asphalt with drying oil, soft asphalt and coal, can be used to protect metal in case the black colour is acceptable. Bituminous coatings are very useful for protecting underground tanks and pipes. Bituminous aluminium paint is used for protection against corrosive atmosphere on the sea shore. Ocean going steel ships need antifouling paint over the anticorrosive primer. Red lead or zinc yellow is used in the primer. This paint is generally known as *marine paint*, *silicon paint* is used as *heat resisting paint.* Its first use was as an insulation varnish for electric motors. Traffic paint is quick drying and therefore the traffic inconvenience is little. Rough texture and absence of gloss helps to increase visibility, one may add tiny glass beads for better visibility during night.

Chemical coatings are usually produced either by chemical dipping in suitable solutions or by a chemical reaction. Surface protective films are made by chemical surface treatments or dips. These coatings provide for corrosion protection either as a final finish or as a base for another coating like painting, lacquering, etc. Chromate coatings are normally used on magnesium alloys by merely dipping the metal, e.g. copper, brass, bronze etc. in a chromate solution. The chromate coating serve the dual purpose: *(i)* Chemically polishing the surface and *(ii)* providing a thin film moderately protective in nature. Chromate coatings are applied to a variety of non-ferrous metals, e.g. Zn, Cd, Mg, Al, Cu, Ag, Be. In a chromate coating, the article to be coated is placed in a solution containing primarily a source of Cr^{+6} ions, e.g., $Na_2Cr_2O_7$, and a mineral acid, H_2SO_4. The coating is formed rapidly and the thickness of the film is about 0.5 micron. Usually, the chromatic films are yellow in colour. Normally, the upper service temperature for coating is ~65°C and we have to take care not to disturb it mechanically during the first 24 hours. Probably, the mechanism of chromate coating is related to corrosion of the metal and simultaneous partial reduction of Cr^{+6} to an amorphous, hydrated layer of chromium chromate. This layer contains insoluble Cr^{+3}, and slightly soluble Cr^{+6}, which can leach out and act as an inhibitor to heal scratches formed during service. We may note that chromates are generally more corrosion resistant than phosphates but at the same time are not as effective for holding oils, waxes, etc. Chromates are less porous than phosphates and oxides, and are often used to seal these types of coatings. Now a days, chromates are increasingly used on aluminium as an alternate to anodizing.

Anodising, i.e., anodic coatings is a chemical treatment applied to aluminium and its alloys so that a greater thickness of the protective film of aluminium oxide on the surface is produced. Such reactions take place in the electrolyte when the base metal is made an anode. The electrolyte consists of about 5 to 10% solution of chromic acid or oxalic acid or H_2SO_4. Anodised coatings can be obtained upto 25 micron in thickness. Obviously, these coatings are about 100 times thicker than natural oxide films and improve resistance to mechanical injury to metals and corrosion resistance of metals. These types of coatings can also be produced on Zn and magnesium. Anodising decreases the porosity and increases corrosion resistance of the film and also give high electrical insulation to the base metal. In order to give decorative effects, this may be coloured with organic dyes and inorganic pigments.

(v) Inhibitors: These are chemical compounds added to electrolytes, antifreeze mixtures and corrosive solutions to retard or slow down a chemical reaction. Sometimes the inhibitor plays the role to form an impervious, insulating film of a compound on either the cathode or the anode. When added in small quantities in corrosive conditions, inhibitors decrease the corrosion rate. Inhibitors are generally used in liquids, solids, semi-solids and packing-materials. There are four types of inhibitors: anodic, cathodic, adsorption and vapour phase. Normally, inhibitors are used to produce an oxide film on manganese and aluminium, antifreeze mixtures for refrigerators and automobile radiators, in water treatment, for surface cleaning processes, e.g. pickling, etc. and for protective greases.

Anodic Inhibitors

These retard the anodic reaction and are usually either oxidizing agents (chromates, nitrates, ferric, salts) that promote passivity, or film formers that precipitate insoluble films on the anode area, e.g. alkalies, silicates, phosphates, benzoates, etc. Here oxygen works as the anodic inhibitor and forms a protective film by forming ferric hydroxide by the oxidation of ferrous hydroxide. We may note that anodic inhibitors are often termed dangerous, particularly the oxidizing types, as they must completely protect the anodic areas to be effective.

When a small anode is not protected, it will corrode fastly as the large cathode-to-anode ratio (corrosion rate is a function of current density). Oxidizing agents may also depolarize the cathode reaction and accelerate corrosion before they have passivated the entire anode area.

Cathodic Inhibitors

These inhibitors prevent oxygen absorption at the cathode area i.e. retard the cathodic process and also the evolution of hydrogen. Primarily cathode inhibitors are of following three types:

(i) Oxygen Scavengers: These remove oxygen from the solution, and also prevent it from depolarizing the cathode. Two common scavengers are sodium sulphite and hydrazine:

$$2Na_2SO_3 + O_2 \rightarrow 2Na_2SO_4$$

$$2N_2H_2 + O_2 \rightarrow 2N_2 + 2H_2O$$

These scavengers are found to be most effective in neutral and slightly acid solutions, but are ineffective in strongly acid solutions where primary cathode reaction is of hydrogen evolution. Sodium sulphite and hydrazine are widely used in closed systems, especially in boiler waters.

(ii) The evolution of hydrogen poisons ions of some of the heavy metals, e.g. arsenic, antimony, bismuth, retard hydrogen depolarization, reaction:

$$2H^+ + 2^- \rightarrow H^2$$

It is found that these inhibitors are useful in low PH solutions but are ineffective when the major cathode reaction is oxygen reduction, e.g. neutral solutions. If the metal is susceptible to hydrogen embrittlement then they are detrimental, because they increase the transfer of hydrogen into the metal as opposed to letting it escape from the solution as a gas.

(iii) Films, which are insoluble and form on the cathode and reduce its effective area also work as cathodic inhibitors. For example, the addition of calcium bicarbonate to iron in neutral or slightly alkaline water converts calcium carbonate insoluble.

In comparison to anodic inhibitors, cathodic inhibitors are not dangerous, because they do not stimulate corrosion if present in sufficient quantity. Moreover, cathodic inhibitors are less effective than anodic inhibitors.

(c) Adsorption Inhibitors These are used to control corrosion by influencing either the anode or cathode reaction, or both. They are usually organic compounds. The commonly used organic inhibitors are amines, hetrocyclic nitrogen compounds, substituted urea and thiourea mercaptans, sulfides and heavy metal soaps. Colloidal inhibitors (organic), e.g. waste sulfite-liquors, alkaline tennates, tannins, starch and glucosades have been found effective when microorganisms grow on metals. Adsorption inhibitors adsorb on the cathode and anode areas and stifle the reaction. Usually, these inhibitors are used in acid solutions (metal-pickling operations, acid storage, etc.). In automobile radiators sodium benzoates act as effective inhibitor.

(d) Vapour-Phase Inhibitors These have high vapour pressures that enable them to spread into the atmosphere, then adsorb into the metal surfaces in the vicinity. They are similar in character to the adsorption inhibitors and are used to avoid corrosion in natural atmosphere in enclosed spaces, e.g. metal parts in storage or during shipment. In the latter case, one may place the inhibitor beside the metal article in powder form, or can be sprayed onto the metal in solution form. Inhibitors of organic type used in vapour phase have been developed particularly for protection during storage, and packing. Dicyclohexylammonium nitrite (*Dichan*) is found to be the most effective vapour phase corrosion inhibitor. Typical examples of vapour-phase inhibitors are amine salts, morpholine and sodium benzoate.

Inhibitors are normally used in closed systems such as automobile radiators and steam boilers.

(vi) Cathodic Protection: This is one of the most effective means of corrosion prevention based on the electrochemical principle. It can be used for all types of different forms of corrosion discussed in this chapter, and may, in some situations, completely stop corrosion. We know that oxidation or corrosion of a metal Me occurs by the following generalized reaction

$$\mathrm{Me} \rightarrow \mathrm{Me}^{n+1} + n\mathrm{e}^- \tag{29}$$

We may note that cathodic protection simply involves supplying, from an external source, electrons to the metal to be protected, making it a cathode; the reaction (29) is thus forced in the reverse (or reduction) direction.

We have read that in electrochemical corrosion there is always a flow of current between the anode and cathode. The anode is corroded and at the same time the cathode is protected. Obviously, corrosion is reduced or eliminated by making the whole surface cathodic to an extraneous surface which acts as anode. One can achieve this by impressed voltage or by the attachment of a *sacrificial anode* so called as the anode metal, which is periodically replaced. This phenomenon of *galvanic protection* is called as cathodic protection. The process of *galvanizing* is simply one in which a layer of zinc is applied to the surface of steel by hot dipping. In the atmosphere and most aqueous environments, zinc is anodic to and will thus cathodically protect the steel if there is any surface damage (Fig. 12.10). This protecting metal (Zn) is known as the sacrificial anode. We may note that any corrosion of the zinc coating will proceed at an extremely slow rate because the ratio of the anode-to-cathode surface area is quite large. Obviously, cathodic protection is the most effective method of corrosion control. We can say that it is the only method which can completely prevent corrosion. The objects to be protected are either buried in the earth or immersed in water, e.g. underground and underwater pipelines, piers, jetties, foundations, ships and offshore structures. A piece of aluminium, zinc or magnesium is attached to the article to be protected making electrical contact with it. The impressed voltage can be externally applied for the protection of structures. In the artificial galvanic cell thus created, the added electrode piece is the anode, which is gradually destroyed. Galvanized iron is a familiar example of cathodic protection. We may note that for the protection of large surfaces, it is essential to insert a number of zinc, magnesium or aluminium anodes distributed at different points. The additional anodes will establish a potential with the metal to be protected and thus helps to prevent its corrosion.

There is also another method of cathodic protection. In this method the source of electrons is an impressed current from an external dc power source as shown in Fig. 12.11 for an underground tank. The negative terminal of the power source is connected to the structure to be protected. The other terminal is joined to an inert anode usually graphite, which is, in this case, buried in the soil. Obviously, the high-conductivity backfill material provides good electrical contact between the anode and surrounding soil. In order to complete the electrical circuit, a current path exists between the cathode and anode through the intervening soil, completing the electrical circuit.

Cathodic protection is especially useful in preventing corrosion of water heaters, underground water tanks and pipes, and marine equipment. Magnesium anodes have been mostly used for cathodic protection of domestic and industrial water heaters. To make cathodic protection very effective, coverings of plastics, paints and enamels are applied. One must take care that anode area is small as compared to the cathode, polarization does not occur and at the same time metals widely separated in the galvanic series are avoided.

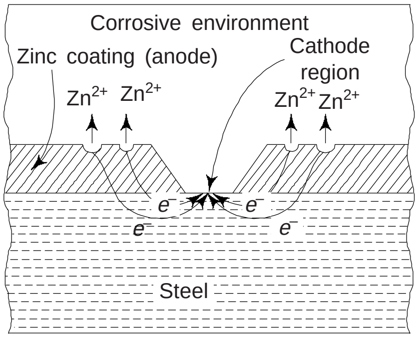


Fig. 12.10 **Galvanic protection of steel as provided by a coating of zinc**

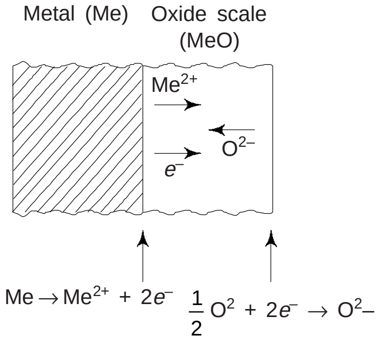


Fig. 12.11 **The representation of processes involved in gaseous oxidation at a metal surface**

11. CORROSION MONITORING AND MANAGEMENT

There are wide range of corrosion monitoring and measurement techniques as follows:

(i) Direct Method:

(a) Visual/Optical Measurement This method requires the corroded material to be subjected to visual inspection and physical measurement. This is only possible in those areas where access is available and when plant is shut-down.

(b) Coupon Testing This method is used in industry. First, the coupons are measured carefully and their physical dimensions are also recorded. The exposed area of coupons to corrosive environment is also recorded.

One can obtain the rate of corrosion by putting the coupons under actual operating conditions preferably the C.W. outlet of critical equipments. One will have to inspect these samples at regular intervals and record the results. The loss of weight, after removal of corrosion products is determined and expressed in mills/year. However, this process is a long and tedious, but gives quite reliable results than many of the laboratory tests.

Using the following relation, the average thickness of corroded layer per year can be calculated and usually expressed in mills/year. Weight and specific gravity of the metal are known.

$$P = \frac{(W_1 - W_2)\,K}{ASt} \tag{30}$$

where P is rate of corrosion (mills/year), W_1 is weight of specimen before corrosion (gm), W_2 is weight of specimen after corrosion (gm), A is exposed area of specimen (sq. cm), t is time of exposure (in hours), S is specific gravity of metal (gm/cm^3) and K is constant, whose value is 3.45×10^6 (mills/year).

Size of the corrosion coupon is: length 75-100 mm, width 10-13 mm and thickness 0.8 to 1.6 mm. One will have to take care to clean the coupons thoroughly. Normally, the coupons are kept in actual operating media for about 1 to 3 months. Coupons are then dried in paper towels, examined and recorded by photograph or description of the appearance prior to removing the deposits and weight loss calculations. The limitations of this method are:

(i) Test location of coupons is limited.
(ii) In preparing coupons and evaluating corrosion rate, time is involved.
(iii) It does not allow continuous monitoring.
(iv) One cannot evaluate the effect of varying operating conditions.

(ii) Electrochemical Method-Linear Polarisation Resistance (LPR): LPR is a widely used electrochemical technique and can be used ‘on line’ to provide an instantaneous or continuous measurement of the corrosivity of the process stream. This is widely used in the oilfield particularly in water systems.

In order to determine the corrosion current density an hence rate, a small potential or current perturbation is applied. To a close approximation, for a corroding electrode, the applied potential and current density are linearly related. One can relate their ratio ‘LPR’ or polarisation resistance (R_p) to the corrosion current density through the following relation:

$$I_{\text{corr}} = \frac{\text{Constant}}{R_p} \tag{31}$$

We may note that the relation (31) only holds if the electrode is polarised well from the corrosion potential and a linear Elogi.

(iii) Electrical Resistance (ER) Equipment: This operates on the basic principle that electrical resistance increases as the cross-sectional area of a metallic conductor decreases. The probe is the heart of an ER equipment, which is introduced in the corrosive environment and with which corrosion measurements are made. The probe acts as an in situ sensor that accumulates the corrosion history of the environment and displays or records the corrosion information on an instrument.

The basic function of an ER instrument is to compare the resistance of an element that is exposed to corrosion with the resistance of an element that is protected from corrosion. If the resistance of an exposed element is measured directly, the data will be affected by element thickness and temperature. With this method, however temperature changes, that equally affect both elements, are cancelled out and the data obtained are inversely proportional to the thickness of the exposed element. This is why, the ER probe is often referred to as an electrical coupon.

We may note that there is an important difference between ER and coupon system, i.e., the coupon has to be removed from the pipe to measure the corrosion rate whereas the probe remains in the system at all times. To know the corrosion rate, electrical resistance changes are converted into thickness changes which can be plotted against time to know the corrosion rate.

Example 1 *Calculate the distance at which the magnesium anode capable of giving 2 MA to be installed for the protection of 320 mm outside diameter pipe, for current density of 10^{-2} MA/m^2.* [Diploma]

Solution Surface area of 1 m pipe 320 mm outside diameter pipe

$$= \pi DL = \frac{22}{7} \times 0.32 \times 1\ m^2 = 1.005\ m^2$$

$$= 199\ m$$

Example 2 *A ship hull requires a current density of 15 MA/metre*^{2} *for cathodic protection. Magnesium is used as the sacrificial anode. Calculate the quality of magnesium required per square meter of the hull surface for a design life of 10 years.* [BE]

Solution 0.0243 kg, i.e., 1 mole of magnesium will be used up as anode for two farad charge, i.e., 2 × 96490 C (A-s), as magnesium is divalent. Now, the charge required per m^2 of hull surface for a design life of 10 years = $15 \times 10^{-3} \times 10 \times 365 \times 24 \times 3600$ A-s. Therefore, the amount of magnesium required.

$$= \frac{0.0243 \times 15 \times 10^{-3} \times 10 \times 365 \times 24 \times 3600}{2 \times 96490}$$

$$= 0.6\ kg/m^2.$$

SUGGESTED READINGS

1. D. Talbot and J. Talbot, *Corrosion Science and Technology*, CRC Press, Boca Raton, FL (1998).
2. P.A. Scheweitzer, *Corrosion and Corrosion Protection Hand-book*, 2nd ed., Marcel Dekker, New York (1989).
3. H.H. Uhlig and R.W. Revie, *Corrosion and Corrosion Control*, 3rd ed., NACE International, Houston, TX, 1995.

REVIEW QUESTIONS

1. Briefly explain the difference between oxidation and reduction electro-chemical reactions. Which reaction occurs at the anode and which at the cathode? [Diploma]

2. Explain the term corrosion? What are the different types of corrosion? [AMIE]

3. What are the various mechanisms of corrosive attack, explain in brief. [BE]

4. Enlist various methods of corrosion. How it is controlled in actual practice. [BE]

5. What is galvanic series of metal and alloys? Why is this series different then electromotive force series? [BE]

6. What is electrochemical corrosion? Explain corrosion mechanisms involving hydrogen evolution and oxygen absorption. [M.Sc. (MS)]

7. Explain the mechanism of electrochemical corrosion. What are the various factors which influence the corrosion of iron and steel. [Diploma]

8. What is chemical corrosion? Explain Galvanic series. Explain cathodic protection and coating for prevention of corrosion. [AMIE]

9. (a) Explain the term corrosion and describe the mechanism of corrosion. (b) Explain different methods used to protect steel against corrosion. [AMIE]

10. Explain the corrosion mechanism. What are the different methods of treating corrosion? [AMIE]

11. Explain dry and wet corrosion. Explain the various methods for corrosion protection. [AMIE]

12. What is the meaning of pitting corrosion. [AMIE]

13. Discuss the mechanism involved in the following types of corrosion: (i) Intergranular corrosion (ii) Stress corrosion (iii) Erosion corrosion. [AMIE]

14. Explain the effect of pH value on corrosion. [AMIE]

15. Explain microbiological corrosion. [AMIE]

16. Show how rust, $Fe(OH)_3$, can form in a galvanic cell, where iron is the anode and the cathode reaction is

$$O_2 + 2H_2O + 4e^- \rightarrow 4OH^- \qquad \text{[BE]}$$

17. Why is tin preferred to zinc on a steel food can? [Diploma]

18. Why is enamelling preferred to paint on a steel refrigerator body? [Diploma]

19. What are inhabitrs of corrosion? Distinguish between inorganic and organic inhabitors. [BE]

20. Discuss the following:
(a) Galvanic cell corrosion *(b)* Methods of corrosion prevention. [AMIE]

21. *(a)* Briefly explain the major types of corrosion of metallic materials
(b) Describe the techniques generally used to control corrosion metals. [AMIE]

22. Discuss the effect of chromium content on the rate of corrosion in steel. [AMIE]

23. Explain dry and net corrosion. Explain the various methods for corrosion protection. [AMIE]

24. Write short notes on
(a) Pitting corrosion
(b) Stress corrosion
(c) Inhibitors
(d) Coatings
(e) Microbiological corrosion
(f) Galvanic cell corrosion
(g) Chemical corrosion
(h) Erosion corrosion
(i) Intergranular corrosion
(j) Protective coatings
(k) Dry and wet corrosion.

PROBLEMS

1. Calculate the voltage at 25°C of an electrochemical cell consisting of pure cadmium immersed in a 2×10^{-3} M solution of Cd^{2+} ions, and pure iron in a 0.4 M solution of Fe^{2+} ions.
[*Ans.* +0.031 V] [BE]

2. Write the spontaneous electrochemical reaction in problem 1. [*Ans.* $Fe^{2+} + Cd \rightarrow Fe + Cd^{2+}$]

3. What is the potential difference that will be observed, when a standard zinc electrode is connected to the calomel reference electrode whose potential is +0.334 V? [*Ans.* 1.094 V] [BE]

4. An electrochemical cell is composed of pure copper and pure lead electrodes immersed in solutions of their respective divalent ions. For a 0.6 M concentration of Cu^{2+}, the lead electrode is oxidized yielding a cell potential of 0.507 V. Calculate the concentration of Pb^{2+} ions if the temperature is 25°C. [*Ans.* $[Pb^{2+}] = 2.5 \times 10^{-2}$ M] [BE]

SHORT QUESTION-ANSWERS

1. What is metallic corrosion?

Ans. It is ordinarily electrochemical, involving both oxidation and reduction reactions.

2. What do you understand by oxidation?

Ans. This is the loss of the metal atom's valence electrons; the resulting metal ions may either go into the coordination solution or form an insoluble compounds.

3. The magnitude of the electric potential that is established between anode and cathode indicates?

Ans. Driving force for the corrosion reaction.

4. The standard electrode potential measures?

Ans. The tendency of a metal to go into solution.

5. What are the different conditions under which a galvanic cell may be formed?

Ans. *(i)* Due to a two phase structure *(ii)* Due to concentration differences in the electrolyte or *(iii)* due to stress differences.

6. What passivation refers?

Ans. It refers to the formation of a thin, high resistance oxide layer which reduces the corrosion current by two or three orders of magnitude.

7. What are the main methods of corrosion prevention?

Ans. *(i)* Designing to avoid contact between dissimilar metals, *(ii)* use of inhabitors, *(iii)* metallic and non-metallic coatings, *(iv)* cathodic protection with a sacrificial anode or with an impressed voltage and *(v)* passivation by alloy additions.

8. What is the unit of rate of corrosion?

Ans. The rate of corrosion may be expressed as corrosion penetration rate, i.e. the thickness loss of material per unit time. Mills/year and millimeters/year are the common units for this parameter. Alternatively, rate is proportional to the current density associated with the chemical reaction.

9. What are the different forms of metallic corrosion?

Ans. *(i)* Uniform attack, *(ii)* galvanic corrosion, *(iii)* crevice corrosion, *(iv)* pitting, *(v)* intergranular corrosion, *(vi)* selecting leaching, *(vii)* erosion-corrosion, and *(viii)* stress corrosion.

10. Is oxidation of metallic materials by electrochemical action is possible?

Ans. Yes, in dry gaseous atmospheres.

OBJECTIVE QUESTIONS

1. Which one is the wrong anode-cathode combination
(1) Zinc-Iron (2) Nickel-Titanium
(3) Iron-Tin (4) Silver-zinc (4)

2. The metals which do not form adherent oxide film on surface are
(1) Gold and Silver (2) Iron and Steel
(3) Copper and Aluminium (4) Nickel and Titanium (1)

3. Which of the following elements added to iron to improve its oxidation resistance
(1) Zinc (2) Magnesium
(3) Chromium and aluminium (4) None of the above (3)

4. The corrosion by the mechanism of oxygen absorption can occur if
(1) Electrolyte is neutral or alkaline
(2) Electrolyte is neutral or acidic
(3) Electrolyte is neither neutral and nor acidic
(4) None of the above (1)

5. In order to form a protective oxide layer, the ratio of the volume of oxide formed to that of metal consumed should be
(1) greater than one (2) less than one
(3) much greater than one (4) none of the above (1)

6. The potential of a galvanic cell of copper (potential of +0.34 Volt) and aluminium (potential of –1.66 Volt) is
(1) +1.5 Volt (2) +1.66 Volt (3) –1.32 V (4) +2 Volt (4)

7. In the corrosion process by evolution of hydrogen
(1) the cathodes have larger area than anodes
(2) the anodes have larger area than cathodes
(3) the areas of anodes and cathodes are same
(4) none of the above (2)

8. The drawback of inorganic inhabitors is
(1) they are ineffective in alkaline environment
(2) they are ineffective in acidic environment
(3) before they become active, they have an induction time
(4) none of the above (2)

9. Pore-free coating is needed when
(1) coating and the protected metal have the same potential
(2) with respect to protected metal coating is base
(3) with respect to protected metal coating is noble
(4) none of the above is correct (3)

10. The highest percentage of chromium that can be added to steel is usually
(1) 25 (2) 18 (3) 15 (4) 10 (2)

11. Carbon percentage in cutley steel is
(1) 1.35 to 1.65 (2) 0.76 to 0.80 (3) 0.25 to 0.30 (4) 0.5 to 1.0 (3)

12. The stainless steels owe their resistance to the presence of
(1) Carbon (2) Sulphur (3) Manganese (4) Chromium (4)

13. In parabolic rate of oxidation, the oxide thickness is proportional to
(1) $t^{1/3}$ (2) $t^{1/2}$ (3) t^2 (4) t^3 (2)

13

Thermal and Optical Properties of Materials

SECTION A: THERMAL PROPERTIES

1. INTRODUCTION

The study of the response of a material to the application of heat is very important. When a solid absorbs energy in the form of heat, its temperature rises and its dimensions increases. The absorbed heat energy by the specimen may be transported to cooler regions if temperature gradient exist, and ultimately, the specimen may melt. Heat capacity, thermal expansion and thermal conductivities are few very important properties that are often critical in the practical and engineering applications of solids.

The theory of *lattice vibrations*, i.e. study of phonons and the *band theory of electrons* provides a useful insight for understanding the broad differences in thermal behaviour among the various classes of solid materials. In this section a brief account of thermal properties is presented.

2. HEAT CAPACITY

When a solid material is heated, it experiences an increase in temperature signifying that some energy has been absorbed. *Heat capacity* of a solid material is a property that is indicative of material's ability to absorb heat from the external surroundings; it represents the amount of energy required to produce a unit temperature rise. Mathematically, heat capacity (C) of a body is defined as

$$C = \frac{dQ}{dT} \tag{1}$$

where dQ refers to the quantity of heat transferred to the body and dT is the rise in temperature. Ordinarily, heat capacity (C) is specified per mole of material (e.g., J/mol-K, or cal/mol-K). Heat capacity per unit mass of a homogeneous substance is known as *specific heat* (c). [$c = C/M$, where M is the mass]. Specific heat has various units (J/kg-K, cal/gm-K). Specific heat is really a measure of the number of degrees of freedom of a system.

One can measure heat capacity of a body in a two ways, according to the environmental conditions accompanying the transfer of heat. One is the heat capacity at constant volume, $C_v = (\partial Q/\partial T)_v$ and the other is at constant external pressure, C_p. The magnitude of C_p is always greater than C_v; however, this difference is very slight for most solid materials at room temperatures and below. The heat capacity at constant volume C_v is the most useful quantity than the heat capacity at constant pressure C_p, because the former can be obtained immediately from the energy of the system.

The heat capacity depends upon the mass of the body, its chemical composition, thermodynamic state and the kind of process employed to transfer the heat.

Vibrational Heat Capacity

The principal mode of thermal energy assimilation in most solids is by the increase of vibrational energy of the atoms. We know that atoms in solid materials are constantly vibrating at very high frequencies and with relatively small amplitudes. We may note that the vibrations of atoms are not independent of one another, however, the vibrations of adjacent atoms are coupled by virtue of the atomic bonding. These vibrations of atoms in a solid are coordinated in such a way that travelling *lattice waves* are produced. These lattice waves may be though of as elastic waves or simply sound waves, having short wavelengths and very high frequencies, which propagate through crystal at the velocity of sound. The vibrational thermal energy for a material consists of a series of these *elastic waves*, which have a range of distributions and frequencies. Only certain energy values are allowed, i.e., the energy is said to be *quantized*, and a single quantum of vibrational energy is called a *phonon*. Obviously, phonon is a quantum of sound; the smallest unit of energy of a vibration corresponding to a sound wave. A phonon is analogous to a *photon* or light quanta. On occasion, the lattice vibrational waves themself are termed phonons.

The concept of phonon is particularly useful in describing the properties of solids where the idea of a gas of phonons can be used to understand the heat capacity. Phonons in a solid have a mean free path due to collisions among themselves via. the phonon-phonon interaction and also due to collisions with impurity atoms. In insulators heat is transmitted by phonons and the magnitude of the thermal conductivity is determined by the phonon mean free paths. The electron-phonon interaction is important in giving an attractive interaction between pairs of electrons in some solids that leads to *super-conductivity* (see chapter 16).

A phonon is a packet of sound waves which behaves like a massless particle with energy $h\nu$ and momentum h/λ, where h is Planck's constant, ν is the frequency of the vibration, and λ is the wavelength. The product of $\lambda\nu = V$, where V is the appropriate velocity of sound. We may note that this particle like aspect is only important at the atomic level because h, and hence $h\nu$, is very small. For everyday phenomena, it is irrelevant.

The thermal scattering of free electrons during electronic conduction in solids is by these lattice vibrational waves, and these elastic waves also participate in the transport of energy during thermal conduction.

The measured heat capacity of solids is usually made at some constant pressure P, such as atmospheric pressure, and is represented by C_p. The theoretical heat capacity is most often calculated for constant volume V, and is denoted by C_v. The difference $C_p - C_v$ is essentially the heat per degree required to expand the solid against its internal elastic force. The difference is given by

$$C_p - C_v = \frac{\alpha^2 TV}{\beta} \tag{2}$$

where α is the temperature coefficient of volume expansion (at constant pressure), T the temperature in °K, V is the volume of the solid and β the isothermal compressibility. The quantities represented by the symbols C_p and C_v are often referred to loosely as specific heats, although they are really heat capacities.

The heat capacity per k mol of a substance is called *molar heat capacity*, and it is obtained by multiplying C_v (or C_p) by the molecular weight.

Figure 13.1 shows the experimental variation of molar specific heat of few typical solids as a function of temperature. We may note that this behaviour is characteristic of all observed solids, the only difference being the temperature at which C_v starts to drop off. We may note the following facts from these curves:

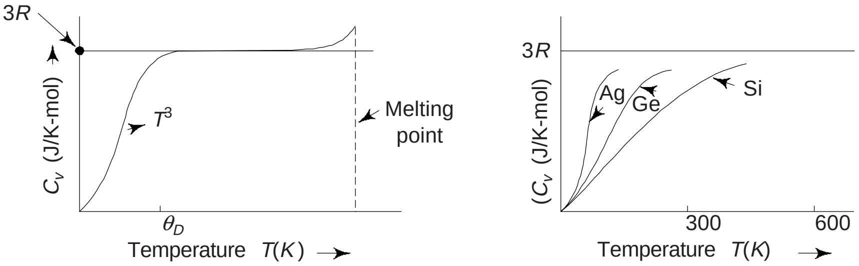
(i) At elevated temperatures C_v is flat, i.e. independent of temperature and equal to $3R$ per k-mol for all solids where R is gas constant. This is *Dulong-Petit law*. It holds to just below the melting point, at which C_v starts to rise.

(ii) At low temperatures C_v begins to fall-off and approches zero as $T \to 0$.

(iii) The approach to zero near $T = 0$ is by a T^3 law, i.e.

$$C_v = AT^3 \tag{3}$$

where A is temperature independent constant. Above what is called the *Debye temperature*, θ_D, C_v levels off and become essentially independent of temperature at a value $\approx 3R$. Thus even though the total energy



(a) (b)

Fig. 13.1 Specific heat of solids (a) General dependence on temperature (b) Experimantal curves for Ag, Ge and Si

of the material is increasing with temperature, the quantity of energy required to produce a one-degree temperature change is constant. The value of θ_D is below room temperature for many solid materials, and 25 J/mol-K (6 cal/mol-K) is a reasonable room temperature approximation for C_v. *Table* 13.1 presents experimental specific heats and other thermal properties for some materials.

Table 13.1 Thermal properties for some substances

Material	c_p *(J/kg-K)*	α_l $[(°C)^{-1} \times 10^{-6}]$	k *(W/m-K)*	L $[\Omega\text{-W}/(K)^2 \times 10^{-8}]$
		Metals		
Aluminium	900	23.6	247	2.20
Copper	386	17.0	398	2.25
Gold	128	14.2	315	2.50
Iron	448	11.8	80	2.71
Nickel	443	13.3	90	2.08
Silver	235	19.7	428	2.13
Tungsten	138	4.5	178	3.20
1025 Steel	486	12.0	51.9	—
316 Stainless steel	502	16.0	15.9	—
Brass (70Cu-30Zn)	375	20.0	120	—
Kovar (54Fe–29Ni-17Co)	460	5.1	17	2.80
Invar (64Fe–36Ni)	500	1.6	10	2.75
Super Invar (63Fe–32Ni–5Co)	500	0.72	10	2.68
		Ceramics		
Alumina (Al_2O_3)	775	7.6	39	—
Magnesia (MgO)	940	13.5[a]	37.7	—
Spinel ($MgAl_2O_4$)	790	7.6[a]	15.0[b]	—
Fused silica (SiO_2)	740	0.4	1.4	—
Soda-lime glass	840	9.0	1.7	—
Borosilicate (Pyrex) glass	850	3.3	1.4	—
		Polymers		
Polyethylene (high density)	1850	106–198	0.46–0.50	—
Polypropylene	1925	145–180	0.12	—
Polystyrene	1170	90–150	0.13	—
Polytetrafluoroethylene (Teflon)	1050	126–216	0.25	—

(*Contd.*)

Table 13.1 (*Contd.*)

Material	c_p *(J/kg-K)*	α_l $[(°C)^{-1} \times 10^{-6}]$	*k* *(W/m-K)*	*L* $[\Omega\text{-}W/(K)^2 \times 10^{-8}]$
Phenol-formaldehyde, phenolic (Bakelite)	1590–1760	122	0.15	—
Nylon 6,6	1670	144	0.24	—
Polyisoprene	—	220	0.14	—

[a] Value measured at 100°C.

[b] Mean value taken over the temperature range 0–1000°C.

Other Heat Capacity Contributions

There also exist other energy-absorptive mechanisms that can add to the total heat capacity of a solid. In most instances, however, these are minor relative to the magnitude of vibrational, i.e. phonon contribution. There is an *electronic contribution* in that electrons absorb energy by increasing their kinetic energy. However, this is possible only for free electrons-those that have been excited from filled states to empty states above the Fermi energy. In metals, only electrons at states near the Fermi energy are capable of such transitions, and these represent only a very small fraction of total number. We may note that an even smaller proportion of electrons experiences excitations in *insulating* and *semiconducting* materials. Obviously, this electronic contribution is ordinarily insignificant, except at temperatures near 0 K.

Moreover, in some materials other energy-absorptive processes occur at specific temperatures for e.g., the randomization of electron spins in a ferromagnetic material as it is heated through its curie temperature T_C. A large spike is produced on the heat capacity vs. temperature curve at the temperature of this transformation.

Now, we evaluate C_v theoretically, and compare the value so obtained with experimental results. Let us consider a crystal composed of N-atoms, which are held together in a periodic array. We may assume these atoms to be free to vibrate about their equilibrium sites subject to constraining force which to a first approximation obey *Hooke's law*.

3. THEORETICAL MODELS

3.1 Dulong-Petit Law (Classical Model)

Dulong and Petit observed in 1819 that, although the specific heats of the solid elements at room temperature differ widely from one another, the atomic heats are nearly all the same, the values being about 26J/°C (6.3 cal/°C). According to kinetic theory of gases, the thermal energy of an ideal monoatomic gas is the same as the kinetic energy. From this, it was deduced that the atomic heat of such a gas is $3R/2$ (R = 8.3 J/°C = 2 Cal/°C). The thermal energy of a solid, however, is the energy of the harmonic motion of the atoms, and this, on the average, is half kinetic and half potential. From this it was supposed that $3R/2$ is the atomic heat arising from the mean kinetic energy, and $3R/2$ that arising from the mean potential energy, yielding a total atomic heat of 3R or 24.9 J/°C (= 6 cal/°C) This result is true only at room temperature and above. Obviously, the classical model fails to explain the observed fall in C_v at low temperatures.

3.2 Einstein Theory

Einstein in 1907 sought to show that the observed failure of the classical model, which gives $C_v = 3R$ for the atomic heat, could be explained in terms of the quantum hypothesis.

A so-called Planck oscillator can absorb or emit radiation only in integral amounts $\hbar\omega$, where $\hbar = h/2\pi$, and ω is the natural angular frequency of the oscillator. The temperature is introduced by considering the mean value of the energy ε of such an oscillator, i.e.

$$\bar{\varepsilon} = \frac{\hbar\omega}{\left[\exp(\hbar\omega/kT) - 1\right]} \tag{4}$$

Einstein's theory assumes that each atom of the solid oscillates with the same frequency ω_E, (Einstein's frequency) and that this the frequency observed in infrared absorption studies in crystals. Each atom vibrates in three dimensions and therefore has the mean energy $3\bar{\varepsilon}$. The energy E of the solid is $3N\,\bar{\varepsilon}$, if it contains Avogadro's number of atoms N. The quantum hypothesis then leads to

$$E = \frac{3N\hbar\omega_E}{\left[\exp\left(\hbar\omega_E/kT\right)-1\right]} \tag{5}$$

A parameter called the Einstein's characteristic temperature θ_E is defined by equating one quantum of energy $\hbar\omega_E$ to the classical energy kT of an oscillator and denoting the particular value of T obtained in this manner by θ_E. According to Einstein, the thermal energy Q of the solid is just the energy E of vibration, so that

$$C_v = \frac{dQ}{dT} = \frac{dE}{dT}$$

This yields the Einstein's formula of specific (atomic) heats,

$$C_v = \frac{3Ry^2\,e^y}{(e^y-1)^2} \tag{6}$$

where $y = \hbar\omega_E/kT = \theta_E/T$ and $Nk = R$. A plot of $C_v/3R$ versus T/θ_E is shown in Fig. 13.2. At $T = \theta_E$, the value of $C_v/3R$ is 0.92, which means that, at this temperature C_v has 92% value of the Dulong-Petit value. Above this temperature, C_v approaches $3R$ with increasing temperature. Below this temperature, C_v decreases to zero, practically vanishing at $T < 0.1\,\theta_E$. Einstein's theory thus concludes that C_v is temperature-dependent. Furthermore, the observation that $C_v/3R = 0.31$ for diamond at $T = 331$ K is explained by stating that diamond has a value of θ_E equal to about 1800 K, which corresponds to an infrared wavelength of 11 μm.

Einstein's theory fails at low temperatures. Experimentally, it was found that C_v remains appreciable still at $T < 0.1\,\theta_E$ for all substances examined, whereas in accordance with Einstein's theory C_v practically vanishes below $T = 0.1\,\theta_E$. However, experimental data reveals that C_v approaches zero at 0 K.

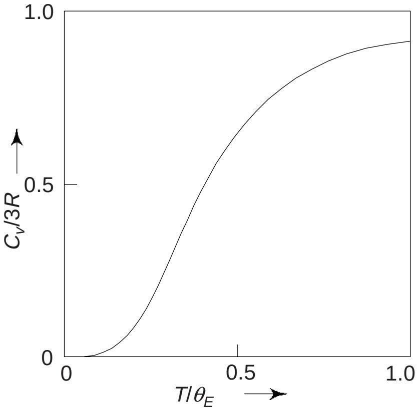


Fig. 13.2 Einstein's specific heat curve

3.3 Debye Theory

The next advance in the theory of specific heat began with the suggestion of Madelung and Sutherland that the Einstein frequency is equivalent not only to the infrared absorption frequency of the crystal but also to the frequency of the shortest lattice vibration which can propagate through crystal. This wave has a wavelength of about twice the interatomic distance. Since waves of longer wavelength can also propagate through the crystal, Madelung made the further suggestion that a whole spectrum of frequencies should be used in computing C_v rather than just the single frequency ω_E.

Debye made two assumptions: *(i)* that the solid is a continuous medium and *(ii)*, that the total number of waves is equal 3N, where N is the number of atoms in the crystal. This assumption implies that the solid is not really continuous after all and that the shortest permissible wavelengths are those of about two interatomic distances. The restriction is expressed mathematically as

$$\int_0^{w_D} \rho(\omega)\, d\omega = 3\text{N} \tag{7}$$

where $\rho(\omega)\, d\omega$ is the number of waves (or modes) with frequencies between ω and $\omega + d\omega$ is solid. The energy associated with each of these waves is that of a Planck oscillator, so that one obtains for the total energy, E,

$$E = \int_0^\infty \frac{\rho(\omega)\,\hbar\omega d\omega}{e^{\hbar\omega/kT} - 1} \tag{8}$$

$\rho(\omega)$ is computed in a manner analogous to that employed in black body radiation, resulting in

$$\rho(\omega) = \frac{V}{2\pi^2}\left(\frac{1}{v_l^3} + \frac{2}{v_t^3}\right)\omega^2 \tag{9}$$

where v_l is the velocity of longitudinal sound waves, v_t that of transverse waves and V the volume of the solid.

The Debye frequency ω_D is the maximum allowable frequency. Thus for $\omega > \omega_D$, $\rho(\omega)$ is zero and the value of integral in Eq. (7) above this limiting frequency is zero. This allows the upper limit in Eq. (8) to be replaced by ω_D. It is customary to replace the Debye frequency ω_D by the Debye characterisitc temperature θ_D, defined by the relation

$$k\theta_D = \hbar\omega_D \tag{10}$$

Finally Debye obtained the expression for specific heat as

$$C_v = 9R\left(\frac{T}{\theta_D}\right)^3 \int_0^{x_D} \frac{e^x\, x^4}{(e^x - 1)^2}\, dx \tag{11}$$

with $x = \hbar\omega/k_B T$. As T approaches infinity, i.e. for $x_D << 1$, $C_v \approx 3R$. We thus obtain the Debye classical T^3 law results. At low temperatures for $T << \theta_D$, the upper limit goes to infinity. Then the integral in Eq. (11) is equal to $4\pi^4/15$. Hence

$$C_v = \frac{12\pi^4}{5} R\left(\frac{T}{\theta_D}\right)^3 \tag{12}$$

Thus C_v varies as T^3 as observed experimentally. Figure 13.3 shows a plot of $C_v/3R$ versus T/θ_D as given by the Debye theory. The experimental results for many materials fit a single curve.

The courses of the two curves shown in Figs. 13.1 and 13.3 are quite similar for T above about 0.2 θ_D. The critical test distinguishing between two theroies must therefore be made at temperatures below about 0.1 θ_D, where the Debye T^3 should hold. This T^3 law failed for metals. The reason for this failure is now understood, for Sommerfeld's theory of metals shows that the conduction electrons can make an important contribution to the heat capacity.

According to Sommerfeld, there must be a linear term in the temperature in the expression for C_v in order to account for the electron contribution. Thus, we obtain

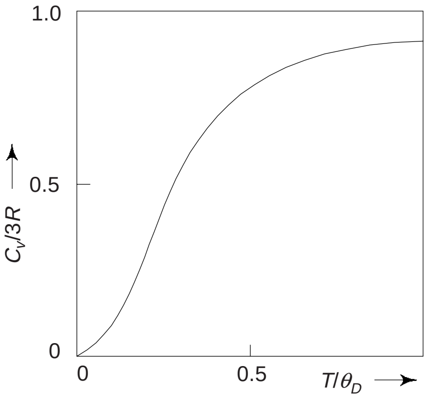


Fig. 13.3 Debye specific heat curve

$$C_v = \gamma T + \left(\frac{12\pi^4 R}{5}\right)\left(\frac{T}{\theta_D}\right)^3 \tag{13}$$

The coefficient γ in the electron term is sometimes called the Sommerfeld gamma. To analyse low temperature C_v data for metals, C_v/T versus T^2 is plotted. According to Eq. (13) this should give a straight line of slope $12\pi^4 R/5\theta_D^3$ and of intercept γ on the C_v/T axis.

4. THERMAL EXPANSION

Most solid materials expand upon heating and contract on cooling. A solid material which expands equally in three mutually orthogonal directions is called *thermally isotropic*.

The change in length with temperature for solid material can be expressed as

$$\frac{l_1 - l}{l} = \alpha (T_1 - T) \tag{13a}$$

or

$$\frac{\Delta l}{l} = \alpha \Delta T \tag{13b}$$

where l and l_1 represent, respectively initial and final lengths with the temperature change from T and T_1. The parameter α is called the *linear coefficient of thermal expansion.* It is a material property that is indicative of the extent to which a solid material expands upon heating, and has units of reciprocal temperature [(°C)$^{-1}$]. No doubt, heating and cooling affects all dimensions of a body, with a resultant change in volume. Volume changes with temperature can be expressed as

$$\frac{\Delta V}{V} = \gamma \Delta T \tag{14}$$

where ΔV and V are the volume change and the original volume, respectively, and γ is the volume coefficient of thermal expansion. In many materials, the value of γ is anisotropic; i.e., it depends on the crystallographic direction along which it is measured. For materials in which the thermal expansion is isotropic, γ is approximately 3α.

The thermal expansion of the solid materials has its origin in the lattice vibrations and these vibrations increases as the temperature rises.

From atomic point of view, thermal expansion is reflected by an increase in the average distance between the atoms. One can best understand this phenomenon with the help of the *potential energy-versus-interatomic spacing curve* for a solid material (Fig. 13.4(a)). From this we can show that the thermal energy increases from level 1 to level 2, the amplitude increases from r_1 to r_2. We may note the form of this curve is represented by

$$U = -\frac{A}{r^m} + \frac{B}{r^n} \tag{15}$$

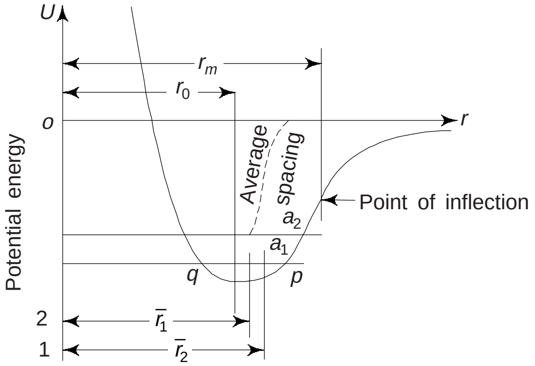


Fig. 13.4(a) **Potential energy-versus-interatomic distance curve for solid materials exhibiting the increase in interatomic separation with rising temperature**

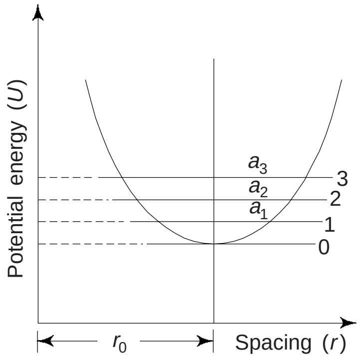


Fig. 13.4(b) **For a symmetric potential energy vs. interatomic distance curve, there is no increase in interatomic separation with rise in temperature (average spacing r_0 will remain same)**

According to the definition of thermal expansion (α), we have

$$\alpha = \frac{1}{l}\frac{dl}{dT} \tag{16}$$

where l is the length of the solid material and T the temperature. $dl/l = d\varepsilon$ is the incremental strain, where ε is strain. Thus

$$\alpha = \frac{d\varepsilon}{dT} \tag{17}$$

The strain is referred to equilibrium spacing r_0. Now, the average strain in the spacing of two atoms is

$$\bar{\varepsilon} = \frac{\bar{r} - r_0}{r_0} \tag{18}$$

where $\bar{r}$ is average spacing (Fig. 13.4a). The maximum and minimum strains in the same spacing, corresponding to two extreme positions p and q are

$$\varepsilon_p = \frac{r_p - r_0}{r_0} \quad \text{and} \quad \varepsilon_q = \frac{r_q - r_0}{r_0} \tag{19}$$

If the oscillations are approximated by sinusoidal curve, then the average strain is obtained as

$$\varepsilon_a = \frac{|\varepsilon_p| + |\varepsilon_q|}{2} \tag{20}$$

However,

$$\bar{\varepsilon} = \frac{m + n + 3}{4}\varepsilon_a^2 \tag{21}$$

Thermal expansion is really due to the asymmetric curvature of this potential energy trough, rather than the increased atomic vibrational amplitudes with rising temperature. If the potential energy curve were symmetric (Fig. 13.4b), there would be no net change in interatomic reparation and, consequently, no thermal expansion.

Thermal coefficient of expansion of solid materials is related to the specific heat and the melting point of the solid. Empirical data reveal that the total volume change of elements on heating between absolute zero and melting point temperature is approximately constant. Obviously, the metals with higher melting point will have lower coefficient of expansion. This means that the metals which have deep potential wells or energy curves more symmetric about the central line passing through minimum energy point will expand less with rise in temperature. Obviously, such metals will have greater bond energy and hence will melt at higher temperature. Figure 13.5 shows the volume expansivity of few metals as a function of temperature. We note that metal like tin has a low melting point (M.P.) and expands rapidly with rise in temperature. On the otherhand a metal like platinum having high M.P. has a low rate of thermal expansion.

It is observed that most of the commercial oxides exhibit uniform thermal expansion, except for silica (SiO_2) and Zirconia (ZrO_2) (Fig. 13.6). These oxides show irregular expansion due to polymorphic transformations. In comparison to metals, organic polymers such as plastics and rubbers have several-fold higher thermal coefficients of expansion because of lower softening temperatures. This makes it difficult to mould parts of these materials to close tolerances ±0.025 mm. Such tolerances are very common in metals. This is why the synthetic resins used as cements find limited applications. By the addition of filler materials of lower expansivity such as glass fibre, asbestos and inorganic powders like alumina, the thermal expansivity of plastics may be reduced. One can obtain the intermediate values of expansion coefficients of filler and matrix materials in cloth reinforced plastics. The expansion in such materials takes place in the plane of the reinforcement. Though proper compounding with filters, it is possible to produce certain plastics, e.g. phenolic and polysters having coefficients of expansion almost approaching to the values for metals.

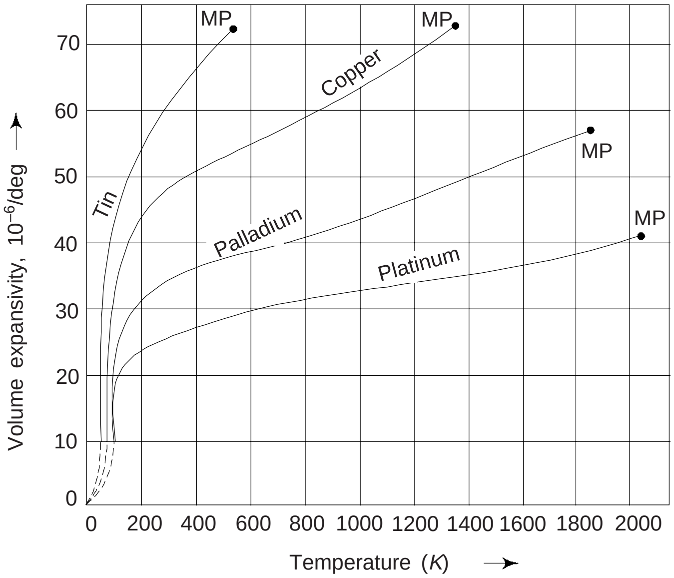


Fig. 13.5 Volume expansivity of few metals with temperature (MP → melting point)

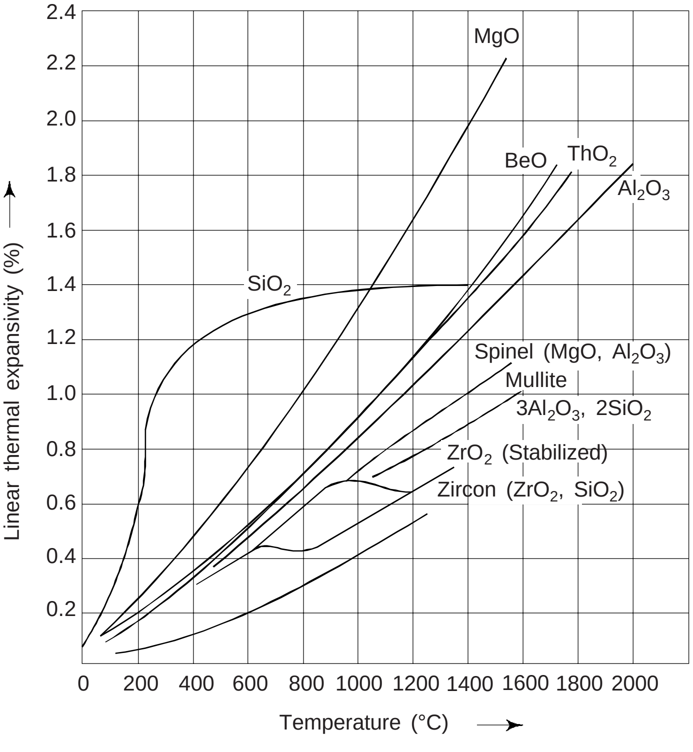


Fig. 13.6 Thermal expansion curves for some refractory oxides

Concluding, we can say that for each class of materials (metals, ceramics, and polymers), the greater the atomic bonding energy, the deeper and more narrow potential energy trough. As a result, the increase in inter-atomic separation with a given rise in temperature will be lower, giving a small value of linear coefficient of thermal expansion (Table 13.1).

Metals

α-for some of the common metals range between about 5×10^{-6} to 25×10^{-6}/°C. For some engineering applications, a high degree of dimensional stability with temperature fluctuations is essential. This has resulted in the development of a family of Fe-Ni, and Fe-Ni-Co alloys that have $\alpha \sim 1 \times 10^{-6}$/°C. One such alloy, namely *Kovar* (Table 13.1), has been prepared to have expansion characteristics close to those of borosilicates (or Pyrex) glass; when joined to Pyrex and subjected to temperature variations, one finds that thermal stresses and possible fracture at the junction are avoided. Other low expansion alloys are Invar and Super-invar (Table 13.1).

Ceramic

In many ceramic materials relatively strong interatomic bonding forces are found. This is reflected in comparatively low α; values typically range between about 0.5×10^{-6}/°C and 15×10^{-6}/°C. For non-crystalline ceramics and also those having cubic crystal structures α is isotropic. Otherwise, it is anisotropic; and, in reality, some ceramic materials, upon heating, contract in some crystallographic directions while expanding in others. α for inorganic glass depends upon the composition. α for fused silica (high purity SiO_2 glass) has a small value $\sim 0.4 \times 10^{-6}$/°C. One can explain this by a low atomic packing density such that interatomic expansion produces relatively small macroscopic dimensional changes.

Ceramic materials that are subjected to temperature changes must have coefficients of thermal expansion that are relatively low, and also isotropic. Otherwise, these brittle materials may experience fracture as a consequence of non-uniform dimensional changes in what is called *thermal shock*.

Polymers

Upon heating, some polymeric materials experience very large thermal expansions upon heating as indicated by the values of α that range from $\sim 50 \times 10^{-6}$ to 400×10^{-6}/°C. Linear and branched polymers show highest α-values because the secondary intramolecular bonds are weak, and moreover, there is a minimum of cross linking. The magnitude of expansion coefficient diminishes with increased crosslinking. The lowest α values are found in thermosetting network polymers, e.g. phenol-formal-dehyde in which the bonding is almost entirely covalent.

5. THERMAL CONDUCTIVITY (K)

This is the phenomenon by which heat is transported from high-to-low temperature regions of a substance. The property that characterizes the ability of a material to transfer heat is termed as the *thermal conductivity*. One can best define it in terms of the expression

$$\theta = -K\,\frac{dT}{dx} \tag{22}$$

where $\theta \rightarrow$ *heat flux* or heat flow, per unit time per unit area (area being taken as that perpendicular to the flow direction), $\frac{dT}{dx} \rightarrow$ *temperature gradient* through the conducting medium and $K \rightarrow$ the thermal conductivity (Fig. 13.7).

The units of Q and K are W/m^2 and W/(m-°K) [k cal/(m-s-k)], respectively. We may note that Eq. (22) is valid only for steady-state heat flow, i.e., for situations in which the heat flux does not change with time. Minus sign in Eq. (22) indicates that the minus sign in the expression indicates that the direction of heat flow is from hot to cold, or down the temperature gradient.

5.1 Mechanisms of Heat Conductions

In solid materials heat is transported by both *lattice vibration waves* (phonons) and *free electrons.* The total thermal conductivity is the sum of the two contributions,

$$K = K_l + K_e \tag{23}$$

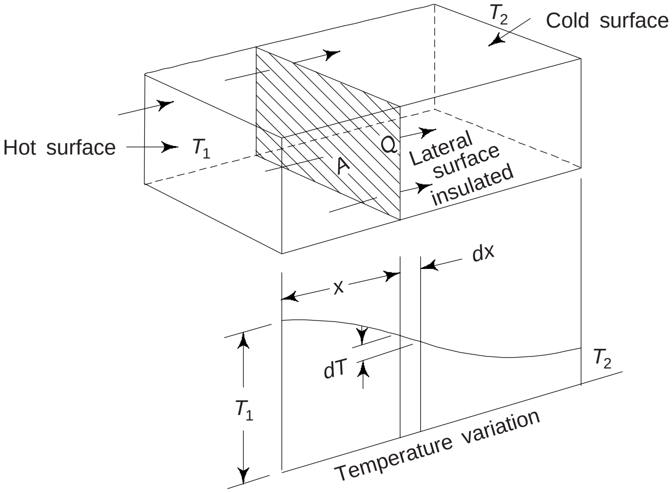


Fig. 13.7 Thermal conductivity (K), temperature gradient $\left(\frac{dT}{dx}\right)$ and flow of heat (Q)

where K_l and K_e represent the lattice vibration and electron thermal conductivities, respectively, usually one or the other predominates. In insulating materials, the conduction is by lattices waves; in pure metals, the lattice contribution is negligible and the heat conduction is primarily due to electrons. In many alloys, impure metals, and semiconductors, both conduction mechanisms contribute.

The thermal energy associated with phonons or lattice waves is transported in the direction of their motion. The K_l contribution results from a net movement of from high-to-low temperature regions of a material body across which temperature gradient exists.

Free or conducting electrons participate in electronic thermal conduction. After gaining kinetic energy in a hot region free electrons then migrate to colder parts of the body, where some of this kinetic energy is transferred to the atoms themselves (as vibrational energy) as a consequence of collisions with phonons or other imperfections in the crystal. We may note that the relative contribution of K_e to the total thermal conductivity increases with increasing free electron concentrations, since more free electrons are available to participate in this heat transference process.

Metals

In pure metals, the electronic component accounts for nearly all the heat conducted, while the lattice component, in most cases, is negligible. The electronic thermal conductivity is related to the electric conductivity through the mean free path. In certain temperature regions the value of the mean free path for both thermal and electrical conduction can be assumed to be the same. The thermal conductivity of a metal can be expressed by a relation given below

$$K = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 \frac{ne^2\,\lambda}{2mv_0} \tag{24}$$

where λ is mean free path, $k \rightarrow$ Boltzman constant, $m \rightarrow$ electron mass, $v_0 \rightarrow$ initial velocity of electron and $e \rightarrow$ electronic charge. Similar expression is used for the electrical conductivity. The relation of heat conductivity (K) and electrical conductivity (σ) is

$$\frac{K}{\sigma} = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 T \tag{25}$$

or

$$\frac{K}{\sigma T} = L \tag{26}$$

where $L = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 = 2.44 \times 10^{-8}\ \Omega\text{–W/K}^2$ is constant, usually known as *Lorentz constant*. This constant is independent of temperature and the same for all metals if the heat energy is transported entirely by free electrons. The relation (26) is known as *Wiedemann-Franz law* and the ratio $\frac{K}{\sigma T} = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2$ is known as Wiedemann-Franz ratio. Here the free electron theory of metals is assumed. The thermal conductivities of several of the common metals generally range between about 20 and 400 W/m-K. Figure 13.8 shows thermal conductivities versus temperature curves for some metals.

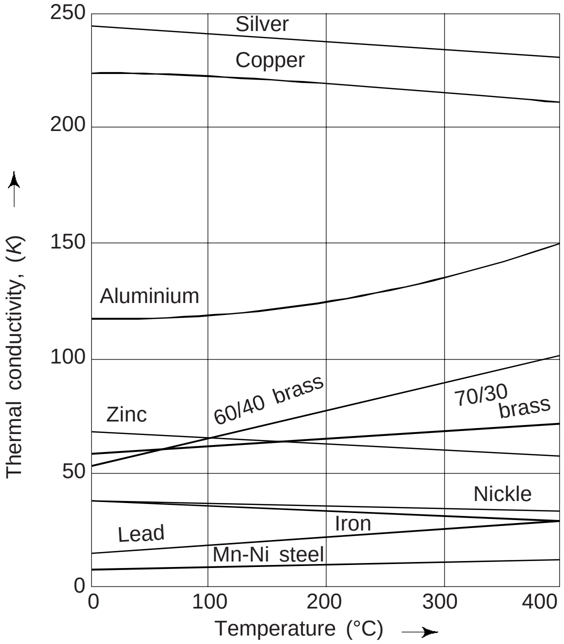


Fig.13.8 Dependence of thermal conductivity on temperature for some metals

The change in thermal conductivities of pure metals, e.g. silver, copper and aluminium and brass are only nominal. However, at lower temperatures, ranging from 0 to 100 K, the thermal conductivity increases as temperature is reduced until a pack is reached after which the conductivity reduces to zero at 0 K (Fig. 13.9).

Scattering of electrons at grain boundaries and impurity or foreign particles reduces the thermal conductivity. This is why the single crystals and pure metals have higher thermal conductivity than polycrystalline metals and alloys respectively.

Alloying metals with impurities results in a reduction in a thermal conductivity (K). The impurity atoms, especially if in solid solution, act as scattering centers, lowering the efficiency of electron motion. Figure 13.10 displays this effect for Cu-Zn alloys. Also, stainless steels which are highly alloyed, become relatively resistive to heat transport.

In the case of semiconductors, such as high purity Ge and Si, phonons are primarily responsible for the thermal conduction. The form of thermal conductivity vs. temperature curves is similar to that for single crystals of dielectric materials.

Ceramics

Materials which do not have free electrons (non-metals), heat is transported by lattice vibrations. The transfer of heat through this mechanism is mush lower and slower and materials which do not have free

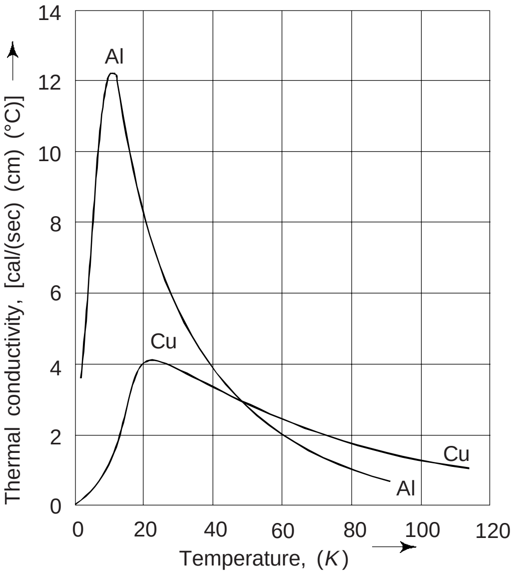


Fig. 13.9 **Thermal conductivity of Al and Cu at low temperatures**

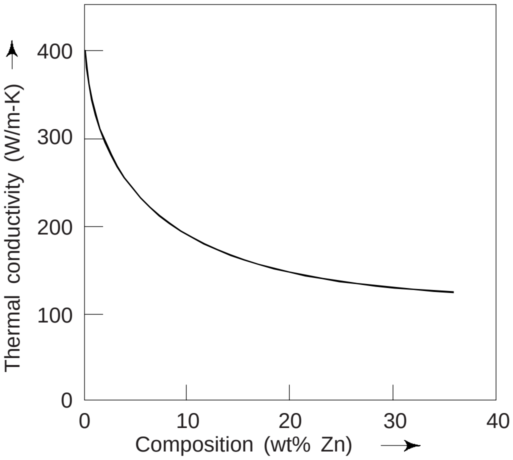


Fig. 13.10 **Thermal conductivity versus composition for Cu-Zn alloys curve**

electrons can serve as *thermal insulators,* e.g. ceramics. In these materials the heat is transmitted through *elastic thermal waves.* We can represent the wave characteristic by particle characteristic if the solid is assumed to be built of phonon gas. With this assumption, the thermal conductivity of such materials can be expressed as

$$K = \frac{1}{3} c v \lambda \tag{27}$$

where $c \rightarrow$ heat capacity of phonons, $v \rightarrow$ average particle velocity of phonons and $\lambda \rightarrow$ mean free path of phonons.

We may note that an *ideal crystal* will have infinite conductivity. In an ideal crystal lattice the atomic vibrations are harmonic. Any number of the lattice waves may be superimposed without interference. This is equivalent of phonons tranvelling over long distances without scattering, making mean free path λ, large.

In actual crystals, the mutual interference of waves is likely to occur resulting in anharmonicity of lattice vibration. Impurity atoms and defects like dislocations and vacancies also interfere with waves or movement of phonons causing scattering. This scattering and hence anharmonicity of lattice vibration diminishes with decreasing temperature. This causes thermal conductivity to rise with decreasing temperature. The thermal conductivity of amorphous materials depends mainly on c as the mean free path and particle velocity are not affected by rise in temperature.

The room temperature thermal conductivities for ceramics range between approximately 2 and 50 W/m-K. For amorphous materials, glasses and polycrystalline materials, λ for the phonons is generally small, of the order of the distance between atoms, and is independent of temperature. The thermal conductivities for these materials is much less than that for the insulating materials above; for e.g. quartz (fused) glass, sintered alumina.

The thermal conductivity of most ceramic materials normally diminishes with increasing temperature, at least at relatively low temperatures (Fig. 13.11). Figure indicates that the thermal conductivity begins to increase at higher temperatures, which is due to radiant heat transfesr; significant quantities of infrared radiant heat may be transported through a transparent ceramic materials. The efficiency of this process increases with temperature.

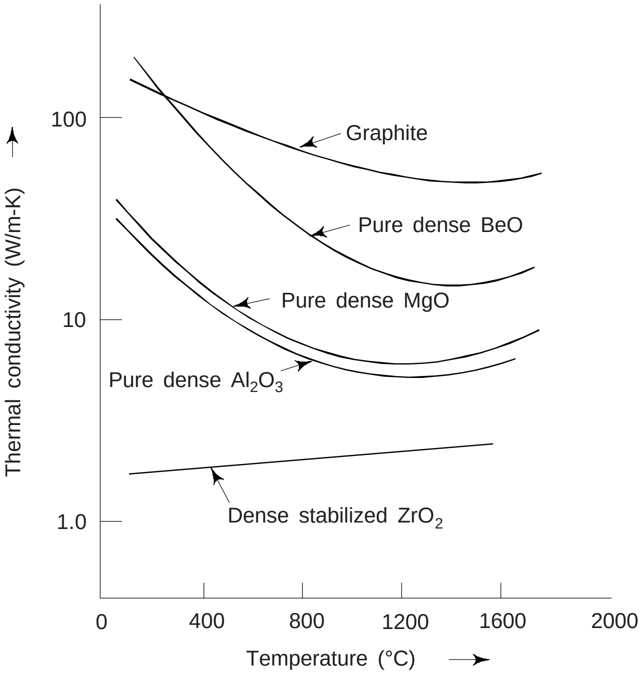


Fig. 13.11 Thermal conductivity vs. temperature curves for several ceramic materials

It is interesting to note that *porosity* in ceramic materials may have a dramatic influence on thermal conductivity; increasing the pore volume will, under most circumstances, result in a reduction of thermal conductivity. Interestingly, many ceramics that are used for thermal insulations are porous. Heat transfer across pores is ordinarily slow and inefficient. Internal pores normally contain still air, which has an extremely low thermal conductivity (~0.02 W/m-K). Moreover, gaseous convection within the pores is also comparatively ineffective.

Polymers

For most polymers, thermal conductivity are on the order of 0.3 W/m-K (Table 13.1). Energy transfer for these materials is accomplished by the vibration and rotation of the chain molecules. The magnitude of the thermal conductivity depends on the degree of crystallinity; a polymer with a highly crystalline and ordered structure will have a greater conductivity than the equivalent amorphous material. This is because of more effective coordinated vibration of the molecular chains for the crystalline state.

Polymers are often utilized as thermal insulators due to their low thermal conductivities. One can further enhance the insulating properties of polymers like ceramics by the introduction of small pores, which are ordinarily introduced by foaming during polymerization (see chapter 17). Foamed polystyrene (styroform) is usually used for drinking cups and insulating chests.

Superconductors

In superconductors at temperature below critical temperature, T_C, the electronic conduction is reduced; at sufficiently low temperatures, the thermal conductivity becomes entirely due to lattice waves and is similar to the form of the thermal conductivity of an insulating material. In Fig. 13.12 the temperature dependence of thermal conductivity of lead is shown; here K_n is the thermal conductivity with the material in the normal state, and K_s in the superconducting state.

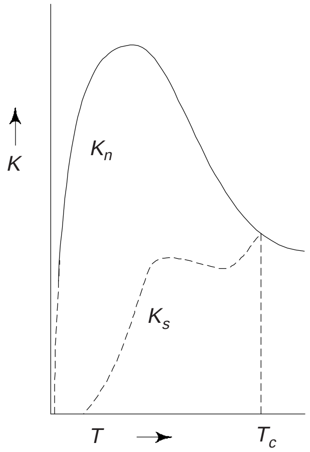


Fig. 13.12 Temperature dependence of thermal conductivity of Pb. K_n → normal state thermal conductivity K_s → superconducting state thermal conductivity

6. REFRACTORIES

The ability of a material to withstand the action of heat without appreciable deformation or softening is called as *refractories*. Obviously, this property is measured by the temperature at which the refractory material softens. Usually, the softening or fusion temperature of refractories is determined under load.

Refractories are capable of withstanding high temperatures. Commercially available most refractories are complex solids. They are usually made of Si, Al, Mg, Ca and Zr. As impurities, other elements may also present. In addition to oxides refractories the present day industries mostly depend upon crystalline oxides, carbides, borides and sulfides. Combination of these materials with metals have resulted into *ceramics*. Now, we will study various refractory materials:

(i) Ceramics: The salient properties of *refractory* ceramics include the capacity to withstand high temperatures without melting or decomposing, and the capacity to remain unreactive and inert when exposed to severe environments. In addition, the ability of ceramics to provide insulation is often an important factor. Refractory materials are marketed in a variety of forms but *bricks* are found to be most common. Typical applications of these refractories include furnace linings for metal refining, glass manufacturing, metallurgical heat treatment and power generation.

The performance of a refractory cermics, to a large extent, depend on its composition. on this basis, One can have several classifications namely, fireclay, silica, basic and special refractories, compositions for a number of commercial refractories are given in *Table* 13.2. For several commercial materials, the raw ingradients consist of both large (or grog) particles and fine particles, which may have different compositions. Normally, for several commercial materials, the raw ingradients consists of both large (or grog) particles and fine particles, which may have different compositions.

Table 13.2 Five common ceramic refractories

	Composition (wt%)							*Apparent*
Refractory type	*Al_2O_3*	*SiO_2*	*MgO*	*Cr_2O_3*	*Fe_2O_3*	*CaO*	*TiO_2*	*porosity (%)*
Fireclay	25-45	70-50	0-1	—	0-1	0-1	1-2	10-25
High-alumina fireclay	90-50	10-45	0-1	—	0-1	0-1	1-4	18-25
Silica	0.2	96.3	0.6	—	—	2.2	—	25
Periclase	1.0	3.0	90.0	0.3	3.0	2.5	—	22
Periclase-chrome ore	9.0	5.0	73.0	8.2	2.0	2.2	—	21

Porosity is one microstructural variable that must be controlled to produce a suitable refractory brick. With porosity reduction, strength, load bearing capacity, and resistance to attack by corrosive materials all increase. At the same time, thermal insulation characteristics and resistance to thermal shock are diminished. Obviously, the optimum porosity depending on the conditions of service have to be achieved.

Fireclay Refractories

The primary ingredients for these are high purity fireclays, alumina and silica mixtures usually containing 25 and 45 Wt% alumina. Over this composition range the highest temperature possible without the formation of liquid phase is 1587°C. Below this temperature the equilibrium phases present are mullite and silica (cristobalite). The presence of a small amount of a liquid phase may be allowable without compromising mechanical integrity during refractory service use.

Fireclay bricks are used basically in furnace construction, to confine hot atmospheres, and to thermally insulate structural members from excessive temperatures.

Silica Refractories (Acid Refractories)

The primary ingradient for these is silica. These materials have high-temperature load-bearing capacity, are widely used in arched roofs of steel-and glass making furnaces; for these applications, temperatures ~1650°C may be realized. Under these conditions some small portion of this brick will actually exist as a liquid.

These refractory materials are also resistant to slags that are rich in silica, called acid slags and are used usually as containment vessels for them.

Basic Refractories

The refractories that are rich in periclase, or magnesia (MgO), are called basic and they may also contain Ca, Cr, and Fe compounds. The presence of silica is deleterious to their high-temperature performance. These refractories are especially resistant to attack by slags containing high concentrations of Mgo and CaO. These find extensive use in some steel-making open hearth furnaces.

Special Refractories

There are yet other ceramic materials, e.g. relatively high-purity oxide materials that are used for rather specialized refractory applications. Many of these oxides may be produced with very little porosity. Included in this group of oxides are alumina, silica, magnesia, beryllia (BeO), Zirconia (ZrO_2), and mullite ($3Al_2O_3$-$2SiO_2$). Others include carbide compounds, in addition to carbon and graphite. Carbon and graphite are very refractory, but find limited applications as they are susceptible to oxidation at temperatures in excess of about 800°C. Silicon carbide (SiC) has been used for electrical resistance heating elements, as a crucible material, and in internal furnace components. These specialized refractories are relatively costly.

7. THERMAL STRESSES

These are the stresses induced in a material body as a result of changes in temperature. These stresses can lead to fracture or undesirable plastic deformation.

To understand the origins and nature of thermal stresses, let us consider a homogeneous and isotropic solid rod that is heated or cooled uniformly; i.e., no temperature gradients are imposed. For free expansion or contraction, the rod will be stress free. If however, axial motion of the rod is restrained by rigid end supports, thermal stresses will be introduced. The magnitude of the stress σ resulting from a temperature change from T to T_f is expressed as

$$\sigma = E\ \alpha\ (T - T_f) = E\ \alpha\ \Delta T \tag{28}$$

where α is the linear coefficient of thermal expansion and E is the modulus of elasticity. Upon heating ($T_f > T$), the stress ($\sigma < 0$) is compressive, as the rod expansion has been constrained. However, if the rod specimen is cooled ($T_f < T$), a tensile stress ($\sigma > 0$) will be imposed. Moreover, the stress in Eq. (28) is the same as the stress that would be required to elastically compress (or elongate) the rod back to its original length after length after it has been allowed to expand (or contract) freely with the $T - T_f$ temperature change.

We may note that thermal stresses may be established as a result of *temperature gradients* across a body, which are frequently caused by rapid heating or cooling, in that the outside changes temperature more rapidly than the interior; differential dimensional changes serve to restrain the free expansion or contraction of adjacent volume elements within the specimen. For example, upon heating, the exterior of a specimen is rather hotter and, therefore, will have expanded more than the interior regions. Obviously, compressive surface stresses are induced and are balanced by tensile interior stresses. However, the interior-exterior stress conditions are reversed for rapid cooling such that the surface is put into a state of tension.

8. THERMAL FATIGUE

The change of temperature imposed upon a body of the specimen produces alternate thermal gradients and hence changing stresses. These changing stresses have the effect of creating thermal fatigue in the specimen. The stresses generated by changing temperature are often high in magnitude creating plastic deformation. Under such conditions, the material fails due to low cycle thermal fatigue. The failure of the material is further hastened because of bad effects of temperature on mechanical strength.

9. THERMAL SHOCK

Thermal shock is generation of sudden and high stresses due to sudden and severe changes in temperature. Thermal shocks created by cooling are much more dangerous than those from heating as this process creates tensile stresses on the surface. Surface of any material is weaker than inside surface of the material for various reasons. For ductile metals and polymers, alleviation of thermally induced stresses may be accomplished by plastic deformation. However, the non-ductility of most ceramics enhances the possibility of brittle fracture from these stresses. Brittle materials are weak in tension.

The capacity of a material body to withstand the thermal stresses brought about by sudden and severe temperature changes is termed its *thermal shock resistance* (TSR) and usually denoted by P. P will be function of elastic properties and thermal conductivity of the material and expressed as

$$P \ (TSR) \approx \frac{\sigma_f K}{E\alpha} = \frac{\sigma_u}{E\alpha}\sqrt{\frac{K}{\rho C}} \qquad (29)$$

where σ_f is high fracture strength, $\sigma_u \rightarrow$ ultimate tensile strength, $E \rightarrow$ modulus of elasticity, $\alpha \rightarrow$ coefficient of thermal expansion, $\rho \rightarrow$ density and $C \rightarrow$ specific heat.

One may prevent the thermal shock by altering the external conditions to the degree that cooling or heated rates are reduced and temperature gradients across a body are minimized. Modification of the thermal and/or mechanical characteristics in Eq. (29) may also enhance the thermal shock resistance of a material. The introduction of some relatively large pores or ductile second phase may also help to improve the thermal shock characteristics of a material, i.e., both serve to impede the propagation of thermally induced cracks.

Usually, there is a necessity to remove thermal stresses in ceramic materials as a means of improving their mechanical strengths and optical characteristics. This may be achieved by an annealing heat treatment.

10. MELTING POINT (M.P.)

Every solid substance has a definite temperature at which it changes from the solid to the liquid state at a given superincumbent pressure, which is called its melting point. Melting point temperature remains constant until the whole of the solid body is melted provided the pressure remains constant, although heat is supplied to the body all the time. The temperature of the body will rise only after its complete melting. The M.P. is different for different materials and for each material it slightly varies when the superincumbent pressure varies.

We may note that the *normal melting point* of a substance is a definite temperature at which it melts at a pressure of one atmosphere. However, the alloys have a range of melting and not a fixed melting temperature. But the eutactic alloys have fixed melting point-temperature.

When heat is added to a solid, its thermal energy increases until the atoms (or molecules on its surface begin to break away from the equilibrium positions. The amplitude of thermal vibration at which this takes place is associated with interatomic spacing at which the bonding force is maximum, for if the atoms can be separated to that point, no further increase in force is required to separate them further. One can ensure that there is definite amplitude and consequently a definite melting temperature for each type of atom or each element.

The temperature at which melting of a solid material begins depends on the amount of thermal energy required, which in turn, depends on the characteristics of the interatomic and intermolecular bonds. Obviously, materials having stronger bonds exhibit higher melting points. Using this restriction, we can place materials in order of decreasing melting points as follows: covalent, ionic, metallic, molecular. We may note that the abruptness of melting depends on the uniformity of size of particles and the degree of order in their arrangements.

We have mentioned that melting temperature and thermal expansion of a solid depend on the bonds between the atoms or molecules of a solid. Thus we can expect that there must exist some relation among them. Solids can be divided into classes, and for each class there is a characteristic amount of thermal expansion at which melting beings. We have for a given class of material.

$$\alpha\, T_m = \text{Constant} \tag{30}$$

Obviously, two materials of a given class that have same α, will therefore, have almost the same melting point.

Example 1 *A brass rod is to be used in an application requiring its ends to be held rigid. The rod is stress free at 20°C. Determine the maximum temperature to which the rod may be heated without exceeding a compressive stress of 172 MPa. Given, modulus of elasticity = 100 GPa for brass and $\alpha = 20 \times 10^{-6}/°C$.* [BE]

Solution We have

$$\sigma = E\ \alpha\ (T - T_f)$$

Given $T = 20°C,\ \alpha = 20 \times 10^{-6}/°C$ and $\sigma = -\ 172$ MPa

$$\therefore \qquad T_f = T - \frac{\sigma}{E\alpha} = 20° - \frac{-172\ \text{MPa}}{(100 \times 10^3\ \text{MPa}) \times 10^{-6}/°\text{C}}$$

$$= 20°\text{C} + 86°\text{C} = 106°\text{C}$$

SECTION B: OPTICAL PROPERTIES

1. OPTICAL PROPERTIES

Optical properties of a material is meant a material's response to exposure to *electromagnetic radiation* and, in particular, to *visible light.* All electromagnetic radiation traverses a vacuum at the same velocity, that of light ($c = 3 \times 10^8$ m/s). This velocity, c is related to the electric permittivity of a vacuum ε_0 and the magnetic permeability of a vacuum μ_0 through

$$c = \frac{1}{\sqrt{\varepsilon_0 \, \mu_0}} \tag{1}$$

The frequency v and the wavelength λ of the electromagnetic radiation are related as

$$c = \lambda v \tag{2}$$

Sometimes it is more convenient to view electromagnetic radiations from a quantum-mechanical picture, in that the radiation, rather than consisting of waves, is composed of packets of energy, which are called *photons.* The energy E of a photon is said to be *quantized*, and defined by

$$E = hv \tag{3}$$

where h (= 6.63×10^{-34} J-s) is Planck's constant. Obviously, photon energy is proportional to the frequency (v) of the radiation, or inversely proportional to the wavelength (λ).

2. INTERACTIONS OF LIGHT WITH SOLIDS

When light passes from one medium into another, severals things may happen. Some of the incident light radiation may be *transmitted* through the medium, some will be *absorbed*, and some will be reflected at the interface between the two media (e.g., air and solid). The intensity I_0 of the incident beam on the surface of the solid medium must be equal to the sum of the intensities of the transmitted (I_T), absorbed (I_A) and reflected beams (I_R), i.e.

$$I = I_T + I_A + I_R \tag{4}$$

Radiation intensity, expressed in W/m^2, corresponds to the energy being transmitted per unit time across a unit area that is perpendicular to the direction of propagation. Equation (4) can also be expressed as

$$T + A + R = 1 \tag{5}$$

where T, A and R, respectively, the *transmissivity* ($T = I_T/I_0$), *absorptivity* ($A = I_A/I_0$) and *reflectivity* ($R = I_R/I_0$), or the fractions of incident light that are transmitted, absorbed, and reflected by a material; their *sum must equal unity*, since all the incident light on the solid surface is either transmitted, absorbed, or reflected.

Transparent Materials: Those materials which are capable of transmitting light with relatively little absorption and reflection are called transparent materials. One can easily through them. The materials through which light is transmitted diffusely are termed as *translucent*, i.e. light is scattered within the interior of these materials, to the degree that objects are not clearly distinguishable when viewed through a specimen of the material. Those materials which are impervious to the transmission of visible light are termed *opaque*.

Bulk materials are opaque throughout the entire visible spectrum, i.e., all light radiation is either absorbed or reflected. On the other hand, electrically insulating materials can be made to transparent. Moreover, some semiconducting materials are transparent where as others are opaque.

3. ATOMIC AND ELECTRONIC INTERACTIONS

The optical phenomena that occur in solid materials involve interactions between the electromagnetic radiation and atoms, ions, and/or electrons. Two of the most important of these interactions are: *(i) electronic polarization* and *(ii) electron energy transitions.*

One component of an electromagnetic wave is simply a rapidly fluctuating electric field $\vec{E}$. For the visible range of frequencies, this electric field interacts with the electron cloud surrounding each atom within its path in such a way as to induce electronic polarization, or to shift the electron cloud relative to the nucleus of the atom with each change in direction of electric field component. Two major consequences of this *electronic polarization* are: *(i)* some of the radiation energy may be absorbed, and *(ii)* light waves are retarded in velocity as they pass through the medium. The second consequence is manifested as *refraction*.

The absorption and emission of electromagnetic radiation may involve electron transitions from one energy state to another. Let us consider an isolated atome the electron energy diagram for which is represented in Fig. 13.13. An electron may be excited from an occupied state E_1 to a vacant and higher lying one, denoted by E_2, by the absorption of a photon of energy. The change in energy experienced by the electron, $E_2 - E_1 = \Delta E$, depends on the frequency of radiation as follows:

$$E_2 - E_1 = \Delta E = h\nu \tag{6}$$

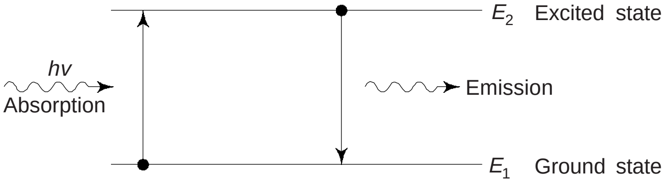


Fig. 13.13 Atomic transitions

We may note that the energy states for the atom are discrete and therefore, only specific $\Delta E_s'$ exist between the energy levels; thus, only photons of frequencies corresponding to the possible ΔE's for the atom can be absorbed by electron transitions. Moreover, all of a photon's energy is absorbed in each excitation event.

The stimulated electron cannot remain in an *excited state* indefinitely; after a short time, electron falls or decays back into its *ground* or *lower energy state* or *unexcited level*, with a reemission of electromagnetic radiation. Several decay paths are possible. However, there must be a conservation of energy for absorption and emission electron transitions.

Obviously, the optical characteristics of solid materials that relate to absorption and emission of electromagnetic radiation are explained in terms of *electron band structure* and the principles relating to electron transitions.

4. OPTICAL PROPERTIES OF METALS

Metals are opaque because the incident radiation having frequencies within the possible range excites electrons into unoccupied energy states above the Fermi energy (Fig. 13.14a); as a consequence, the incident radiation is absorbed, in accordance with Eq. (6). Total absorption within the metal remains within a very thin outer layer, usually less than 0.1 μm; thus only metallic films thinner than 0.1 μm are capable of transmitting visible light.

A metal can absorb all frequencies of visible light because of the continuously available empty electron states, which permit electron transitions as shown in Fig. 13.14(a). In fact, metals are opaque to all electromagnetic radiations on the low end of the frequency spectrum, from radio waves, through infrared, the visible, and into about the middle of the ultraviolet (UV) radiation. However, metals are transparent to high-frequency (X-and γ-rays) radiations.

Most of the absorbed radiation from metal surface is reemitted in the form of visible light of the same wavelength, which appears as reflected light (Fig. 13.14(b)). The reflectivity for most metals is between 0.90 and 0.95. Small amount of energy from electron decay is dissipated as heat.

Since metals are opaque and highly reflective, the perceived colour is determined by the wavelength distribution of the radiation and not absorbed. When a bright silvery appearance is exposed to white light radiations, we note that the metal is highly reflective over the entire range of visible spectrum, i.e., for the reflected beam, the composition, of these reemitted photons, in terms of frequency and number is approxi-

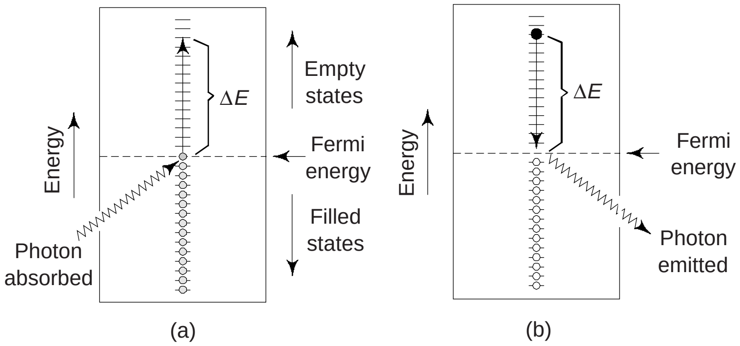


Fig. 13.14(a) **Mechanism of photon absorption for metallic materials in which an electron is excited into a higher energy unoccupied state. ΔE is equal to the energy of the photon absorbed, (b) Reemission of photon of light by the direct transition of an electron from a high to a low energy state**

mately the same as for incident beam of light. This reflective behaviour is exhibited by A1 and Ag. Copper and gold appear red-orange and yellow, respectively, because some of the energy associated with light photons having short wavelengths is not reemitted as visible light.

5. OPTICAL PROPERTIES OF NON METALS

Non metallic materials, by virtue of their electron energy band structures may be transparent to visible light. Obviously, in addition to reflection and absorption, refraction and transmission phenomena also important for non-metallic materials.

(i) Refraction: Light that is transmitted into the interior of transparent materials experiences a decrease in velocity, and, as a result, is bent at the interface. This phenomenon is called refraction. The *index of refraction* (n) is defined as

$$n = \frac{c}{v} \tag{7}$$

where c is the velocity of light in vacuum and v the velocity of light in a medium given by

$$v = \frac{1}{\sqrt{\varepsilon\mu}} \tag{8}$$

where ε and μ are, respectively, the permittivity and permeability of the particular material. We have

$$n = \frac{c}{v} = \frac{\sqrt{\varepsilon\mu}}{\sqrt{\varepsilon_0\,\mu_0}} = \sqrt{\varepsilon_r\,\mu_r} \tag{9}$$

where ε_r and μ_r are the dielectric constant and the relative magnetic permeability, respectively. Since most substances are only magnetic, μ_r, ≅ 1, and

$$n = \sqrt{\varepsilon_r} \tag{10}$$

Obviously, for transparent materials, there is a relation between the index of refraction and the dielectric constant. Since the phenomenon of refraction is related to electronic polarization at the relatively high frequencies for visible light; thus, the electronic component of the dielectric constant, may be determined from the measurement of n using Eq. (10).

We have seen that the retardation of electromagnetic radiation in a medium results from electronic polarization, and thus the size of the constituent atoms or ions has a considerable influence on the magnitude of this effect, i.e., generally the larger an atom or ion, the greater will be the electronic polarization, the slower the velocity *v*, and the greater the index of refraction *n*. For a typical soda lime galss, $n \approx 1.5$. Additions of large Ba and Pb ions (as BaO and PbO) to a glass will enhance *n* significantly, e.g. highly leaded glasses containing 90 Wt% PbO have $n \approx 2.1$.

For glasses and crystalline ceramics (having cubic crystal structures), *n* is independent of crystallographic direction, i.e. it is isotropic. On the other hand, non-cubic crystals have an anisotropic *n*, i.e., *n* is greatest along the direction that have the highest density of ions. Table 13.1 provides *n* values for several glasses, polymers and transparent ceramics. For crystalline ceramics in which *n* is anisotropic, average values are cited.

Table 13.1 Refractive indices for some transparent ceramics, polymers etc

Material	n_{av} *(Average value of Index of refraction)*
Ceramics	
Silica glass	1.458
Borosilicate (pyrex) glass	1.47
Soda-lime galss	1.51
Quartz (SiO_2)	1.55
Dense optical flint glass	1.65
Spinel ($MgA1_2O_4$)	1.72
Periclase (MgO)	1.74
Corundum (Al_2O_3)	1.76
Polymers	
Polytetrafluoroethylene	1.35
Polymethyl methacrylate	1.49
Polypropylene	1.49
Polyethylene	1.51
Polystyrene	1.60

(ii) Refraction: When light radiation enter from one medium into another medium having a different *n*, some of light is *scattered* at the interface between the two media even if both the media are transparent. The reflectivity (*R*) of incident light that is reflected at the interface of the two media,

or

$$R = \frac{I_R}{I_0} \tag{11}$$

where I_o and I_R are the intensities of the incident and reflected beams of light, respectively. If the light is normal (or perpendicular) to the interface, then we have

$$R = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 \tag{12}$$

where n_1 and n_2 are the indices of refraction of the two media. When the incident light is not normal to the interface, *R* will depend on the angle of incidence. When light is transmitted from vaccum or air ($n \approx 1$) into a solids (of refractive index n_s), then

$$R = \frac{(n_s - 1)^2}{(n_s + 1)^2} \tag{13}$$

Obviously, the higher the index of refraction of a solid, the greater is its reflectivity *R*. The reflectivity for typical, silicate glasses is ~0.05. Just as the incident *n* of a solid depends on the wavelength (λ) of light so also does the *R* vary with λ. One can significantly minimise the reflection losses for lenses and other optical instruments by coating the reflecting surface with very thin layers of dielectric materials, e.g. MgF_2.

(iii) Absorption: Non metallic materials may be transparent or opaque to visible light. If they are transparent, they often appear coloured. Basically, light radiations in these materials are absorbed by two mechanisms which themself also influence the transmission characteristics of these non metals. One of these mechanism is *electronic polarization*. Absorption of light radiation by this mechanism is important only at light frequencies in the vicinity of the relaxation frequency of the constituent atoms. The second mechanism involves *valence band-conduction band electron transitions*. This mechanism depends on the electron energy band structure of the material.

Absorption of a photon of light may occur by the promotion or excitation of an electron from the nearly filled valence band, across the band gap, and into an empty state within the conduction band (Fig. 13.15a); a free electron in the conduction band and a hole in the valence band are created. The energy of excitation ΔE is related to the absorbed photon energy through Eq. (6). However, these excitations with the accompanying absorption can take place only if the photon energy $hv > E_g$ (bang gap energy), or in terms of wave length $\frac{hc}{\lambda} > E_g$. The maximum band gap energy $(E_g)_{\max}$ for which absorption of light (minimum wavelength $\lambda_{\min} \approx 0.4\ \mu\text{m}$) is just

$$(E_g)_{\max} = \frac{hc}{\lambda_{\min}} = \frac{4.13 \times 10^{-15}\ \text{eV-s} \times 3 \times 10^{-8}\ \text{m/s}}{4 \times 10^{-7}\ \text{m}} = 3.1\ \text{eV}$$

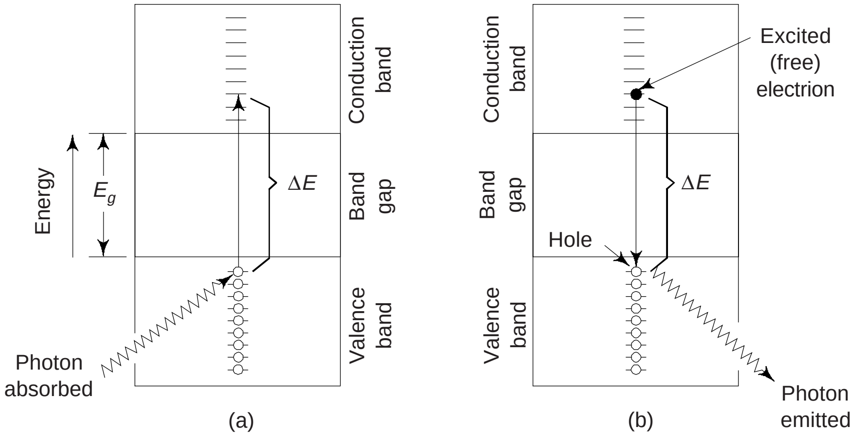


Fig. 13.15(a) Photon absorption mechanism for non-metallic materials in which an electron is excited across the band gap, leaving behind a hole in the valence band ($\Delta E > E_g$). (b) Emission of a photon of light by a direct electron transition across the band gap

Obviously, no visible light is absorbed by non-metallic materials having E_g greater than about 3.1 eV. These materials, if of high purity, will appear transparent and colourless.

Now, the maximum wavelength for visible light $\lambda_{\max}$, is about 0.7 μm. The minimum band gap energy $(E_g)_{\min}$ for which there is absorption of visible light is

$$(E_g)_{\min} = \frac{hc}{\lambda_{\min}} = \frac{4.13 \times 10^{-15}\ \text{eV-s} \times 3 \times 10^{8}\ \text{m/s}}{7 \times 10^{-7}\ \text{m}} = 1.8\ \text{eV}.$$

This result show that all visible light is absorbed by valence band to conduction band electron transitions for those semiconducting materials that have band gap energies less than about 1.8 eV. Obviously, these materials are opaque. This Clearly reveals that only a portion of the visible spectrum is absorbed by materials whose band gap energies range between 1.8 and 3.1 eV. Consequently, these materials appear coloured.

Depending on the magnitude of E_g, every non-metallic material becomes opaque at some wavelength, e.g. Diamond having $E_g = 5.6$ eV, is opaque to radiation having $\lambda \sim 0.22$ μm.

Interactions with light radiation can also occur in dielectric solid materials having wide band gaps, involving other than valence-band-conduction-band electron transitions. When impurities or other electrically active defects are present, electron levels within the band gap may be introduced such as the donor and acceptor levels, except that they lie closer to the centre of the band gap. As a result of electron transitions involving these levels within the band gap, light radiation of specific wavelengths may be emitted.

The intensity of net absorbed radiation depends on the character of the medium as well as the path length within. The intensity of transmitted or non-absorbed radiation I_T' decreases continuously with distance x that the light traverses;

$$I_T' = I_0' \exp(-\beta x) \tag{14}$$

where I_0' is intensity of the non reflected incident radiation and β, the *absorption coefficient* (in mm^{-1}), is characteristic of the particular material. Moreover, β varies with the wavelength of the incident radiation. x is measured from the incident surface into the material. Materials are found to be highly absorptive with large β values.

(iv) Transmission: One may apply the phenomena of absorption, reflection, and transmission to the passage of light through a transparent solid (Fig. 13.16). For an incident beam of intensity I_0 that impinges on the front surface of the given specimen of the material of thickness l and absorption coefficient β, the transmitted intensity at the back face of the specimen I_T is

$$I_T = I_0 (1 - R)^2 \exp(-\beta l) \tag{15}$$

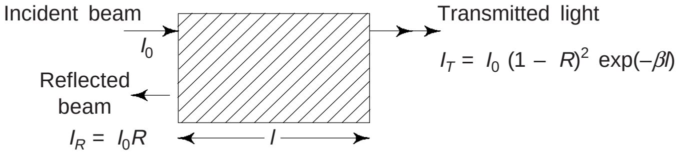


Fig.13.16 The transmission of light through a transparent medium. There is reflection at front and back faces of the medium and also absorption within the medium

where R is the reflectance. It is assumed that the same medium exist outside both front and back faces of the specimen. For light having wavelength of 0.4 μm, the fractions transmitted, absorbed, and reflected are approximately 0.90, 0.05, and 0.05 respectively, whereas for $\lambda = 0.55$ μm, one obtains the respective fractions as 0.50, 0.48 and 0.02.

(v) Colour: As a consequence of specific wavelength ranges of light that are selectively absorbed, transparent materials appear coloured and the colour discerned is a result of the combination of wavelengths that are transmitted. The material appears colourless, if absorption is uniform for all visible wavelengths, e.g. high purity inorganic glasses and high-purity and single-crystal diamonds and sapphire. The colour depends on the frequency distribution of both transmitted and reemitted light beams. High-purity and single-crystal aluminium oxide or sapphire is colourless, whereas Ruby, which has a brilliant red colour, is simply sapphire to which 0.5 to 2% of chromium oxide (Cr_2O_3) is added. The Cr^{3+} ion substitutes for the $A1^{3+}$ ion in the Al_2O_3 crystal structure and, furthermore, introduces impurity levels within the wide energy band gap of the sapphire. Inorganic glasses are coloured by incorporating transition or rare-earth ions while the glass is still in the molten state. The representative colour-ion pairs includes Cu^{2+}, blue-green; Co^{2+}, blue-violet; Cr^{3+}, green; Mn^{2+}, yellow; and Mn^{3+}, purple. These coloured glasses finds wide uses as glazes, decorative coatings on ceramic ware.

(vi) Opacity and Translucency in Insulators: The extenst of translucency and opacity for inherently transparent dielectric materials mostly depends on their internal reflectance and transmittance-characteristics.

Several dielectric materials that are intrinsically transparent may be made translucent or even opaque because of interior reflection and refraction. As a result of multiple scattering events, a transmitted light beam is deflected in direction and appears diffuse. When the scattering is so extensive that virtually none of the incident beam is transmitted, undeflected, to the back surface, opacity results.

There are several different sources which may cause this internal scattering. Normally, polycrystalline specimens in which the index of refraction (n) is anisotropic appear translucent. A diversion in the incident beam takes place due to reflection and refraction occuring at grain boundaries. This results from a small difference in the index of refraction between adjacent grains which do not have the same crystallographic orientation.

Scattering of light also takes place in *two-phase materials.* In these materials one phase is finely dispersed within the other. When there is a difference in the refractive index for the two phases, the beam dispersion occurs across phase boundaries and greater this difference, the more efficient is scattering.

Many ceramic pieces, as a consequence of fabrication or processing contain some residual porosity in the form of finely dispersed pores. These pores also scatter light radiation effectively.

The degree of translucency is influenced primarily by the extent of crystallinity for *intrinsic polymers* (without additives and impurities). As a result of different indices of refraction, some scattering of visible light occurs at the boundaries between crystalline and amorphous regions. This degree of scattering for highly crystalline specimens is extensive and this leads to translucency, and, in some cases, even opacity. We may note that highly amorphous polymers are completely transparent.

6. APPLICATIONS OF OPTICAL PHENOMENA

(i) Luminescence: Light emission that cannot be attributed merely to the temperature of the emitting body is termed luminescence, i.e. there are substances which are capable of absorbing energy and reemitting visible light. This phenomenon is called luminescence. Various types of luminescence are often distinguished according to the source of energy which excite the emission. When the light energy emitted results from a chemical reaction, such as in the slow oxidation of phosphorus at ordinary temperatures, the emission is called *chemiluminescence.* When the luminescent chemical reaction occurs in a living system, such as in the glow of the firefly, the emission is called *bioluminescence.* In these examples part of the energy of a chemical reaction is converted into light. There are also types of luminescence that are initiated by the flow of some form of energy into the body from the outside. According to the source of the exciting energy, these luminescences are termed as *cathodoluminescence* if the energy comes from electron bombardment, *radioluminescence* or *roentgenoluminescence* if the energy comes from x-rays or from γ-rays; *photoluminescence* if the energy comes from ultraviolet, visible or infrared radiation; and *electroluminescence* if the energy comes from the application of an electric field. By attaching a suitable prefix to the word luminescence, similar designations may be coined to characterize luminescence excited by other agents. The classification of luminescence phenomena into different categories is only a matter of convenience and not of any fundamental distinction. Furthermore, luminescence is classified according to the magnitude of the delay time between *absorption* and *reemission* events. If reemission occurs for times much less than 1 second, the phenomenon is called *fluorescence;* for longer times, it is called *phosphorescence*. A number of materials can be made to *fluoresce* or *phosphoresce*, e.g. some sulphides, oxides, tungstates, and a few organic materials. Normally, pure materials do not display these phenomena, and to induce them, impurities in controlled concentrations must be added.

There are number of commercial applications of luminescence. Fluorscent lamps consist of glass housing, coated on the inside with specially prepared tungstates or silicates. Within the tube from a mercury glow discharge ultraviolet light is generated, which causes the coating to fluorsce and emit white light. The picture viewed on the screen of a television is the product of luminescence. The inside portion of the screen is coated with a material that fluoresces as an electron beam inside the picture tube very rapidly traverses the screen. One can also detect x-rays and γ-rays making use of the fact that certain phosphorus emit visible light or glow when introduced into a beam of radiation that is otherwise invisible.

Some PN rectifying junctions (see chapter 15) may also be used to generate visible light during the process of electroluminescence. When a forward-biased potential is applied across the device, electrons and holes will annihilate each other within the recombination region according to the relation,

$$\text{electron} + \text{hole} \rightarrow \text{energy } (\Delta E).$$

Under some circumstances the energy produced will appear as visible light. Such diodes are called *light emitting diodes* (LEDs), which are used for digital display. The characteristic colour of an LED depends on the particular semiconducting material that is used.

When a luminescent system provided with a special configuration is excited, or “pumped,” with sufficient intensity of excitation to cause an excess of excited atoms over unexcited atoms (a so-called *population inversion*), it can produce a *laser* (Light Amplification By Stimulated Emission of radiation). This laser emission is a coherent stimulated luminescence, in contrast to the incoherent spontaneous emission from most luminescent systems as they are ordinarily excited and used.

(ii) Photoconductivity: The increase in electrical conductivity caused by the excitation of additional free charge carries by light of sufficiently light energy in semiconductors and insulators is called photoconductivity. Effectively a radiation-controlled electrical resistance, a photoconductor can be used for a variety of light and particle detection applications, as well as a *light-controlled switch.* Other major applications in which photoconductivity plays a central role are *television-cameras* (vidicons), normal silver halide emulsion photography, and the very large field of electro-photographic reproduction. The phenomena related to photoconductivity have also played a large part in the understanding of electronic behaviour and crystalline imperfections in a variety of different materials.

Sunlight may be directly converted into electrical energy in solar cells, which also employ semiconductors. The operation of these devices is, in a sense, the reverse of that for LED. A P-N junction is used in which photo excited electrons and holes are drawn away from the junction, in opposite directions, and become part of an external current.

(iii) Laser: The word ‘laser’ is an acronym for light amplification by stimulated emission of radiation. It is a source of highly pure intense beam of light. Laser action can occur in atoms, molecules in gases, liquids, solids, and flames, and ions. The emitted wavelength by laser source covers the ultraviolet to the radio frequency regions, the output power varying form a few mW to MW. Some lasers emit *continuous* waves while others, in pulses. *Table* 13.2 lists several common lasers and their characteristics.

Table 13.2 Characteristics and applications of few different types of lasers

Laser	*Type*	*Common wavelengths (μm)*	*Max. output power (W)*[a]	*Applications*
He-Ne	Gas	0.6328, 1.15, 3.39	0.0005-0.05 (CW)	Line-of sight communications, recarding/playback of holograms
CO_2	Gas	9.6, 10.6	500-15,000 (CW)	Heat treating, welding, cutting, scribing, marking
Argon	Gas ion	0.488, 0.5145	0.005-20 (CW)	Surgery, distance measurements, holography
HeCd	Metal vapor	0.441, 0.325	0.05-0.1	Light shows, spectroscopy
Dye	Liquid	0.38-1.0	0.01 (CW) 1×10^6 (P)	Spectroscopy, pollution detection
Ruby	Solid state	0.694	(P)	Pulsed holography, hole piercing
Nd-YAG	Solid state	1.06	1000 (CW) 2×10^8 (P)	Welding, hole piercing, cutting
Nd-Glass	Solid state	1.06	5×10^{14} (P)	Pulse welding, hole piercing
Diode	Semiconductor	0.33-40	0.6 (CW) 100 (P)	Bar-code reading, CDs and video disks, optical communications

[a] “CW” denotes continuous; “P” denotes pulsed.

The phenomena involved in the laser action encompass interactions between atoms and molecules and electromagnetic fields. The salient features of laser radiation are: *(i)* directionality, *(ii)* high intensity, *(iii)* monochromaticity and *(iv)* coherence.

Laser applications are diverse. Since lasers beams may be focused to produce localized heating, they are used in some surgical procedures and for cutting, welding, and machining metals. Complex operations in surgery are performed with laser beams; such operations are less painful, bloodless, and quick. Lasers are also used in communication, computer circuitary, biology, meterology, military warfare, holography, etc. Laser beams being highly directional, large distances can be measured accurately with their help, e.g. the distance of the moon from the earth has been measured with an error of only 15 cm in 384000 km. Lasers have opened up new researches on the interaction of matter with light and in *non-linear optics.*

(iv) Optical Fibres in Communication: An optical-fibre is a very thin and flexible medium of cylindrical shape. The three principal sections of a fibre are: *(i)* the *core*, *(ii)* the *cladding*, and *(iii)* the *jacket*. The innermost section, made of glass or plastic, is referred to as *core*. A glass or plastic coating surrounding the core is known as the *cladding.* The optical properties of the cladding are different from those of the *jacket*. It is made of plastic or polymer and other materials, and protects the structure from moisture, abrasion, mechanical shocks, and other environmental hazards.

The actual working structure of the optical fibre is the core. The light entering the core at a suitable angle propagates through it, suffering a number of total internal reflections at the core-cladding interface. Optical fibres serve as cables for communications to carry optical signals from one end to the other over a wide band. A schematic diagram showing the components of an optical fibre communication system is shown in Fig. 13.17. With the development of optical fibre technology, the field of communications has recently experienced a great revolution. Now, all telecommunications are transmitted via optical fibre rather than through conventional copper wires. Signal transmission through a metallic wire conductor is electronic, whereas using otpically transparent fibres, signal transmission is *photonic*, i.e. it uses photons of electromagnetic or light radiation. Use of optical fibres in communications has improved speed of transmission, information density, and transmission density, with a reduction in error rate; moreover, there is no electromagnetic interference with fibre optics.

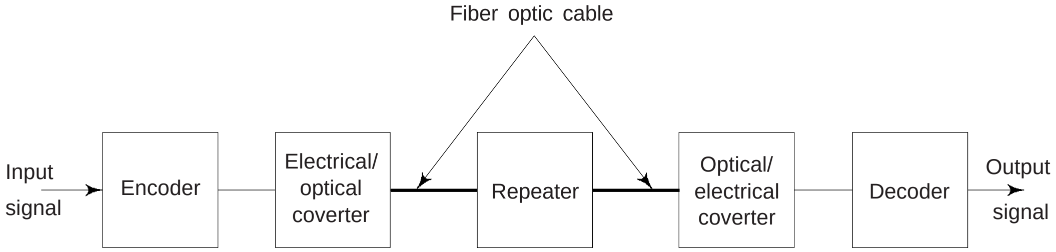


Fig. 13.17 The components of an optical fibre communications system

High purity silica glass is used as the fibre material. Fibre diameters normally range between about 5 and 100 μm. The optical fibres are relatively flaw free and, thus, remarkably strong; during production the continuous fibres are tested to ensure that they meet minimum strength standards.

Exceptionally pure and high-quality optical fibres are fabricated using advanced and sophisticated processing techniques. Impurities and other defects that absorb, scatter and thus attenuate the light beam must be eliminated.

SUGGESTED READINGS

1. J. Ziman, ‘*The Thermal Properties of Materials’*, Scientific American, Vol. 217, No. 3, Sept. 1967, pp. 180-188.

2. R.M. Rose, et al., '*The Structure and Properties of Materials*', Vol. IV, Electronic Properties, Wiley, Newyork, 1966.
3. S.L. Kakani and C. Hemrajani, '*Solid State Physics*', 4th ed. Sultan Chand, New Delhi (2004).
4. A. Javan, '*The Optical Properties of Materials*', Scientific American, Vol. 217, No. 3, Sept. 1967, pp. 238-248.
5. J.M. Rowell, '*Photonic Materials*', Scientific American, Vol. 255, No. 4, Oct. 1986, pp. 146-157.

REVIEW QUESTIONS

1. When heated, a material expands, why?

2. Briefly explain why C_v rises with increasing temperature at temperatures near 0 K.

3. Briefly explain why C_v becomes virtually independent of temperatures at temperatures far removed from 0 K. [B.E.]

4. What is the mechanism through which heat is transmitted in a solid body? How do the impurities affect thermal conductivity? [M.Sc. (M.S.)]

5. How does porosity affect the thermal insualting properties an insulator? [Diploma]

6. What are the refractory materials? Describe the use of these materials as insulating materials. [B.E.]

7. A bimetallic strip is constructed from strips of two different metals that are bonded along their lengths. Explain how such a device may be used in thermostat regulate temperature. [B.E., B.Sc. (Hons)]

8. Briefly explain thermal expansion using the potential energy-versus-interatomic spacing cruve. [Diploma]

9. Explain why, on a cold day, the metal door handle of an automobile feels cooler to the touch than a plastic steering wheel, even though both are at the same temperature. [B.E.]

10. The thermal conductivity of a single crystal specimen is slightly greater than a polycrystalline one of the same material, why? [B.E.]

11. The thermal conductivity of a plain carbon steel is greater than for a stainless steel, why? [B.E.]

12. Why the thermal conductivities are higher for crystalline than non-crystalline ceramics? [M.Sc. (M.S.)]

13. Discuss how thermal stresses are generated and how they affect the material in service. [B.E.]

14. How specific heats of different materials behave, explain? [Diploma]

15. Explain in brief, why metals are typically better thermal conductors than ceramic materials.

16. Explain thermal insulation and thermal expansion.

17. For some ceramic materials, thermal conductivity first decreases and then increases with rising temperature, why? [B.E.]

18. *(a)* Explain in brief why thermal stresses may be introduced into a structure by rapid heating or cooling *(b)* What is the nature of surface stresses for cooling? *(c)* What is the nature of surfae stresses for heating? Is thermal shock more likely to occur on rapid heating or cooling for a ceramic material? [B.E.]

18. What measures you will take to reduce the likelihood of thermal shock of a ceramic piece? [Diploma]

19. Distinguish between materials that are opaque, translucent, and transparent in terms of their appearance and light transmittance. [REC]

20. Explain the phenomenon of electronic polarization by electromagnetic radiation. What are the two consequences of electronic polarization in transparent materials? [B.E.]

21. Explain, why metals are opaque to electromagnetic radiation having photon energies within the visible region of the spectrum? Why are metals transparent to high-frequency X-ray and γ-ray radiation? [REC]

22. Whether a material can have an index of refraction less than unity, explain. [Diploma]

23. Describe in brief the phenomenon of dispersion in a transparent medium.

24. How reflection losses of transparent materials are minimized by thin surface coatings, explain? [Diploma]

25. Describe the three absorption mechanisms in non-metallic materials.

26. Why the magnitude of absorption coefficient depends on the wavelength of the radiation?

27. What determines the characteristics colour of *(a)* a metal and *(b)* a transparent non-metal? [B.Sc. (Hons)]

28. Why some transparent materials appear coloured while others are colourless? [B.E.]

29. Describe the phenomena of luminescence, what is the distinction between fluorescence and phosphorescence? [Diploma]

30. Why amorphous polymers are transparent, while predominantly crystalline polymers appear opaque, or, at best, translucent, explain?

PROBLEMS

1. An aluminium wire of length 10 m is cooled from 38°C to –1°C. Calculate the change in length. [*Ans.* –9.2 mm]

2. To what temperature must a cylindrical rod of tungsten 15.025 mm in diameter and a plate of 1025 steel having a circular hole of diameter 15 mm have to be heated for the rod to just fit into the hole? Initial temperature is 25°C. [*Ans.* 222.4°C] [B.E.]

3. A copper wire is stretched with a stress of 70 MPa at 20°C. Keeping the length constant, to what temperature must the wire be heated to reduce the stress to 35 MPa? [*Ans.* 39°C] [BE, REC]

4. Calculate the velocity of light in diamond, which has a dielectric constant ε_r of 5.5 (at frequencies within the visible range) and a magnetic susceptibility of -2.17×10^{-5} [*Ans.* 1.28×10^8 m-s^{-1}] [M.Sc. (M.S.)]

5. The fraction of non reflected radiation that is transmitted through a 10 mm thickness of a transparent material is 0.9. What fraction of the light will be transmitted, if the thickness is increased to 20 mm? [*Ans.* $I_T'/I_0' = 0.81$] [BE, REC]

6. The transmissivity T of a transparent material 20 mm thick to normally incident light is 0.85. If the index of refraction of this material is 1.6, find the thickness of material that will yield a transmissivity of 0.75. Consider all reflection losses. [*Ans.* 67.3 mm] [B.E.]

OBJECTIVE QUESTIONS

1. Which one of the following statement is correct?
- (1) Metals with high melting point have high coefficient of thermal expansion.
- (2) There is no relation between coefficient of thermal expansion and melting point of a material.
- (3) Thermal coefficient of expansion for all metals is same.
- (4) Metals with low melting point have high coefficient of thermal expansion. (4)

2. Linear thermal expansivity of all refractory oxide materials
- (1) remain constant with increasing temperature
- (2) decreases steadily with increase in temperature
- (3) increases steadily with increase in temperature
- (4) decreases steadily with increase in temperature except for ZrO_2 and SiO_2. (4)

3. The magnitude of the stress σ resulting from a temperature change from T_o to T_f is

(1) $\sigma = E\,\alpha\,(T_o - T_f)^2$ (2) $\sigma = E\,\alpha^2\,(T_o - T_f)$

(3) $\sigma = E\ \alpha\,(T_o - T_f)$ (4) $\sigma = E\ \alpha\sqrt{(T_0 - T_f)}$ (3)

4. Wiedemann-Franz law is
(1) $L = K\ \sigma\ T$ (2) $L = K/\sigma\ T$
(3) $L = K\big/\sqrt{\sigma T}$ (4) $L = K\big/\sqrt{\sigma T}$ (2)

5. At low temperatures the relationship between C_v and the absolute temperature is
(1) $C_v = AT$ (2) $C_v = AT^{1/2}$ (3) $C_v = AT^2$ (4) $C_v = AT^3$ (4)

6. For materials in which the thermal expansion is isotropic, the volume coefficient of thermal expansion α_v is approximately
(1) equal to the linear coefficient of thermal expansion α_l
(2) $\alpha_v \approx 2\alpha_l$ (3) $\alpha_v \approx 3\alpha_l$
(4) none of the above (3)

7. If K_l and K_e represent the lattice vibration and electron thermal conductivities, then the total conductivity (K) of a solid material is
(1) $K = K_l - K_e$ (2) $K = \sqrt{K_l K_e}$
(3) $K = K_l + K_e$ (4) $K = K_l/K_e$ (3)

8. Thermal conductivity of
(1) glass increases while that of graphite decreases with temperature.
(2) glass and graphite remains unaffected by the change in temperature
(3) glass and graphite increase with rise in temperature
(4) glass increases while that of graphite decreases with temperature (4)

9. Alumina doped with magnesia will have reduced thermal conductivity as its structure becomes
(1) free from pores (2) crystalline
(3) amorphous (4) none of the above (1)

10. Thermal insulators transmit heat through
(1) conduction, convection and radiation (2) conduction only
(3) conduction and convection only (4) conduction and radiation only (1)

11. The relation between velocity of light c, electric permittivity of a vacuum ε_0 and the magnetic permeability μ_0 is
(1) $c = \dfrac{1}{\varepsilon_0\,\mu_0}$ (2) $c = \dfrac{1}{\sqrt{\varepsilon_0\,\mu_0}}$
(3) $c = \sqrt{\varepsilon_0\,\mu_0}$ (4) $c = (\varepsilon_0\mu_0)^{3/2}$ (2)

12. The transmissivity (T), absorptivity (A) and reflectivity (R) for a solid medium are related by
(1) $T + A - R = 1$ (2) $T + A + R = 1$
(3) $T - A - R = 1$ (4) $T + A + R = 0$ (2)

13. The index of refraction of a material is
(1) $n = \dfrac{c}{v}$ (2) $n = \sqrt{\varepsilon_0\,\mu_0}$
(3) $n = \sqrt{\varepsilon_r / \mu_r}$ (4) $n = \dfrac{1}{\sqrt{\varepsilon_r\,\mu_r}}$ (1)

SHORT QUESTION-ANSWERS

1. What is thermal conduction?

Ans. The transport of thermal energy from high to low temperature regions of a material is termed thermal conduction. For steady state heat transport, the flux is proportional to the temperature gradient along the direction of flow; the proportionality constant is called the thermal conductivity.

2. How heat is transported is solid materials?

Ans. For solid materials, heat is transported by free electrons and by vibrational lattice waves or phonons.

3. Why ceramics and polymers are poor thermal conductors?

Ans. Free electron concentrations in these materials are low and phonon conduction predominates.

4. What is the effect of thermal stresses?

Ans. Thermal stresses in a body are introduces as a consequence of temperature changes. These may lead to fracture or undesirable plastic deformation.

5. What are the prime sources of thermal stresses?

Ans. There are two prime sources of thermal stresses:

(i) restrained thermal expansion (or contraction) and *(ii)* temperature gradients established during heating or cooling.

6. What is a thermal shock?

Ans. It is the fracture of a body resulting from thermal stresses induced by rapid temperature changes. Ceramic materials being brittle, are especially susceptible to this type of failure.

7. What are the possible interactive phenomena which takes place during the interaction of a solid material with electromagnetic radiation having wavelength within the visible region of the spectrum?

Ans. Possible interactive phenomena include: refraction, reflection, absorption, and transmission of incident light.

8. Why metal appears opaque?

Ans. Metal appears opaque as a result of the absorption and then reemisison of light radiation within a thin outer surface layer. The preceived colour of a metal is determined by the spectral composition of the reflected light.

9. Why some light absorption occurs in even transparent materials?

Ans. As a consequence of electronic polarization.

14

Electrical and Magnetic Properties of Materials

1. INTRODUCTION

The electrons in the outermost shell of atoms controls the electrical and magnetic behaviour of substances. Recent developments in the field of electronic structure of substances have accelerated the growth of number of useful solid state devices. The prime objective of this chapter is to explore the electrical and magnetic properties of materials.

2. ELECTRICAL CONDUCTION

One of the most important characteristics of a solid material is the ease with which it transmits an electric current. Ohm's law relates the current I (i.e., the time rate of charge passage, Q/t, $Q \rightarrow$ Charge and $t \rightarrow$ time) to the applied voltage V as follows:

$$V = I\,R \tag{1}$$

where R is the resistance of the material through which the current I is passing. The units of V, I and R are Volt, ampere and Ohm respectively. The electrical resistivity (ρ) of solids is probably the most important of all physical properties. The value of electrical resistance is influenced by specimen configuration, and for many materials is independent of current. However, the resistivity (ρ) of material is independent of specimen geometry but it is related to R through the relation

$$\rho = \frac{RA}{l} \tag{2}$$

where l is the distance between the two points of the specimen at which the voltage is measured, and A is the area of cross-section perpendicular to the direction of the current. The units of ρ are Ohm-meters (Ω – m). From Eqs. (1) and (2),

$$\rho = \frac{VA}{Il} \tag{3}$$

Many factors influence the value of ρ for a given material. Values of resistivity of common materials at 20°C are given in Table 14.1 and few engineering materials are given in Table 14.1(a).

Table 14.1 Resistivity of some common materials at 20°C

Material	*Resistivity (ρ) (Ω – m)*	*Material*	*Resistivity (ρ) (Ω – m)*
Silver	1.6×10^{-8}	Germanium	10^{-5} to 0.6
Aluminium	2.66×10^{-8}	Silicon	10^{-5} to 2.5×10^{3}

(*Contd.*)

Table 14.1 (*Contd.*)

Material	*Resistivity (ρ) (Ω – m)*	*Material*	*Resistivity (ρ) (Ω – m)*
Iron	9.1×10^{-8}	PVC	1.0×10^{10}
Copper	1.67×10^{-8}	Bakelite	1.0×10^{11}
Nickel	13.3×10^{-8}	Mica	1.0×10^{11}
Carbon-steel	1.7×10^{-7}	Glass	1.0×10^{16}
Polythene	1.0×10^{16}	Stainless steel	7.0×10^{-7}
Graphite	1.4×10^{-7}	Steatite Porcelain	1.0×10^{13}
		Alumina	10^{11}
		Diamond	10^{12}

Table 14.1(a) Room temperature electrical resistivity for few engineering materials

Material	*Electrical Resistivity (Ω – m)*
Steel alloy 1020 (annealed)	1.60×10^{-7}
Steel alloy 4140 (quenched and tempered)	2.20×10^{-7}
Steel alloy 4340 (quenched and tempered)	2.48×10^{-7}
Stainless steel alloy 440 A (annealed)	6.0×10^{-7}
Stainless steel alloy 17-7 PH (annealed)	8.3×10^{-7}
Gray irons	
• Grade G 1800	15.0×10^{-7}
• Grade G 3000	9.5×10^{-7}
• Grade G 4000	8.5×10^{-7}
Ductile irons	
• Grade 60-40-18	5.5×10^{-7}
• Grade 80-55-06	6.2×10^{-7}
• Grade 120-90-02	6.2×10^{-7}
Aluminium alloy 7075 (T 6 treatment)	5.22×10^{-8}
Aluminium alloy 356.0 (T 6 treatment)	4.42×10^{-8}
Copper alloy C 3600 (free cutting brass)	6.6×10^{-8}
Copper alloy C 71500 (copper-nickel, 30%)	37.5×10^{-8}
Magnesium alloy A Z 91 D	17.0×10^{-8}
Titanium alloy Ti – 5Al – 2.5Sn	15.7×10^{-7}
Nickel 200	0.95×10^{-7}
Inconel 625	12.90×10^{-7}
Monel 400	5.47×10^{-7}
Haynes alloy 25	8.9×10^{-7}
Invar	8.2×10^{-7}
Super invar	8.0×10^{-7}
Kovar	4.9×10^{-7}
Lead-tin-solder (60 Sn – 40 Pb)	1.50×10^{-7}

The resistivity of some widely used metal and their alloys along with their temperature coefficient are given in Table 14.2.

Table 14.2 Electrical resistivity of some metals and their alloys

Metals and alloys	*Resistivity (ρ) at 20°C (Ω-m)*	*Temperature coefficient α/°C*
• Copper, annealed	1.67×10^{-8}	4.29×10^{-3} (0-100°C)
• Copper, reduced 75% by cold drawing	1.71×10^{-8}	—

(*Contd.*)

Table 14.2 *(Contd.)*

Metals and alloys	*Resistivity (ρ) at 20°C (Ω-m)*	*Temperature coefficient α/°C*
• Cartridge brass annealed 70% Cu and 30% Zn	6.20×10^{-8}	1.48×10^{-3} (20°C)
• Aluminium, annealed	2.65×10^{-8}	4.29×10^{-3} (20°C)
• Iron, annealed	9.71×10^{-8}	6.57×10^{-3} (20°C)
• Constantan 55% Cu and 45% Ni	49×10^{-5}	0.02×10^{-3} (25°C)
• Manganin 84% Cu, 12% Mn and 4% Ni	44×10^{-7}	0.009×10^{-3} (25°C) -0.42×10^{-3} (100°C)
• Nichrome 80% Ni and 20% Cr	108×10^{-5}	0.14×10^{-3} (0-500°C)

3. ELECTRICAL CONDUCTIVITY (σ)

Sometimes, electrical conductivity (σ) is used to specify the electrical character of a material. Electrical conductivity is simply the reciprocal of resistivity (ρ), i.e.

$$\sigma = \frac{1}{\rho} \tag{4}$$

Electrical conductivity is indicative of the ease with which a material is capable of conducting an electrical current. The units of σ are reciprocal Ohm-meters [$(\Omega\text{-m})^{-1}$] or mho/m. It is also expressed in Siemens/m.

When an electric field E is applied to a conductor an electric current begins to flow and the current density by Ohm's law is

$$J = \sigma E \tag{5}$$

The conductivity may be defined as the movement of electrical charge from one point to another and it depends on the number of charge carriers (n), the charge per carrier (e) and the mobility of carriers (μ), i.e.

$$\sigma = ne\mu \tag{6}$$

The unit of mobility (μ) is $\text{m}^2\ \text{V}^{-1}\text{s}^{-1}$. Mobility is an important term in the study of semiconductors.

One can obtain the expression (6) from first principle as follows:

Let E be the electric field applied to a conductor, e be the charge on the electron and m the mass of the electron. The electrons move in a specific direction under the influence of the electric field. The directional motion of a free electron is called a *drift*. The average velocity gained during this drift motion is termed as *drift velocity*.

As electron is being negatively charged particle, the force acting on it under the electric field intensity E is,

$$F = -eE \tag{7}$$

The electron drift is in a direction opposite to that of the applied field. During the accelerated motion, the electron collides with the defects in the lattice. As a result of the consequence scattering, the electron loses the velocity it gained from the electric field. The effect of the crystal lattice may be reduced considerably due to a retarding force (may be due to damping). This force is proportional to the velocity v and mass m of the electron. The retarding force is represented as $-\alpha mv$, where α is a constant. We can write the equation of motion of the electron as

$$m\frac{dv}{dt} = -eE = -\alpha mv \tag{8}$$

or $$dv = -\frac{eE}{m}\,dt$$

or $$v = -\frac{eE}{m}\ t + \text{constant} \tag{9}$$

If the average time between collisions is 2τ then, during this time, the electron is acted upon by a force F given by Eq. (7). But at $t = 0$, $v = 0$ (immediately after each collision) and hence the integeration constant in (9) is zero. Thus (9) takes the form

$$v = -\frac{eE}{m}\tau$$

Obviously, the mean velocity = $-eE\ \tau$/m, where τ is called the relaxation time and is time interval in which there is unit probability of a collision.

When there is a collision of lattices (resistance), the current density J due to n electrons per unit volume of charge e and drift velocity v is expressed as

$$J = nev \tag{10}$$

$$= ne\left(\frac{eE\tau}{m}\right) \tag{11}$$

$$= \sigma E \tag{12}$$

$$\therefore \qquad \sigma = \frac{J}{E} = \frac{ne^2\ \tau}{m} = \frac{ne(e\tau)}{m} \tag{13}$$

The velocity in a unit electric field, i.e. v/E is the mobility (μ) of the electron. Hence

$$\mu = \frac{v}{E} = \frac{1}{E/v}\left(\frac{Ee\tau}{m}\right) = \frac{e\tau}{m}$$

Using Eqs. (10) and (12), one obtains

$$J = ne\mu E$$

$$\text{and} \qquad \sigma = ne\mu \tag{14}$$

Equation (14) is of immense importance for solid materials. From (14), we note that the electrical conductivity depends on two factors: *(i)* the number n of charge carriers per unit volume and *(ii)* their mobility, μ.

Solid materials exhibit an amazing range of electrical conductivities, extending over 27 orders of magnitude; probably no other physical property experiences this breadth of variation. In fact, one way of classifying solid materials is according to the ease with which they conduct an electric current. Within this classification scheme, one can classify materials into three broad categories:

(i) Conductors (ii) Semi conductors and *(iii) Insulators*

Metals and their alloys are good conductors. Copper, silver and gold are among the best conductors of electricity, followed by aluminium, iron and nickel. Some semimetals, e.g. graphite also fall in this group. According to free electron theory, when the outer orbit of an atom has less than one half of the maximum 8 electrons, the material is usually a metal and a good conductor of electricity. Metals have conductivities on the order of 10^7 $(\Omega\text{-m})^{-1}$. The electrical resistivity of conductor range from 10^{-9} to 10^{-4} (Ω-m).

Semiconductors are materials which behave are insulators at 0 °K but a significant rise in electrical conductivity is observed as the temperature rises. At room temperature, the electrical conductivity of semiconductor falls between that of a conductor and insulator. The electrical conductivity semiconductors ranges from about 10^5 to 10^{-7} $(\Omega\text{-m})^{-1}$, as compared with the conductivity on the order of 10^7 $(\Omega\text{- m})^{-1}$ for good conductors and minimum conductivity of 10^{-15} $(\Omega - \text{m})^{-1}$ for good insulators. Semiconductors form the base materials for a number of electronic devices. Germanium and silicon are the widely used common semiconductors. According to free electron theory, when the outer orbit of an atom has exactly one-half the maximum eight electrons, the material has both metal and non-metal properties and usually exhibit semiconducting properties. We must note that the electrical resistivity of a semiconductor is usually strongly dependent on temperature.

There are also materials with very low conductivities, ranging between 10^{-10} and 10^{-20} $(\Omega\text{-m})^{-1}$. These are electrical *insulators*. Mica, PVC, rubber, porcelain and bakelite are few examples of insulators. The resistivity range of insulators extends from 10^4 to 10^{17} (Ω-m). According to electron theory, when the outer orbit of an atom has more than one half of the maximum eight electrons, the material is usually a non-metal and non-conductor.

A brief comparison between metals, semiconductors and insulators is presented in *Table* 14.3.

Table 14.3 A comparison between conductors, semiconductors and insulators

Conductors	*Semiconductors*	*Insulators*
1. Conductivity decreases with increase in temperature upto nearly zero value	• Conductivity increases with increase in temperature. Conductivity is particularly sensitive to impurity type and content.	• Conductivities increases with increase in temperature
2. Conductivity of metals is of the order of 10^7 $(\Omega\text{-m})^{-1}$	• Conductivity of semiconductors range from 10^{-6} to 10^4 $(\Omega\text{-m})^{-1}$	• Conductivity of insulators range between 10^{-10} to 10^{-20} $(\Omega\text{-m})^{-1}$
3. Electrical resistivity is very low and range from 10^{-9} to 10^{-4} (Ω-m)	• Resistivity is normally high and range between 10^{-3} to 10^3 (Ω-m)	• Resistivity is very high and range between 10^4 to 10^{17}(Ω-m)
4. Temperature coefficient of resistance is not constant	• Temperature coefficient of resistance is negative	• Negative resistance temperature coefficient. Probably with the rise in temperature some electrons reach to the conduction band.
5. At low temperatures they exhibit semiconductivity. At very low temperatures, some elements and their alloys exhibit infinite conductivity, i.e. super conductivity	• At low temperature semiconductors become *dielectrics* (insulators)	• No change in the properties of the insulators observed
6. Have unfilled overlapping energy bands	• Have filled energy bands and small forbidden zones	• There is a large energy gap in between valence and conduction band
7. Current carriers in conductors are free electrons which exist whether external field is applied or not	• Current carriers are originated due to absorption of electrical, radiant or thermal energy from external source. Electrons and holes, both serve as current carriers	• Energy required for electrons to cross the energy gap between conduction band and valence band is very large and hence no conduction

4. ELECTRONIC AND IONIC CONDUCTION

We have seen that an electric current results from the motion of electrically charged particles in response to the forces that act on them from an externally applied field. Positively charged particles are accelerated in the electric field direction, and negatively charged particles in the direction opposite to the electric field direction. A current arises within most solid materials due to flow of electrons and this is termed as *electronic conduction*. In addition, for *ionic* materials a net motion of charged ions is possible that produces a electric current; such is termed *ionic conduction.*

5. BAND STRUCTURE IN SOLIDS

A solid may be thought of as consisting of a large number, say N, of atoms initially separated from one another, which are subsequently brought together and bonded to form the ordered atomic arrangement exhibited by the crystalline material. At relatively large separation distances, each atom is independent of all the other atoms as will have the atomic energy levels and electron configuration as if isolated. However, as the atoms in a solid come within the close proximity of one another, electrons are acted upon, or perturbed, by the electrons and nuclei of adjacent atoms. This influence is such that each distinct atomic state may split into a series of closely spaced electron states in the solid, to form what is termed an *electron energy band*. The extent of splitting of atomic states depends on interatomic separation (Fig. 14.1) and begins with the outermost electron shells, since they are the first to be perturbed as the atoms coalesce. Within each band, the energy states are discrete, yet the difference between adjacent states is exceedingly small. At the equilibrium spacing, band formation may not occur for the electron subshells nearest the nucleus (Fig. 14.2(b)). Moreover, gaps may exist between adjacent bands, as also shown in Fig. 14.2(a), normally energy lying within these band gaps are not available for electron occupancy. The conventional way of representing electron band structure is shown in Fig. 14.2(a).

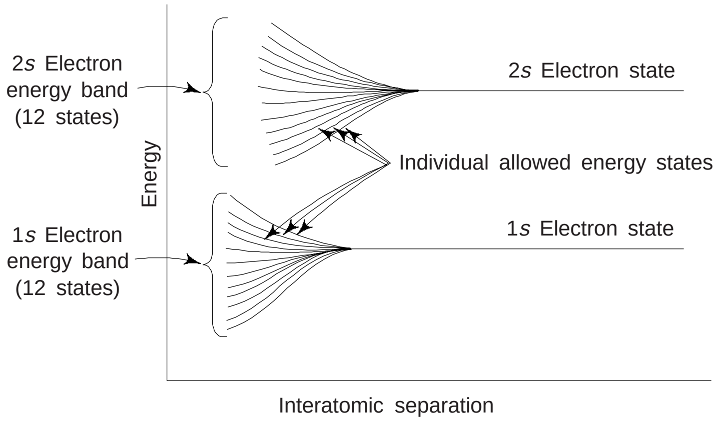


Fig. 14.1 A schematic plot of electron energy vs. interatomic separation of an aggregate of 12 atoms, i.e. for $N = 12$. Upon close approach, each of the 1*s* and 2*s* atomic states splits to form an electron energy band consisting of 12 states

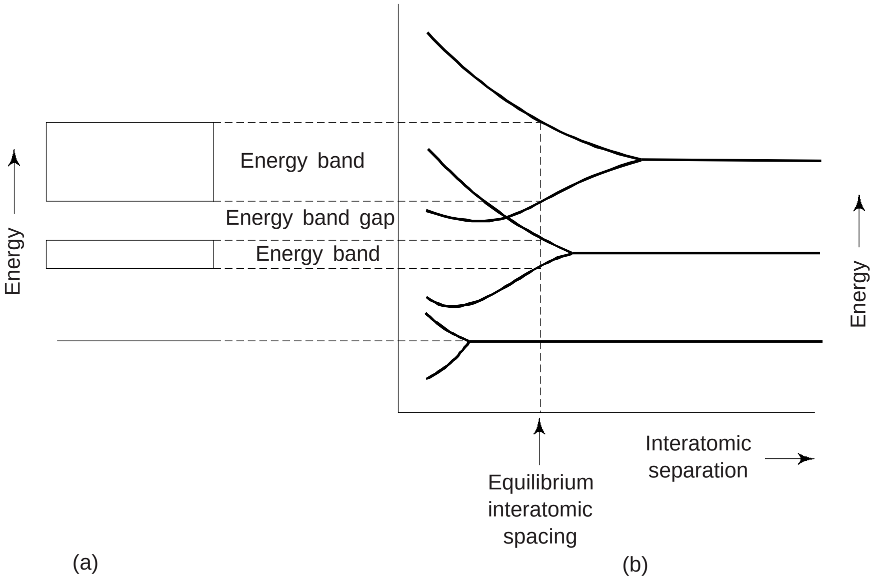


Fig. 14.2 (a) The conventional representation of the electron energy band structure for a solid material at the equilibrium interatomic separation. (b) Electron energy vs. interatomic separation for an aggregate of atoms, exhibiting how the energy band structure at the equilibrium separation in (a) is generated

The number of states within each band will equal the total of all states contributed by the N atoms, e.g. a s-band will consists of N-states, and a p-band of 3 N states. With regard to occupancy, each energy state with accommodate two electrons, which must have oppositely directed spins. Moreover, bands will contain the electrons that resided in the corresponding levels of the isolated atoms; for e.g., a 4s energy band in the solid will contain those isolated atom's 4s electrons. Of course, there will be empty band and, possibly, bands that are only partially filled.

The electrical properties of a solid material are a consequence of its *electron band structure*, i.e. the arrangement of outermost electrons in bands and the way in which they are filled with electrons. A distinctive band structure type exists for metals, for semiconductor, and for insulators.

The band structure in a solid determines whether the solid is an insulator or a conductor or a semiconductor. The bands are filled upto a certain level by the electrons within each atom. The highest band in which electrons are still predominantly attached to their atoms are found is called *valence band*. This is the band in which the valence (outermost) electrons from each atom will be located. These are the electrons that are the possible carriers of electricity. However, in order for an electron to conduct, it must get up to slightly higher energy so that it is free of the grip of its atom. At 0 K four different types of band structures are possible. In the first (Fig. 14.3(a)), one outermost band is only partially filled with electrons. The energy corresponding to the highest filled energy state of 0 K is called the Fermi energy E_F, as indicated in future. This energy band structure is typified by some metallic conductors in particular those that have a single s valence electron, e.g. copper, etc. Each copper atom has one 4s electron. However, for a solid comprised of N atom, the 4s band is capable of accommodating 2N electrons. Obviously, only half the available electron positions within this 4s band are filled.

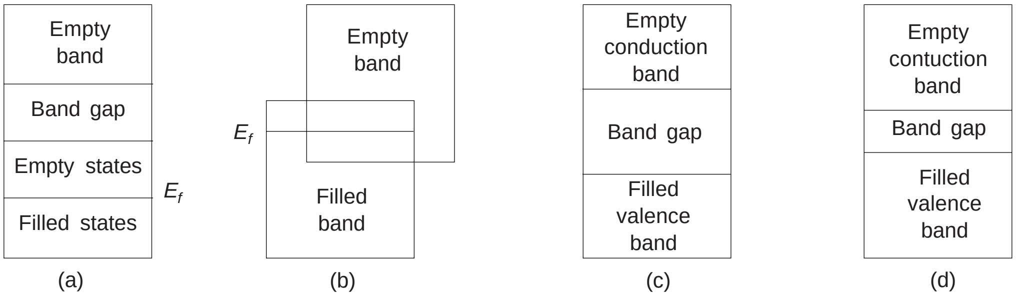


Fig. 14.3 Various possible electron band structure in solids at 0 K. (a) and (b) for conductors. Band structure (a) is found in metals such as copper, in which there are available electron states above and adjacent to filled states, in the same band. Band structure (b) is for metals such as magnesium, where in there is an overlap of filled and empty outer bands. (c) The electronic band structure characteristic of insulators; the band gap > 2 eV or more. (d) The electron band structure found in semiconductors; E_g < 2 eV

For the second band structure, also found in metallic conductors (Fig. 14.3(b)), there is an overlap of an empty band and a filled band. Magnesium has this type of band structure. Each isolated Mg has two 3s electrons. However, when a solid is formed, the 3s and 3p bands overlap. In this instance and at 0K, the Fermi energy, E_F is taken as that energy below which, for N atoms, N states are filled, two electrons per state. Therefore, the band theory tells us that we have a conductor, when

(i) the *valence band* is not filled, so electron can move to higher states in the valence band and be free (energy gaps between valence band and conduction band are very small), or

(ii) when there is no energy gap between the valence band and the conduction band, i.e. there is overlapping, so electrons can easily make the transitions from the valence to the conduction bands.

The final two band structures (Fig. 14.3(c) and (d)) are similar; one band (the *valence band*) that is completely filled with electrons that is separated from an empty *conduction band* and an *energy band gap* lies between them. When energy band gap is of the order of 1 eV (for Si, E_g = 1.12 eV; and for Ge, E_g = 0.72 eV). These materials with narrow band gap are called *semiconductors*. These are discussed in chapter 15.

When the energy band gap is relatively wide ~ 5 eV or even more, the materials are called insulators. The band theory of solids tells us that an insulator is a material in which the valence band (VB) are filled and the energy band gap E_g between VB and conduction band (CB) is too large so that valence electrons cannot jump at normal temperatures from VB to CB. An insulator does not conduct at 0 K or even at room temperature because there are no conduction electrons in it. However, an insulator may conduct if its temperature is very high or if a high voltage is applied across it. This is known as *breakdown* of an insulator.

6. CONDUCTION IN TERMS OF BAND AND ATOMIC BONDING MODELS

We have seen that as far as solid state theory is concerned only the upper energy bands (valence bands) are of interest, since electrons at lower levels practically do not take part in interactions among atoms. This means that only electrons with greater than the Fermi energy (The energy corresponding to the highest filled electron filled state at 0 K in metal is called Fermi energy and designated by E_F) may be acted on and accelerated in the presence of an electric field. These are the electrons that participate in the conduction process, which are termed *free electrons*. As mentioned earlier that another charge entity called a *hole* is found in semiconductors and insulators. Holes have energies less than Fermi energy, E_F and also participate in electronic conduction. Thus the ensuing discussion reveals, the electrical conductivity is a direct function of the number of free electrons and holes. We may also note that the distinction between conductors, semiconductors and insulators lies in the number of these free electron and hole charge carriers.

Metals

For an electron to become free, it has to be excited or promoted into one of the vacant or empty and available energy states above Fermi energy, E_F (Fig. 14.3(a) and (b)). In Figs. 14.3(a) and 14.3(b), there are vacant energy states adjacent to the highest filled state at E_F. Obviously, very little energy is required to promote electrons into the low-lying empty states (Fig. 14.4). Generally, the energy supplied by an electric field is sufficient to excite large numbers of electrons into these conducting states.

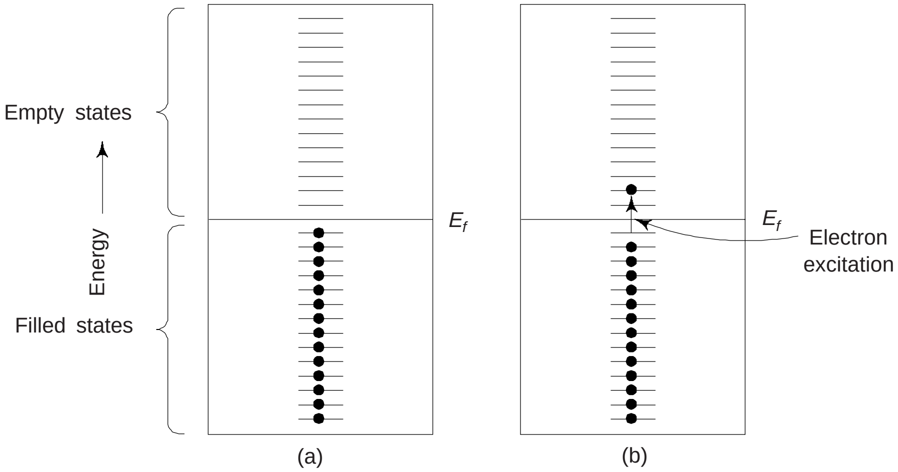


Fig. 14.4 Occupancy of electron states (a) bafore and (b) after an electron excitation for a metal

In a metallic bonding model it is assumed that all the valence electrons have freedom of motion and form an '*electron gas*', which is uniformly distributed throughout the lattice of ion cores. Although these electrons are not locally bound to any particular atom, nevertheless, they must experience some excitation to become conducting electrons that are free in true sense. Obviously, only a fraction are excited, this still gives rise to a relatively large number of free, i.e. conducting electrons and, consequently, metals exhibit a high conductivity.

Semiconductors and Insulators

We have seen that for insulators and semiconductors, empty states adjacent to the top of the filled valence band are not available. To become free, we have to promote electrons across the energy band gap and into empty states at the bottom of the conduction band. This is possible only by supplying to an electron the difference in energy between these two states, which is approximately equal to the band gap energy, E_g (Fig. 14.5).

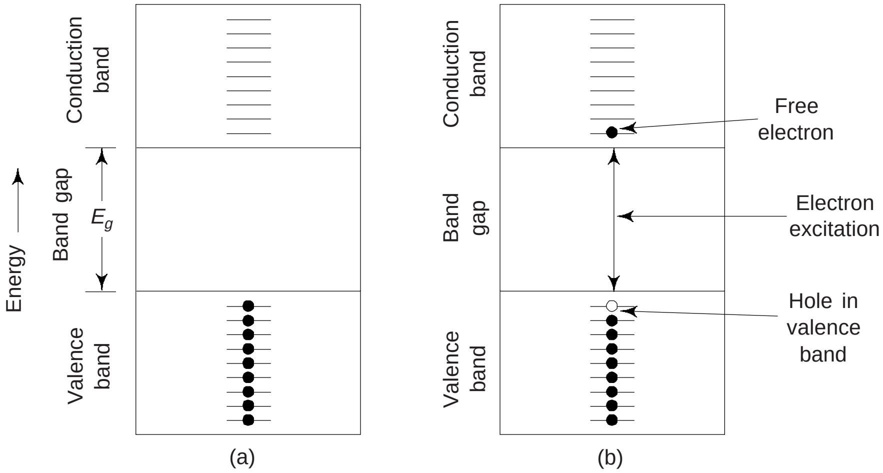


Fig. 14.5 Occupancy of electron states for an insulator and or semiconductor (a) before and (b) after an electron excitation from the valence band into the condution band, in which both a free electron and hole are generated

The number of electrons excited thermally, i.e., by heat energy into the conduction band depends on the energy band gap width as well as temperature. At a given temperature, the larger the band gap energy E_g, the lower is the probability that a valence electron will be promoted into an energy state within the conduction band; this results in very few conduction electrons. This means that larger the band gap, the lower is the electrical conductivity at a given temperature. Obviously, the distinction between semiconductors and insulators lies in the width of the band gap; for semiconductors it is narrow, whereas for insulators it is relatively wide.

With the rise in temperature, the thermal energy that is available for electron excitation increases. Obviously, more electrons are promoted into the conduction band which gives rise to an enhanced conductivity.

Let us now examine the conductivity of insulators and semiconductors in the light of atomic bonding models. For electrically insulating materials, interatomic bonding is ionic or strongly covalent. Obviously, the valence electrons are tightly bound to or shared with the individual atoms. This means that these electrons are highly localized and are not in any sense free to wander throughout the crystal. In case of semiconductors, the bonding is covalent (or predominantly covalent) and relatively weak. This means that the valence electrons in semiconductors are not as strongly bound to the atoms. Consequently, these electrons are more easily freed by thermal excitation than they are for insulators.

7. ELECTRICAL RESISTIVITY OF METALS

Most metals are extremely good conductors of electricity. Room temperature conductivities of few common metals are given in Table 14.3. Metals have high conductivities due to the large number of free electrons that have been excited into empty states above the Fermi energy. Obviously, *n* has a large value in the conductivity expression (14).

Table 14.3 Room-temperature electrical conductivity for few metals and alloys

Metal or alloy	*Electrical conductivity* $(\Omega\text{-m})^{-1}$
Copper	6.0×10^7
Silver	6.8×10^7
Gold	4.3×10^7
Aluminium	3.8×10^7
Iron	1.0×10^7
Platinum	0.94×10^7
Brass (70 Cu – 30 Zn)	1.6×10^7
Plain carbon steel	0.6×10^7
Stainless steel	0.2×10^7

Let us now discuss conduction in metals in terms of the resistivity (the reciprocal of conductivity).

The crystalline defects serve as scattering centers for conduction electrons in metals and increase in their number raises the resistivity, i.e. lowers the conductivity. The concentration of these imperfections depends on temperature, composition, and the degree of cold work of a metal specimen. It has been observed experimentally that the total resistivity of a metal is the sum of the contributions from the thermal vibrations, impurities and plastic deformation; i.e., the scattering mechanism act independently of one another. Mathematically, we can write this as follows:

$$\rho_{\text{total}} = \rho_t + \rho_i + \rho_d \tag{15}$$

where ρ_t, ρ_i and ρ_d are the individual thermal, impurity and deformation resistivity contributions, respectively. Equation (15) is sometimes called as *Matthiessen's rule*. Figure 14.6 shows a plot of resistivity versus temperature for copper and several copper-nickel alloys in annealed and deformed states. Figure also show the influence of each ρ variable on the total resistivity. The additive character of the individual resistivity contributions is demonstrated at –100°C.

7.1 Factors Affecting Resistivity

(i) Influence of Temperature: Any rise in temperature of a conductor (which contains small amounts of impurities) increases thermal agitation of the metallic ions as they vibrate about their mean position. This reduces the mean free path and restricts the free movement of electrons, thus reducing the conductivity of the metal, i.e. this increases the resistivity of metal. For the pure metal and all the copper-nickel alloys shown in Fig. 14.6, the resistivity rises linearly with temperature above about –200°C. Thus

$$\rho_t = \rho_o + aT \tag{16}$$

where ρ_o and a are constants for each particular metal.

(ii) Influence of Impurities: Another factor which reduces the mean free path of electrons is the impurity or solute atoms. The solute atoms provide the breakage in the regular crystalline structure, thus presenting an obstacle in the movement of electron waves. A solid solution alloy will always have lower conductivity than its pure components though both individual components have higher conductivity than the alloy. For additions of a single impurity that forms a solid solution, the impurity resistivity ρ_i is related to the impurity concentration C_i in terms of the atom fraction (at %/100) as follows:

$$\rho_i = AC_i\,(1 - C_i) \tag{17}$$

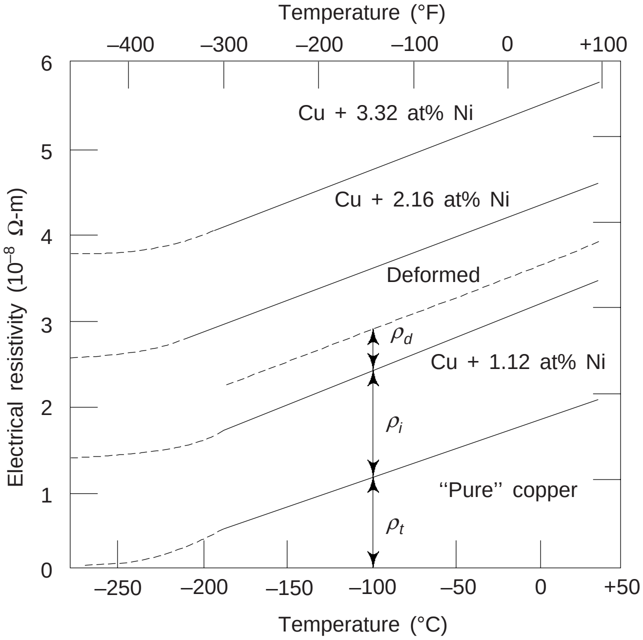


Fig. 14.6 **The electrical resistivity vs. temperature curves for Cu and three copper-nickel alloys, one of which has been deformed. The contributions to the resistivity due to thermal, impurity and deformation are shown at –100°C**

where *A* is a composition-independent constant that is a function of both the impurity and host metals. The influence of Ni impurity additions at about room temperature resistivity of Cu is shown in Fig. 14.7, upto 50 Wt% Ni; over this composition range Ni is completely soluble in Cu.

We may note that Ni atoms in Cu act as scattering centres, and increasing the concentration of Ni in Cu results in the enhancement of resistivity.

One can use the rule of mixtures expression for a two-phase alloy consisting of α and β phases to approximate the resistivity as follows:

$$\rho_i = \rho_\alpha V_\alpha + \rho_\beta V_\beta \tag{18}$$

where the *V*'s and ρ's represent volume fractions and individual resistivities for the respective phases.

(iii) Influence of Plastic Deformation: Plastic deformation also raises the electrical resistivity as a result of increased number of electron-scattering dislocations. The effect of plastic deformation on resistivity is also shown in Fig. 14.6.

(iv) Effect of Pressure: At room temperature the general behaviour of ρ of metal is to decrease initially with increasing pressure and it may pass through a minimum. The initial decrease is due to the effect of pressure in reducing the amplitude of lattice vibrations. The subsequent increase is probably due to modification of the electron band structure which leads to increased phonon scattering.

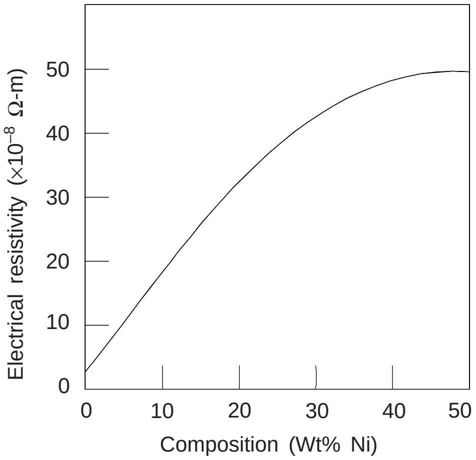


Fig. 14.7 **Room temperature electrical resistivity vs. composition for Cu-Ni alloys**

8. ELECTRICAL CHARACTERISTICS OF ALLOYS USED FOR COMMERCIAL PURPOSES

Copper is the most widely used metallic conductor due to its electrical and other properties. For many electrical applications, oxygen-free-high-conductivity (OFHC) copper having extremely low oxygen and other impurity contents is produced. Aluminium, having about one-half conductivity that of copper, is also frequently used as an electrical conductor. Silver has higher conductivity than either Cu or Al, however, its use is restricted due its very high cost compared to Cu and Al.

On occassion, it is necessary to improve the mechanical strength of a metal alloy without impairing significantly its conductivity. Both *solid solution alloying* and *cold working* improve strength at the expense of conductivity, and thus a trade off must be made for these two properties. Most often strength is enhanced by introducing a second phase that does not have so adverse an effect on conductivity, e.g. copper-beryllium alloys are precipitation hardened, but even so, the conductivity is reduced by about a factor of 075 over high purity Cu.

9. MECHANISMS OF STRENGTHENING IN METALS

We often need to design alloys having high strengths yet some ductility and toughness; ordinarily, ductility is sacrified when an alloy is strengthened. Almost all strengthening techniques rely on this simple principle: *restricting* or *hindering dislocation motion renders a material harder and stronger*. The strengthening mechanisms for single phase metals are:

(i) Strengthening by Grain-size Reduction: A fine grained material (one that has small grains) is harder and stronger than one that has coarse grained, since the former has a greater total boundary area to impede dislocation motion. For many materials, the yield strength σ_y varies with grain size in accordance with

$$\sigma_y = \sigma_o + k_y d^{-1/2} \tag{19}$$

where d is the average grain diameter, and σ_o and k_y are constants for a particular material. Equation (19) is termed the *Hall Petch equation*. We may note that Eq. (19) is not valid for both very large (i.e., coarse) grain and extremely fine grain polycrystalline materials.

The grain size reduction improves not only strength but also the toughness of many alloys.

(ii) Solid-Solution Strengthening: Another technique to strengthen and harden metals is alloying with impurities that go into either substitutional or interstitial solid solution. This is called solid-solution strengthening. Alloying elements or impurities increase the resistivity of materials due to the presence of solute atoms in solid solutions. The solute atoms interfere with the motion of electrons. In a Cu-Ni alloy, the Cu atoms do not impede the motion of free electrons but Ni atoms impede their motion. We may note that high-purity metals are always softer and weaker than alloys composed of the same base metal. Increasing the concentration of the impurity results in an attendant increase in tensile and yield strength. Alloys are stronger than pure metals because impurity atoms that go into solid solution normally impose lattice strains on the surrounding host atoms.

(iii) Strain Hardening or Work Hardening or Cold Working: This is the phenomenon whereby a ductile metal becomes harder and stronger as it is plastically deformed. Sometimes it is also called *work hardening*, or, because the temperature at which deformation takes place is "cold" relative to the absolute melting temperature of the metal, *cold working.* Most metals strain harden at room temperature.

Cold work increases the resistivity of the metal as it changes the crystal structure and the localised strain interfaces with the movement of electron. However, subsequent annealing decreases the resistivity. Hand drawn copper wire has therefore high resistivity than annealed copper.

It is sometimes convenient to express the degree of plastic deformation as *percent cold work* than as strain. Percent cold work is

$$\%\ CW = \left(\frac{A_o - A_d}{A}\right) \times 100 \tag{20}$$

where $A_o \rightarrow$ original area of cross-section that experiences deformation, and $A_d \rightarrow$ the area after deformation.

Strain hardening is often utilized commercially to enhance the mechanical properties of metals during fabrication procedures.

We may note that deformation and strengthening of multiphase alloys are more complicated, involving concepts which are beyond the scope of this book.

Example 1 *Calculate the resistance of an aluminium wire 100 m long and having cross-sectional area of 3 sq. mm at 20°C. Given, the resistivity of Al at 20°C = 2.66 × 10^{-8} Ω-m.* [AMIE]

Solution Given

$$l = 100 \text{ m}$$
$$\rho = 2.66 \times 10^{-8} \ \Omega - \text{m}$$
$$A = 3 \times 10^{-6} \text{ sq. m}$$
$$R = \frac{\rho l}{A}$$
$$= \frac{2.66 \times 10^{-8} \times 100}{3 \times 10^{-6}}$$
$$= 88.67 \times 10^{-2} \ \Omega.$$

Example 2 *The resistivity of pure copper is 1.56 micro-ohm-cm. An alloy of copper containing 2 atomic percent nickel has a resistivity of 4.06 micro-ohm-cm. An alloy of copper, containing 1 atomic percent silver, has a resistivity of 1.7 micro-Ohm-cm. Find the resistivity of a copper alloy containing 1 atomic percent nickel and 3 atomic percent silver.* [Diploma]

Solution Resistivity of pure Cu = 1.56 micro-ohm-cm. Resistivity of Cu containing two atomic percent Ni = 4.06 micro-ohm-cm

$$\therefore \quad \text{Increase due to one atomic \% Ni} = \frac{4.06 - 1.56}{2}$$
$$= 1.25 \text{ micro-ohm-cm}$$

Now, resistivity of copper, containing one atomic percent silver = 1.7 micro-ohm-cm

$$\therefore \quad \text{Increase in resistivity due to one atomic \% Ag} = 1.7 - 1.56 \text{ micro-ohm-cm}$$
$$= 0.14 \text{ micro-ohm-cm}$$

Resistivity of copper alloy containing one atomic percent Ni and 3 atomic percent

$$\text{Ag} = 1.56 + 1.25 + 3 \times 0.14$$
$$= 1.56 + 1.25 + 0.42$$
$$= 3.23 \text{ micro-ohm-cm}$$

Example 3 *The resistivity of pure copper at room temperature is 1.8 × 10^{-8} Ω-m. The resistivity of Cu – 4% Ni alloy at room temperature is 7 × 10^{-8} Ω-m. Show that the resistivity due to impurity scattering per % of Ni in the Cu lattice is 1.3 × 10^{-8} Ω-m.* [Diploma]

Solution Let us take that the resistivity of copper at 0 K is negligible. Now, 1.8×10^{-8} Ω-m is the thermal part of resistivity of copper as well as of the alloy between 0 K and room temperature.

$$\therefore \quad \text{Impurity scattering per \% of Ni is then} = \frac{(7.0 \times 1.8) \times 10^{-8}}{4} = 1.3 \times 10^{-8} \ \Omega\text{-m}$$

10. INSULATORS

We have read that in an ideal insulator, all valence electrons are occupied in bond formation and almost no free electrons are available for electrical conduction. Thus an insulator (electrical) is a non-metallic material that has a filled valence band at 0 K and relatively wide energy band gap. Consequently, the room temperature electrical conductivity is very low, less than about 10^{-10} (Ω-m)$^{-1}$. A good insulator may have a resistivity as high as 10^{14} (Ω-m). Common electrical insulating materials are polyethylene, bakelite, lucite, mica, PVC, rubber, porcelain etc. Most polymers and ionic ceramics are insulating materials at room temperatures. Thermal agitation and imperfections adversely affect the insulating properties of a material but, as is real materials, a few free electrons are always there to conduct electricity.

Non-metallic crystals are held together by ionic and covalent bonds and valence electrons are much more closely associated with their atoms than in the metallic bond. Generally, insulators are those materials which have the electrons completely filled in the *Brillouin Zone* and also at the same time have wide energy gaps.

The conductivity of insulators increases with the rise in temperature because more and more conducting electrons are released. However, the materials which are insulators at 0 K but develop significant conductivities at room temperature become semiconductors. Room temperature conductivities of few selected non-metallic materials are given in *Table* 14.4. Of course, many insulating materials are used on the basis of their ability to insulate, and thus a high electrical resistivity is desirable.

Table 14.4 Room temperature conductivities for some selected non-metallic materials

Material	*Electrical conductivity* $[(\Omega - m)^{-1}]$	*Material*	*Electrical conductivity* $[(\Omega - m)^{-1}]$
Graphite	$3 \times 10^4 - 2 \times 10^5$	POLYMERS	
CERAMICS		Phenol-formal-dehyde	10^{-9}-10^{-10}
Concrete (Dry)	10^{-9}	Polymethyl-methacrylate	$< 10^{-12}$
Soda lime glass	10^{-10}-10^{-11}	Nylon 6, 6	10^{-12}-10^{-13}
Porcelain	10^{-10}-10^{-12}	Polystyrene	$< 10^{-14}$
Borosilicate glass	$\sim 10^{-13}$	Polyethylene	10^{-15}-10^{-17}
Aluminium oxide	$< 10^{-13}$	Polytetra fluoroethylene	$< 10^{-17}$
Fused silica	$< 10^{-13}$		

One can classify the requirements of good insulating materials as electrical, mechanical, thermal and chemical. Electrically the insulating material should have high resistivity to reduce the leakage current and high dielectric strength so that it may withstand higher voltage without being punctured or broken down. Moreover, the insulator should have small dielectric loss.

Insulators are used on the basis of volume and not weight and hence a low density is preferred. A uniform viscosity and for liquid insulators ensures uniform thermal and electrical properties.

Liquid and gaseous insulators are used also as coolants, e.g. transformer oil, hydrogen and helium are used for insulation and cooling purposes. Obviously, for such materials good thermal conductivity is a desirable property. To prevent mechanical damage, the insulator should also have small thermal expansion. Moreover, it should be non-ignitable or if ignitable it should be self extinguishable.

Chemically, the insulating materials should be resistant to oils, liquid gas fumes, acids and alkalies. Moreover, the insulator should not absorb water particles as water lowers the insulation resistance and the dielectric strength.

Insulators should have certain mechanical properties depending on the use to which they are put. Thus when insulators are used for electric machine insulation, they should have sufficient mechanical strength to withstand vibration. In such cases, good heat conducting property is also desirable.

In making dielectric capacitors insulating materials with large electronic and ionic polarizabilities and therefore large permittivity are used. Titanium oxide with a permittivity ~100 is a good example of such a material. We may note that the use of molecules with a permanent dipole moment is not desirable because of possibility of large dielectric losses at high frequencies.

Properties of the some of the insulating materials are given in Table 14.5.

Table 14.5 Properties of few insulating materials at room temperature

Material	*Dielectric constant* (ε_r)	*Dielectric strength (kV/mm)*	*Density* ($\times 10^{-3}$) (kg/m^3)	*Resistivity* (ρ) (Ω-m)
Mica	4-5.5	60-125	2.5-2.7	10^{14}-10^{19}
Glass	4-10	20-30	2.2-4.0	10^{13}-10^{16}
Ebonite (RP)	4-4.5	25	1.3	1×10^{20}
Bakelite	4-4.6	10-40	1.2	—
Asbestos	—	2	2.3-2.6	2×10^{7}
Rubber (soft)	2.6-3.0	15-25	1.7-2.9	4×10^{15}
Silk (natural)	4.5	—	—	—
Polystyrene	2.2-2.8	25-50	1.05-1.65	$5 \times 10^{13} - 5 \times 10^{19}$
Polyvinyl chloride (PVC) resin	3.1-3.5	50	1.38	—

11. DIELECTRICS

These are the materials or insulators which have the unique characteristic of being able to store electric charge. The electrons in these substances are localized in the process of bonding the atoms together. Obviously, covalent or ionic bonds, a mixture of both, or Vander Waals bonding between closed shell atoms give rise to solids or gases which exhibit dielectric or insulating properties. Dielectric materials may be gases, liquids or solids with the exception of air which is the insulating material between the bare conductors of the overhead electric grid system. Liquid dielectrics are used mainly as impregnants for high voltage paper insulated cables and capacitors as filling and cooling media for transformers and circuit breakers. Most common properties of dielectric materials are: *(i)* dielectric constant *(ii)* dielectric strength *(iii)* insulation resistance *(iv)* surface resistivity *(v)* loss factor *(vi)* tangent of loss factor in terms of a capacitor or phase difference *(vii)* polar and non-polar materials.

Materials, which are capable of separating electrical conductors circuit breakers, e.g. silicon, oils, liquid dielectrics have high dielectric constant, high resistance, high dielectric strength when moisture and impurities are removed from them. They have high temperature dissipation capacity and least dielectric losses. Their dielectric constant is greater than one. Following materials are important from engineering point of view:

Mica is the widely used insulating material in switch gears armature windings, electrical heating devices like iron, hot plates etc. It is also used in capacitors for high frequency application. Mica is an inorganic compound of silicates of aluminium, soda potash and magnesia. It is crystalline in nature and can be easily split into very thin flat sheets. The two important types of mica are: *(i)* muscovite and *(ii)* phlogopite. Mica has a good dielectric strength and mechanical strength. Its dielectric constant varies between 5 and 7.5, loss tangent between 0.0003 and 0.015 and dielectric strength between 700 and 1000 kV/mm.

Asbestos is also used as an insulator in the form of paper, tape, cloth and board. Asbestos is widely used in panel boards, insulating tubes and cylinders in the construction of air cooled transformers. Asbestos is an inorganic material, which is used to designate a group of naturally occurring fibre material. Asbestos has good dielectric and mechanical properties.

Ceramics are generally non-metallic inorganic compounds, e.g. silicates, aluminates, oxides, carbides, borides, nitrides and hydroxides. Ceramics used as dielectrics may be broadly described as alumina, porcelains, ceramics, titanates, etc. These have excellent dielectric and mechanical properties. The dielectric constant of most commonly used ceramics varies between 4 and 10. These are used in switches in plug holders, thermocouples, cathode heaters, vacuum type ceramic metal seals etc. Ceramic capacitors may be operated at high temperatures and can be moulded into any shape and size.

Electric grade ceramics are used for the manufacturing of insulators, terminal blocks, plates, frames, coils, etc. They must have low losses, good insulating properties and high strength.

Insulators for operation at low frequencies are made of electric-grade porcelain which possesses fairly good electric properties. A drawback is that it has high losses which increase sharply on heating above 200°C, and a low strength.

Insulating parts for operation at high frequencies are mostly made of steatite, a talc base material. Steatites contain no harmful impurities and retain their properties at temperatures upto 100°C. They can be pressed easily, shrink on burning by only 1-2%, and are suitable for making parts in which dense but porous structure and accurate dimensions are essential. In contrast to other kinds of ceramics, steatite can be cut quite easily (after burning). Among their drawbacks are cracking under the action of rapidly varying temperatures and some difficulties the burning involves.

Glass is an inorganic insulating material, which comprises of complex system of oxides. Silica (SiO_2) is the most essential constituent of many commercially used glasses. It is fused with alkali (like potash, soda etc.) and some base (like lime, lead oxide etc.). The silica glass (having 100% SiO_2) is the best insulating material. The dielectric constant of glass varies between 3.7 and 10 and loss tangent between 0.0003 and 0.01 and dielectric strength between 2.5 and 50 kV/mm. Glass is used in electric bulbs, X-ray tubes, mercury switches, electronic valves as insulating material. It is also used in capacitors as dielectric material.

Resins are organic polymers and may be natural or synthetic. The synthetic resins are produced artificially. The commonly used synthetic resins are polyethylene, polystyrene, polyvinyl chloride, acrylic resins, teflon, nylon, etc. These have good dielectric and mechanical properties. The dielectric constant of resins varies between 2 and 4.5, the loss tangent between 0.0002 to 0.04 and dielectric strength is quite high. These are used in transformers, high frequency capacitors. These are also used as a dielectric material in d.c. capacitors.

Rubber

These are organic polymers and may be natural or synthetic. The natural rubber obtained from rubber tree has limited applications due to its resistance to low and high temperatures. The synthetic rubbers are produced artificially by *copolymerisation* of isobutylene and isoprene. These have good electrical and thermal properties. The dielectric constant of rubber varies between 2.5 and 5, and loss tangent between 0.01 and 0.03. Rubber's are used as an insulating material for electric wires, tapes, cables, coatings, motor windings, transformers, etc.

Gaseous dielectrics, e.g. nitrogen, hydrogen, etc. have dielectric constant 1.0.

The best dielectrics are polystyrene, polycarbonate, polyethylene, polymide, mineral oil, pure alumina and pure silica.

11.1 Dielectric Constant

In an electric field, the charge density D is directly proportional to the applied field, i.e.

$$D = \varepsilon E \tag{21}$$

where ε is the dielectric constant or permittivity of the material placed between the electrodes. For vacuum,

$$D = \varepsilon_o E \tag{21(a)}$$

where $\varepsilon_o = 8.854 \times 10^{-12}$ farad/m is called the absolute permittivity.

One can define the relative permittivity (ε_r) of a dielectric as:

(i) The ratio of electric field density produced in the dielectric medium (E_o) to that produced in vacuum (E) by the same electric field strength, i.e.

$$\varepsilon_r = \frac{E_0}{E}$$

ε_r is also called as macroscopic dielectric constant.

(ii) The ratio of capacitance of a condenser containing a given dielectric to the same condenser with vacuum as dielectric.

Thus relative permittivity can be express as

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \tag{22}$$

Now, charge density = $\varepsilon_o \varepsilon_r E = \varepsilon_o E + \varepsilon_o (\varepsilon_r + 1) E$

where $\varepsilon_r - 1 = \chi$ is called electric susceptibility. χ = Bound charge density/free charge density.

Permittivity is influenced by permeability (μ), stress distribution, temperature and frequency.

11.2 Phase Difference and Dielectric Loss

A dipole tends to align itself along the direction of applied electric field and for ac fields tends to follow the field and be in a phase with it. However, the interaction of this dipole with other dipoles in the medium prevents this and this leads to dielectric loss. This loss appears as heat. This dielectric loss is connected with ε_r'', the imaginary part of the dielectric constant. The rate of loss of energy in unit volume of the material is obtained as

$$W = \frac{1}{2} \varepsilon_0 \varepsilon_r'' \omega E_0^2 \tag{23}$$

Obviously, the energy loss is proportional to ε_r''. Usually the dielectric loss is expressed in terms of a quantity called *'the loss tangent'*, tan δ. This is defined as $\tan \delta = \varepsilon_r''/\varepsilon_r'$. The angle δ is the compliment of angle ϕ, i.e. $90 - \phi = \delta$, where ϕ is the angle between the applied field and the resultant current vector (Fig. 14.8).

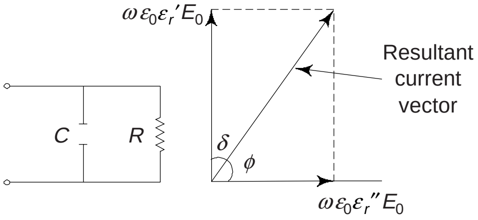


Fig. 14.8 **Vector relationship between the field vector E_0, the current vector $\omega \varepsilon_0 \varepsilon_r' E_0$ which leads the field by 90° and the current vector $\omega \varepsilon_0 \varepsilon_r'' E_0$ which is in phase with the field. The angle δ is shown**

From alternating current studies, we have

$$\text{Power} = V_R I_R \cos \phi$$

where V_R and I_R are rms values of voltage and current, and for a circuit containing capacitance C and resistance R. We have

$$\tan \delta = \frac{1}{\omega CR} \tag{24}$$

Using (22), one obtains for a dielectric

$$W = \frac{1}{2} \varepsilon_0 \varepsilon' \tan \delta \, \omega E_0^2 = \frac{1}{2} \frac{E_0^2}{\rho} \tag{25}$$

where ρ is the resistivity. Thus

$$\tan \delta = \frac{1}{\omega \varepsilon_0 \varepsilon' \rho} \tag{26}$$

The energy losses in a dielectric material are due to: *(i)* ionization, *(ii)* leakage current, *(iii)* polarization, and *(iv)* structural inhomogeneity.

Ionization losses occurs in gases and solids having pores with entrapped gases. With the rise in field strength applied to a gas, a stage is reached when the gas molecules gets ionized due collisions. This leads to enhanced conduction leading to dielectric losses. At low or ordinary electric fields, the conductivity is low and hence the loss is also small (Eq. (25)). For example, for $\rho \approx 10^{16}$ Ω–m, $\varepsilon_r' \approx 1$ and at 100 Hz, $\tan \delta < 2 \times 10^{-8}$. In case of solids with gas inclusions the loss tangent increases with voltage from the voltage required for the ionization of gas molecules to the value of voltage at which ionization is complete. When such inclusions are high, there are chances of the failure of the dielectric.

Due to the leakage current, the losses in highly conducting liquids and solids are high (Eq. (24)). In case of neutral liquids the losses will be small as their conductivity is low. Purity of oil is quite important as otherwise dielectric losses increase due to conduction. Pure transformer oil has tan $\delta \approx 0.0005$.

Dipole losses in the radiofrequency region are usually due to dipole rotation or ions jumping from one equilibrium position to another. The dielectric losses associated with ions, the frequencies of which fall in the infrared region, are usually referred to as *infrared absorption*. Similarly, the losses in the optical region, associated with the electrons, are referred to as *optical absorption*.

11.3 Dielectric Strength

As the voltage applied across a dielectric is increased, dielectric loses its insulation property at a certain voltage called the *breakdown voltage* (V_{br}). The corresponding field strength (E_{br}) is given by

$$E_{br} = \frac{V_{br}}{t} \tag{27}$$

Here t is the thickness of the material. This voltage, V_{br} per unit thickness of the material is called the *dielectric strength* and usually expressed in Volts/mil or kV/millimeter or mega Volt/m. The breakdown of a dielectric is due to the collision of accelerated electrons or ions with molecules. At relatively high fields, the electrons in the dielectric gain enough energy to knock other charged particles and make them available for conduction.

Dielectric failure involves deterioration and electric breakdown or cascading. Resistance to electronic breakdown in a dielectric is *intrinsic dielectric strength* of the material. Intrinsic breakdown in a dielectric is caused by imperfection. When the applied potential across the dielectric becomes high enough, a few electrons are broken loose at points where their bonds are strained by the presence of imperfections. Upon being freed, these electrons are accelerated rapidly through the material. The dielectric material loses completely its insulating capacity as a result of this process.

Obviously, dielectric strength is structure sensitive property of the material. We may note that the actual thickness of the material affects the breakdown potential (V_{br}) per unit thickness, i.e. dielectric strength. Usually thicker materials possess lower dielectric strength than thinner materials.

The fundamental breakdown mechanisms in a solid are: *(i) intrinsic breakdown (ii) thermal breakdown (iii) discharge breakdown* and *(iv) electrochemical breakdown.*

11.4 Power Factor

The ratio of power loss in a material to the product of applied voltage and current is called power factor. The power factor and loss angle depend upon the nature of the material, applied voltage, humidity, temperature and frequency.

11.5 Polar and Non-Polar Materials

Permanent dipole moments exist in some molecules by virtue of an asymmetrical arrangement of positively and negatively charged regions. Such molecules are termed polar molecules. For example, in the HCl molecule (Fig. 14.9a), the electron of hydrogen atom spends more time moving around the Cl atom than around the H atom. Therefore the centre of negative charges does not coincide with that of positive charges, and the molecule has a dipole moment directed from the Cl atom to the H atom. That is, we may write H^+Cl^-. The electric dipole of HCl molecule is $p = 3.43 \times 10^{-30}$ C-m. In the CO molecule Fig. 14.9 (b), the charge distribution is only slightly asymmetric and the electric dipole moment is relatively small, about 0.4×10^{-30} C-m, with the carbon atom corresponding to the positive and the oxygen atom to the negative end of the molecule.

Polar molecules can also induce dipoles in adjacent non polar molecules, and a bond will form as a result of attractive forces between the two molecules. Moreover, the magnitude of this bond will be greater than for fluctuating induced dipoles.

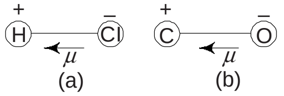


Fig. 14.9 Polar diatomic molecules

Vander Waals forces will also exist between adjacent polar molecules. We may note that the associated bonding energies are significantly greater than for bonds involving induced dipoles.

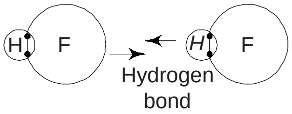
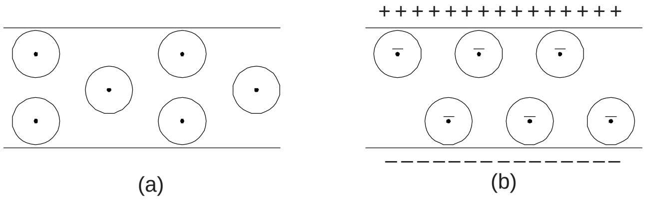


Fig. 14.10 Representation of hydrogen bonding in HF molecule

The strongest secondary bonding type, the hydrogen bond, is a special case of polar molecule bonding. We find that it occurs between molecules in which hydrogen is covalently bonded to fluorine (as in HF molecule), oxygen (as in H_2O molecule), and nitrogen (as in NH_3 molecule). For each H–F, H–O or H–N bond, the single hydrogen electron is shared with the other atom. Obviously, the hydrogen end of the bond is essentially a positively charged bare proton that is unscreened by any electrons. This highly positively charged end of the molecule is capable of a strong attractive force with the negative end of an adjacent molecule (Fig. 14.10). Clearly, this single proton forms a bridge between two negatively charged atoms. We may note that the magnitude of the hydrogen bond is generally greater than that of the other types of secondary bonds and may be as high as 0.52 eV/molecule (= 51 kJ/mol). We find that the melting and boiling temperatures for hydrogen fluoride and H_2O are abnormally high in light of their low molecular weights, as a consequence of hydrogen bonding.

Polarization

When a material is placed in an electric field (Fig. 14.11), e.g. between the plates of a condenser, the field strength of charged particles within the material interact with the electric field. If the material is a conductor, some of the free electrons simply move to the side nearest the positive electrode until they totally counteract the applied electric field. Obviously, no field is left within the material. The displacement of charged particles occurs almost simultaneously, bringing about equilibrium.



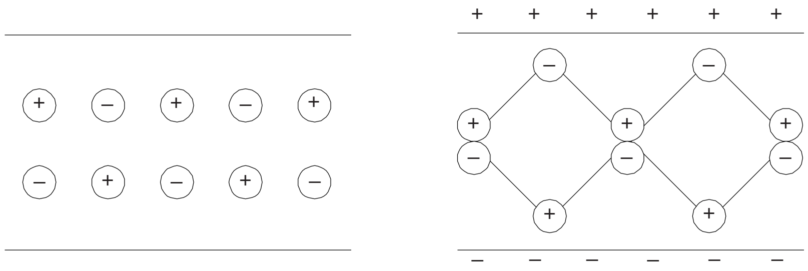
(a) (b)

Fig. 14.11 Electronic polarization (a) In the absence of field (b) In the presence of electric field, the formation of induced dipole

If the material is non-conducting or insulator (dielectric), electrons can only be displaced locally as they are bound to the individual atoms. This local displacement of electrons, however, is sufficient to polarize the material. The negative electron cloud is displaced in each atom relative to the positive nucleus, thereby creating a small induced dipole whose negative pole is toward the positive side of the electric field. All dielectric materials are subjected to such *electronic polarization* (Fig. 14.11).

Induced polarization also occurs in ionic materials. In an ionic crystal, for e.g. the negative ions are attracted towards the positive side and vice versa as illustrated in Fig. 14.12. NaCl is a good example of ionic polarization. Its interionic separation is 0.1 nm and dipole moment is of the order of 10^{-29} C-m.

There are many molecular structures which have permanent electronic dipoles due to asymmetric nature of their atomic structures. Few examples of such molecules are: H_2O, HCl, polyvinylchloride, etc. Some molecules also contain permanent charges which produce dipole moments in the individual molecules. There are many polymers of this type (polymers) and glasses contain dipoles. The permanent molecular dipoles in such materials can rotate about their axis of symmetry to align with an applied field which exert a torque in them. This is known as *orientation polarization.* This type of polarization is rarely perfect as thermal fluctuations do not permit all the dipoles to align with the applied field. With the rise in temperature



(a) (b)

Fig. 14.12 Ionic Polarization (a) In the absence of external field (b) Ions displacement towards opposite charged plates in an external field

the degree of polarization decreases. In the process of orientation polarization carriers collect at the boundaries between the constituent materials resulting in a local distortion of the applied field. This is called as *Space-Charge polarization*.

The dipole moment of a single pair of charges is given by

$$\mu = qd \tag{28}$$

where q is one of the charges and d is the distance between charges.

Non-Polar Materials

Most of the hydrocarbons are non-polar. In non-polar materials the molecules, which are usually diatomic and composed of two atoms of the same type, may be represented as a positive nuclei of charge q surrounded by a symmetrically distributed negative electron cloud of charge $-q$. In the absence of an applied electric field the centres of gravity of the positive and negative charge distributions coincide. When the molecules are subjected to an external electric field the positive and negative charges experience electric forces tending to move them apart in the direction of external electric field. The distance moved is very small ($\sim 10^{-10}$ m) since the displacement is limited by restoring forces which increases with increasing displacement. The centres of positive and negative charges no longer coincide and the molecules are said to be polarized.

11.6 Insulation Resistance

This separates a number of a conductors at different electrical potentials and does not allow a large flow of electric current between them. The difference of potential may cause leakage of current along two paths, i.e. *(i)* over the surface of insulation, and *(ii)* through the solid material of the insulators.

We may note that the resistance offered along the said two paths is not the same due to the nature of the material. The former is due to the surface resistivity of the material, whereas the latter is the result of the volume resistivity of the material. The combined effect of two is called the insulation resistance of the material.

We may define the surface resistivity as the resistance between the two opposite edges of a square of unit area of insulation surface. The volume resistivity is defined as the resistance presented to the flow of an electric current by a material of unit cross section and of unit length at 0°C.

The resistive properties of real dielectrics are:

(i) Under the action of a potential, mobility with which any conducting species moves.

(ii) Temperature variation of conductivity to state why the conductivity of a dielectric varies in exponential manner with temperature.

(iii) To analyse the time response of dielectrics, frequency or time variation of conductivity.

(iv) While designing an equipment, the knowledge of breakdown or the highest voltage a dielectric will withstand is of interest. Usually insulators breakdown at 10^6 V/m to 10^9 V/m dc at 20°C.

11.7 Classification of Dielectrics

One can classify dielectric as under following three groups:

Class	*Effect of electric field*
Simple dielectrics	Creates dipoles
Paraelectrics	Orients dipoles
Ferroelectrics	Orients domains of aligned permanent dipoles

11.8 Ferroelectricity

The group of dielectric materials called *ferroelectrics* exhibit spontaneous polarization i.e., polarization in the absence of an electric field. In a sense, ferroelectrics are the electric analog of the ferromagnets, which may display permanent magnetic behaviour. In ferroelectrics, the polarization can be changed and even reversed by an external electric field. The reversibility of the spontaneous polarization is due to the fact that the structure of a ferroelectric crystal can be derived from a non-polarized structure by small displacement of ions. In most ferroelectric crystals, this non polarized structure becomes stable if the crystal is heated above a critical temperature, the *ferroelectric Curie temperature* (T_c); i.e. the crystal undergoes a phase transition from the polarized phase (ferroelectric phase) into an unpolarized phase (Paraelectric phase). The change of the spontaneous polarization at T_c can be continuous or discontinuous. The T_c of different types of ferroelectric crystals range from a few degrees absolute to a few hundred degrees absolute.

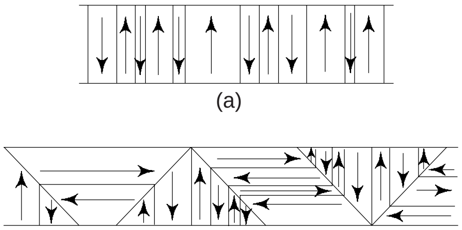
From a practical standpoint ferroelectrics can be divided into two classes: *(i)* In this *first class* of ferroelectrics, polarization can occur only one crystal axis, e.g. Rochelle salt, KH_2PO_4, $(NH_4)_2\ SO_4$, guanidine aluminium sulphate hexahydrate, glycine sulphate, colemanite and thiourea, *(ii)* In this *second class* of ferroelectrics, spontaneous polarization can occur along several axes that are equivalent in the paraelectric phase, e.g. $BaTiO_3$-type (or perovskite type) ferroelectrics, $Cd_2Nb_2O_7$, PbN_2O_6, certain alums, such as methyl ammonium alum and $(NH_4)_2 - Cd_3(SO_4)_3$.

One can also place ferroelectrics as *proper ferroelectrics* and *improper ferroelectrics*. In proper ferroelectrics, for e.g., $BaTiO_3$, KH_3PO_4, and Rochelle salt, the spontaneous polarization is the order parameter. In improper ferroelectrics, the spontaneous polarization can be considered a by-product of another structural phase transition. The properties of some representative ferroelectrics are given in *Table* 14.6.

Table 14.6 Properties of some representative ferroelectric crystals

Group	*Crystal*	*Curie temperature* T_c *(°K)*	P_s $C/m^{-2} \times 10^{-2}$	*At T (°K)*
Ilmenites and Perovskites	GeTe	670	—	—
	$LiNbO_3$	1480	71	296
	$KNbO_3$	710	30	600
	$BaTiO_3$	393	26	300
	$SrTiO_3$	32	3	4.2
KDP type	KH_2PO_4 (KDP)	123	4.7	100
	KD_2PO_4	213	5.5	100
	$Rb_2H_2PO_4$	147	5.6	90
	KH_2AsO_4	97	5.0	78
TGS type	$(NH_2CH_2COOH)_3 \cdot H_2SO_4$ (Triglycine sulphate)	322	2.8	275
Rochelle salt type	$NaKC_4H_4O_6 \cdot 4H_2O$ (Rochelle salt)	296 (upper) 255 (lower)	0.25	275

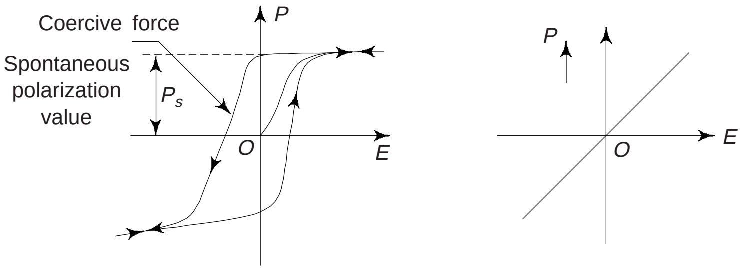
The spontaneous polarization in ferroelectrics can occur in at least two equivalent crystal direction, thus a ferroelectric crystal consists in general of regions of homogeneous *domains*. Ferroelectrics of the first class consist of domains with parallel and antiparallel polarization (Fig. 14.13(a)), whereas ferroelectrics of the second class can assume much more complicated domain configurations (Fig. 14.13(b)). The region between two adjacent domains is called a *domain wall*. Within this wall, the spontaneous polarization changes its direction.


(a)

(b)

Fig. 14.13 Domain configurations (simplified) encountered in ferroelectric. (a) First class (b) Second class

When an electric field is applied to a specimen of a ferroelectric crystal, the polarization first rises rapidly with applied field to a value above which the dependence is linear. Linear extrapolation to zero field gives, the *saturation* or *spontaneous polarization*. On subsequently reducing the field to zero, remanent (residual) polarization remains. The negative field to reduce the polarization to zero is called the *coercive field*. Obviously, a ferroelectric crystal can be switched by the application of an electric field and a *hysteresis loop* is associated with the switching. The existence of a dielectric hysteresis loop in a dielectric material implies that the substance possesses a spontaneous polarization, P_s and the value of spontaneous polarization (depending upon the shape of hysteresis loop) depends upon a number of factors such as the dimensions of the specimen, the temperature, the texture of the crystal, the thermal and electrical properties of the crystal. A typical hysteresis loop is shown in Fig. 14.14. For most ferroelectrics, the values of P_s are between 10^{-7} and 10^{-4} C/cm^2. In non ferroelectric dielectrics, electric fields between 10^5 and 10^8 V/cm would be necessary in order to achieve such large polarizations.



(a) (b)

Fig. 14.14 Net polarization *P* of a ferroelectric crystal versus externally applied electric field. (a) Hysteresis (b) *P*– *E* relation above curie temperature

The hysteresis loop of a ferroelectric material changes its shape as the temperature is increased. The height and width decrease with increase of temperature. At a certain temperature known as *ferroelectric Curie temperature* (T_c), the loop merges to a straight line and the ferroelectric behaviour of the material disappears. The electric susceptibility χ_E in the *paraelectric phase* (above curie temperature, T_c) is related to the temperature by the Curie-Weiss law

$$\chi_E = \frac{C}{T - T_C} \tag{29}$$

where C is Curie constant and $T_c \rightarrow$ Curie-Weiss temperature.

As a rule, the dielectric constant ε measured along a ferroelectric axis increases in the paraelectric phase when the Curie Temperature, T_c is approached. In many ferroelectrics, this increase can be approximated by Curie-Weiss law,

$$\varepsilon = \frac{C}{T - T_0} \qquad 29(a)$$

where T_0 is equal or somewhat smaller than T_c. For $BaTiO_3$, this law holds unaltered upto frequencies of 2.4×10^{10} Hz. Ferroelectric crystals will have an extremely large dielectric constant.

One of the most common ferroelectrics is the barium titanate ($BaTiO_3$). The spontaneous polarization in this crystal is a consequence of the positioning of the Ba^{2+}, Ti^{4+}, and O^{2-} ions within the unit cell (Fig. 14.15). The Ba^{2+}, ions are located at the corners of the unit cell, which is of tetragonal symmetry (a cube that has been elongated slightly in one direction). The dipole moment results from the relative displacements of the O^{2-} and Ti^{4+} ions from their symmetrical positions as shown in the side view of the unit cell. We can see that O^{2-} ions are located near, but slightly below, the centers of each of six faces, whereas the Ti^{4+} ion is displaced upward from the unit cell corner. Obviously, a permanent ionic dipole moment is associated with each unit cell. However, when $BaTiO_3$ is heated above its *ferroelectric* T_c (= 120°C), the unit cell becomes cubic, and all ions assume symmetric positions within the cubic unit cell; the material now has a perovskite crystal structure, and ferroelectric behaviour ceases. Ceramic materials and especially ferroelectrics are the most important ones among dielectrics. Ceramics may have very diversified electric properties (Table 14.7) and are almost insusceptible to ageing and heating.

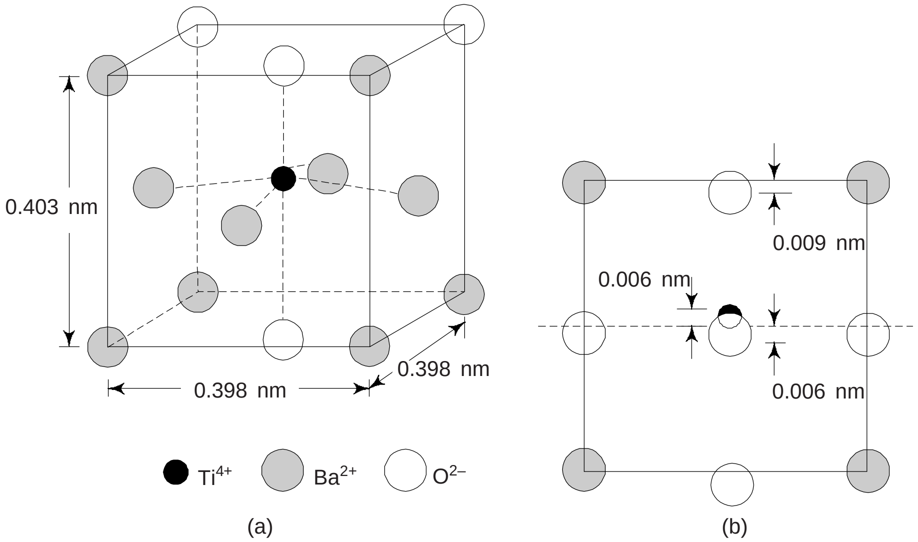


Fig. 14.15 **$BaTiO_3$ unit cell (a) in an isomeric projection, and (b) looking at one face, exhibiting the displacement of Ti^{4+} and O^{2-} ions from the center of the face**

Table 14.7 Properties of dielectrics

Material	ε	*tan* $\delta \times 10^4$	ρ_v, (Ω-m)	E_{br} (MV/m)
Low frequency dielectrics				
Ceramics				
• electric grade porcelain	6.5	200	10^{11}	25
• ferroelectric ceramics T-7500	7500	300	10^7	2.5
• ferroelectric ceramins VK-1 for variconds	$(20\text{-}100) \times 10^3$	1800	10^6	2.5
• piezoelectric ceramics T-1700	1700	300	10^6	2.5
Plastics				
• polyvinyl chloride	4-13	200	10^{12}	16-30
• powdered phenolic plastic with wood dust	11-5	1100-2700	10^5-10^9	10-11
• same with minear filler	10-6	500-1000	10^{10}-10^{11}	18
• epoxy plastics	4-2.6	400-500	10^9-10^{12}	16-20

(*Contd.*)

Table 14.7 (*Contd.*)

Material	ε	*tan* $\delta \times 10^4$	ρ_v, (Ω-m)	E_{br} (MV/m)
High frequency dielectrics				
Ceramics				
• ultraporcelain	8.5	2-10	10^{10}	20
• alumina	9.5	2	10^{13}	15
• steatite	6.5	3-9	10^{9}	20
• celsian ceramics	7.5	3	10^{10}	35
• ticond T-150	150	3	10^{8}	10
• thermocond T-20	20	3	10^{8}	10
Plastics				
• polyethylene	2.4	2-5	10^{13}	20
• polystyrene	2.5	3-5	10^{13}	30
• polytetrafluoroet hylene	2	2	10^{15}	20
Glass	3.5-16	18-175	10^{11}-10^{16}	30-100
Glass ceramics	5-7	3-20	10^{12}-10^{16}	28-48

Spontaneous polarization of this group of materials results as a consequence of interactions between adjacent permanent dipoles where in they mutually align, all in the same direction. For e.g., with $BaTiO_3$, the relative displacements of O^{2-} and Ti^{4+} ions are in the same direction for all the unit cells within some volume region of the specimen. At room temperature, ε for $BaTiO_3$ may be as high as 5000. Consequently, capacitors made from these ferroelectric materials can be significantly smaller than capacitors made from other dielectric materials. This is why that these materials are widely used in the manufacture of miniatured capacitors.

11.9 Electrostriction

This is a mechanical deformation which always accompanies polarization in a dielectric. An electric field polarizes any material by inducing dipole moments. This displacement of charges from their equilibrium positions alters the mechanical dimensions of a solid; it causes electrostriction. However, mechanical stress applied to a neutral material cannot induce dipole moments; i.e., *electrostriction has no inverse*. If a mechanical distortion creates a voltage, the effect must be caused by permanent dipole moments anchored in the structure without a centre of symmetry.

11.10 Piezoelectricity or Pressure Electricity

Electricity, or electric polarity, resulting from the application of mechanical pressure on a dielectric crystal is called piezoelectricity or pressure electricity. The application of a mechanical stress produces in certain dielectric (electrically non-conducting) crystals an electric polarization (electric dipole moment per cubic meter) which is proportional to this stress. If the crystal is isolated, this polarization manifests itself as a voltage across the crystal, and if the crystal is short-circuited, a flow of charge can be observed during loading. Conversely, application of a voltage between certain faces of the crystal produces a mechanical distortion of the material. This reciprocal relationship is referred to as the piezoelectric effect. The phenomenon of generation of a voltage under mechanical stress is referred to as the *direct piezoelectric effect*, and the mechanical strain produced in the crystal under electric stress is called the *converse piezoelectric effect*. The piezoelectric effect is demonstrated in Fig. 14.16.

Piezoelectricity occurs only in insulating materials. Only few ceramic materials exhibit this property. Piezoelectric materials are used extensively in *transducers* for converting a mechanical strain into an electrical signal. Such devices include microphones, phonograph pickups, vibration-sensing elements, and the like. The converse effect, in which a mechanical output is derived from an electrical signal input, is also widely used in such devices as sonic and ultrasonic transducers, headphones, loudspeakers, and cutting

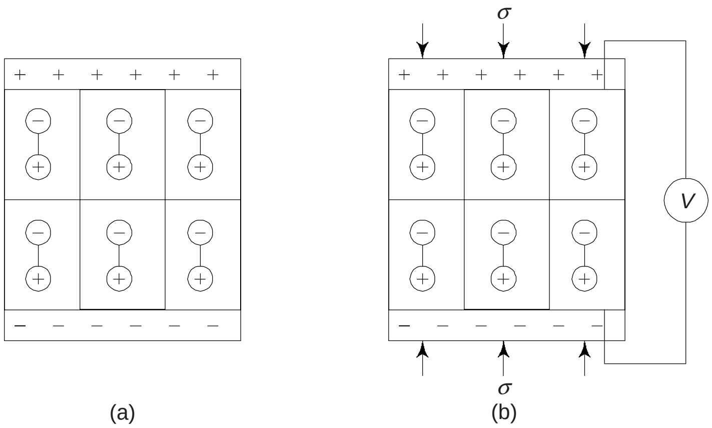


Fig. 14.16 Piezoelectric effect (a) dipoles within a piezoelectric material (b) a voltage is generated when the material is subjected to a compressive stress

heads for disk recording. Both the direct and converse effects are employed in devices in which the mechanical resonance frequency of the crystal is of importance. Such devices include electric wave filters and frequency control elements in electronic oscillator circuits.

The necessary condition for the piezoelectric effect is the absence of symmetry in the crystal structure. Of the 32 crystal classes, 21 lack a centre of symmetry, and with the exception of one class, all of these are piezoelectric. Piezoelectric materials include titanates of barium and lead, zirconate ($PbZrO_3$), ammonium dihydrogen phosphate ($NH_4H_2PO_4$), natural quartz. As stated earlier, this property is a characteristic of materials having complicated crystal structure with a low degree of symmetry. One may improve the piezoelectric behaviour of a polycrystalline specimen by heating above its Curie temperature and then cooling to room temperature in strong electric field.

Piezoelectrics are required to have high piezoelectric modulus and low losses. Ferroelectric ceramics have the piezoelectric modulus around 6×10^{-10} C/N, which is one or two order of magnitude higher than that of quartz. The structure of piezoelectric ceramics is a solid solution of barium titanate, barium niobate, lead niobate or lead titanate.

Piezoelectric materials have the following important applications: *(i)* frequency resonators *(ii)* gramophone pickups *(iii)* filters *(iv)* ultrasonic flaw detectors *(v)* underwater sonar transducers and *(vi)* air transducers (ear-phones, microphones, hearing aids, etc.).

11.11 Uses of Dielectrics

We have seen that dielectric materials are electrically insulative, yet susceptible to polarization in the presence of an electric field. This polarization phenomenon accounts for the ability of the dielectrics to increase the charge storing capability of capacitors. Now, we can summarize the main uses of dielectrics as follows: *(i)* piezoelectric and electro-optic devices *(ii)* in capacitors, resistors and strain gauges *(iii)* thermionic valves, radiation detectors, electric devices, dielectric amplifier *(iv)* dielectrics are usually used as ordinary insulators in power cables, signal cables, electric motors, etc. *(v)* dielectrics are used in transformers and various form of switchgear and generators where the dissipation problem of heat is active, and a common way of getting rid of it is to insulate with a transformer oil, i.e. mineral oil.

Example 4 *Show that the relative dielectric constant of a barium titanate crystal, which, when inserted in a parallel plate condenser of area 10 mm × 10 mm and distance of separation of 2 mm, gives a capacitance of 10^{-9} F is 2259.* [M.Sc.(M.S.)]

Solution We have $C = 10^{-9}$ F

$d = 2 \times 10^{-3}$ m

$$\varepsilon_o = 8.854 \times 10^{-12}$$
$$A = 10 \times 10^{-3} \times 10 \times 10^{-3} = 10^{-4} \text{ m}^2$$
$$C = \frac{\varepsilon_r \varepsilon_0 A}{d}$$
$$\therefore \quad \varepsilon_r = \frac{Cd}{\varepsilon_0 A}$$
$$= \frac{10^{-19} \times 2 \times 10^{-3}}{8.854 \times 10^{-12} \times 10^{-4}}$$
$$= 2259$$

Example 5 *Using the given data for $BaTiO_3$ crystal, calculate the polarization. The shift of the titanium ion from the body centre is 0.06 Å. The oxygen anions of the side faces shift by 0.06 Å, while the oxygen anions of the top and bottom faces shift by 0.08 Å, all in a direction opposite to that of the titanium ion.* [B.E]

Soluiton Let us calculate the dipole moments due to the effective number of each type of ion in the unit cell of $BaTiO_3$ as:

(i) dipole moment due to two O^{2-} ions on the four side faces

$$\mu_1 = Qd = 2 \times 2 \times 1.6 \times 10^{-19} \times 0.06 \times 10^{-10} \text{ C-m}$$

(ii) dipole moment due to one O^{2-} on top and bottom

$$\mu_2 = Qd = 1 \times 2 \times 1.6 \times 10^{-19} \times 0.08 \times 10^{-10} \text{ C-m}$$

(iii) dipole moment due to one Ti^{4+} ion at body centre

$$\mu_3 = Qd = 1 \times 4 \times 1.6 \times 10^{-19} \times 0.06 \times 10^{-10} \text{ C-m}$$

$\therefore$ Total dipole moment = 1.02×10^{-29} C-m

Now, the polarization is the total dipole moments per unit volume. Now, ignoring the effect due to barium ions, one obtains

$$P = \frac{\mu}{A} = \frac{1.02 \times 10^{-29}}{4.03 \times (3.98)^2 \times 10^{-30}}$$
$$= 0.16 \text{ C-m}^2.$$

12. MAGNETISM

This comprises those physical phenomena involving magnetic fields and their effects upon materials. Magnetic fields may be set up on a macroscopic scale by electric currents or by magnets. On atomic scale, individual atoms cause magnetic fields when their electrons have a net magnetic moment as a result of their angular momentum. A magnetic moment arises whenever a charged particle has an angular momentum. It is the cooperative effect of the atomic magnetic moments which causes the macroscopic magnetic field of a permanent magnet. However, the underlying principles and mechanisms that explain the magnetic phenomenon are complex and subtle, and their understanding has eluded physicists until relatively recent times. Several of our modern technological devices rely on magnetism and magnetic materials; these include electric motors, electric power generators and transformers, components of sound and video reproduction systems, telephones, radio, television, computers, etc.

Well known examples of magnetic materials which exhibit magnetic properties are: iron, some steels and the naturally occurring mineral lodestone. The important facts about the magnetic materials are:

(i) There are some materials which exhibit magnetic properties even without the application of any magnetic field and become more magnetic when a weak magnetic field is applied to them.

(ii) There are many other materials which lose their initially strong magnetism when heated above a certain critical temperature and become comparatively weakly magnetised.

(iii) There are some materials which show a magnetic response in a direction opposite to that of any externally applied field.

(iv) Magnetic materials are all media capable of being magnetized in a magnetic field, i.e. of creating their own magnetic field. According to their magnetic properties, such materials are divided into three principal groups: *diamagnetic*, *paramagnetic* and *ferromagnetic materials*. From the applications point of view all the magnetic materials can be placed under two groups: *(i)* Soft and *(ii)* Hard magnetic materials.

Iron, some steels, and the naturally occurring mineral lodestone are well-known examples of materials which exhibit magnetism. Ferro and ferrimagnetic materials are the most important magnetic materials from the point of view of practical applications. We will first consider the terms and definitions used in magnetism.

11.1 Terminology

(i) Magnetic Induction or Magnetic Flux Density (B): In the presence of magnetic field in vacuum, the magnetic induction (B) is related to the field strength H (in units of A–m^{-1}) as follows:

$$B = \mu_o H \tag{30}$$

where μ_o is called the *permeability of free space*. B is expressed in units of *tesla* or Weber per square metre. The units of permeability are

$$\mu_o = \frac{B}{H} = \frac{\text{Wb} - \text{m}^2}{\text{A} - \text{m}^{-1}} \qquad (\because \text{ Wb} - \text{m}^2 = \text{NA}^{-1}\text{m}^{-1})$$

$$= H - m^{-1} \text{ (H = Wb} - \text{A}^{-1} \text{ where H} \rightarrow \text{Henry)}$$

In SI units, the permeability of free space (μ_o) has a value of $4\pi \times 10^{-7}$ H-m^{-1}.

(ii) Magnetic Field (H): It is said to occupy a region when the magnetic effect of an electric current or of a magnet upon a small test magnet which is brought in the vicinity is detectable. Magnetic field strength is denoted by H. When a magnetic material is placed in a magnetic field H, it becomes magnetized, i.e., it becomes itself a magnet. Magnetic field strength (H) is expressed in units of A-m^{-1}.

It the magnetic field is applied to a solid medium, the magnetic induction in the solid is given by a relationship

$$B = \mu \text{H} \tag{31}$$

where μ is the permeability of the solid material through which the magnetic lines of force pass. In general μ is not equal to μ_0. The ratio μ/μ_0 is the relative permeability of the medium and designated by μ_r. Mathematically,

$$\mu_r = \frac{\mu}{\mu_0} \tag{32}$$

(iii) Magnetization: This may be defined as the process of converting a non-magnetic bar into a magnetic bar. This term is almost analogous to the polarization in dielectric materials. The flux density

$$\begin{aligned} B &= \mu H = \mu_o \mu_r H \\ &= \mu_o \mu_r H + \mu_o H - \mu_o H \\ &= \mu_o H + \mu_o H (\mu_r - 1) \qquad (33) \\ &= \mu_o H + \mu_o M \\ &= \mu_o (H + M) \qquad (34) \end{aligned}$$

where $M = \mu_o (H + M)$ is called the magnetization of solid and expressed in ampere/metre. From the above relation, we find that if a magnetic field is applied to a material, the magnetic flux density is equal to the effect on vacuum and on the material. The magnetization (M) may thus be defined as the magnetic dipole moment per unit volume of the bar.

(iv) Magnetic Susceptibility (χ): The magnitude of the magnetization, M is proportional to the applied field as follows:

$$M = \chi H \tag{35}$$

where χ is called the magnetic susceptibility, which is *unitless*. The magnetic susceptibility and the relative permeability are related as follows:

$$\chi = \mu_r - 1 = M/H \tag{36}$$

We may note that B, M and H are vectors. Magnetic units and conversion factors for SI and CGS – emu systems are given in Table 14.8. We may note that magnetic units may be a source of confusion because there are really two systems in common use. The ones used thus far are SI [rationalized MKS (metre-kilogram-second)]; the others come from the CGS-emu (centimetre-gram-second-electromagnetic unit) system.

Table 14.8 Magnetic units and conversion factors for the SI, CGS and emu systems

Quantity	*Symbol*	*SI units*		*CGS-emu unit*	*Conversion*
		Derived	*Primary*		
• Magnetic induction (flux density)	B	tesla (Wb/m^2)	kg/s–C	gauss	$1\ \frac{\text{Wb}}{\text{m}^3} = 10^4$ gauss
• Magnetic Field Strength	H	$\frac{\text{amp}-\text{turn}}{\text{m}}$	C/m–s	oersted	$\frac{1\,\text{amp}-\text{turn}}{n}$ $= 4\pi \times 10^{-3}$ oersted
• Magnetization	M (SI) I (CGS – emu)	$\frac{\text{amp}-\text{turn}}{\text{m}}$	C/m–s	maxwell/cm^2	$\frac{1\,\text{amp}-\text{turn}}{\text{m}} = 10^{-3}$ maxwell/cm^2
• Permeability of a vacuum	μ_o	henry/m	kg–m/C^2	unitless	$4\pi \times 10^{-7}$ henry/m = 1emu
• Relative permeability	μ_r (SI) μ'(CGS-emu)	unitless	unitless	unitless	$\mu_r = \mu'$
• Susceptibility	χ (SI) χ' (CGS-emu)	unitless	unitless	unitless	$\chi = 4\pi\chi'$

Note: units of Weber (Wb) are volt-seconds, units of heary are Webers per ampere

(v) Magnetic Dipoles: Magnetism is *dipolar*, i.e. magnetism is characterized by having two opposite poles: north (N) and south (S). Magnetic dipoles are found to exist in magnetic materials, which, in some respects are analogous to electric dipoles. The strength of a magnetic dipole is measured by the product of the pole strength and the distance between the poles. This is called *magnetic moment.* Magnetic dipoles are influenced by magnetic fields and within a magnetic field, the force of the field itself exerts a torque that tends to orient the dipoles with the magnetic field. One source of magnetism in an atom is the orbital motion of electrons. Each electron revolving around the nucleus in a atom constitutes a circulating electric charge or current and thus produces a small magnetic field. Moreover, each spinning electron on its axis also can be conceived as a circulating charge and also produced a small magnetic fields. Many times it is convenient to think of magnetic forces in terms of fields. Imaginary lines of force may be drawn to indicate the direction of the force at positions in the vicinity of the field source.

11.2 Origins of Magnetic Moments

The macroscopic magnetic properties of a substance are a consequence of *magnetic moments* associated with individual electrons. Each electron in an atom has magnetic moments that originate from the following two sources:

(i) orbital magnetic moment of electrons
(ii) spin magnetic moment of electrons.

We may note that permanent magnetic moments can also arise from *spin magnetic moment of the nucleus*. Of the three, spin dipole moments of electrons are important in most magnetic materials. Magnetic moments associated with *(a)* an orbiting electron and *(b)* a spinning electron is shown in Fig. 14.17. Electron in an atom is continuously orbiting around the nucleus; being a moving charge, an electron may be considered to be a small current loop, generating a very small magnetic field, and having a magnetic moment along its axis of rotation (Fig. 14.17(a)). Moreover, each electron may also be thought of as spinning around an axis; the other magnetic moment originates from this electron spin which is directed along the spin axis as shown in Fig. 14.17(b). We may note that spin magnetic moments may be only in an "up" direction or in an antiparallel "down" direction. Obviously, each electron in an atom may be thought of as being a small magnet having permanent orbital and spin magnetic moments. The net magnetic moment due to electron spin in a sodium atom is one unit, called a *Bohr magneton* μ_B, which is of magnitude 9.27×10^{-24} A–m^2. For each electron in an atom the spin magnetic moment is $\pm\ \mu_B$ (+sign for spin up and–sign for spin down). The orbital magnetic moment contribution is equal to $m_l\mu_B$, m_l being the magnetic quantum number of the electron.

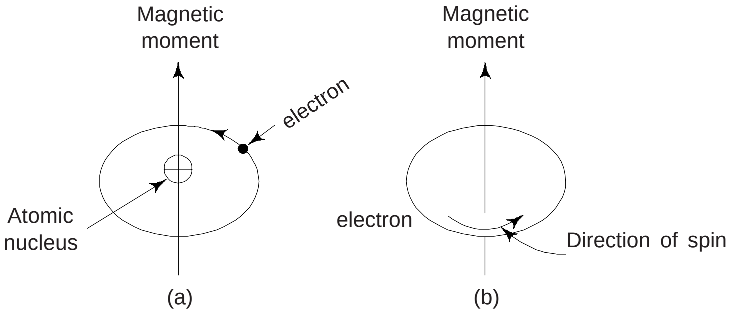


Fig. 14.17 Magnetic moment associated with (a) an orbiting electron and (b) a spining electron

Magnetic properties of a material originate due to the imbalance of spin orientation in atoms. A number of atoms or molecules have paired and unpaired electrons. We may note that there is one electron in the outermost *s*-shell which is unpaired and align itself in an applied field, which gives rise to magnetism. This reveals that the magnetic moment of an atom in the solid state is due only to an incomplete inner shell. It is interesting to note that the electrons in the incomplete inner shells of transition elements have their spins aligned in the same direction, or one may say that the electrons arrange themselves among the energy levels to give the maximum possible total spin angular momentum. This is quite consistent with Pauli exclusion principle. Thus the net magnetic moment, then for an atom is just the sum of the magnetic moments of each of the constituent electrons, including both orbital and spin contributions, and taking into account moment cancellation. For an atom having completely filled electron shells or subshells, when all electrons are considered, there is total cancellation of both orbital and spin moments. This is why the materials composed of atoms having completely filled electron shells are not capable of being permanently magnetized. This type of materials include the inert gases. (He, Ne, Ar, etc.) as well as some ionic materials.

One can obtain the atomic moments of 3rd transition elements in the solid state by adding the spin magnetic moments of the electrons in the unfilled shell.

11.3 Classification of Magnetic Materials

There are three classes into which all the magnetic materials may be grouped according to their magnetic behaviour, although there is some overlap among groups:

1. Diamagnetic substances (Fig. 14.18a)

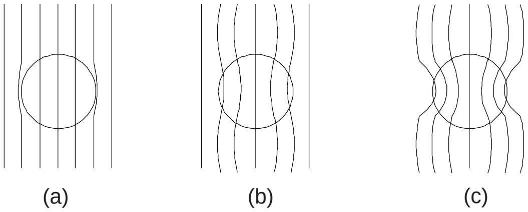


Fig. 14.18 **(a) Diamagnetic solid: Lines of forces are slightly repulsed. Relative permeability is less than 1 (b) Paramagnetic solid: Lines of forces are attracted weakely. Relative permeability is slightly greater than 1. (c) Ferromagnetic solid: Lines of forces are attracted very strongly. Relative permeability is much greater than 1**

2. Paramagnetic substances (Fig. 14.18b)
3. Ferromagnetic substances (Fig. 14.18c)

In addition, *antiferromagnetism* and *ferrimagnetism* are considered to be the subclasses of ferromagnetism.

(i) Diamagnetism: Diamagnetism is a very weak form of magnetism exhibited by substances with a negative magnetic susceptibility ($\chi = \mu_o M/B$), i.e., by substances which magnetize in a direction opposite to that of an applied magnetic field. A diamagnetic substance has a magnetic permeability less than 1, and is repelled when placed near a magnet. The examples are organic solids like naphthalene, benzene, etc.; metals like silver, gold and copper; atoms with rare gas configurations like, A, He, Ne, etc. The magnetization of diamagnetic substances is associated with the currents induced on application of a magnetic field. According to Lenz's law, the flow of an induced current is in such a direction as to oppose the change of flux of inducing field; this accounts for the negative susceptibility. The diamagnetic susceptibility is invariably small, of the order of 10^{-5} cm^3/mole. When placed between the poles of a strong electromagnet, diamagnetic materials are attracted toward regions where the field is weak.

The atomic magnetic dipole configuration for a diamagnetic material with and without an external field is shown in Fig. 14.19. The arrows in the figure represent atomic dipole moments. Figure 14.20 shows the dependence of B on the external field H for a material which exhibits diamagnetic behaviour.

Susceptibilities of several diamagnetic materials are given in Table 14.9.

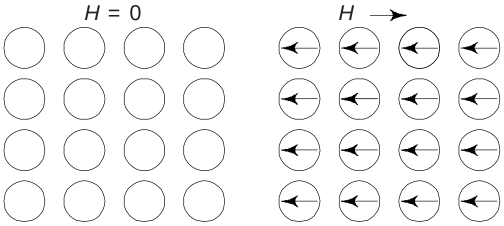


Fig. 14. 19 **Schematic illustration of the atomic dipole configuratin for a diamagnetic material with and without a magnetic field. No dipoles exists in the absence of an external field, whereas in the in the presence of a field, dipoles are induced that are aligned opposite to the field direction**

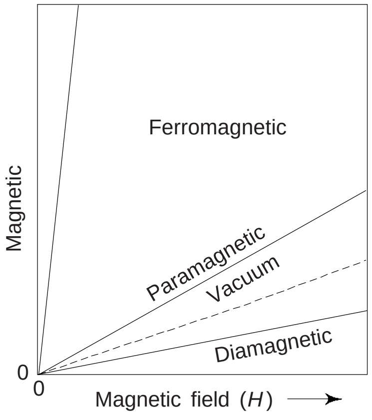


Fig. 14.20

Table 14.9 Magnetic susceptibilities for diamagnetic materials at room temperature

Material	*Susceptibility, χ (volume) (SI units)*
Copper	-0.96×10^{-5}
Gold	-3.44×10^{-5}
Mercury	-2.85×10^{-5}
Silicon	-0.41×10^{-5}
Silver	-2.38×10^{-5}
Zinc	-1.56×10^{-5}
Aluminium oxide	-1.81×10^{-5}
Sodium Chloride	-1.41×10^{-5}

Diamagnetism is found in all materials; but because it is so weak, it can be observed only when other types of magnetism, i.e. para or ferromagnetism are totally absent. The condition for pure diamagnetism is that all electronic spins be paired and all orbital moments either be zero or effectively cancel one another. Nearly all molecules with an even number of electrons satisfy this condition; an important exception is O_2. The condition is also satisfied by most non-metallic solids, except compounds containing atoms with incomplete innershell electron groups, such as the transition rare-earth, and actinide elements.

The diamagnetic response of a substance is small; only a very small fraction of the applied magnetic field is shielded from the interior of the substance by the induced diamagnetic currents. There is one case, however, in which the inducing field is completely shielded (except for small surface effects). This is the perfect diamagnetism exhibited by superconductors and is known as Meissner effect (see chapter 16). We may note that diamagnetism is of no practical importance.

(ii) Paramagnetism: A property exhibited by substances which, when placed in a magnetic field, are magnetized parallel to the field to an extent proportional to the field (except at very low temperatures or in extremely large magnetic fields). Paramagnetic materials always have permeabilities greater than 1, but the values are in general not nearly so great as those of ferromagnetic materials.

For some solid materials, each atom possesses a permanent dipole moment by virtue of incomplete cancellation of electron spin and/or orbital magnetic moments. In the absence of an external magnetic field, the orientations of these atomic magnetic moments are random, such that a piece of material possesses no net macroscopic magnetization. These atomic dipoles are free to rotate, and *paramagnetism* results when they preferentially align, by rotation, with an external field as illustrated in Fig. 14.21. These magnetic dipoles are acted on individually with no mutual interaction between adjacent dipoles. In as much as the dipole align with the external field, they enhance it, giving rise to a relative permeability μ_r that is greater than 1, and to a relatively small and positive susceptibility. Susceptibilities for few paramagnetic substances are given in Table 14.10. Susceptibilities for paramagnetic substances range from about 10^{-5} to 10^{-2}. A schematic *B-H* curve for a paramagnetic substance is also shown in Fig. 14.20.

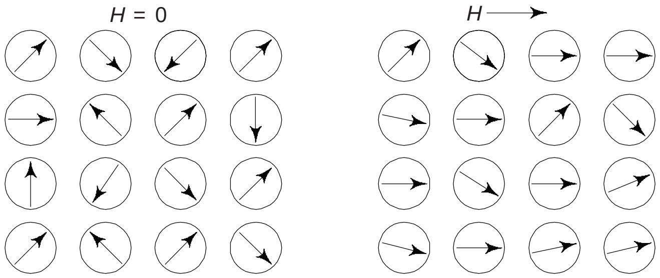


Fig. 14.21 Atomic dipole configuration with and without an external magnetic field for a paramagnetic material

Table 14.10 Magnetic susceptibilities for paramagnetic materials at room temperature

Materials	*Susceptibility χ (volume) (SI units)*
Aluminium	2.07×10^{-5}
Chromium	3.13×10^{-4}
Chromium chloride	1.51×10^{-3}
Manganese sulphate	3.70×10^{-3}
Molybdenum	1.19×10^{-4}
Sodium	8.48×10^{-6}
Titanium	1.81×10^{-4}
Zirconium	1.09×10^{-4}

The following types of substances are paramagnetic:

(a) All atoms and molecules which have an odd number of electrons. According to quantum mechanics, such a system cannot have a total spin equal to zero; therefore, each atom or molecule has a net magnetic moment which arises from the electron spin angular momentum. Examples are organic free radicals and gaseous nitric oxide.

(b) All free atoms and ions with unfilled inner electron shells and many of these ions when in solids or in solution. Examples are transition, rare-earth and actinide elements and many of their salts. This includes ferromagnetic and antiferromagnetic materials above their transition temperatures.

(c) Several miscellaneous compounds including molecular oxygen and organic biradicals.

(d) Metals, in this case, the paramagnetism arises from the magnetic moments associated with the spins of the conduction electrons and is called the *Pauli paramagnetism.*

Relatively few substances are paramagnetic. Aside from the Pauli paramagnetism found in metals, the most important paramagnetic effects are found in the compounds of the transition and rare-earth elements which have practically filled 3*d* and 4*f* electron shells respectively.

Most paramagnetic substances at room temperature have a static susceptibility which follows a Langevin-Deby law (Eq. (37)),

$$\chi = \frac{\mathrm{Np}^2\,\mu_B^2}{3kT + N\alpha} \tag{37}$$

where N is the number of magnetic dipoles per unit volume, p is the effective magneton dipoles per μ_B is the Bohr magneton, k is Boltzmann's constant, T is absolute temperature and α is the temperature independent contribution of Van-Vleck paramagnetism.

Both paramagnetic and diamagnetic substances are considered to be non magnetic because they exhibit magnetization only in the presence of an external field. Moreover, for both, the flux density B within them is almost the same as it would be in vacuum.

(iii) Ferromagnetism: A property exhibited by certain metals, alloys and compounds of the transition (iron groups as BCC α ferrite and cobalt and nickel), rare earth metals such as gaddolinium (Gd), and actinide elements in which, below a certain temperature called the *Curie temperature* (Eq. 38),

$$\chi = \frac{C}{T - \theta} \tag{38}$$

We must note that the general behaviour of the susceptibility of ferromagnetic materials above the Curie temperature, T_c, follows the Curie-Weiss law (Eq. (38)). The behaviour is followed in the region well above the ferromagnetic curie temperature T_c. The paramagnetic curie temperature θ is usually slightly greater than the temperature of transition T_c. Comparison of T_c and θ for three ferromagnetic metal is given below comparison of θ and T_c.

Comparison of θ and T_c

Parameter	*Fe*	*Co*	*Ni*
θ(K)	1093	1428	650
T_c(K)	1043	1393	631

In a region just a fraction of degree above the "critical point", or Curie temperature T_c, the susceptibility is found to approximate the following relation,

$$\chi = \frac{C'}{(T - T_c)^{\gamma}}$$

with γ generally very close to 1.33. We may note that the theory is extremely complicated and not entirely satisfactory], the atomic magnetic moments tend to line up in a common direction. All ferromagnetic materials exhibit paramagnetic behaviour above their ferromagnetic Curie point. The Curie temperature makes a transition between order and disorder of the alignment of the atomic magnetic moments. Some materials exhibit a special form of ferromagnetism below the Curie temperature called *ferrimagnetism*. The magnitude of the paramagnetic susceptibility is determined by the Curie constant C (Eq. 38). A typical value of C is 0.2 K/cm^3 for iron. Below the Curie point, the static susceptibility is not usually defined for a ferromagnetic substance, since the ferromagnet may have a finite magnetization in zero applied field. Magnetic susceptibilities as high as 10^6 are possible for ferromagnetic substances. Consequently, $H << M$, and from Eq. (34), we write

$$B \cong \mu_o M \tag{39}$$

Some physical properties of ferromagnetic elements are summarized in Table 14.10.

Table 14.10 Some selected physical properties of ferromagnetic elements

Element	*Electronic configuration*	*Crystal structure*	*Magnetization at 0 K (Amp/m)*	*Ferromagnetic curie temp. T_c (K)*	*Melting temp. (K)*
Fe	$3d^6\ 4s^2$	bcc	1.7×10^6	1043	1810
Co	$3d^7\ 4s^2$	hcp	1.4×10^6	1404	1750
Ni	$3d^8\ 4s^2$	fcc	0.48×10^6	632	1732
Gd	$4f^7\ 5d^1\ 6s^2$	hcp	5.66×10^6	290	1586

Permanent magnetic moments in ferromagnetic materials result from atomic magnetic moments due to electron-spin-uncancelled electron spins as a consequence of the electron structure. There is also an orbital magnetic moment contribution that is small in comparison to the spin moment. Moreover, in a ferromagnetic material, coupling interactions cause net spin magnetic moments of adjacent atoms to align with one another, even in the absence of an external field (Fig. 14.22).

The characteristic property of a ferromagnet is that, below the curie temperature, it can possess a spontaneous magnetization in the absence of an applied magnetic field. Upon application of a weak magnetic field, the magnetization increases rapidly to a high value called the *saturation* magnetization, which is a general function of temperature (Eq. 40),

$$\frac{M}{M_{\text{sat}}} = \frac{T\alpha}{\theta} \tag{40}$$

where M_{sat} is saturation value of magnetization, θ is the paramagnetic Curie temperature and $\alpha = \mu_o \mu_B\ \gamma M/k_B T$,

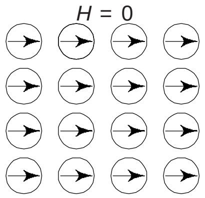


Fig. 14.22 Mutual alignment of atomic dipoles for a ferromagnetic material, which exist even in the absence of external field (H = 0)

where γM is the Weiss field. Obviously, for a given temperature T, Eq. (40) is a plot of M/M_s versus α represent a straight line with a slope equal to T/θ. For $T \geq \theta$, there is no spontaneous magnetization.

The task of a successful theory of ferromagnetism are to account for the spontaneous magnetism below the Curie point, the temperature dependence of the saturation magnetization, and the nature of the magnetization process, or magnetization curve.

Weiss Theory

The Weiss molecular field theory of ferromagnetism represents the first realistic attempt to account for the properties of a ferromagnet. This theory rests on two hypotheses:

(i) Below the Curie point, a ferromagnetic substance is composed of small, spontaneously magnetized regions called *domains*. The total magnetic moment of the material is the vector sum of the magnetic moments of the individual domains. It is now known that these assumed domains really exist and are usually 0.1 and 0.01 cm across.

(ii) Each domain is spontaneously magnetized because a strong molecular (magnetic) field tends to align the individual atomic magnetic moments within the domain.

The consequence of these assumptions is that, while each domain is spontaneously magnetized, the directions of magnetization of the domains do not coincide; therefore the overall magnetization of the sample may be much smaller than if it were composed of a single domain. Application of a relatively weak field (~100-1000 Oersteds) and (often very much less) is sufficient to align the directions of the magnetization of the domains, thereby achieving a large magnetization.

The second hypothesis of the Weiss theory leads to the existence of a Curie temperature below which a domain may be spontaneously magnetized in the absence of an applied magnetized field. According to the theory, the effective field acting on any atomic magnetic moment within the domain may be written as,

$$H = H_o + \lambda M \tag{42}$$

where H_o is an externally applied magnetic field and λM (M = Magnetization) is the Weiss molecular field whose order of magnitude in Fe is 10^7 Oersteds.

There is an overall general agreement between theory and experimental results of temperature dependence of spontaneous magnetization for Fe, Ni and Co. The low temperature behaviour, however, is better described by *magnon theory*. The theory satisfactorily describes the temperature dependence of susceptibility in the paramagnetic region provided the temperature is well above the Curie temperature. The Curie temperature determined from the theory of spontaneous magnetization differs by a few degrees from the experimentally determined value for the paramagnetic region.

Heisenberg Exchange Interaction Theory

The Heisenberg theory of ferromagnetism treats the origin of the Weiss molecular field on an atomic basis. It may be remarked that the ordinary dipole-dipole interactions among atomic magnetic moments are much too small to account for the Weiss field. The foundation of the Heisenberg theory is the Pauli exclusion principle. Due to Heisenberg exchange interaction two neighbouring spins in a solid are coupled together with an exchange energy given by

$$E_{\text{exch}} = -2\,J\ \vec{S}_i \cdot \vec{S}_j \tag{42}$$

where $\vec{S}_i$ and $\vec{S}_j$ are the spin angular momentum vectors of the two electrons i and j, and J is the so called exchange integral between the two electrons. The exchange integral decreases rapidly with distance between the electrons and depends in a complicated way upon the spatial distribution (wave function) of the electrons. It is extremely difficult to compute. The strength of J can be estimated from the Curie temperature T_c. If there are Z nearest neighbours to a central i^{th} spin, the exchange energy for this spin is

$$E_{\text{exch}} = -2\sum_{j=1} J_{ij}\ \vec{S}_i \cdot \vec{S}_j \ \approx -2\ ZJS^2 \tag{43}$$

where we have assumed that all the J_{ii} are equal to J and the spins are parallel. This exchange energy must be equal to kT_c since at T_c the magnetic order is destroyed. Thus

$$J \approx \frac{k_B T_C}{2 Z S^2} \tag{44}$$

If the exchange integral is positive, the parallel arrangement is favoured, and if J is large enough, ferromagnetism should result. If J is large and negative, antiferromagnetism or ferrimagnetism supposedly arises. The order of magnitude of J given by $J \sim kT_c \sim 10^{-20}$ Joule.

There seems to be no question that the Heisenberg theory correctly accounts for the tendency of the electrons in the same ferromagnetic atom to exhibit parallel spins. However, whether or not, it leads to the correct explanation of the interatomic alignment of spins is still a subject of much controversy. In insulators it usually proves possible to express the coupling between atomic spins by Eq. (42), provided J is interpreted as an effective exchange integral. In metals the problem is much more complicated. There is little doubt, however that the basic Heisenberg idea is correct and that the molecular field arises from the interplay between electrical forces and the effects of Pauli exclusion principle.

Crystalline Anistropy Energy

This accounts for the experimental fact that ferromagnets tend to magnetize along certain crystallographic axis, called directions of easy magnetization. For example, a single crystal of Fe, which is made up of a cubic array of iron atoms, tends to magnetize in the direction of the cube edges. It requires about 1.4×10^4 J/m^3 (at room temperature) to move the magnetization into a hard direction along a cubic body diagonal.

The Heisenberg exchange energy, Eq. (42) is isotropic and cannot account for the observed anisotropy, which probably has its origin in a complicated interplay of spin-orbit coupling, crystalline electric fields, and overlap of orbital wave functions. Anisotropy energy depends on the strain of the crystal, giving rise to *magnetostriction*, i.e., changes in length of a substance when it is magnetized.

Ferromagnetic Domains

Small regions of spontaneous magnetization, formed at temperatures below the Curie point, are known as domains. Domains originate in order to lower the magnetic energy as illustrate in Fig. 14.23. In Fig. 14.23(b) it is shown that two domains will reduce the extent of the external magnetic field, since the magnetic lines of force are shortened. On further subdivision (Fig. 23(c)), this filled is still further reduced.

Another way to describe the energy reduction is to note that the interior demagnetizing fields, coming from surface poles, are much smaller in the long, thin domain of Fig. 14.23(c) than in the "fat" domain of Fig. 14.23(a).

The question arises as to how long this subdivision process continues. With each subdivision there is a decrease in field energy, but there is also an increase in Heisenberg exchange energy, since more and more

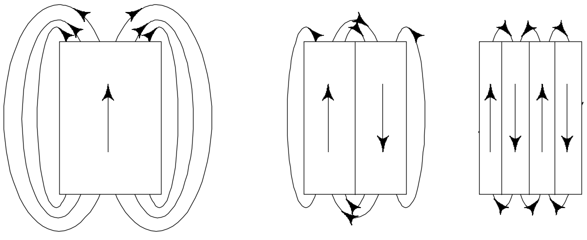


Fig. 14.23 Lowering of magnetic field energy by domains. (a) Lines of force for a single domain. (b) Shortening of lines of force by division into two domains. (c) Reduction of field energy by further subdivision

magnetic moments are aligning antiparallel. Finally a state is reached in which further subdivision would cause a greater increase in exchange energy than it would cause decrease in field energy, and the ferromagnet would assume this state of minimum total energy.

Bloch Wall

Also because of exchange energy, the reversal of magnetization between domains does not occur abruptly but takes place gradually through a transition zone called the Bloch Wall (Fig. 14.24). In iron the wall is ~50 nanometers thick and has the total energy $\sim 2 \times 10^{-3}$ J/m^2.

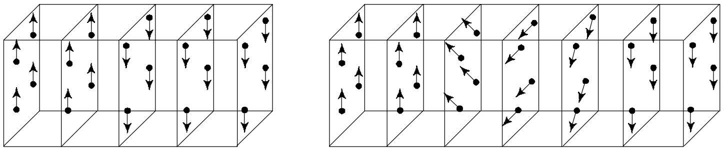

(a) (b)

Fig. 14.24 Lowering of exchange energy by the transition zone known as Bloch wall. The reversal of magnetization between domains does not take place abruptly as is shown but by degrees as is illustrated in Fig. 24(b)

Domain Model

The orientation of domains in a crystal is primarily determined by the need to minimize the magnetic energy (Fig. 14.22). It is possible to eliminate all surface magnetic poles by forming flux closure domains (Fig. 14.25). Here the normal component of magnetization is continuous across all domain boundaries. The demagnetizating fields are zero everywhere, except for a trivial effect of surface poles in the Bloch Walls. In a uniaxial crystal, that is, a crystal with a single easy direction, an arrangement as shown in Fig. 14.25(b) will be preferred since it has a lower density of magnetization normal to easy direction, or in the hard direction. Even in cubic crystals, in which all directions of magnetization in Fig. 14.25 may be easy, Fig. 14.25(b) will be preferred because of magnetostriction. In iron, for example, each domain increases in length along the direction of magnetization by a fraction $\sim 2 \times 10^{-5}$. Thus the domains of Fig. 14.25 can be fitted smoothly together only by straining them elastically against this magnetostriction, and the required strain energy will be smaller in Fig. 14.25(b) than in Fig. 14.25(a).

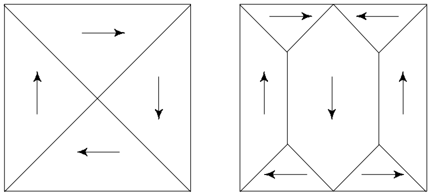
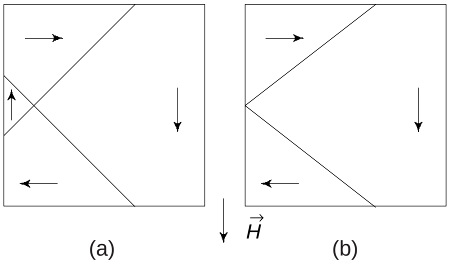

(a) (b)

Fig. 14.25 Flux closure domains (single crystal). (a) Large domains at right angles. (b) Reduction in their size, causing reduction of anisotropy energy of uniaxial crystals or of strain energy of cubic crystals

If an external magnetic field is applied to a ferromagnetic specimen, two things can happen, according to the domain model: *(i)* the domains parallel or nearly parallel to $\vec{H}$ can grow in size at the expense of the antiparallel domains shown in Fig. 14.26, or (b) the magnetic moment of the domains can rotate into the field direction. In either case the sample acquires a magnetization, which increase as the field increase, until all the domains are parallel to the applied field, at which point the material is said to be saturated and the magnetization is equal to $\overrightarrow{M_s}$.



(a) (b)

Fig. 14.26 Growth of domains parallel to an applied magnetic field $\vec{H}$

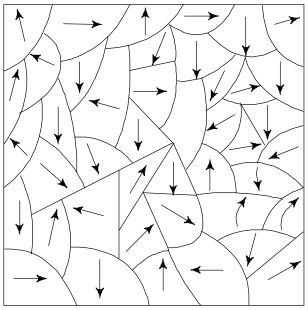


Fig. 14.27 Orientation of domains in a polycrystal

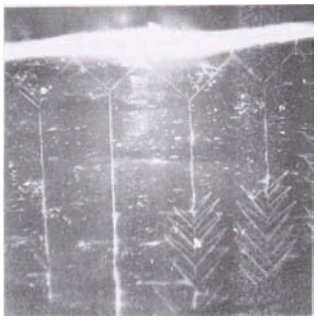


Fig. 14.28 Bitter powder patterns on a (100) surface of silicon-iron

In polycrystals the domain structure is more complicated (Fig. 14.27), depending upon such variables as grain orientation and grain boundaries. It is possible, however, for domains to cross grain boundaries.

On minimizing the total contribution from *(i)* magnetic, or demagnetizing energy *(ii)* anisotropy *(iii)* magnetostriction *(iv)* elastic strain, and *(v)* Bloch wall energy, it is found that, depending upon the composition and shape of the crystal, the theoretical domain thickness should vary from about 0.1 to 0.001 cm.

Direct experimental evidence of the existence of domains is furnished by Bitter powder patterns (Fig. 14.28). It is also possible to observe domains by means of transmission or reflection of electrons in an electron microscope. The Kerr magnetooptic effect (rotation of plane of polarization of light reflected from a magnetic surface) has also been used to study domains. These two methods are particularly useful if the Bloch Walls are extremely thick (because of small magneto crystalline anisotropy) and the resultant Bitter powder patterns very blurred.

(iv) Antiferromagnetism: This originates when the spin moments of the neighbouring atoms are ordered in an antiparallel arrangement (Fig. 14.29) or when the exchange integral is negative. This property is exhibited by some metals, alloys and salts of transition elements in which the atomic magnetic moments, at sufficiently low temperatures, form an ordered array which alternates or spirals so as to give no net total moment in zero applied field. The most direct way of detecting such arrangements is by means of neutron diffraction. A crystal exhibiting antiferromagnetism may be considered to be consisting of two independent sublattices *A* and *B*, one of which is spontaneously magnetized in the opposite direction. This type of magnetism was first observed in the crystals of MnO. Manganese oxide (MnO) is a ceramic material that is ionic in character having both Mn^{2+} and O^{2-} ions. No net magnetic moment is associated with the O^{2-} ions, since there is a total cancellation of both spin and orbital moments. However, the Mn^{2+} ions possess a net magnetic moment that is predominantly of spin origin. These Mn^{2+} ions are arrayed in the crystal structure of MnO crystal such that the moments of adjacent Mn^{2+} ions are antiparallel (Fig. 14.29). This is why that the opposing magnetic moments cancel one another, and as a consequence, the solid as a whole possess no net magnetic moment.

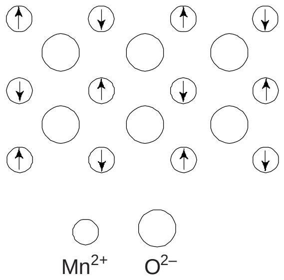


Fig. 14.29 The antiferromagnetism in manganese oxide

When a magnetic field is applied, a small magnetization appears in the direction of the field which increases further with temperature. Such a behaviour is typical of an antiferromagnetic material. The magnetization becomes the maximum at a critical temperature T_N, called the *Neel temperature*, which is analogous to the Curie temperature in the paramagnetic or ferromagnetic substances. This is the transition temperature below which the spontaneous magnetic ordering takes place. Above this temperature, the magnetization, the magnetization decreases continuously which is indicative of the paramagnetic state of the material. The variation of susceptibility with temperature is governed by

$$\chi = C/(T + \theta) \tag{45}$$

where $\theta = T_N$ is called the Neel temperature. Some representative antiferromagnets are given in Table 14.11.

Table 14.11 Some representative antiferromagnets

Substance	*Crystal type*	*T_N (K)*	*Paramagnetic curie temperature θ (K)*
MnF_2	Rutile	67	80
MnO	NaCl	122	610
FeO	NaCl	198	507
$KMnF_3$	Perovskite	88	158
$CuCl_2 \cdot 2H_2O$	Orthorhombic	4.3	4.5
CrSb	NiAs	723	550
Cr_2O_3	Al_2O_3	307	485
$ZnFe_2O_4$	Spinel	9	—
EuTe	NaCl	7.8	6
MnTe	Hexagonal close-packed	403	690

(v) Ferrimagnetism: In some compounds, e.g. in some ceramics, etc., the constituent atoms may be antiferromagnetically coupled but with different magnetic moments. This would give rise to a net magnetic moment in each coupling and the some of the moments of all the coupling could result in magnetization which is comparable in order of magnitude to ferromagnets. This phenomenon is termed as ferrimagnetism. The distinction between the two lies in the source of the net magnetic moments. This type of magnetism occurs in materials such as ferrites which are basically the oxides of various metal elements. These ionic materials may be represented by the chemical formula MFe_2O_4, in which M represents any one of several metallic elements. The prototype ferrite is Fe_3O_4 or $FeO \cdot Fe_2O_3$, the mineral magnetite, sometimes also called as lodestone.

The formula for Fe_3O_4 may be written as $Fe^{2+}\ O^{2-} - (Fe^{3+})_2\ (O^{2-})_3$ in which the Fe ions exist in both +2 and +3 valence states in the ratio of 1 : 2. A net spin magnetic moment exists for each Fe^{2+} and Fe^{3+} ion, which corresponds to $4\mu_B$ and $5\mu_B$ respectively for the two ion types (Fig. 14.30). Moreover, O^{2-} ions are magnetically neutral. We may note that there are antiparallel spin-coupling interactions between the Fe ions, similar in character to antiferromagnetism. However, the net ferrimagnetic moment arises from the incomplete cancellation of spin moments. Net magnetic moments for six cations are given in Table 14.12.

Table 14.12 Net magnetic moments for cations

Cation	*Net magnetic moment(μ_B)*
Fe^{3+}	5
Fe^{2+}	4
Mn^{2+}	5
CO^{2+}	3
Ni^{2+}	2
Cu^{2+}	1

One can also produce cubic ferrites having other compositions by adding metallic ions that substitute for some of the iron in the crystal structure. Again from the ferrite chemical formula, $M^{2+}O^{2-} - (Fe^{3+})_2\ (O^{2-})_3$, in addition to Fe^{2+}, M^{2+} may represent divalent ions such as Ni^{2+}, Mn^{2+}, Co^{2+} and Cu^{2+}, each of which possesses a net spin magnetic moment different from $4\mu_B$ (Table 14.12). Obviously, by adjustment

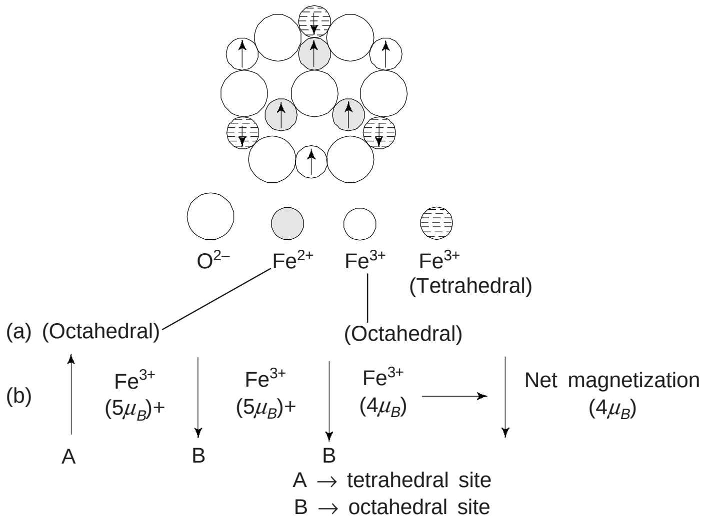


Fig. 14.30 **(a) The spin magnetic moment configuration for Fe^{2+} and Fe^{3+} ions in Fe_3O_4. (b) Magnetizations of the individual sites and the net magnetization in one formula unit of Fe_3O_4**

of composition, ferrite compounds having a range of magnetic properties may be produced, e.g., nickel ferrite has the formula $NiFe_2O_4$. Mixed ferrites may also be produced containing mixtures of two divalent metal ions such as (Mn, Mg) Fe_2O_4 in which Mn^{2+} : Mg^{2+} ratio may be varied.

Other than cubic ferrites, ceramic materials are also ferrimagnetic. These include the hexagonal ferrites and garnets. Hexagonal ferrites have a crystal structure similar to the inverse spinel, with hexagonal symmetry rather than cubic. We may write the chemical formula for these materials as $AB_{12}O_{19}$, in which A is divalent metal such as Ba, Pb or Sr, and B is a trivalent metal such as Al, Fe, chromium and gallium. $PbFe_{12}O_{19}$ and $BaFe_{12}O_{19}$ are two most common examples of hexagonal ferrites.

The most versatile of ferrimagnetic systems are the rare earth iron garnets. The garnet unit cell has three sets of inequivalent magnetic sites, differing in their coordination to neighbouring oxygen ions. Obviously, garnets have very complicated crystal structure, which may be represented by the general formula $M_3Fe_5O_{12}$; where M represents a rare earth ion such as samarium, europium, gadolium or yttrium. Rare-earth ions of various species have widely different magnetic moments, exchange interactions, and crystalline anisotropies. By a suitable choice of rare-earth ions, it is possible to design ferrimagnetic systems with prescribed magnetization and temperature behaviour. Yttrium iron garnet ($Y_3Fe_5O_{12}$), sometimes denoted by YIG, is the most common garnet material.

Ferromagnetic materials which finds wide applications in electrical engineering have a disadvantage, that they have low electrical resistivity. The laminations used for electrical machines have a resistivity ~14 × 10^{-4} Ω-m whereas the highest value obtainable in ferromagnetic alloys is less than 10^{-2} Ω-m. Obviously, this disadvantage of the ferromagnetic materials limit their application in high frequency alternating current applications, high eddy current losses and poor magnetic utilization of metals occur in sheets even at low frequencies. The saturation magnetization for ferrimagnetic materials are not as high as for ferromagnets. Moreover, ferrites being ceramic materials, are good electrical insulators. For some magnetic applications, such as high-frequency transformers, a low electrical conductivity is most desirable, because this prevents eddy currents in cores of coils.

11.4 Magnetostriction

The change of length of a ferromagnetic material when it is magnetized is called magnetostriction. More generally, it is the phenomenon that the state of strain of a ferromagnetic sample depends on the direction and extent of magnetization. The phenomenon has an important application in magnetostriction transducers.

Magnetostriction results from the dependence of the crystalline anisotropy energy upon the state of strain of the crystalline lattice. If the crystal deforms (for e.g., suffers a change in length), the anisotropy energy may be lowered more than the elastic energy is raised. Thus a strained state will be favoured.

The total energy of a ferromagnetic substance depends upon the state of strain and the direction of magnetization through three contributions. The first two consist of the crystalline anisotropy energy of the unstrained lattice plus a correction which takes into account the dependence of the anisotropy energy on the state of the strain. The third contribution is that of the elastic energy, which is independent of magnetization direction and is a minimum in the unstrained state. The state of strain of the crystal will be that which makes the sum of the three contributions to the energy a minimum. The result is that, when magnetized, the lattice is always distorted from the unstrained state, unless there is no anisotropy.

Since spontaneous magnetization occurs below the curie temperature (T_c), there will always be a spontaneous lattice distortion which depends on magnetization direction in the ferromagnetic state. In Ni, the lattice spacing parallel to the magnetization is always smaller than the lattice spacing perpendicular to the magnetization.

The magnetostriction coefficient (λ) is defined as the fractional change (extension or contraction) in length, $\delta l / l$ associated with the change in magnetization from 0 to saturation. The saturation magnetostriction is denoted by λ_s. All domains are oriented in the direction of magnetization, when the material is saturated. In this situation, the expansion in the direction of magnetization is the same at saturation for a single domain or for a number of domains. λ_s is positive in some materials and negative in others. A negative value of λ_s shows that there is contraction in the direction of magnetization and vice versa. We may note that any elastic deformation, whether expansion or contraction, causes an increase in strain energy.

Figure 14.31 shows the magnetostriction versus applied field curves for Ni, Co, Fe and 45 permalloy. From figure it is clear that the dimensions tend to increase in the direction of H and $\perp^r$ to H for 45 permalloy and Fe whereas for Co and Ni the opposite is true. Ni has the highest negative magnitude and permalloy (45 Ni) has one of the highest positive magnetostrictions.

When a high permeability (soft) magnetic material is required, the magnetostriction should be small in order that anisotropy not be induced by lattice distortions. Magnetostriction appears to be a major source of transformer hum. As the silicon content of soft magnetic steel is increased toward 6.5%, the magnetostriction disappears. Unfortunately, the metal becomes the excessively brittle.

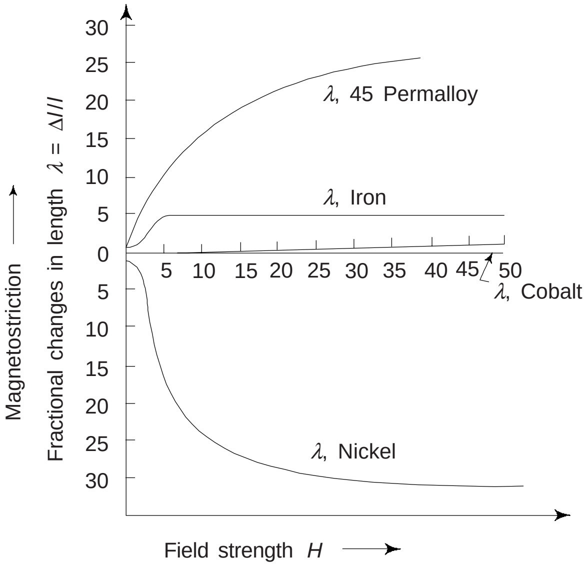


Fig. 14.31 Magnetostriction of some common materials exhibiting expansion and contraction

Magnetostriction is an important property. This effect is exploited in transducers used for the reception and transmission of high frequency sound vibrations. Ni is often used for this application.

Magnetostriction or mechanical deformation of magnetic domains has also other important effects. When a ferromagnetic material is strained, we find that the domains tend to realign themselves into positions of lower energy. Because of this the permeability of the material is changed and it becomes easier or more difficult to magnetize. Realignment puts the domains in a better position to be polarized and permeability is increased in materials having positive magnetorestriction, the opposite is true for materials with negative mechanical deformation.

We may note that due to thermal agitation magnetostriction is decreased, because at higher temperatures, the domains are hindered in their reorientation.

11.5 The Influence of Temperature on Magnetic Behaviour

Magnetic characteristics of materials can also be influenced by temperature. We have read that raising the temperature of a solid results in an increase in the magnitude of the thermal vibrations of atoms. Since the atomic magnetic moments are free to rotate and hence with rise in temperature, the increased thermal motion of the atoms tends to randomize the directions of any moments that may be aligned.

For ferromagnetic, antiferromagnetic, and ferrimagnetic materials, the atomic thermal motions counteract the coupling forces between the adjacent atomic dipole moments, thereby causing some dipole misalignment, regardless of whether an external magnetic field is present. This results in a decrease in the saturation magnetization for both ferro-and ferrimagnets. The saturation magnetization is a maximum at 0 K, at which temperature the thermal vibrations are a minimum. With rise in temperature, the saturation magnetization (M_s) diminishes gradually and then abruptly drops to zero at what is called the *Curie temperature* (T_c). Figure 14.32 shows the magnetization-temperature curves for Fe and Fe_3O_4. The mutual spin coupling forces are completely destroyed at T_c, such that above T_c both ferro-and ferromagnetic substances are paramagnetic. The magnitude of Curie temperature varies from material to material, e.g. for Fe, Co, Ni and Fe_3O_4 the respective values of T_c are 768°C, 1120°C, 335°C and 585°C.

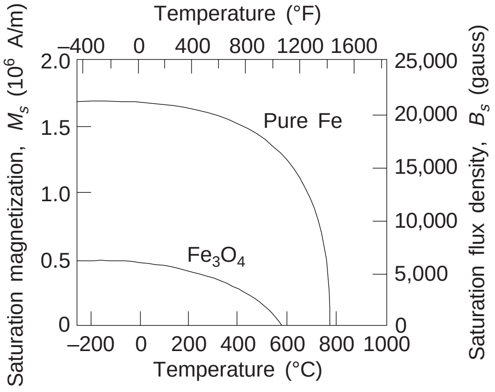


Fig. 14.32 Saturation magnetization (M_s) versus temperature curves for Fe and Fe_3O_4

We may note that antiferromagnetism also affected by temperature. At Neel temperature (T_N), this behaviour vanishes and above T_N, antiferromagnetic materials also become paramagnetic.

11.6 Domains and Hysteresis

At a temperature below T_c, any ferro-or ferrimagnetic material is composed of small-volume regions in which there is a mutual alignment in the same direction of all magnetic moments (Fig. 14.33). Such a region is called domain, and each one domain is magnetized to its saturation magnetization. Adjacent domains are separated by *domain boundaries or walls*, across which the direction of magnetization gradually changes as shown in Fig. 14.34. As mentioned earlier, domains are normally microscopic in size, and for a polycrystalline specimen, each grain may consist of more than a single domain. Obviously, in a macro spiece of specimen, there will be large number of domains, and all may have different magnetization orientations. Thus for the entire solid, the magnitude of the M field is the vector sum of the magnetizations of all the domains, each domain contribution being weighted by its volume fraction. We may note that for an unmagnetized specimen, the appropriately weighted vector sum of the magnetizations of all the domains is zero.

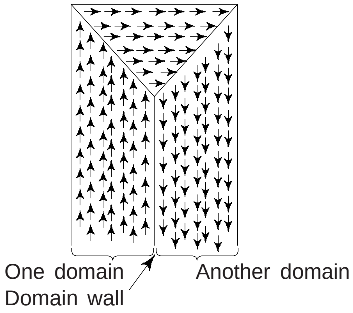


Fig. 14.33 **Domains in a ferro-or ferrimagnetic material. The arrows represent atomic magnetic dipoles. Within each domain, all dipoles are aligned, whereas the direction of alignment varies from one domain to another domain**

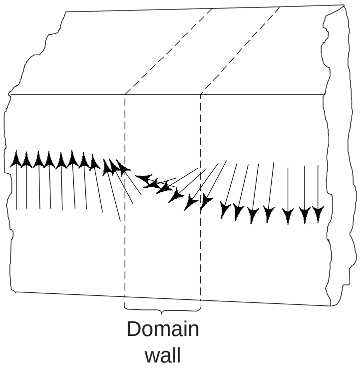


Fig. 14.34 **The gradual change in magnetic dipole orientation across a domain wall or boundaries**

When the ferromagnetic sample that is initially demagnetized is subjected to a continuously increasing magnetizing force *H*, the relation between *H* and flux density *B* is shown by the normal magnetization curve *Oab* (Fig. 14.35). The point *a* indicates the magnetic condition as the increasing magnetic intensity has reached H_1. If *H* is increased to a maximum value H_2 and then decreased again to H_1, the decreasing flux density does not follow the path of increase, but decrease at a rate less than that at which it rose. The lag in the change of *B* behind the change of *H* is called *hysteresis*. If the value of *H* is further reduced from H_1 to zero, *B* is not reduced to zero but to a value B_r. The specimen has retained a permanent magnetism. This ordinate B_r is called the *retentivity* or *remanence*. The value of *B* may be reduced to zero at *e* by reversing the direction of *H* and increasing its value of H_c. This value of H_c is called the *coercive force* or *coercivity*.

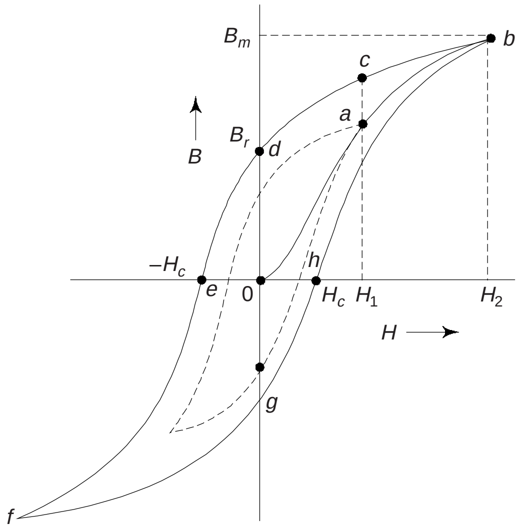


Fig. 14.35 **Magnetic flux density (*B*) versus the magnetic field strength (*H*) for a ferromagnetic material that is subjected to forward and reverse saturations (points *b* and *f*). Hysteresis loop *b d e f g h b*. The remanence B_r and the coercive force H_c are also shown**

As *H* is increased in the negative direction, the magnetization proceeds along the curve of Fig. 14.35 until at *f* the values of *B* and *H* are the same as those at *b*, but opposite in direction. When reverse changes in *H* are made, the magnetization changes along the curve *fghb*. This entire loop *bdefghb* is called a *hysteresis loop*. If the hysteresis loop starts from another point on the normal magnetization curve, such as a, there will be a smaller hysteresis loop entirely within the larger, such as the dotted loop of the Fig. 14.35. These curves sometimes called *B-H* curves and are used to describe magnetic materials. We may note that it is possible to reverse the direction of the magnetic field at any point along the curve and generate other hysteresis loops. To demagnetize the ferromagnetic or ferrimagnetic samples one has to repeatedly cycle it in *H* field that alternates direction and decreases in magnitude. Magnetization curves are generally obtained from long, thin samples to avoid demagnetization. The demagnetization effects are extremely important in ferromagnetic resonance.

We may note that ferromagnetic materials in common use are polycrystalline; i.e., a piece of the material consists of tremendous number of single crystals of random orientation. In each single crystal, there are many domains. The magnetization of the whole body consists of the magnetization of the various single crystals within the body.

In magnetizing the core, energy must be supplied. As H and B increase along *fghb* (Fig. 14.35), the energy gain is proportional to the area under that portion of the curve. Along the path *bcdef*, there is a loss in energy proportional to the area under *bcdef*. The net loss in energy per cycle per unit volume is given by Eq. (46) where

$$W = \oint dW = \oint HdB \tag{46}$$

the integral is taken around the closed loop. But $\oint HdB$ is the area of the hysteresis loop, and the energy loss per unit volume per cycle is equal to the area of the hysteresis loop. This energy is converted into heat. If B is expressed in teslas and H in Amp/m, the energy loss is in J/m^3 (cycle).

Steinmetz found an empirical relation between the energy loss per unit volume per cycle W and the maximum value B_m of the flux density during the cycle given by

$$W = \eta\ (B_m)^n \tag{47}$$

where η is called the *Steinmetz coefficient*. Steinmetz found a value of about 1.6 for the exponent n for many materials, but it varies from about 1.5 to 2.5 for others.

In alternating-current machinery, masses of iron are in fields that are constantly reversing. Therefore, the iron is constantly being carried around hysteresis paths, and there is an energy loss per cycle that depends upon the hysteresis loop for the particular iron that is used. This hysteresis loss results in undesirable heating of the iron as well as waste of energy. Obviously, the hysteresis curves are quite important in determining the quality of a magnetic material and selecting the material for a particular application. In accordance with their H_c and W values, ferromagnetic materials are classified as *soft* and *hard*.

11.6 Soft Magnetic Materials

Soft magnetic materials are characterized by their low loss and high permeability. There are a variety of alloys used with various combinations of magnetic properties, mechanical properties, and cost (Table 14.13). There are seven major groups of commercially important materials: iron and low carbon steels, iron-silicon alloys, iron-aluminium-silicon alloys, nickel-iron alloys, iron-cobalt alloys, ferrites and amorphous alloys.

Table 14.13 Some properties of selected soft magnetic materials

Material	*Composition (wt%)*	*Initial relative permeability* μ_i	*Saturation flux density* B_8 *[tesla (gauss)]*	*Hysteresis loss/cycle [J/m^3 (erg/cm^3)]*	*Resistivity* ρ *(Ω-m)*
Commercial iron ingot	99.95Fe	150	2.14 (21,400)	270 (2700)	1.0×10^{-7}
Silicon-iron (oriented)	97Fe, 3Si	1400	2.01 (20,100)	40 (400)	4.7×10^{-7}
45 Permalloy	55Fe, 45Ni	2500	1.60 (16,000)	120 (1200)	4.5×10^{-7}
Supermalloy	79Ni, 15Fe, 5Mo, 0.5Mn	75,000	0.80 (8000)	—	6.0×10^{-7}
Ferroxcube A	48MnFe$_2$O$_4$ 52ZnFe$_2$O$_4$	1400	0.33 (3300)	~40 (~400)	2000
Ferroxcube B	36NiFe$_2$O$_4$ 64ZnFe$_2$O$_4$	650	0.36 (3600)	~35 (~350)	10^7

The behaviour of soft materials is controlled by the pinning of domain walls at heterogeneities such as grain boundaries and inclusions. In addition, eddy-current loss is minimized through alloying additions which increase the electrical resistivity. Initial permeability, important in electronic transformers and inductors, is improved by minimizing all sources of magnetic anisotropy. A high maximum permeability, necessary for motors and power transformers, is increased by the alignment of the anisotropy, for e.g., through development of crystal texture or magnetically induced anisotropy. For this reason the relative area within the hysteresis loop must be small; it is characteristically thin and narrow (Fig. 14.36).

The saturation field or magnetization is determined only by the composition of the material. For e.g., in cubic ferrites, substitution of a divalent metal ion such as Ni^{2+} for Fe^{2+} in $FeO – Fe_2O_3$ will change the saturation megnetization. However, susceptibility and coercivity (H_c), which also influence the shape of the hysteresis curve, are sensitive to structural variables rather than to composition. We may note that a low value of coercivity corresponds to the easy movement of domain walls as the magnetic field changes magnitude and/or direction. Structural defects, e.g. particles of a non magnetic phase or voids in the magnetic material tend to restrict the motion of domain walls, and thus increase the coercivity (H_c). Obviously, a soft magnetic material must be free from such structural defects.

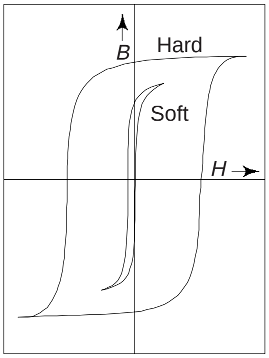


Fig. 14.36 Magnetization curves for soft and hard magnetic materials

Electrical resistivity is the another property which need to be considered. In addition to hysteresis energy losses, energy losses may result from electrical currents that are induced in a magnetic material by a magnetic field that varies in magnitude and direction with time; these are called *eddy currents*. These energy losses in soft magnetic materials can be minimized by increasing the electrical resistivity. This is achieved in ferromagnetic materials by forming solid solution alloys; Fe-Si and Fe-Ni alloys, etc. The ceramic ferrites are commonly used for applications requiring soft magnetic materials because they are intrinsically insulators. However, their applicability is somewhat limited, in as much as they have relatively small susceptibilities.

The class of alloys used in largest volume is by far Fe and 1-3.5% Si-Fe for applications in motors and large transformers. In these applications the cost of the material is often the dominant factor, with losses and excitation power secondary but still important. A major improvement in these alloys occurred around 1940 with a development of a {110} <001> crystal texture in 3.2% Si-Fe which greatly reduced losses and increased permeability. Since then the losses have decreased steadily, roughly logarithmatically, through alloying, texture improvements, decreasing the thickness of strip and application of stressed insulating coatings.

Many special alloys find their use in special devices designed to exploit unusual properties. These include materials with high saturation induction B_s (the magnetic induction at very large values of H), usually Co-Fe alloys; and alloys with high μ, most often Ni-Fe alloys. These alloys are the mainstay of the telecommunication industry. Some of the devices require high initial permeability (μ at very low fields) and other may depend on high values of $\mu_{\max}$ (the maximum permeability).

In power equipment the major consideration is the power lost in the magnetic circuit under operating conditions. The total core loss P_c consists of the hysteresis loss P_h and eddy current loss P_e. Thus

$$P_c = P_h + P_e = \text{const} \times f + \text{const} \times \frac{B^2 f^2}{\rho} \tag{48}$$

where f is the frequency and ρ is the electrical resistivity of the material.

The hysteresis loss may be kept low by using a material with a narrow hysteresis loop (one with low coercive force H_c, the value of H required to reduced B to zero) and minimizing mechanical strain after the final stress-relief anneal. Eddy-current loss may be reduced by breaking up eddy-current paths, for e.g. by using laminations rather than solid cores.

Audio-frequency devices require very thin laminations or resin-bonded alloy powder cores. The very high electrical resistivity ($1 - 10^6$ Ω-cm) qualifies many ferrites for very high frequency applications, for e.g., in cores and transformers.

It is interesting to note that the hysteresis characteristics of soft magnetic materials may be enhanced for some applications by an appropriate heat treatment in the presence of a magnetic field. Using such a technique, one may produce a square hysteresis loop, which is desirable in some magnetic amplifier and pulse transformer applications. Moreover soft magnetic materials are used in switching circuits.

11.7 Hard Magnetic Materials

Hard magnetic materials are utilized in permanent magnets, which have a high resistance so that they strongly resist demagnetization once magnetized. In view of hysteresis behaviour, a *hard magnetic material* or permanent magnet has a high remanence, coercivity, and saturation flux density, as well as a low initial permeability, and high hysteresis energy losses. Figure 14.36 show a comparison of the hysteresis characteristics of hard and soft magnetic materials.

The important characteristics relative to applications for hard magnetic materials are the coercivities H_c and what is termed the "energy product" $(BH)_{\max}$. The energy product $(BH)_{\max}$ corresponds to the area of the largest B-H rectangle that one can construct within the second quadrant of the hysteresis curve (Fig. 14.37). The units of $(BH)_{\max}$ are kJ/m^3 (MGOe: 1 MGOe = 10^6 gauss-oersted = 7.96 kJ/m^3). We may note that the value of $(BH)_{\max}$ is representative of the energy required to demagnetize a permanent magnet, i.e., the larger $(BH)_{\max}$ the harder is the magnetic material in terms of its magnetic characteristics. Although the overall quality of a permanent magnet depends on $(BH)_{\max}$, but the design considerations, high H_c, high residual induction B_T (the magnetic induction when H is reduced to zero) and reversibility of permeability μ may also be controlling factors.

We may note that hysteresis behaviour of a material is related to the ease with which the magnetic domain boundaries move; by impending domain wall motion, the coercivity and susceptibility are enhanced, such that a large external field is required for demagnetization. Moreover, these characteristics are interrelated to the microstructure of the material. To understand the relation between the resistance to demagnetization, i.e., the coercivity H_c, and the metallurgical microstructure, it is essential to understand the mechanisms of magnetization reversal. The two major mechanisms are reversal against a *shape anisotropy* and reversal through nucleation and growth of reverse magnetic domains against *crystal anisotropy*.

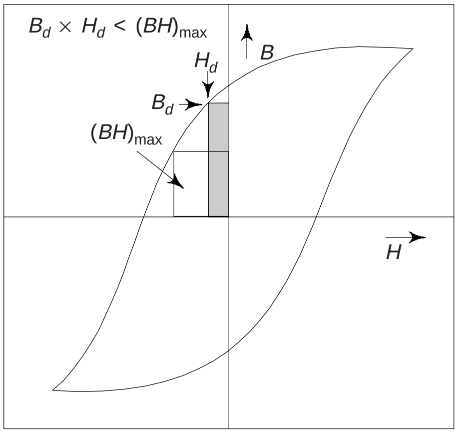


Fig. 14.37 **Magnetization curve displaying hysteresis. Within the second quadrant, two *B-H* energy product rectangles are drawn, the area of rectangle labelled $(BH)_{\max}$ is the largest possible, which is greater than the area defined by $B_d - H_d$**

Conventional Hard Magnetic Materials

Hard magnetic materials can be studied under two main categories: *Conventional* and *high energy hard materials*. The conventional hard magnetic materials have $(BH)_{\max}$ values that range between about 2 and 80 kJ/m^3 (0.25 and 10 MGOe). Ferromagnetic materials—magnet steels, cunife (Cu-Ni-Fe) alloys, alnico (Al-Ni-Co) alloys as well as hexagonal ferrites ($BaO\text{-}6Fe_2O_3$) are conventional hard magnetic materials. Some of the critical properties of few of these hard magnetic materials are presented in Table 14.14.

Table 14.14 Representative properties of few hard magnetic materials

Material	*Composition (wt%)*	*Remanence B_r [tesla (gauss)]*	*Coercivity H_c [amp-turn/m (Oe)]*	*$(BH)_{max}$ [kJ/m^3 (MGOe)]*	*Curie temperature T_c [°C (°F)]*	*Resistivity ρ (Ω-m)*
Tungsten steel	92.8 Fe, 6 W, 0.5 Cr, 0.7 C	0.95 (9500)	5900 (74)	2.6 (0.33)	760 (1400)	3.0×10^{-7}
Cunife	20 Fe, 20 Ni, 60 Cu	0.54 (5400)	44,000 (550)	12 (1.5)	410 (770)	1.8×10^{-7}
Sintered alnico 8	34 Fe, 7 Al, 15 Ni, 35 Co, 4 Cu, 5 Ti	0.76 (7600)	125,000 (1550)	36 (4.5)	860 (1580)	—
Sintered ferrite 3	$BaO\text{-}6Fe_2O_3$	0.32 (3200)	240,000 (3000)	20 (2.5)	450 (840)	$\sim 10^4$
Cobalt rare earth 1	$SmCo_5$	0.92 (9200)	720,000 (9,000)	170 (21)	725 (1340)	5.0×10^{-7}
Sintered neodymium iron-boron	$Nd_2Fe_{14}B$	1.16 (11,600)	848,000 (10,600)	255 (32)	310 (590)	1.6×10^{-6}

The hard magnet steels are generally alloyed with tungsten and/or chromium. These two elements readily combine with carbon in the steel to form tungsten and chromium carbide precipitate particles, which are especially effective in obstructing domain wall motion under the proper heat-treating conditions. An appropriate heat treatment for the other metal alloys forms extremely small single domain and strongly Fe-Co particles within a non magnetic matrix phase.

High-Energy Hard Magnetic Materials

Recently, some intermetallic compounds having a variety of compositions have been developed. These permanent magnetic materials have $(BH)_{\max}$ in excess of about 80 kJ/m^3. The two that have found commercial exploitation are $SmCo_5$ and $Nd_2Fe_{14}B$ (Table 14.14).

Samarium-Cobalt Permanent Magnets ($SmCo_5$)

$SmCo_5$ is a member of a group of alloys that are combinations of Co or Fe and a light rare-earth element; a number of these alloys exhibit high-energy, hard magnetic behaviour, but $SmCo_5$ is the only one material that is of commercial significance. $(BH)_{\max}$ for $SmCo_5$ materials [between 120 and 240 kJ/m^3 (15 and 30 MGOe)] are quite high in comparison to conventional hard magnetic materials (Table 14.14); in addition, they have relatively large H_c (coercivities). One uses powder metallurgical techniques to fabricate $SmCo_5$ magnets. First, the appropriately alloyed material is ground into a fine powder and then the powder particles are aligned using an external magnetic field and thereafter pressed into the desired shape. The piece is then sintered at an elevated temperatures, followed by another heat treatment which improves the magnetic properties of the material.

Neodymium-Iron-Boron Permanent Magnets ($Nd_2Fe_{14}B$)

Sm is a rare and relatively expensive material. Moreover, the price of Co is variable and its sources are unreliable. This is why, the $Nd_2Fe_{14}B$ alloys have become the materials of choice for a large number and wide diversity of applications requiring hard magnetic materials. We may note that H_c and $(BH)_{\max}$ of these materials rival those of $SmCo_5$ alloys (Table 14.14).

The magnetization and demagnetization behaviour of these hard magnetic materials is a function of domain wall mobility, which, in turn, is controlled by the final microstructure-i.e., the size, shape and the orientation of the crystallites and grains, as well as the nature and distribution of any second-phase particles that are present. We may note that microstructure will depend on how the material is processed. There are two different processing techniques available for the fabrication of $Nd_2Fe_{14}B$ magnets: *powder metallurgy* (sintering) and *rapid solidification* (melt spinning). The powder metallurgical method is similar to that described for the $SmCo_5$ materials. In case of rapid solidification techniques, the alloy, in molten form, is quenched very rapidly such that either an amorphous or very fine grained and thin ribbon is produced. This ribbon material then pulverized, compacted into the shape as per requirement, and subsequently heat treated. Rapid solidification is the more involved of the two fabrication processes; nevertheless, it is continuous, whereas powder metallurgy is batch process, which has its inherent disadvantages.

Hard materials are used in motors, loudspeakers, meters, holding devices and in a host of different devices in a variety of technological fields. Permanent magnets are far superior to electromagnets in that their magnetic fields are continuously maintained and without the necessity of having to expand electrical power. Moreover, no heat is generated during operation. This is why permanent magnets are preferred in motors. Motors with permanent magnets are much smaller than their electromagnet counterparts and are utilized extensively in fractional horsepower units. Few familiar motor applications include: in cordless drills and screw drivers; in automobiles (starting, window winter, wiper, washer and fan motors); in clocks and in audio and video recorders. There are other common devices, e.g. speakers in audio systems, light-weight earphones, hearing aids, and computer peripherals etc. that employ these hard magnetic materials.

11.8 Magnetic Storage

Recently, magnetic materials have become increasingly important in the field of *information storage*; in fact, magnetic recording has become virtually the universal technology for the storage of electronic information. The audio tapes, VCRs, computer hard disks, floppy disks, credit cards, etc. are few examples of magnetic storage. In computers, semiconductor elements serve as primary memory and magnetic disks are used for secondary memory because they are capable of larger quantities of information and also at a lower cost. The recording and television industries rely heavily on magnetic tapes for the storage and reproduction of audio and video sequences.

Computer bytes, sound, or visual images in the form of electrical signals are recorded on very small segments of the magnetic storage device-a tape or disk. Transference to and retrieval from the magnetic tape or disk is accomplished by means of an inductive read-write head, which consists basically of a wire coil wound around a magnetic material core into which a gap is cut. Using electrical signal, data are introduced (or "written") within the coil, which generates a magnetic field across the gap. This magnetic field in turn magnetizes a very small area of the disk or tape within the proximity of the head. Upon removal of the magnetic field, the magnetization remains; i.e. the signal has been stored. Figure 14.38 shows the essential features of this recording process.

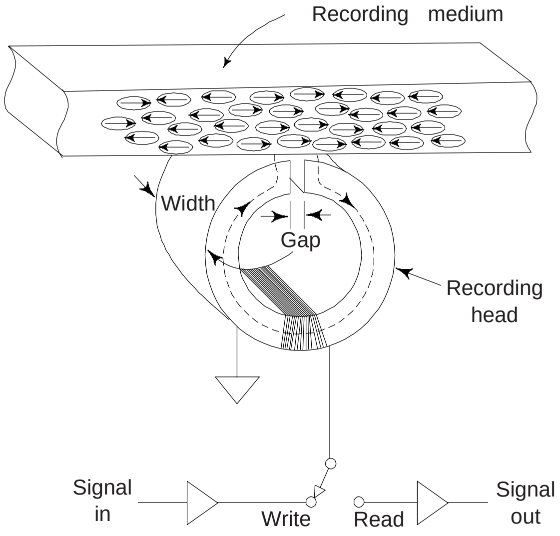


Fig. 14.38 Using magnetic storage medium, how information is stored and retrieved

We may note that the same head may be utilized to retrieve (or "read") the stored information. A voltage is induced when there is a change in the magnetic field as the tape or disk passes by the head coil gap. One can amplify this and then converted back into its original form or character (Fig. 14.38).

More recently, hybrid heads that consist of an inductive-write and a magnetoresistive read head in a single unit have been introduced. In the magnetoresistive head, the electrical resistance of the magnetoresistive thin film element is changed as a result of magnetic field changes when the tape or disk passes by the read head. We may note that higher sensitivities and higher data transfer rates make magnetoresistive very attractive.

Magnetic media are of two principal types: *particulate* and *thin film*. First media consist of very small needle-like or acicular particles, normally of $\gamma-$ Fe_2O_3 ferrite or CrO_2; these are applied and bonded to a polymeric film (for magnetic tapes) or to a metal or polymer disk. During manufacture, these particles are aligned with their long axis in a direction that parallels the direction of motion past the head. Each particle is a single domain that may be magnetized only with its magnetic moment lying along this axis. Corresponding to the saturation magnetization in one axial direction, and its opposite, two magnetic states are possible. These two states make possible the storage of information in digital form, as 1's and 0's. In one system, a 1 is represented by a reversal in the magnetic field direction from one small area of storage medium to another as the numerous acicular particles of each such region pass by the head. A lack of reversal between adjacent region is indicated by a 0.

The *thin-film storage technology* provides higher storage capacities at lower costs. It is used mainly on rigid disk drives and consists of multilayered structure. A magnetic thin film layer is the actual storage component. This thin film is normally either a CoPtCr or CoCrTa alloy, with a thickness of between 10 and 50 nm (100 and 500 Å). A substrate layer below and upon which the thin film resides is pure Cr or Cr alloy. The thin film is itself is polycrystalline, having an average grain that is typically between 10 and 30 nm (100 and 300 Å). Within the thin film each grain is a single magnetic domain and it is highly desirable that grain shape and size be relatively uniform. For magnetic storage disk where these thin films are employed, the crystallographic direction of easy magnetization for each grain is aligned in the direction of disk motion (or the direction opposite). We may note that the mechanism of magnetic storage within each of these single domain grains is the same for the needle shaped particles, i.e. the two magnetic states correspond to domain magnetization in one direction or its antiparallel equivalent.

We may note that the storage density of thin films is greater than for particulate media because the packing efficiency of thin-film domains is greater than for the acicular particles; particles will always be separated with void space in between. The known storage densities for particulate media are of the order of 1×10^8 bit/in^2 (1.5×10^5 bit/mm^2). For thin films, storage densities are approximately an order of magnitude greater, i.e. $\sim 5 \times 10^{10}$ bit/in^2 (8×10^7 bit/mm^2).

The hysteresis loops for these magnetic storage media should be relatively large and square. Obviously, these characteristic ensure that storage will be permanent, and in addition, magnetization reversal will result over a narrow range of applied field strengths. For particular recording media, saturation flux density normally ranges from 0.4 to 0.6 tesla. For thin films, B_s lie between 0.6 and 1-2 tesla. H_c values are typically in the range of 1.5×10^5 to 2.5×10^5 A/m, i.e. 2000 to 3000 °C.

Example 6 *The saturation magnetization of BCC Iron is 1750 kA/m. Show that the net magnetic moment per iron atom in the crystal is 2.2. Given lattice parameter of BCC iron = 2.87Å* [B.E., M.Sc.(M.S.)]

Solution Volume of unit cell of BCC iron = $(2.87)^3 \times 10^{-30}$ m^3

Number of atoms in the unit cell = 2

$$\therefore \quad \text{Net magnetic moment per atom} = 1750 \times 1000 \times (2.87)^3 \times 10^{-30} \times 1/2$$

$$= 2.068 \times 10^{-23} \text{ A-m}^2$$

$$\text{The magnetic moment (in units of } \mu_B) = \frac{2.068 \times 10^{-23}}{9.273 \times 10^{-24}} = 2.2$$

Example 7 *The density of nickel is 8.90* $\times 10^3$ *kg/m*^{3}. *Avogadro's number* $N_A = 6.023 \times 10^{23}$ *atoms/mol. Atomic weight of Ni is 58.71 gm/mol. Calculate (i) the saturation magnetization (ii) the saturation flux density.* [B.E., Diploma]

Solution

(i) $M_s = 0.60\ \mu_B N$

and N (number of atoms/m^3) $= \dfrac{\rho N_A}{A}$

$$= \frac{8.90 \times 10^6\ \text{gm/m}^3 \times 6.023 \times 10^{23}\ \text{atoms/mol}}{58.71\ \text{gm/mol}}$$

$$= 9.13 \times 10^{28}\ \text{atoms/m}^3$$

$$\therefore \qquad M_s = \left(\frac{0.60\ \text{Bohr magneton}}{\text{atom}}\right)\left(\frac{9.27 \times 10^{-24}\ \text{A-m}^2}{\text{Bohr magneton}}\right)\left(\frac{9.13 \times 10^{28}\ \text{atoms}}{\text{m}^3}\right)$$

$$= 5.1 \times 10^5\ \text{A/m}$$

(ii) Saturation flux density,

$$B_s = \mu_o M_s$$

$$= \left(\frac{4\pi \times 10^{-7}\ H}{m}\right)\left(5.1 \times 10^5\ \frac{A}{m}\right)$$

$$= 0.64\ \text{tesla.}$$

Example 8 *Each cubic unit cell of* Fe_3O_4 *contains 8* Fe^{2+} *and 16* Fe^{3+} *ions and the unit cell edge length is 0.839 nm. Calculate the saturation magnetization of* Fe_3O_4. [B.E., M.Sc. (M.S.)]

Solution

$$M_s = N' \mu_B \qquad \text{(i)}$$

where N' is the number of Bohr magneton per unit cell n_B divided by the unit cell volume, V_c, i.e.

$$N' = \frac{n_B}{V_C} \qquad \text{(ii)}$$

We may note that the net magnetization results from Fe^{2+} ions only. Since there are 8 Fe^{2+} ions per unit cell and 4 Bohr magnetons per Fe^{2+} ion, therefore, $n_B = 32$. Since unit cell is a cube, and therefore $V_c = a^3$, a being the unit cell edge length. Therefore,

$$M_s = \frac{n_B \mu_B}{a^3}$$

$$= \frac{\dfrac{(32\ \text{Bohr magneton})}{\text{unit cell}}\ \dfrac{9.27 \times 10^{-24}\ A\text{-}m^2}{\text{Bohr magneton}}}{(0.839 \times 10^{-3}\ m)^3/\text{unit cell}}$$

$$= 5 \times 10^5\ \text{A/m.}$$

Example 9 *In a 440 V, 50 Hz transformer, the total iron loss is 2300 W. When the applied voltage is 220 V at 25 Hz, the total iron loss is 750 W. Show that the eddy current loss at the normal voltage and frequency is 1600 W.* [BE]

Solution We know that the hysteresis loss (P_h) and the induced *emf* loss (P_e) are proportional to the frequency. P_e is proportional to the square of the induced *emf*. We have

$$P_e + P_h = 750\ \text{W} \quad \text{(at 25 Hz)}$$

At 440 V and 50 Hz, we have

$$4P_e + 2P_h = 2300 \text{ W} \quad \text{(at 50 Hz)}$$

or

$$2p_e = 800 \text{ W}$$

∴ The eddy current loss at the normal voltage and frequency = $4P_e$ = 1600 W.

SUGGESTED READINGS

1. S.L. Kakani and C. Hemrajani, *Solid State Physics*, Sultan Chand, New Delhi-2 (4th ed. 2004).
2. L. Solymar and D. Walsh, *Electrical Properties of Materials* Oxford University Press, New York (6th ed., 1998).
3. C. Kittel, *Introduction to Solid State Physics*, Wiley, New York (7th ed., 1995).
4. R.H. Bube, *Electrons in Solids*, Academic Press, San Diego (3rd ed. 1992).
5. D. Jiles, *Introduction to Magnetism and Magnetic Materials*, Nelson Thornes, Cheltenham, UK (1998).
6. R.S. Teeble and D.J. Craik, *Magnetic Materials*, Wiley-Interscience, New York (1969).

REVIEW QUESTIONS

1. Distinguish between a metal, a semiconductor and an insulator on the basis of their energy band structure.

2. *(a)* Explain electrical conductivity on the basis of energy band diagram.
   (b) What is the importance of a energy band diagram and how does it compare the conductor and insulator? [AMIE]

3. What is an insulator? What is the reason that a material behaves an insulator? [AMIE]

4. Describe an expression for the conductivity of conductors from the first principle. [AMIE]

5. Discuss factors that affect the electrical resistance of materials. [AMIE]

6. Explain what insulators, conductors and semiconductors are? Give examples of each. [AMIE]

7. Define a dielectric. Explain the common properties and uses of dielectrics. [AMIE]

8. Describe with examples the three categories of dielectrics. [AMIE, Diploma]

9. Describe the characteristic properties of ferroelectric materials.

10. Explain the meaning and origin of piezoelectricity. Justify the statement that "all ferroelectric crystals are piezoelectric, but all piezoelectric crystals are not necessarily ferroelectric".

11. How do temperature and impurities affect electrical resistivity of metals? [AMIE]

12. Distinguish between the characteristics of dia, para, ferro, ferri and antiferro magnetism. Give an example of each type of material. Comment on the temperature variation of susceptibility of all types of materials.

13. Classify the magnetic materials. [AMIE]

14. Describe briefly paramagnetism, diamagnetism and ferromagnetism. [Diploma]

15. Explain: *(i)* Diamagnetism *(ii)* paramagnetism *(iii)* ferromagnetism *(iv)* antiferromagnetism and *(v)* ferrimagnetic materials. [AMIE]

16. Compare properties of diamagnetic, paramagnetic and ferroelectric materials. [AMIE]

17. Draw *B-H* curve for ferromagnetic materials and explain it. [MSU, AMIE]

18. Explain the principle of a ferrimagnetism with a diagram. [AMIE]

19. Describe the effects of magnetostriction on the properties of some materials. [BE]

20. How can a soft material be converted into a hard one? Distinguish between the two types, giving examples. What are their main applications? [BHU (Elect.), AMIE]

21. Write short note on the following
   (i) Ferroelectricity *(ii)* Piezoelectricity
   (iii) Electrostriction *(iv)* Dielectrics

(v) Ferrites *(vi)* Soft and hard magnets
(vii) B-H curve for ferromagnetic materials *(viii)* Hard magnetic materials
(ix) Magnetization *(x)* Magnetic domains
(xi) Curie point *(xii)* Ferromagnetism
(xiii) Uses of dielectric materials *(xiv)* Soft ferrites
(xv) Antiferromagnetism *(xvi)* Ferrimagnetism [AMIE]

22. Write whether *True* or *False* and justify your answer:
(a) Presence of phosphorus increases conductivity of copper.
(b) Increase in temperature of a conductor increases resistivity.
(c) Elevated service temperature decreases the dielectric strength of insulating materials.
(d) Conductivity of metals decreases at elevated temperature.
(e) For permanent magnetic materials, the product (*BH*) should be maximum possible.
[MSU, Diploma]

23. Explain, why:
(a) In actual practice, several thin layers of dielectric are used in capacitors instead of a single thick layer?
(b) Oxygen free high conductivity copper is specified for bus bars?
(c) Manufactures usually specify maximum safe operating temperature on electric motors?
(d) In case of dielectric materials, energy required for an electron to cross the gap is large?
(e) The breakdown voltage does not increase in proportion to the increase in thickness of dielectrics?
[Diploma, M.Sc. (M.S.)]

24. Which of the two elements, cobalt and gadolium, has the higher saturation magnetization at *(i)* 0 K, and *(ii)* 300 K? [B.E.]

[Hint:

(i) The thermal energy at 0 K is kT. This tends to randomize the spins and hence thermal energy is zero. Obviously, all the spins remain aligned in both Co and Gd. The net magnetic moment of Gd is 7/atom and that of Co is only 1.7 per atom. This means Gd will have higher saturation magnetization. In fact, the actual values will depend on the number of atoms per unit volume in the two cases.

(ii) Curie temperature of Gd is 289 K. Obviously, Gd at 300 K will be paramagnetic and will have negligible magnetization as compared to Co, which has a much higher curie temperature.]

25. In $NiFe_2O_4$, the ferric ions are antiferromagnetically coupled. The magnetization is due to the Ni ions. When Zn is added to $NiFe_2O_4$, the magnetization increases, even though the zinc ions are not ferromagnetic. Explain? [M.Sc. (M.S.), Diploma]

[**Hint:** Looking to the crystal structure of $NiFe_2O_4$ crystal, we note that it has inverse spinel structure and all the Ni^{2+} and half of the Fe^{3+} ions in the octahedral sites. The remaining other half of the Fe^{3+} are in tetrahedral sites. The ferric ions in the two sites are antiferromagnetically coupled. When Zn is substituted for the ferric ions in the tetrahedral sites, the antiferromagnetic coupling will not exactly cancel because the number of ferric ions in the octahedral sites will be more. Obviously, this would increase the magnetization, as there is an extra contribution from the uncoupled ferric ions of the octahedral sites.]

26. What is the distinction between electronic and ionic conduction? [B.E.]

27. How does the electron structure of an isolated atom differ from that of a solid material? [AMIE]

28. If a material is cooled through its melting temperature at an extremely rapid rate, it will form a non-crystalline solid, i.e., a *metallic glass* (Metallic glasses are formed from reasonably inexpensive metals like Fe, Ni, Co and have a number of superior properties for technical applications compared to those possessed by the crystalline solids, for e.g., the filaments of metal glasses are stronger and stiffer and hence offer potential application as lost cost high performance structural reinforcement elements).

29. *(a)* Explain the two sources of magnetic moments for electrons.
(b) Do all electrons have a net magnetic moment? Why or why not?
(c) Do all atoms have a net magnetic moment? Why or why not? [B.E., Diploma]

30. There is associated with each atom in paramagnetic and ferromagnetic materials a net magnetic moment. Explain why ferromagnetic materials can be permanently magnetized whereas paramagnetic ones cannot. [AMIE]

31. Mention the major similarities and differences between ferromagnetic and ferrimagnetic materials. [BHU(BE)]

32. What is the difference between the *spinel* and *inverse spinel* crystal structures? [B.E.]
[**Hint:** The cubic ferrites have the *spinel* crystal structure. There are eight occupied tetrahedral (*A*) sites and 16 occupied octahedral (*B*) sites in a unit cube. The lattice constant is about 8 Å. A remarkable feature of the spinels is that all exchange integrals J_{AA}, J_{AB} and J_{BB} are negative and favour *antiparallel alignment* of the spins connected by the interaction.]

33. Mention the differences between hard and soft magnetic materials in terms of both hysteresis behaviour and typical applications. [B.E., Diploma]

34. It is possible, by various means, e.g., alteration of microstructure and impurity additions), to control the ease with which domain walls move as the magnetic field is changed for ferro and ferri magnetic materials. Draw and *B-H* hysteresis loop for a ferromagnetic material, and superimpose on this plot the loop alterations that would occur if domain boundary movement were hindered. [M.Sc. (M.S.), B.E.]

PROBLEMS

1. The resistance of a given sample of copper of diameter 0.1 mm and length 0.2 m is 0.439 Ω. Calculate its resistivity and conductivity. [*Ans.* 1.72×10^{-8} Ω – m, 5.81×10^{7}/Ω – m] [B.E.]

2. At room temperature the electrical conductivity and the electron mobility for Cu are $6 \times 10^7/(\Omega\text{-m})$ and 0.0030 $m^2/(V - s)$, respectively. *(a)* Find the number of free electrons per m^3 for Cu at room temperature . *(b)* What is the number of free electrons per Cu atom? Given density 8.9 gm/cm^3.

$$\left[(a)\ n = 1.25 \times 10^{29}/\text{m}^3,\ (b)\ 1.48\ \frac{\text{free electrons}}{\text{atom}} \right]$$ [AMIE, B.E.]

3. Show that the percentage increase in resistivity of *(i)* Cu *(ii)* nichrome on heating from 300 K to 1000 K will be 300% and 7% respectively. [AMIE]

4. Determine the power loss due to hysteresis in transformer core of 0.01 m^3 volume of 50 Hz frequency. The area of the loop is 600 J/m^3. [AMIE]

5. The unit cell of magnetite is cubic with a lattice parameter of 8.37Å. There are 16 ferric ions and 8 ferrous ions in the unit cell. Show that the saturation magnetization of magnetite is 510 kA/m. [B.E.]

6. The electrical resistivity of iron increases by a factor of 6, on alloying with 4% Si. Show that the eddy current losses for this change in composition is 83%. [AMIE]

7. A coil of wire 0.20 m long and hanging 200 turns carries a current of 10 Å. *(a)* Find the magnitude of the magnetic field strength *H* *(b)* If the coil is in vacuum then compute the flux density *B* *(c)* If the susceptibility for titanium is 1.81×10^{-4} (SI units), then calculate the flux density inside a bar of titanium that is positioned within the coil *(d)* Calculate the magnetization *M*. (B.E.]

(a) H = 10,000 A-turns/m, *(b)* $B_0 = 1.257 \times 10^{-2}$ tesla,
(c) $B \cong 1.257 \times 10^{-2}$ tesla *(d)* M = 1.81 A/m

8. The chemical formula for manganese ferrite can be expressed as $(MnFe_2O_4)_8$ because there is eight formula units per unit cell. If this material has a saturation magnetization of 5.6×10^5 A/m and density 5 gm/cm^3, find the number of Bohr magnetons associated with each Mn^{2+} ion.
[4.6 Bohr Magnetons/Mn^{2+} ion] [B.E., M.Sc. (MS)]

9. On alloying with 4% silicon, the electrical resistivity of iron increases by a factor of 6. Show that the decrease in eddy current losses for this change is composition is 83%. [B.E.]

OBJECTIVE QUESTIONS

1. Most widely used conducting materials are
(a) gold and silver *(b)* tungsten and platinum
(c) copper and aluminium *(d)* silicon and germanium (c)

2. The SI units of electrical conductivity are
(a) Ω-m *(b)* Ω^{-1}-m^{-1}
(c) kg^{-1} m^{-3} s^{3}A^2 *(d)* Ω/m (c)

3. Ohm's law relates the electric field E, conductivity σ and current density J as
(a) $J = \sigma E$ *(b)* $J = \sigma E^{1/2}$ *(c)* $J = \sigma E^2$ *(d)* $J = E/\sigma$ (a)

4. If e, μ and n respectively represent the charge, mobility and concentration of electrons respectively, then the electrical conductivity of a metal is
(a) $ne\mu$ *(b)* $ne^2\mu$ *(c)* $n^2e^2\mu$ *(d)* ne/μ (a)

5. The Fermi level for copper is 7 eV. The maximum velocity of free electrons at 0 K is
(a) 0 km/s *(b)* 1110 km/s *(c)* 1570 km/s *(d)* 7000 km/s (c)

6. If ρ_o and $\rho(T)$ are the temperature independent and temperature dependent parts of the resistivity of a metal, then the total resistivity is
(a) $\rho = \rho_o - \rho(T)$ *(b)* $\rho = \rho_o + \rho(T)$
(c) $\rho = \rho_o + (\rho(T))^2$ *(d)* $\rho = \rho_o/\rho(T)$ (b)

7. Which of the following metals has the lowest temperature coefficient of resistance?
(a) Silver *(b)* Aluminium *(c)* Gold *(d)* Copper (c)

8. An insulator is a material whose resistivity is approximately within the range
(a) 10^7 Ω-m to 10^{13} Ω-m *(b)* 10^{-3} Ω-m to 10^6 Ω-m
(c) 10^3 Ω-m to 10^{17} Ω-m *(d)* 10^{-9} Ω-m to 10^9 Ω-m (c)

9. The Fermi level E_F depends on the length L of a linear solid as
(a) E_F is independent of L *(b)* $L^{1/2}$ *(c)* L^2 *(d)* L^3 (a)

10. Volt in SI units is
(a) kgm^{2}s^{-3}A^{-1} *(b)* Ω-A *(c)* Ω-A^{-1} *(d)* Nm^{2}s^{-1}A^{-1} (a)

11. If the Fermi energy of a metal is 1.4 eV, the Fermi temperature of metal is approximately
(a) 1.6×10^2 K *(b)* 16 K *(c)* 1.6×10^3 K *(d)* 1.6×10^4 K (d)

12. The composition of manganin alloy is
(a) 60% Cu, 10% Mn, 30% Ni *(b)* 80% Cu, 10% Mn, 10% Ni
(c) 80% Cu, 20% Mn *(d)* 60% Cu, 15% Mn, 25% Ni (d)

13. In order to increase the mechanical strength of an aluminium conductor, one should go for
(a) doping *(b)* cold working
(c) heat treatment *(d)* steel reinforcement (d)

14. Magnetic susceptibility has the dimensions of
(a) Wb-m *(b)* dimensionless *(c)* Wb/m^2 *(d)* Amp/m (b)

15. 1 Bohr magneton is
(a) 9.27×10^{-13} amp-m^2 *(b)* 9.27×10^{24} amp-m^2
(c) 9.27×10^{-24} amp-m^2 *(d)* 9.27×10^{-6} amp-m^2 (c)

16. The units of magnetic permeability are
(a) Wb-A^{-1}-m^{-1} (or H/m) *(b)* Wb/m^2
(c) A/m *(d)* None of these (a)

17. In SI units, tesla is expressed as
(a) kg/m^{2}s^{2}A *(b)* kg/s^{2}A *(c)* NA^{-1} m^{-3} *(d)* none of the above (b)

18. Magnetic induction B and magnetic field intensity H are related by
(a) $B = \mu_o + \mu_o H$ *(b)* $B = \mu_o \mu_r H$
(c) $B = \mu_o H^2$ *(d)* $B = \mu_o + H$ (b)

19. The magnetization (M) of a solid, magnetic induction B and field strength H are related by
(a) $M = (B/\mu_o) - H$ *(b)* $M = 1 + B/H$
(c) $M = 1 - \frac{B}{H}$ *(d)* $M = B - \mu_o H$ (a)

20. Which one of the following material does not have permanent magnetic dipoles?
(a) ferromagnetic *(b)* antiferromagnetic
(c) paramagnetic *(d)* diamagnetic (d)

21. A suitable material for audio and TV-transformers is
(a) pure iron *(b)* Fe and 4% Si alloy
(c) Iron and 30% Ni alloy *(d)* Ferrite (d)

22. The highest measured energy product in a permanent magnet is about
(a) 200 kJ/m^3 *(b)* 300 kJ/m^3 *(c)* 400 kJ/m^3 *(d)* 600 kJ/m^3 (c)

23. In a ferromagnetic material, susceptibility is
(a) very large and negative *(b)* very large and positive
(c) zero *(d)* negative (b)

24. For Cu the resultant spin in Bohr magneton is
(a) 5 *(b)* 9 *(c)* 3 *(d)* 0 (d)

25. Below the ferromagnetic curie temperature, the ferromagnetic material exhibit *B-H* whose form is
(a) straight line *(b)* exponential curve
(c) *B-H* loop *(d)* none of these (c)

26. The temperature below which certain materials are ferromagnetic and above which they are paramagnetic is called
(a) Neel temperature *(b)* Curie temperature
(c) Weiss temperature *(d)* None of these (b)

27. At curie temperature, the spontaneous magnetization for ferromagnetic material is
(a) zero *(b)* infinite
(c) 10 *(d)* none of these (a)

28. At Neel temperature
(a) susceptibility is maximum *(b)* susceptibility is minimum but not zero
(c) susceptibility is zero *(d)* none of these (a)

29. The garnet crystal used in microprocessor has the formula
(a) $Y_3Fe_5O_{12}$ *(b)* $Y_2Fe_3O_5$ *(c)* YFe_3O_5 *(d)* YFe_5O_{18} (b)

30. For paramagnetic materials, μ_r (relative permeability) at room temperature is
(a) 10 *(b)* 5 *(c)* 1 *(d)* zero (c)

31. The susceptibility of a diamagnetic material is about
(a) -10^{-5} *(b)* 10^{-3} *(c)* 10^5 *(d)* 10^7 (a)

SHORT QUESTION ANSWERS

1. How electric current results for most materials?

Ans. For most materials, an electric current results from the motion of free electrons, which are accelerated in response to an applied electric field.

2. On what factor the number of free electrons in a material depend?

Ans. The number of free electrons in most materials depends on the electron energy band structure of the material. An electron energy band is just a series of electron states that are closely spaced with respect to energy, and one such band may exist for each electron subshell found in the isolate atom. By electron energy band structure is meant the manner in which the outermost bands are arranged relative to one another and then filled with electrons.

3. Can we consider the total energy of the free electrons in a solid where the potential field is uniform as their kinetic energy?

Ans. Yes. This energy varies as the square of the wave number k.

4. What is Fermi energy level?

Ans. This is defined as that level which has a 50% probability of occupation by an electron. At 0 K temperature, the Fermi energy level is the highest filled electron energy level.

5. What is the basic requirement for electrical conduction with respect to Fermi energy level in a solid?

Ans. The availability of empty electron states immediately above the Fermi energy level in a solid is a basic requirement for electrical conduction.

6. Where metals and alloys are chosen?

Ans. Metals are chosen where high conductivity and large temperature coefficient of resistance is desired. Alloys are chosen for heating purposes by I^2R dissipation and for low temperature coefficient of resistance.

7. What are the characteristics of dielectric materials?

Ans. Dielectric materials are electrically insulative, yet susceptible to polarization in the presence of an electric field. This polarization phenomenon accounts for the ability of the dielectrics to increase the charge storing capacity of capacitors, the efficiency of which is expressed in terms of a dielectric constant.

8. How polarization takes place in dielectrics?

Ans. Polarization results from the inducement by or orientation with the electric field of atomic or molecular dipoles; a dipole is said to exist when there is a net spatial separation of positively and negatively charged entities.

9. What are the possible polarization types in a dielectric?

Ans. These include ionic, electronic orientation and space charge. However, it is not necessary that all types be present in a particular dielectric. These processes have widely varying relaxation times.

10. Whether the dielectric constant of a material depends on the frequency of the alternating applied field and the temperature?

Ans. Yes.

11. When the energy loss for an dielectric in an alternating field is maximum?

Ans. When the relaxation time of a polarization process matches the period of the alternating field, the electrical energy loss of a dielectric is a maximum.

12. What will happen to the charged stored in a capacitor, if a dielectric material is inserted between the plates of a capacitor?

Ans. The charge stored in the capacitor increases due to polarization in the dielectric.

13. What are the main causes of electric breakdown of a dielectric?

Ans. The main causes of electric breakdown of a dielectric are impurities, surface flaws such as cracks and pores and absorption of moisture and gases on the surface.

14. What are ferroelectric materials?

Ans. These are the materials which may exhibit polarization spontaneously, i.e., in the absence of any external field and have very large dielectric constants. They are useful as piezoelectric crystals and as capacitors.

15. What is piezoelectricity?

Ans. This is the phenomenon whereby polarization is induced in a material by the imposition of external forces.

16. What is the cause of the macroscopic magnetic properties of a material?

Ans. These are as a consequence of interactions between an external magnetic field and magnetic dipole moments of the constituent atoms. Associated with each individual electron are both orbital and spin magnetic moments. The net magnetic moment for an atom is just the sum of the contributions of each of its electrons, where in there will be spin and orbital moment cancellation of electron pairs.

17. How diamagnetism results?

Ans. Diamagnetism results from changes in electron orbital motion that are induced by an external field. All materials are diamagnetic.

18. What are paramagnetic materials?

Ans. These are the materials having permanent atomic dipoles, which are acted on individually and are aligned in the direction of an external field.

19. Why dia and paramagnetic materials are considered to be non magnetic?

Ans. The magnetizations for these types of materials are relatively small and persist only while an applied field is present.

20. How, a large and permanent magnetization is established in ferromagnetic materials?

Ans. Due to bonding forces, the spins of electrons of neighbouring atoms are permanently aligned in ferromagnetic materials. The common ferromagnetic elements are Fe, Co and Ni.

21. What are antiferromagnetic materials?

Ans. Antiparallel coupling of adjacent cation spin moments is found for some ionic materials. Those in which there is total cancellation of spin moments are termed antiferromagnetic.

22. What are ferrimagnetic materials?

Ans. These materials have an antiparallel arrangement of electron spins, but due to incomplete cancellation, there is a net magnetic moment that is comparable in order of magnitude to that of ferromagnetic materials.

23. How the net magnetization results for cubic ferrites?

Ans. The net magnetization for cubic ferrites results from the divalent ions (e.g., Fe^{2+}) that reside on octahedral lattice sites, the spin moments of which are mutually aligned.

24. How ferro and ferrimagnetic materials behave with temperature?

Ans. With rising temperature, increased thermal vibrations tend to counteract the dipole coupling forces in ferro- and ferrimagnetic materials. Consequently, the saturation magnetization gradually diminishes with temperature upto the curie temperature (T_c), at which point it drops to zero; above T_c, these materials are paramagnetic.

25. Explain domain character of ferro and ferrimagnetic materials.

Ans. Below curie temperature (T_c), a ferro-or ferrimagnetic material is composed of domains-small volume regions where in all net dipole moments are mutually aligned and the magnetization is saturated. The total magnetization of the material is just the appropriately weighted vector sum of the magnetizations of all these domains. When an external magnetic field is applied, domains having magnetization vectors oriented in the direction of the magnetic field grow at the expense of domains that have unfavourable magnetization orientations. At total saturation the entire solid material is a single domain and the magnetization is aligned with the field direction. The change in domain structure with increase or reversal of a magnetic field is accomplished by the motion of domain walls.

26. How will you explain both hysteresis and permanent magnetization in terms of domain walls?

Ans. Hysteresis (the log of B field behind the applied H field) and permanent magnetization (or remanence) result from the resistance to movement of domain walls.

27. What are soft magnetic materials?

Ans. Domain wall movement in these materials is easy during magnetization and demagnetization. Consequently they have small hysteresis loops and low energy losses, i.e. low eddy current losses. We may note that by increasing the electrical resistivity of the magnetic medium eddy current losses may be reduced.

28. What are hard magnetic materials?

Ans. Domain wall motion in these materials is much more difficult, which results in larger hysteresis loops; because greater fields are required to demagnetize these materials, the magnetization is more permanent.

15

Semiconductors

1. INTRODUCTION

Semiconductors constitutes a class of materials whose *electrical conductivity* is not as high as that of metals; nevertheless, they have some unique electrical characteristics that render them especially useful. At room temperature, the electrical conductivity of a metal is about 10^8 siemens/m while that of an insulator like diamond is about 10^{-10} siemens/m. The electrical conductivity of a semiconductor lies in the range of 10^5 to 10^{-4} siemens/m. The electrical properties of semiconductors are extremely sensitive to the presence of even minute concentrations of impurities. Semiconductors have the following additional characteristic properties.

(i) A pure (intrinsic) semiconductor has a negative temperature coefficient of resistance, i.e. the resistance of semiconductors decreases with increase in temperature and vice versa. This behaviour of semiconductors is contrary to the positive temperature coefficient of resistance exhibited by a metal.

(ii) Upon irradiation by light, a semiconductor exhibit a photovoltage or a change of resistance.

(iii) With respect to a metal, semiconductors have high thermoelectric power of signs both positive and negative.

(iv) In general there are two types of semiconductors. Those in which electrons and holes are produced by *thermal activation* are called *intrinsic* or pure semiconductors, e.g. Silicon and Germanium. In other type the current carriers, holes or free electrons, are produced by the addition of very small quantities of elements of Group III or V of the Periodic Table, and are known as *extrinsic semiconductors*. Extrinsic semiconductors may be either *n*-or *p*-type depending on whether electrons or holes, respectively, are the predominant charge carriers. Donor impurities introduce excess electrons, whereas acceptor impurities introduce excess holes. The junction between a *p*-type and an *n*-type semiconductor possesses ractification properties.

(v) The electrical conductivity of semiconducting materials is particularly sensitive to impurity type and content, as well as to temperature.

The band gap of a semiconductor is smaller than that of an insulator. At 0 K, semiconductors have a completely filled valence band, separated from an empty conduction band by a relatively narrow forbidden gap, generally less than 2 eV. The band gap in semiconductors lies in the range 0.2 to 2.5 eV, where as the band gap of an insulator like diamond is 6 V. Due to this smaller band gap, electrons are thermally excited from the valence band to conduction band at ordinary temperatures in a semiconductor. This accounts for a larger electrical conductivity of an intrinsic (pure) semiconductor than an insulator. The two elemental semiconductors are Si and Ge, having band gap energies of approximately 1.1 and 0.7 eV, respectively. Both are found in Group IVA of the periodic table and are covalently bonded. Ge was the key semiconducting material for the majority of the early solid state devices. Subsequently, it has been replaced by Si in many applications. Si has a better thermal stability, is readily available and also has an advanced technology. In addition to Ge and Si, a host of compound semiconducting materials like gallium arsenide

(GaAs), indium phosphide (Inp), indium antimonide (InSb), indium arsenide (InAs), lead sulphide (PbS), Cadmium sulphide (CdS), lead telluride (PbTe), Zinc telluride (ZnTel), Mercury (II) indium telluride ($Hg\,In_2Te_4$), Zinc selenide (ZnSe), cadmium selenide (CdSe), Mercury (II) selenide (HgSe), Magnesium antimonide (Mg_2Sb_2), Magnesium iodide (MgI_2), etc., also exhibit semiconducting properties. These semiconductor materials have also been used for several electronic devices, e.g. GaAs is used in transistors, lasers and microwave and millimetre wavedevices, PbS and PbTe are used in infrared detectors, CdS in light meters, CdTe in the detection of nuclear radiation; and so on.

An important subject of scientific and technological interest is *amorphous semiconductors*. In an amorphous substance the atomic arrangement has some short-range but no long range order. The representative amorphous semiconductors are selenium, germanium and silicon in their amorphous states, and arsenic and germanium chalcogenides, including such ternary systems as Ge-As-Te. Some amorphous semiconductors can be prepared by a suitable quenching procedure from the melt. Amorphous films can be obtained by vapour deposition.

Alloys of semiconducting compounds also find wide range of important applications. For example, gallium arsenide phosphide ($Ga\,As_x\,P_{1-x}$) is used in light-emitting diodes (LED); indium gallium arsenide ($In_{1-x}Ga_xAs$) in microwave and optoelectronic devices; cadmium mercury telluride ($Cd_xHg_{1-x}Te$) in infrared detectors; etc.

The nano-sized powders of silicon, silicon nitride (SiN), silicon carbide (SiC) and their thin films have been considered for applications in opto-electronic devices and quantum optic devices. SiC and SiN are also used as advanced ceramics with controlled microstructures because their strength and toughness increase when the grain size diminishes. Nano-structured materials are discussed in chapter 19.

Today no society can be called modern or developed unless it has sizeable electronics industry. And there can be no electronics industry without the semiconductors and related technologies. Semiconductors form the backbone of electronics. Semiconductors affect all walks of life whether it is communications, computers, biomedical, power, aviation, defence, entertainment etc. The transistors, integrated circuits (ICs), lasers and detectors, sensors and other semiconductor devices through the items of daily use touch our life everyday. A brief account of the structure of semiconductors and some simple and popular semiconductor devices is presented in this chapter.

2. INTRINSIC SEMICONDUCTORS

It the electrical conductivity of a semiconductor is entirely due to motion of charge carriers (electrons and holes) which are generated by thermal excitations from the valence band to the conduction band, i.e. as a result of thermal disruption of covalent bonds, the semiconductor is referred to as *intrinsic* (or *pure*) semiconductors. In intrinsic semiconductors, electrons and holes are always equal in number. A semiconductor is not truly intrinsic unless its impurity content is less than 1 part impurity in 100 million parts of semiconductor.

Ge and Si are two important elemental semiconductors. Each atom in a Ge crystal has four valence electrons. The inner ionic core of the Ge atom has a positive charge of +4 units of electronic charge. The four valence electrons are held by covalent bonds with the valence electrons of four nearest neighbour Ge atoms. A simplified two dimensional representation of a Ge crystal at absolute zero temperature is shown in Fig. 15.1 Since the valence electrons bind one atom to another, they are not available for electrical conduction in the absence of any thermal disruption of chemical bonds at 0 K. Obviously, at 0 K a pure or intrinsic semiconductor behaves like an insulator. In the band picture, we can describe the situation by saying that at 0 K the valence band is completely filled and the conduction band is completely empty.

At room temperature, some of the valence electrons have enough thermal energy to break their valence or covalent bonds. These electrons thus become *free* to move at random throughout the crystal. The energy (E_g) needed to break a covalent bond at room temperature is about 1.1 eV in silicon and 0.72 in Ge. Each electron that leaves the covalent bond creates a *vacant site* or *hole* at its original position. A vacancy in the covalent bond caused by a free electron is represented by a tiny circle (Fig. 15.2(a)). This vacancy or

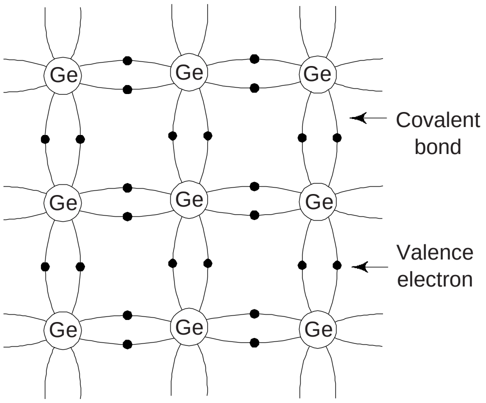


Fig. 15.1 **A simplified representation of the crystalline structure of a semiconductor germanium (Ge) crystal at absolute zero temperature**

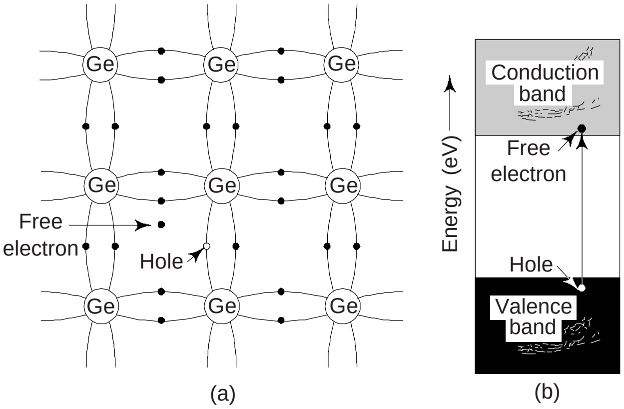


Fig. 15.2 **Creation of an electron-hole pair by the rupture of a covalent bond at room temperature at room temperature (a) Two dimensional lattice of Ge crystal at room temperature (b) energy band picture of creation of electron-hole pair in Ge crystal at room temperature**

deficiency of electron is called a *hole*. In the energy band picture, the electrons are thermally excited from the valence band into the conduction band where they become free (Fig. 15.2(b)). The vacancies thus formed in the valence band are holes. A hole is equivalent to a net positive charge equal to that of electron. Whenever a free electron is generated, a hole is created simultaneously at its original position, i.e. free electrons and holes in a semiconductor are always generated in pairs. Obviously, the concentration of free electrons and holes will always be equal in an intrinsic semiconductor. This type of generation of free electron-hole pairs in semiconductors is termed as *thermal generation*. Since a hole has a positive charge it moves in a direction oppose to that in which an electron moves when an external electric field is applied (Fig. 15.3).

The electrons in the conduction band and hoes in the valence band moves in a random fashion within the crystal due to their thermal energy. When an external field is applied to the semiconductor, a drift velocity is superimposed on the random thermal motion of the charge carriers, i.e., electrons and holes. The drift of the electrons in the conduction band and that of holes in the valence band produce an electric current. The electrons move towards the positive electrode, whereas the holes towards the negative electrode (Fig. 15.3). The currents produced by the movement of electrons and holes in opposite directions and since the electron carries a negative charge and the hole a positive charge. Thus, the conventional current flows within the semiconductor from the positive electrode to the negative electrode. The energy of a hole is measured downward from the top of the valence band.

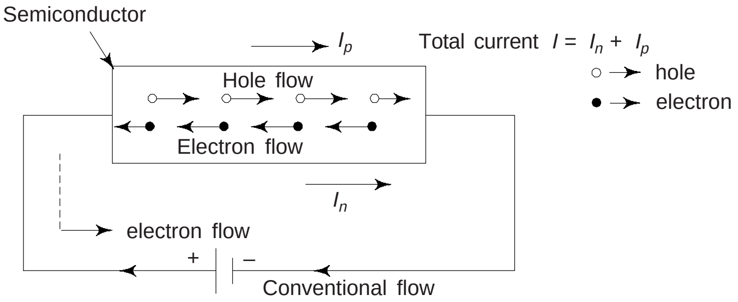


Fig. 15.3 **Conduction in intrinsic semiconductor**

The motion of the electrons in the valence band may be considered to be equivalent to the motion of holes in the opposite direction. Obviously, the holes also contribute to the conductivity. When an electron-hole pair is thermally created, a valence electron in a neighbouring atom can have sufficient thermal energy to jump into the position of the hole and reconstruct the covalent bond. In doing so, the electron leaves a hole in its initial position. Effectively, the hole moves from one position to the other position. Thus the holes move in the direction opposite to that of the valence electrons. That is why a hole behaves like a free positive charge equal in magnitude to the electronic charge. The electrons in the conduction band move more easily than the electrons which cause the motion of holes in the valence band. Hence the contribution to the electric current by the electrons in the conduction band is more than that by the holes in the valence band. The salient features of an intrinsic semiconductor can be summarised as follows:

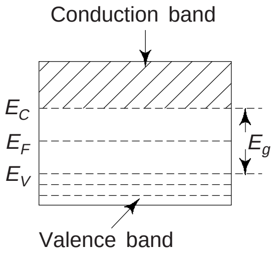


Fig. 15.4 Fermi level for an intrinsic semiconductor

(i) The number of electrons in the condution band is equal to the number of holes in the valence band. In equilibrium, the electron concentration n and hole concentration p are equal, i.e. $n = p = n_i$, where n_i is termed as intrinsic concentration.

(ii) The Fermi level lies in the energy gap exactly between the valence and conduction bands ($E_F = E_g/2$). (Fig. 15.4). The band gaps for some of the elemental and compound semiconductors are given in Table 15.1.

(iii) The contribution of the electrons to the electric current is more than that due to the holes.

(iv) About 1 atom out of 10^3 atoms of an intrinsic semicondutor contributes to the conduction.

(v) An electron and hole can behave as a pair bound to each other. Such a bound pair is usually referred as *exciton.* An exciton is electrically neutral and so does not take part in electrical conduction.

Table 15.1 Band gap energies, electron and hole mobilities and intrinsic electrical conductivities at room temperature for few elemental and compound semiconductors

Material	*Band gap (eV)*	*Electrical conductivity* $[(\Omega\text{-m})^{-1}]$	*Electron mobility* (m^2/V-s)	*Hole mobility* (m^2/V-s)
Si	1.11	4×10^{-2}	0.14	0.05
Ge	0.67	2.2	0.38	0.18
GaP	2.25	—	0.03	0.015
GaAs	1.42	10^{-6}	0.85	0.04
InSb	0.17	$2 \times 10^{+4}$	7.7	0.07
CdS	2.40	—	0.03	—
ZnTe	2.26	—	0.03	0.01

Effective Mass

When an external field is applied to a semiconductor, the charge carriers, i.e., the electrons and the holes, experience forces due to the external applied field and also due to the internal periodic field produced by the crystal. If the external applied is much weaker than the internal field, the effect of the latter is to modify the mass of the carriers in such a way that the carriers respond to the applied field with this modified mass obeying the laws of classical mechanics. This modified mass of the carriers is termed the *effective mass* of the carriers and usually denoted by m^*. m^* is usually different from the electronic mass (m) in vacuum. The effective mass approximation avoids the quantum nature of the problem and allows us to use classical mechanics to study the effect of external fields or forces on the charge carriers, i.e., on the electrons or the holes inside the crystal. Thus the force on an electron of mass m and charge $-e$ in an electric field $\vec{E}$ in terms of effective mass m^* can be written as

$$m^*\vec{a} = -e\vec{E}$$

where $\vec{a}$ is the acceleration.

Recombination of Electrons and Holes

This is the process in which the free electrons in the conduction band jump into the valence band to combine with holes. In this process of recombination the electron-hole (e-h) pair is destroyed. The rate of recombination is approximately proportional to the product of electron concentration and hole concentration. In this process the minimum energy released in the form of electromagentic radiation is equal to the band gap (E_g). We have

$$E_g = hv$$

where h is Planck's constant and v is the frequency of the radiation. The wavelength of the radiation is given by

$$\lambda = \frac{c}{v} = \frac{ch}{E_g}$$

where c ($= 3 \times 10^8$ m/s) is the velocity of light in free space. The generation and recombination of electrons and holes are illustrated in Fig. 15.5.

While some electron-hole pairs are lost due to recombination, new e-h pairs are generated due to thermal excitation. For an intrinsic semiconductor at a constant temperature the rate of combination and the rate of generation of e-h pairs are equal and hence the electron and hole concentrations remain constant at their thermal equilibrium value. With the rise in temperature, the thermal equilibrium value of the electron and the hole concentrations also increases.

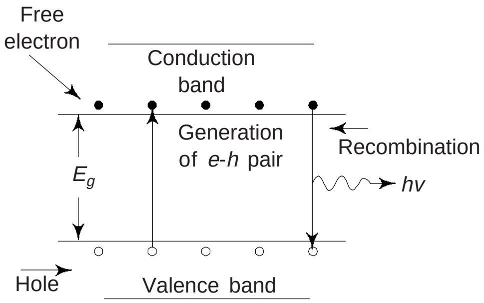


Fig. 15.5 **The generation and recombination of electrons and holes in a semiconductor**

2. (a) Intrinsic Conductivity

Since there are two types of charge carriers (free electrons and holes) in the intrinsic semiconductor, its *specific conductance* is the sum of the conductivities σ_n due to free electrons and σ_p due to holes. Thus the electric conductivity of an intrinsic semiconductor is

$$\sigma_i = \sigma_n + \sigma_p = ne\mu_n + pe\mu_p \quad (1)$$

where μ_e and μ_n are electron and hole mobilities respectively. The mobility μ is defined as the drift velocity per unit electric field. The magnitude of μ_n is always less than μ_e for semiconductors. n and p are concentrations (number of charge carrier per cubic meter) of electrons and holes respectively in a semiconductor. Since for an intrinsic semicondutor

$$n = p = n_i \quad (2)$$

one obtains

$$\sigma_i = n_i e\,(\mu_e + \mu_p) \quad (3)$$

The mobilities depend on temperature as a modest power law, i.e. μ goes as $1/T^{3/2}$

or $$\mu_e = \alpha T^{-3/2} \text{ and } \mu_p = \beta T^{-3/2}$$

Thus $$\mu_e + \mu_p = (\alpha + \beta)\, T^{-3/2} = \gamma\, T^{-3/2}$$

Equation (3) becomes

$$\sigma = \gamma n_i\, T^{-3/2} = \gamma\,(en_i)\, T^{-3/2}$$

Now, $n_i = CT^{3/2} \exp(-E_g/2k_BT)$

with $C = 2\left(\dfrac{2\pi m k_B}{h^2}\right)^{3/2} = \left[\dfrac{2\pi \times 9.1 \times 10^{-31} \times 1.38 \times 10^{-23}}{(6.626)^2 \times 10^{-68}}\right]^{3/2}$

$= 4.83 \times 10^{21}$

Hence $n_i = 4.83 \times 10^{21}\, T^{3/2} \exp(-E_g/2\, k_BT)$

Using this value of n_i, one obtains

$$\sigma_i = \gamma Ce \exp(-E_g/2\, k_BT)$$

or
$$\sigma_i = B \exp(-E_g/2\, k_BT) \tag{4}$$

or
$$\rho_i = \frac{\exp(Eg/2k_BT)}{B} = A \exp(E_g/2\, k_BT) \tag{5}$$

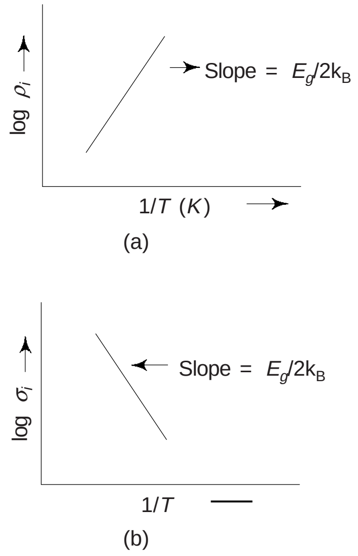
where ρ_i is the *intrinsic resistivity*. Taking logarithms on both sides of Eq. (5),

$$\log \rho_i = \frac{E_g}{2k_BT} + \log A \tag{6}$$

Equation (6) suggests us a method of determining the energy gap (E_g) of an intrinsic semiconductor. Equation (6) explains why the conductivity of intrinsic semiconductor varies exponentially with increase in temperature.

A measurement of intrinsic resistivity of the material at various temperatures helps us to plot the results on a semi logarithmic paper. The slope of the curve is ($E_g/2\, k_BT$) and hence the energy gap can be determined (see Fig. 15.6(a) and (b)).

Figure 15.7 shows the plot of ρ_i vs. 1/T for some intrinsic semiconductors. E_g for Ge and Si turns out to 0.72 eV and 1.2 eV respectively. The ratio of the conductivity of copper to that of Ge is 3×10^7 while the corresponding ratio of conduction electron density is 2×10^9. We must remember that the mobility of electrons in Ge (= 0.4 m^2/V-s) is about 100 times that in copper. The mobility of holes in Ge is ~ 0.2 m^2/(V-s), so that electrons carry about 2/3rd of the current while holes carry the reminder.



(a)

(b)

Fig. 15.6 (a) Log ρ_i vs. 1/T (b) log σ_i vs 1/T

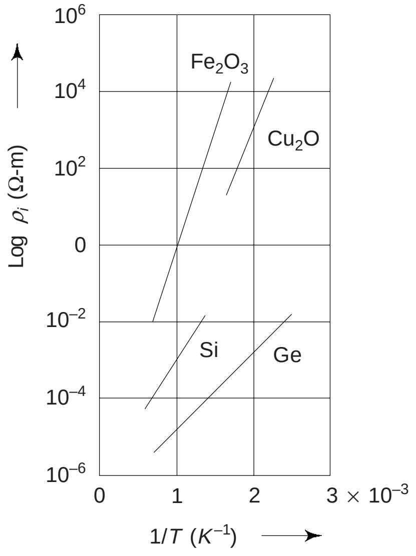


Fig. 15.7 Plot of ρ_i vs. 1/T for some semiconductors in the intrinsic range

Thus the conductivity (or resistivity) of a semiconducting material, in addition to being dependent on electron/and or hole concentrations, is also a function of charge carriers' mobilities, i.e., the ease with which electrons and holes are transported through the crystal. Furthermore, magnitudes of electron and hole mobilities are influenced by the presence of those some *crystalline defects* that are responsible for the scattering of electrons in metals-thermal vibrations (i.e., temperature) and impurity atoms (if any).

3. EXTRINSIC SEMICONDUCTORS

Intrinsic semiconductors are not of much practical use because their conductivity is very low, i.e., only 1 atom in 10^9 contributes to electrical conduction. To achieve an appreciable current density in intrinsic semiconductors, one will have to apply very large electric field, which is not practicable. However, it is found that the presence of impurities even to the extent of 1 in 10^9 alters significantly the conduction of Ge and Si. The impurities that are used have known properties. They can be deliberately introduced in carefully controlled proportions. As a result one obtains a semiconductor of any predetermined conductivity. For instance the conductivity of Si is increased thousand times by the addition of 10 parts per million of boron. The conduction that occurs then is termed as *impurity conduction* and is of paramount importance in the operation of a semiconductor device. This process of adding an impurity to an intrinsic semiconductor is called *doping*. The impurity that is added is called *dopant*. When these impurity or foreign atoms are added into the semiconducting structure, the available quantum states are altered; one or more new energy levels may appear in the band structure of the semiconductor. This introduces significant changes in the properties of semiconductors. The resulting material is called as a *doped*, *impure* or *extrinsic* semiconductor.

Extrinsic semiconductors are of two kinds, one in which the impurity contributes additional electrons in the conduction band and the other in which it contributes additional holes in the valence band. In Ge or Si, addition of group V elements like phosphorous (P), antimony (Sb), and arsenic (As) produces excess free electrons, whereas group III impurity elements like indium (In), boron (B), aluminium (Al), and gallium (Ga) creates excess holes. In a compound semiconductor, e.g. InSb, a group VI element tellurium (Te) generates excess free electrons and group II element zinc (Zn) produces excess holes. We may note that when small amounts (1 part in 10^7 approximately) of tetravalent or pentavalent impurity is added during crystal formation of a semiconductor, the impurity atoms lock into the crystal lattice since they are not greatly different in size from Ge or Si atom, and the crystal is not unduly distorted. Depending upon the impurities, an extrinsic semiconductor can be of two types: *N-type* and *P-type*.

3. (a) N-type Semiconductors

Germanium and Silicon are tetravalent. The impurity atoms may be either pentavalent or trivalent, i.e., from group V and III of the periodic table. If a small quantity of a pentavalent impurity (having 5-electrons in the outermost orbit) like Arsenic (As), Antimony (Sb) or Phosphours (P) is introduced in Germanium, it replaces equal number of Germanium atoms without changing the physical state of the crystal. Each of the four out of five valency electrons of impurity say of Arsenic enters into covalent bonds with Germanium, while the fifth valency electron is set free to move from one atom to the other as shown in Fig. (15.8). The impurity is called *donor* impurity as it donates electron and the crystal is called *N*-type semiconductor. A small amount of Arsenic (impurity) injects billions of free electrons into Germanium thus increasing its conductivity enormously. In *N*-type semiconductors the *majority* carriers of charge are the electrons and holes are *minority* carriers. This is because when donor atoms are added to a semiconductor, the extra free electrons give the semiconductor a greater number of free electrons than it would normally have. And, unlike, the electrons that are freed because of thermal agitation, donor electrons do not

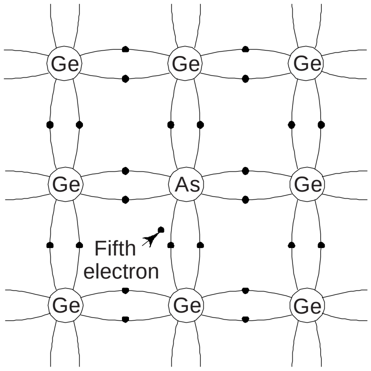


Fig. 15.8 ***N*-type semiconductor**

produce holes. As a result, the current carriers in a semiconductor doped with pentavalent impurities are primarily negative electrons.

The impurity atom has five valence electrons. After donating one electron, it is left with +1 excess charge. It then becomes a positively charged immobile ion. It is immobile because it is held tightly in the crystal by the four covalent bonds as shown in Fig. 15.8.

It is important to understand that in *N*-type semiconductors, although electrons (negative charges) are the majority carriers, but the semiconductor doped with impurity remains electrically neutral. Free electrons and holes are generated in pairs due to thermal energy and negative charge of electrons donated by impurity atoms is exactly balanced by positive charge of the immobile ions. Representation of an *N*-type semiconductor is shown in Fig. 15.9, we have not shown Silicon or Germanium atoms in this figure. One should assume them as a continuous structure over the whole background. The fixed or immobile ions are regularly distributed in the crystal structure. The electrons and the holes, being free to move, are shown randomly distributed at any moment.

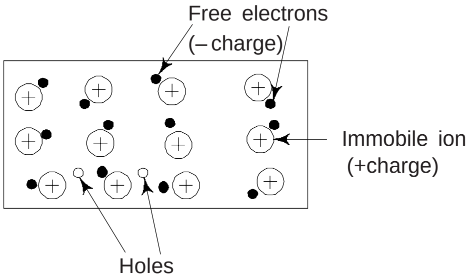


Fig. 15.9 **Representation of an *N*-type semiconductor**

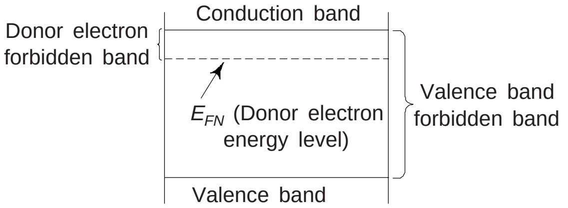


Fig. 15.10 **Excess free electrons in *N*-type semiconductors produce a donor energy level close to the conduction band**

Since *N*-type semiconductors have extra free electrons, and pure semiconductors do not, the energy band diagram for a doped semiconductor is slightly different from that of a pure semiconductor. In effect, another energy level exists; a level for the donor electron, which is closer to the conduction band. The forbidden band for the donor electron is much narrower than the forbidden band for the valence electron; so one can see that it is much easier to cause electron flow in an *N*-type semiconductor (Fig. 15.10).

3. (b) *P*-type Semiconductor

When a trivalent impurity (having 3 electrons in outermost orbit) like Indium (In), Boron (B) or Gallium (Ga) is added in a Germanium intrinsic semiconductor, the impurity atoms will displace some of the Germanium atoms in the crystal during its formation as shown in Fig. 15.11. In this case only three out of the four possible covalent bonds are filled while the fourth bond is vacant and the vacancy acts as a hole. Hence a hole moves relative to the electron in a direction opposite to the direction of electron, when an electron moves from one bond to the other. This trivalent impurity known as the *acceptor* or *P-type* impurity injects into the crystal billions of holes and the majority carriers of the charge are the holes responsible for the conductivity of the crystal. For this reason such crystals are called *P*-type semiconductors or *P*-type crystals. *P*-type semiconductor can be represented as shown in Fig. 15.12.

The energy band diagram of *P*-type semiconductor also differs from that of the pure superconductor. Since there is an extra number of holes, which tend to attract electrons, they aid in starting current flow. As a result, the acceptor energy level is also somewhat higher than that of the valence band. However, it is not as high as the donor level (Fig. 15.13). *P*-type semiconductors will conduct easily than pure semiconductors, but not quite as easy as *N*-type semiconductors.

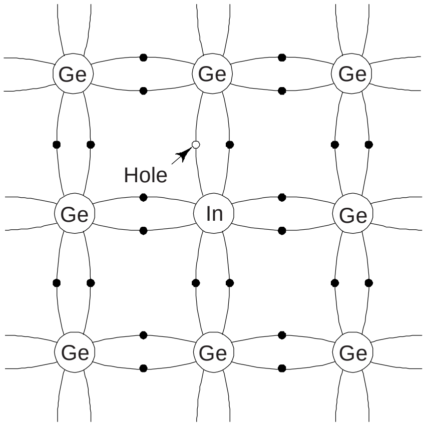


Fig. 15.11 ***P*-type semiconductor**

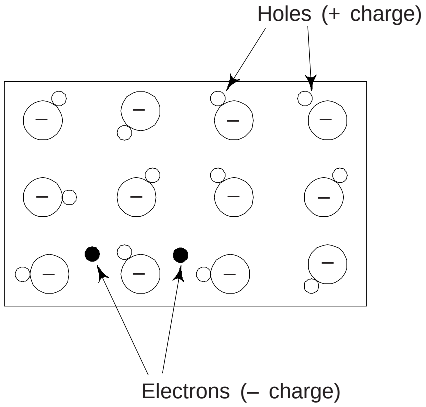


Fig. 15.12 **Representation of a *P*-type semiconductor**

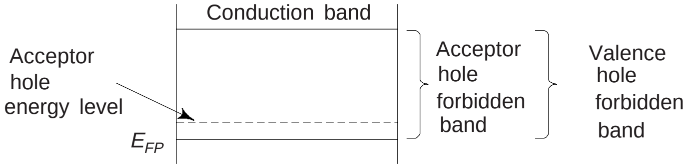


Fig. 15.13 **Excess holes in *P*-type semiconductors introduce an acceptor energy level close to valence level**

3. (c) Effect of Temperature

When the temperature of *N*-type semiconductor is raised, the number of electron-hole pairs due to thermal excitations from the valence band to the conduction band will increase. The number of electrons coming from the donor level will remain constant as the donor atoms are already ionized. Obviously, at very high temperature the concentration of thermally generated free electrons from the valence band will be much larger than the concentration of free electrons contributed by the donors. At this situation the hole and the electron concentrations will be nearly equal and the semiconductor will behave like an intrinsic one. On the basis of the same argument, one can say that a *P*-type semiconductor will also behave as intrinsic semiconductor at very high temperature. Generalizing, we can say that as the temperature of an extrinsic semiconductor increases the semiconductor passes from an extrinsic to an intrinsic one.

3. (d) Effect of Doping on Electrical Conductivity

The electrical conductivity of a semiconductor increases significantly on doping. Let us consider Ge at room temperature (~300 K).

The intrinsic carrier concentration (n_i) for Ge is $7.2 \times 10^{19}/\text{m}^3$. The conductivity of Ge in terms of electron mobility (μ_e) and hole mobility (μ_p) is

$$\begin{aligned}\sigma &= n_i e\,(\mu_e + \mu_p)\\ &= 7.2 \times 10^{19} \times 1.6 \times 10^{-19}\,(0.38 + 0.18)\\ &= 6.68\ \Omega^{-1}\text{m}^{-1}\end{aligned}$$

The values of electron and hole mobilities for Ge are taken from Table 15.1.

On doping Ge with a pentavalent impurity of concentration of 1 ppm, we have

$$N_d = 4.42 \times 10^{28} \times 10^{-6} = 4.42 \times 10^{22}.\ \text{Obviously,}$$

$$N_d \sim 600\ n_i$$

Now, we calculate the hole and electron concentrations in the doped Ge at 300 K, using

$$p_0 = n_i^2/n_0 \quad \text{and} \quad n_0 \approx N_d, \text{ so,}$$

$$n_0 \approx 4.4 \times 10^{22}/\text{m}^3 \quad \text{and} \quad p_0 \approx 1.2 \times 10^{17}/\text{m}^3$$

where N_d is the donor impurity density.

Ignoring the hole contribution to the conductivity of doped Ge, one obtains

$$\sigma = n_0 \, e\mu_n = 4.4 \times 10^{22} \times 1.6 \times 10^{-19} \times 0.39$$

$$= 2760 \ \Omega^{-1}\text{-m}^{-1}$$

This shows that on doping, the electrical conductivity of Ge increases significantly, i.e. by a factor of ~ 400.

3. (e) Charge Carrier Densities in Extrinsic Superconductors

We have seen that the introduction of impurities in pure semiconductors increases the density of one type or another type of change carriers. The product of holes and electrons in a semiconductor is constant depending on the width of energy gap and temperature and hence the introduction of the impurities results in an increase in the density of one type of carrier and a reduction in the density of the other type of carrier. In an extrinsic semiconductor, the carriers introduced by the impurities are called *majority carriers* and the other type are called *minority carriers*. It is important to note that the low value for minority carrier density is due to added recombination.

Let N_d be the donor impurity density, N_a the acceptor impurity density, p the density of holes and n the density of electrons in an extrinsic semiconducting material. From the condition of charge neutrality, we have

$$N_d + p = N_a + n$$

or

$$n = (N_d - N_a) + p$$

$$= (N_d - N_a) + \frac{n_i^2}{n} \quad (\because np = n_i^2)$$

or

$$n^2 - (N_d - N_a)\, n - n_i^2 = 0 \tag{7}$$

Solving the quadratic equation in n, one obtains

$$n = \frac{(N_d - N_a) \pm \sqrt{(N_d - N_a)^2 + 4n_i^2}}{2} \tag{8}$$

or

$$n \approx (N_d - N_a) \quad \text{when} \quad (N_d - N_a) >> n_i$$

Obviously, the electron density (n) in the N-type semiconductor equals the difference in the donor and acceptor impurity densities when they are large compared to the intrinsic density, n_i. Similarly the hole density (p) in a P-type semiconductor is given by

$$p = (N_a - N_d) \quad \text{when} \quad (N_a - N_d) >> n_i$$

3. (f) Donor and Acceptor States

When an impurity atom from group V of the periodic table, say Phosphorous is added to a pure Ge or Si crystal as a pentavalent impurity, these impurity atoms enter the lattice by substitution for normal atoms, and not in interstitial positions. These impurity atoms contribute five electrons per atom to the valence band, i.e. we have an extra electron per impurity atom. These additional electrons (which cannot be accommodated in the valence band of the original lattice) occupy some discrete energy levels just below the conduction band; the separation may be a few tenths of an electron volt. These excess electrons are released by the impurity atoms and excited into the conduction band. The excited electrons then contribute to the electrical conductivity of the semiconductor. Conversely, the impurity may consist of atoms having *fewer* electrons than of a semiconductor [Fig. 15.14(a)]. For the cases in which Si and Ge are the host substances, the impurity atoms could be boron or aluminium, each of which contributes only three electrons. In this

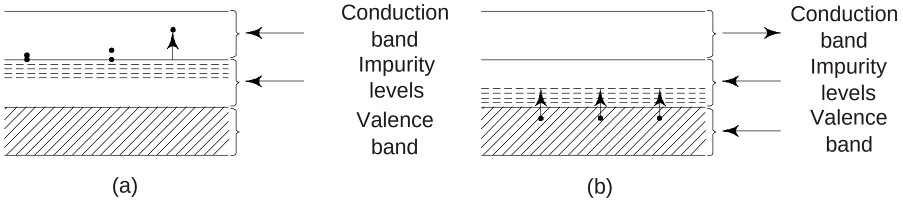


Fig. 15.14 Impurities in a semiconductor (a) Donors, or *N*-type (b) Acceptors, or *P*-type

situation the impurity introduces vacant discrete energy levels, very close to the top of the valence band. Therefore it is easy to excite some of the more energetic electrons in the valence band into the impurity levels. This process produces vacant states, or holes, in the valence band. These holes then act as positive electrons [Fig. 15.14(b)]. We must note that to produce significant changes in the conductivity of a semiconductor, it is sufficient to have about one impurity atom per million semiconductor atoms.

We have already seen that the crystal as a whole remains neutral because the electron remains in the crystal. We have further seen that the band diagram of N-type or P-type semiconductor differs from that of the pure semiconductor. The band diagram for N-type or P-type semiconductor explains clearly why the conduction becomes possible in impure semiconductors at comparatively low temperatures.

3. (g) Fermi Level in Extrinsic Semiconductor

We have read that Fermi Level is situated in the middle of the band gap in an intrinsic semiconductor as the electron and hole densities are equal. When the intrinsic semiconductor is doped, the carrier densities change, consequently the position of the Fermi level also changes. The shift in the position of the Fermi level can easily be related to the majority carrier density in an extrinsic semiconductor if it is assumed that the addition of impurities do not affect the densities of energy states in the conduction and valence bands.

Let N_c and N_v denote the density of states in the conduction band and density of states in the valence band, respectively. We have for an intrinsic semiconductor

and

$$\left.\begin{aligned} n &= N_c \exp\left[\frac{E_{fi} - E_c}{kT}\right] \\ p &= N_v \exp\left[\frac{E_c - E_{fi}}{kT}\right] \end{aligned}\right\} \tag{9}$$

Here E_{fi} is the energy associated with the Fermi level in an intrinsic semiconductor. For an intrinsic semiconductor, we have $n = p$ and therefore from Eq. (9), we have

$$\frac{N_c}{N_v} = \exp\left[\frac{E_c - E_v - 2E_{fi}}{kT}\right]$$

Let E_{fn} be the energy associated with the Fermi level in an N-type semiconductor having an electron density n, we have

$$n = N_c \exp\left[\frac{E_{fn} - E_c}{kT}\right]$$

and

$$p = N_c \exp\left[\frac{E_v - E_{fin}}{kT}\right]$$

$\therefore$

$$\frac{n}{p} = \frac{n^2}{n_i^2} = \frac{N_c}{N_v} \exp\left[\frac{2E_{fn} - E_c - E_v}{kT}\right] = \exp\left[\frac{2(E_{fn} - E_{fi})}{kT}\right]$$

or
$$n = n_i \exp\left[\frac{E_{fn} - E_{fi}}{kT}\right] \tag{10}$$

Similarly, one obtains for a p-type semiconductor

$$p = n_i \exp\left[\frac{E_{fi} - E_{fi}}{kT}\right] \tag{11}$$

Thus the shift in the Fermi level in the n and p type of semiconductor can be expressed as

$$\left.\begin{aligned} E_{fn} - E_{fi} &= kT \ln \frac{n}{n_i} \\ E_{fi} - E_{fp} &= kT \ln \frac{p}{p_i} \end{aligned}\right\} \tag{12}$$

Figure 15.15 represents the shift in the Fermi level in the N-and P-types of semiconductors.

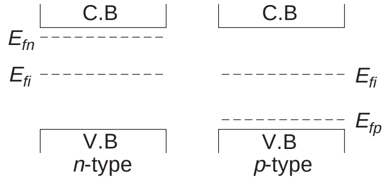


Fig. 15.15 Shift in the Fermi level in the *N*- and *P*-type of semiconductors

3. (h) Thermal Ionization of Extrinsic Semiconductors

When the temperature of an extrinsic semiconductors is raised above 0° K, the impurity atoms get ionized. Due to ionization the donor impurity atoms give rise to electrons in the conduction band and the acceptor impurity atoms give rise to holes in the valence band. These electrons or holes alongwith those generated by intrinsic action, then serve as the current carriers at temperature. Two cases of interest are

(i) Conduction electron concentration is equal to

$$n = n_0 \exp\left[(E_F - E_g)/kT\right] \approx N_d \tag{13}$$

This shows that under the present conditions the concentration of conduction electrons is approximately equal to the concentration of donors. This means that all the donors are ionized. Equation (15.13) suggests that at room temperature the impurity concentration for Si and Ge upto 10^{14} to 10^{16} donors per cc suits this range, i.e., if we dope an intrinsic semiconductor crystal with this concentration of donors, one can certainly predict that one will have ~10^{16} electrons/cc.

(ii) At higher temperature, the carrier concentration is proportional to $\sqrt{N_d}$.

3. (i) Charge Densities in Extrinsic Semiconductors

The density of impurity atoms in N-and P-type materials is so low compared to the density of semiconductor atoms that the rate of thermal pair generation is not affected appreciably by the presence of impurity atoms. In the case of intrinsic semiconductors we have seen that the concentration product

$$np = n_i^2 \tag{14}$$

was a constant at a given temperatures.

Assuming all impurity atoms in extrinsic semiconductors to be ionized at the usual operating temperatures, the free charge densities in impurity material can be based upon N_D, the donor atom density in N-material, or N_A, the acceptor atom density in P-material.

The electrical neutrality of the material demands

$$p = N_D = n + N_A \tag{15}$$

The L.H.S. of equation (15) gives the total positive charge as the sum of holes in the valence bonds and the positive charge associated with the donor atoms that have given up electrons and become positive ions. The R.H.S. of equation (15) sums the negative charge of free electron density n and the negative charge due to electrons held by the ionized acceptor atoms.

Using Eq. (14), one can write

$$n = \frac{n_i^2}{p} \quad \text{and} \quad p = \frac{n_i^2}{n} \tag{16}$$

only donor impurities are introduced in N-type material, so $N_A = 0$. The donor density will be made much larger than the density of intrinsic holes, or

$$N_D >> p$$

and in N-type material the electron density is written from Eq. (15) as

$$n \cong N_D \tag{17}$$

on the similar reasoning for P-type material, we have $N_D = 0$ and $N_A >> n$. One obtains density relations in P-type material as

$$n \cong \frac{n_i^2}{N_A} \tag{18}$$

$$p \cong N_A \tag{19}$$

From the above results we can conclude that the density of majority carrier approximates the impurity atom density at usual ambient temperatures, and the density of minority carriers is reduced below the intrinsic level. This means that the increased electron density in N-type material raises the probability that an electron will meet and recombine with a hole, and so the hole density is decreased to maintain n^2 constant.

4. SEMICONDUCTOR DEVICES

A semiconductor device can be defined as a unit which consists, partly or wholly, of semiconducting materials and which can perform useful functions in electronic apparatus and solid state research. Examples of semiconductor devices are semiconductor diodes (*P-N* junction), transistors, integrated circuits (Ic) etc. Si, Ge and GaAs are most commonly used materials for the fabrication of semiconductor devices. For convenience the properties of these semiconductors are summarized in Table 15.2.

Almost all semiconductor devices are comprised of a single crystal semiconductor incorporating two or more semiconducting regions of different impurity density. The difference in the electric fields and carrier densities associated with differently doped regions, called *junctions*, permit a wide range of essentially non-linear conductivity effects in devices incorporating two, three or more distinct regions. Most semiconductor devices can be understood through the simplest of such junctions, called the *P-N* junction, which is a system of two semiconductors in physical contact, one with excess of electrons (*N*-type) and other with excess of holes (*P*-type).

Table 15.2 Properties of Si and Ge semiconductors

			Value		
Property	*Symbol*	*Unit*	*Germanium (Ge)*	*Silicon (Si)*	*GaAs*
Atomic numbers			32	14	
Atomic weight			72.6	28.08	144.63
Density		kg/m^3	5.32×10^3	2.33×10^3	5.32×10^3
Atom concentration		atoms/m^3	4.4×10^{28}	5×10^{28}	2.21×10^{28}
Relative dielectric constant	ε_r		16	11.8	10.9
Band gap at 0°K	E_{go}	eV	0.785	1.21	1.43
Band gap at 300 K	E_g	eV	0.72	1.1	1.32
Intrinsic Carrier (generation) at 300 K	n_i	$\frac{\text{Carriers}}{\text{m}^3}$	2.5×10^{19}	1.5×10^{16}	9.0×10^{12}

(*Contd.*)

Table 15.2 (*Contd.*)

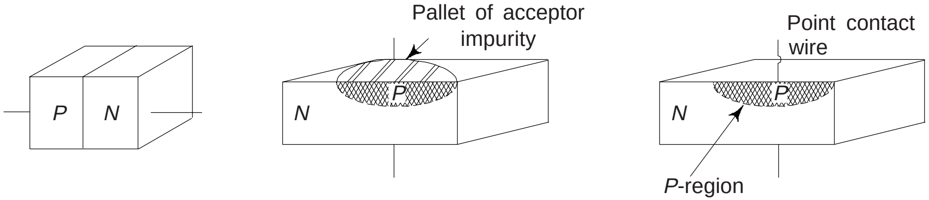
			Value		
Property	*Symbol*	*Unit*	*Germanium (Ge)*	*Silcion (Si)*	*GaAs*
Crystal structure			Diamond	Diamond	Zincblende
Lattice constant	a	Å	5.65748	5.43086	5.6534
Melting point		°C	936	1420	1250
Minority carrier life time		second (s)	$\sim 10^{-3}$	$\sim 2.5 \times 10^{-3}$	$\sim 10^{-8}$
Breakdown Field		V/m	$\sim 10^{7}$	$\sim 3 \times 10^{7}$	$\sim 4 \times 10^{5}$
Diffusion constant	D_n (electrons)	m^2/s	0.009842	0.003367	0.001036
	D_n (holes)		0.004662	0.001295	0.000906
Effective density of states in the conduction band	N_c	m^{-3}	1.04×10^{25}	2.8×10^{25}	4.7×10^{22}
Effective density of states in the valence band	N_v	m^{-3}	6.4×10^{24}	1.02×10^{25}	7.0×10^{22}
Intrinsic conductivity	σ_i	Sm^{-1}	2.2428	0.4325×10^{-3}	1.2832×10^{-6}
Mobility (Drift) at 300 K	μ_n (electron)	$\frac{m^2}{V-s}$	0.38	0.13	0.85
	μ_p (holes)	$\frac{m^2}{V-s}$	0.18	0.05	0.04
Work function	W	Volt (V)	4.4	4.8	4.7
Raman phonon energy		eV	0.037	0.063	0.035

4.1 *P-N* Junction

When a *P*-type semiconductor is brought into contact with *N*-type semiconductor as the process of crystallisation is taking place, the resulting combination is called a *P-N* junction. This junction has important properties and is, in effect, the basis of modern semiconductor theory and practice. Most semiconductor devices contain one or more *P-N* junctions. The most important characteristic of a *P-N* junction is its ability to conduct current in one direction only. In the reverse direction it offers very high resistance.

Formation of a P-N Junction: Figure 15.16 shows three types of such junctions. In Fig. 15.16(a), *P* and *N* regions have been grown into the germanium block by mixing acceptor and donor impurities, respectively, into the single crystal during its formation. This is known as a grown junction. It is worthwhile to mention that the grown type of *P-N* junction is not a sandwich made by attaching a *P* block to a *N* block, but actually consists of *P* and *N* layers in a single piece of germanium.

The diffused junction in Fig. 15.16(b) is made by placing a pellet of acceptor impurity, such as indium, on one face of a wafer of *N*-type germanium and then heating the combination to melt the impurity. Under proper conditions of temperature and time, a portion of the impurity metal will diffuse a short distance into



(a) Grown *P-N* junction

(b) Diffused *P-N* junction

(c) Point contact *P-N* junction

Fig. 15.16 Types of *P-N* Junction

the wafer, thereby creating a region of *P*-type germanium in intimate association with W-type bulk. This is also called an *alloyed junction* or fusion-alloy junction from the fact that a small amount of pellet material alloys with the germanium.

Figure 15.16(c) shows a point contact type. Here, a fine, pointed wire (catwhisker) makes pressure contact with the face of a *N*-type germanium wafer. After assembly, the device is electroformed by passing a high-current surge momentarily across the junction of Wafer and Whisker. The heat generated during the short interval drives a few electrons from the atoms in the region of the point contact, leaving holes and thus converting into *P*-type a small volume of germanium immediately under and around the point.

Silicon *P-N* junction is produced in a similar manner. In most instances, the Silicon has been processed in such a way as to make it *P*-type. To create the junction, as *N*-type material is either inserted at the proper point in the crystal process or (in the diffused, junction process) an *N*-type material is later inserted into the body of *P*-type wafer. Like germanium diodes, silicon diodes also are produced both in junction and point contact types.

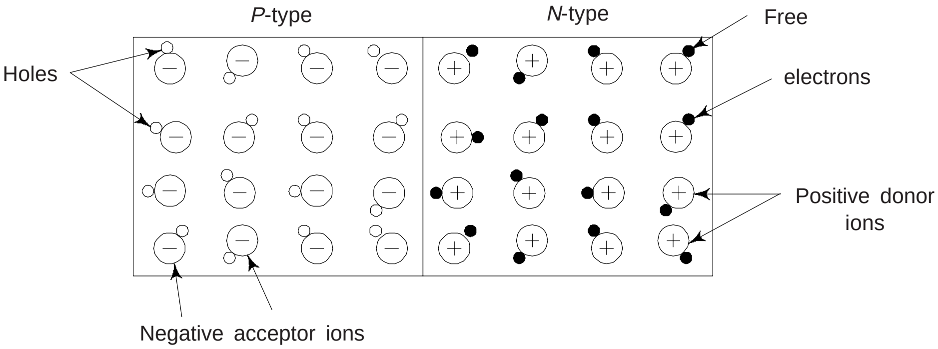


Fig. 15.17 A *P-N* junction when just formed

P-N junction with no external voltage: Figure 15.17 shows a *P-N* junction just immediately after it is formed. There is no external voltage connected to the *P-N* junction. Since *N*-type material has a high concentration of free electrons while *P*-type material has a high concentration of holes, the following processes are initiated:

(i) At the junction, holes from the *P* region diffuse into the *N* region and free electrons from the *N* region diffuse into the *P* region. This process is called *diffusion*. Holes combine with the free electrons in the *N*-region whereas electrons combine with the holes in the *P*-region.

(ii) The diffusion of holes from *P* region to *N* region and electrons from *N* region to *P* region across the junction takes place because they move haphazardly due to thermal energy and also because there is a difference in their concentrations (The *P* region has more holes whereas *N* region has more free elctrons) in the two regions.

(iii) As the free electrons move across the junction from *N*-type to *P*-type, positive donor ions are uncovered, i.e., they are robbed off free electrons. Hence a positive charge is built on the *N*-side of the junction. At the same time, the free electrons cross the junction and uncover the negative acceptor ions by filling in the holes. Therefore, a net negative charge is established on *P*-side of the junction. When a sufficient number of donor and acceptor ions is uncovered, further diffusion is stopped. It is because now, a barrier is set up against further movement of charge carriers. This is called *potential barrier or junction barrier*. The potential barrier is of the order of 0.1 to 0.3 volt. Figure 11.18 shows the electrostatic potential difference across the *P-N* junction. How this potential barrier is developed? When a sufficient number of donor and acceptor ions is uncovered, further diffusion is prevented. It is because now positive charge (ions) on *N*-side repels holes to across from *P*-type to *N*-type and

negative charge (ions) on *N*-side repels free electrons to enter from *N* type to *P* type. Because of this a difference in potential exists between the two sections, which inhibits further electron-hole combinations at the junction, and the Fermi level of the two sides is in the same level as shown in Fig. 15.18(c).

(iv) The region across the *P-N* junction in which the potential changes from positive to negative is called the *depletion region*. The width of this region is of the order of 6×10^{-8} m. Since this region has immobile (fixed) ions which are electrically charged, it is also called as the *space-charge region*. Outside this region on each side of the junction, the material is still neutral.

(v) The potential barrier for a silicon *P-N* junction is about 0.7 V, whereas for a germanium *P-N* junction it is approximately 0.3 V.

The potential barrier discourages the diffusion of majority carriers across the junction. However, the potential barrier helps minority carriers (few free electrons in the *P* region and few holes in the *N* region) to drift across the junction. The minority carriers are constantly generated due to thermal energy. But electric current cannot flow since no circuit has been connected to the *P-N* junction.

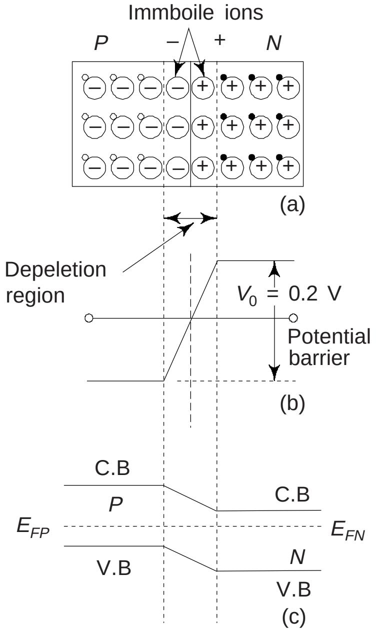


Fig. 15.18 Potential barrier across the *P-N* junction

Forward and Reverse Biasing

(i) Forward Biasing: We have seen that the natural tendency of the majority carriers (free electrons in the *N*-section and holes in the *P*-section) was to combine at the junction. This is how the depletion region and potential barrier were formed. Actually, the combination of electrons and holes at the junction allows electrons to move in the same direction in both the *P* and *N* sections. In the *N*-section, free electrons move toward the junction; in the *P* section, for the holes to move toward the junction, valence electrons move away from the junction. Therefore, electron flow in both the sections is in the same direction. This, of course, would be the basis of current flow.

With the *P-N* junction alone, the action stops because there is no external circuit and because of the potential barrier that builds up. So, for current to flow, a battery can be connected to the diode to overcome the potential barrier. And the polarity of the battery should be such that the majority carriers in both sections are driven toward the junction. When the battery is connected in this way, it provides forward bias, causing *forward or high current* to flow, because it allows the majority carriers to provide the current flow.

To apply forward bias, positive terminal of the battery is connected to *P* type and negative terminal to *N* type as shown in Fig. 15.19. The applied forward potential establishes an electric field which acts against the potential barrier field. Obviously, the resultant field is weakened and the barrier height is reduced at the *P-N* junction, Fig. 15.19(b) and (c). Since potential barrier height is very small (~0.2 V) and hence a small forward voltage is sufficient to completely eliminate the barrier. Obviously, at some forward voltage the potential barrier at the *P-N* junction can be eliminated altogether. Then the junction resistance will become almost zero, and a low resistance path is established for the entire circuit. Thus a large current is generated in the circuit even for the small potential difference applied. Such a circuit is called forward bias circuit and the current is called forward current. The salient features of the forward bias circuit are summarized below:

(i) At some forward voltage, the potential barrier is eliminated altogether.

(ii) The *P-N* junction offers low forward resistance (r_f) to current flow.

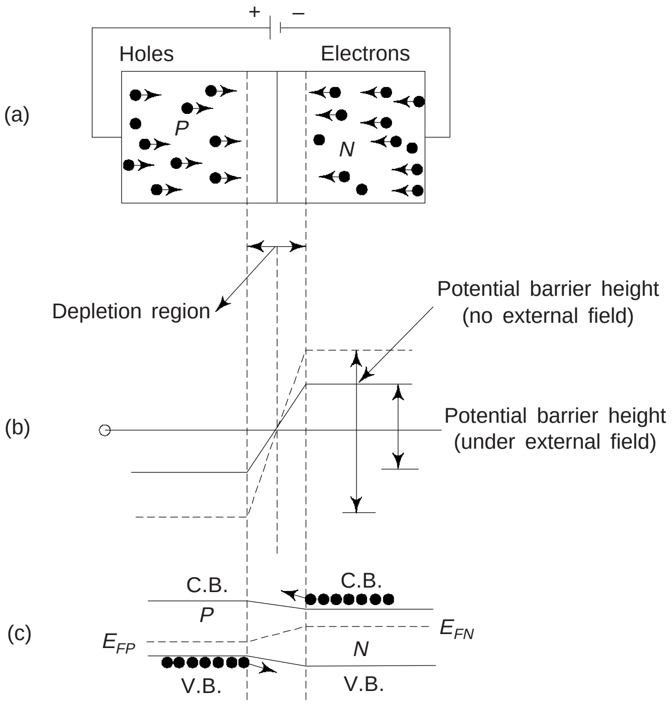


Fig. 15.19 ***P-N*** **Junction showing forward bias**

(iii) The magnitude of current in the circuit due to the establishment of low resistance path depends upon the applied forward voltage, and it reduces as the voltage is increased thereby forward current increases. It is given by

$$I = I_0 \left(e^{v/\eta V_T} - 1\right) \tag{20}$$

where V = applied voltage

$$V_T = \frac{T}{11600} \text{ here } T = \text{Temperature in Kelvin}$$

η = 1 for Ge and 2 for Si

The mechanism of current flow in a forward biased *P-N* junction is as follows:

(i) The free electrons from the negative battery terminal continue to arrive into the *N*-region while the free electrons in the *N*-region move towards the *P-N* junction.

(ii) The electrons travel through the *N*-region as free electrons. Obviously, current in *N*-region of the *P-N* junction is by fee electrons.

(iii) When these free electrons reach the *P-N* junction, they combine with holes and become valence electrons. Since a hole is in the covalent bond and hence when a free electron combines with a hole, it becomes a valence electron.

(iv) Current in the *P* region is by holes. The electrons travel through *P*-region as valence electrons.

(v) These valence electrons after leaving the crystal, flow into the positive terminal of the battery.

The current flow in a forward biased *P-N* junction is illustrated in Fig. 15.20.

(ii) Reverse Biasing: We have seen that for forward current flow, the battery must be connected to drive the majority carriers towards the junction, where they combine to allow electrons to enter and leave the *P-N* junction. If the battery connections are reversed, the potential at the *N* side will draw the free electrons away from the junction, and the negative potential at the *P* side will attract the holes away from the junction.

With this battery connection, then, the majority carriers cannot combine at the junction, and majority current cannot flow. For this reason, when a voltage is applied in this way, it is called reverse bias.

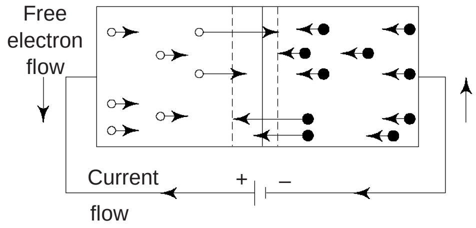


Fig. 15.20 Current flow under forward bias

However, reverse bias can cause a *reverse current* to flow, because minority carriers are present in the semiconductor sections. Remember, that although the *P* section is doped to have excess holes, yet some electrons are freed because of thermal agitation. Also, although the *N* section is doped to have excess free electrons, some electrons are freed to produce holes in the *N* section. The free electrons in the *P* section, and the holes in the *N* section are the minority carriers. Now, with reverse bias, one can see that the battery potentials repel the minority carriers toward the junction. As a result, these minority carriers cross the *P-N* junction in exactly the same way that the majority carriers did with forward bias. However, since there are much fewer minority carriers then there are majority carriers, this minority current, or reverse current as it is usually called, is much less and is the order of μA, with the same voltage than majority or forward, current would be. Reverse bias *P-N* junction is shown in Fig. 15.21. The salient features of reverse biased *P-N* junction are following:

(i) The height of the potential barrier is increased and width of depletion region also increases, Fig. 15.21(b) and (c).

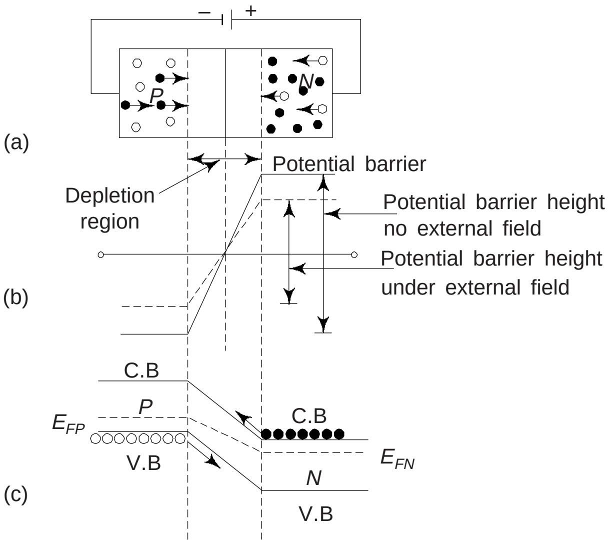


Fig. 15.21 *P-N* Junction showing reverse bias

(ii) The reverse bias *P-N* junction offers very high resistance to current flow. This resistance is called reverse resistance (R_r).

(iii) Due to the establishment of high resistance path, very small current flows in the circuit. This current is usually called as the *leakage current or reverse saturation current.*

The current in the above situation is given by

$$I = I_0(e^{v/\eta V_T} - 1) \tag{21}$$

Since applied voltage *V* is negative and as it increases the first term in R.H.S. reduces very fast and current from equation *(ii)* reduces to $I = (–)\ I_0$ which is quite low and constant.

Volt-Ampere Characteristic of *P-N* Junction

The *P-N* junction can be represented by a symbol of arrow and dash as shown in Fig. 15.22. The arrow head represents the *P*-section of the crystal and shows the direction of flow of holes or conventional current. Since the *P-N* junction diode's resistance changes according to the direction of current flow and hence it is called a *non linear* device. Basically, its non-linearity is dependent on the polarity of the applied voltage.

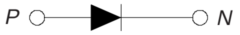


Fig. 15.22 Symbol of the *P-N* junction diode

A graph between the potential difference across the *P-N* junction and the current through the junction is called the *V-I* characteristics of the *P-N* junction diode and is shown in Fig. 15.23.

In the forward direction, though considerably more current flows and the current for the most part, increases linearly as the bias voltage is increased. In the forward direction, then, the *P-N* junction can be considered a linear device over a large portion of its operating curve. The small portion of the curve that is just above zero bias is non linear. This results because both majority and minority current actually comprise the overall current. Since the minority carriers are low energy carriers, majority current starts first, and then as the voltage is raised, minority current joins in, causing a non linear rise in current. But as the voltage is increased further, minority current becomes saturated since there are only few minority carriers. The curve then follows the majority current increase which is linear.

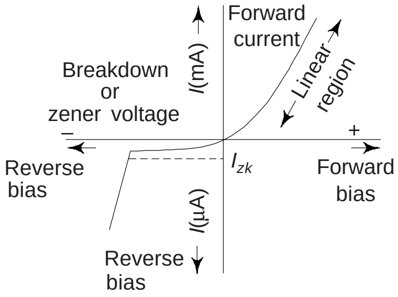


Fig. 15.23 *V-I* Charateristics of a *P-N* junction diode

Because of the non-linearity of the curve, if a very small signal voltage is applied to the diode so that it only operates around the knee, the signal will be distorted. The signals must be large enough so that they operate mostly over the linear part of the curve.

When reverse bias is applied, a slight reverse current flows. This reverse current increases only negligibly as the bias voltage is increased a lot (20 to 25 volts). At this stage current suddenly rises in the reverse direction due to the breakdown of the crystal, i.e., covalent bonds of the crystal are broken in very large number. This breakdown reverse bias is called the Breakdown or *zener voltage*. Diodes are also designed to produce a useful wide range of zener breakdown region, and are used in special voltage-regulating circuits.

Effect of Temperature on PN Junction Diodes

We know that temperature rise boosts the generation of electron-hole pairs in semiconductors and increases their conductivity. From a consideration of the energies of the carriers crossing the depletion region in a *P-N* junction diode, an involved calculation yields the following relation between diode current, voltage and temperature.

$$I = I_0 \exp [V/\eta\, V_T - 1] \tag{22}$$

Here

I = the diode current (forward if positive, reverse if negative)
I_0 = the diode reverse current, also called the reverse saturation current, at temperature T.
V = the diode voltage, positive for forward bias, negative for reverse bias; in volts
η = 1 for germanium and 2 for silicon.
V_T = T/11600, a quantity in volts, dependent upon temperature and is known as volt temperature equivalent.
T = Temperature of the diode junction (°K)

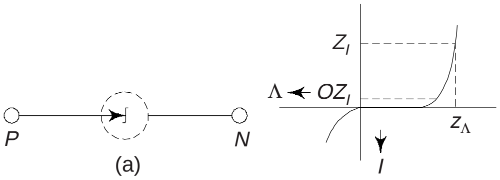
On increasing the temperature the forward characteristic (Fig. 15.23) shifts to left, showing increase in current for same voltage and it shifts to right when temperature is decreased showing biased current.

4.2 Zener Diodes

We have seen that in the breakdown region, large changes in diode current produce only small changes in diode voltage. So a semiconductor *P-N* diode designed to operate in the breakdown, region may be employed as a constant voltage device. The diodes used in such a manner are called *avalanche breakdown* or *zener diodes*.

They are used as a voltage regulator. The voltage source *V* and the resistance *R* are so selected that the diode operates in the breakdown region. The diode voltage in this region which is also the voltage across the load R_L is called *zener voltage* (V_z) and the diode current is called zener current (I_z). As the load current (I_L) or the supply voltage changes, the diode accommodates itself to these changes and maintains nearly constant load voltage (V_z).

The diode will continue to regulate the voltage until the diode current falls to knee current I_{zk} (Fig. 15.24). Depending upon the nature of the semiconductor and its doping, the breakdown voltage in diode ranges from about 3 volt to several hundred volts. The breakdown phenomenon is reversible and harmless so long as the safe operating temperature is maintained.



(a)

Fig. 15.24

The mechanism of diode breakdown at reverse voltage is explained below:

(i) Avalanche Breakdown: In this mechanism, the minority charge carriers (electrons in *P* type and holes in *N* type) acquire sufficient energy from the applied reverse voltage to produce new charge carriers by removing valence electrons from the covalent bonds. The new carriers in turn produce additional charge carriers and the process multiplies to give large reverse currents. The diode is then said to be in the region of *avalanche breakdown*, usually, a junction with a broad depletion layer (therefore a low field intensity) breaks down by this mechanism. With the increase of temperature, the vibrations of the atoms in the crystal increase which increases the possibility of collisions of the charge carriers with the lattice atoms and reduces the possibility for the carriers to gain sufficient energy to start avalanche process. Thus at high voltages, the avalanche process is prominent and does take place to cause diode breakdown. The operating voltages in such diodes range from several volts to a few hundred volts.

(ii) Zener Breakdown: In this mechanism, the breakdown is initiated through a direct rupture of covalent bonds rather than avalanche process due to the existence of strong electric field across depletion layer. A junction having a narrow depletion layer (and hence high field intensity $E = \frac{V_r}{d}$) will break down by this mechanism. An increase in temperature increases the energy of the valence electrons and makes it easier for these electrons to escape from these covalent bonds. Smaller applied voltage is, therefore, required to pull these electrons from the crystal lattice. The *zener effect plays an important role only in diodes with breakdown voltage below about 6 volts.* The zener diode is always used in reverse biased condition.

The range of voltages about the breakdown voltage in which a zener diode conducts in reverse direction is called *tolerance*.

The breakdown voltage of zener diode depends upon operating temperature. It is found to decrease with increase in junction temperature. This is due to the increased reverse current, (i.e. increase in minority carriers) that flows with increasing temperature. The decrease in breakdown voltage is about 2 mV per degree centigrade rise in temperature.

The maximum power which a zener diode can dissipate (or handle) without damage is referred as its *power rating* and denoted by P_{ZM}. Zener diodes for the commercial purposes available in the market have power rating from 1/4 W to more than 50 W.

The opposition offered to the current flowing through the zener diode in the operating reigon is known as *zener resistance* (R_Z) or Zener impedance (Z_Z).

Zener diodes find wide commercial and industrial applications, e.g. voltage stabilizer, meter protection, wave shaping, etc.

Constant Current Diode

These are the diodes that work exactly opposite to Zener diode. These diodes keep the *current constant flowing through them when* the voltage changes, i.e. instead of holding the voltage constant, these diodes hold the current constant. The range of voltage over which a diode can keep the current constant is known as *voltage compliance*. These diodes are optimised for a particular voltage compliance.

4.3 Tunnel Diode

It is a device just like a *P-N* junction, which offers negative resistance under certain bias conditions. The negative resistance of the diode is due to tunneling and it is called a tunnel diode or Esaki diode. It is made very much like an ordinary *P-N* junction diode, except that both the *P* and *N* regions are heavily doped ($1 : 10^4$) thousand times more than ordinary diode. It is used as an active device in electronic circuits in the frequency range of few megahertz. The semiconductors with very high impurity concentration are referred to as degenerate semiconductors. Typical tunnel diode characteristic is shown in Fig. 15.25.

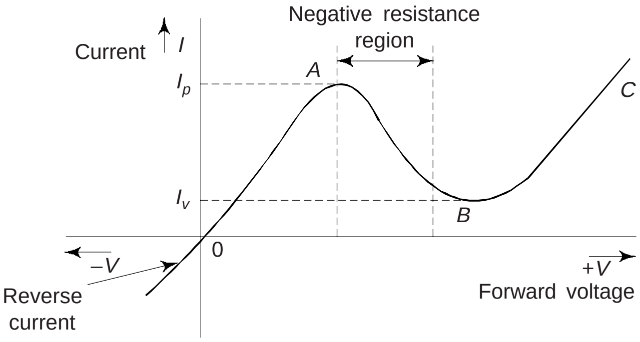


Fig. 15.25 Tunnel diode characteristic

As a result of high impurity levels, the contact potential is high, the depletion layer is very narrow and the Fermi levels lie in the conduction band for *N*-side and in the valence band of the *P* side. Under these circumstances, a very abrupt transition from *P* to *N* type material is achieved within the crystal. Since depletion region is very narrow; this gives rise to extremely large electric fields.

Tunnel Effect: When a *P-N* semiconductor is heavily doped, (it has many majority carriers and ions), and under forward biased the hole and valence electron random drift is heavy. As a result, it is not uncommon for a large number of electrons to fill holes and release energy to only a few other valence electrons. These few electrons, then, have their energy levels raised considerably so that they can cross from the *N* to the *P* section and current increases as shown by *OA* (Fig. 15.25), even with little or no applied voltage. This action which seems to allow a valence electron to cross a potential barrier without enough applied external energy is called the *tunnel effect* because it seems as though the valence electron 'tunnels' through the forbidden band. On further increasing voltage the barrier height decreases and Fermi level E_{FN} is much raised so current decreases as shown by *AB* (Fig. 15.25). This region is called *negative resistance region*. On further increasing voltage the conduction band of *N* type is in level with *C B* of *P* type and free electrons of *N* type easily cross to *P* type thereby current *BC* again increases like ordinary diode as shown by *BC* (Fig. 15.25).

The negative resistance region *AB* (Fig. 15.25) allows the diode to be used as an oscillator. It can also be used as an electronic switch since it has a good response in negative resistance region. By its nature, the tunnel diode has a rather high reverse current, but operation under this condition is not generally used.

Our main interest in the tunnel diode is its application as a very high speed switch. Since tunneling takes place at the speed of light, the transient response is limited only by total shunt capacitance (junction plus stray wiring capacitance) and peak driving current. Switching times of the order of a nanosecond are reasonable and times as low as 50 p. secs have been obtained.

The advantages of a tunnel diode are *(i)* low cost, *(ii)* low noise, *(iii)* high speed, *(iv)* environmental immunity and *(v)* low power.

The disadvantages of the diode are its low output voltage swing and the fact that it is a two terminal device unlike ordinary diode in which current flows only when forward biased. Because of this latter feature, there is no isolation between the input and output and it leads to serious circuit design problems.

However we may have a special type of tunnel diode whose peak value current is of the order of valley current. Such a diode is called *Backward diode*.

4.2 Photodiode

A photodiode is a *P-N* junction diode packed into a transparent plastic packet working under reverse biased condition.

Principle: A very small current flows through a *P-N* junction diode when it is reverse biased. It is because, minority charge carrier takes part in conduction. The number of minority carrier depend upon the working temperature. However when photons (light) of suitable frequency is incident on the junction, the number of charge carrier increases. But this happens only when light radiations touch the junction which is not the case in ordinary diode.

Working: In a photo diode visible light is focussed on the junction through a lens. Light on being incident produces free electrons and holes. Thus number of charge carriers increases and current increases. Generally, a photodiode is optimized for its sensitivity to light. As the light intensity increases more and more charge carriers are generated and reverse bias current increases.

In this respect a photodiode acts as a photo detector, a device which converts incoming light signal into electrical signal.

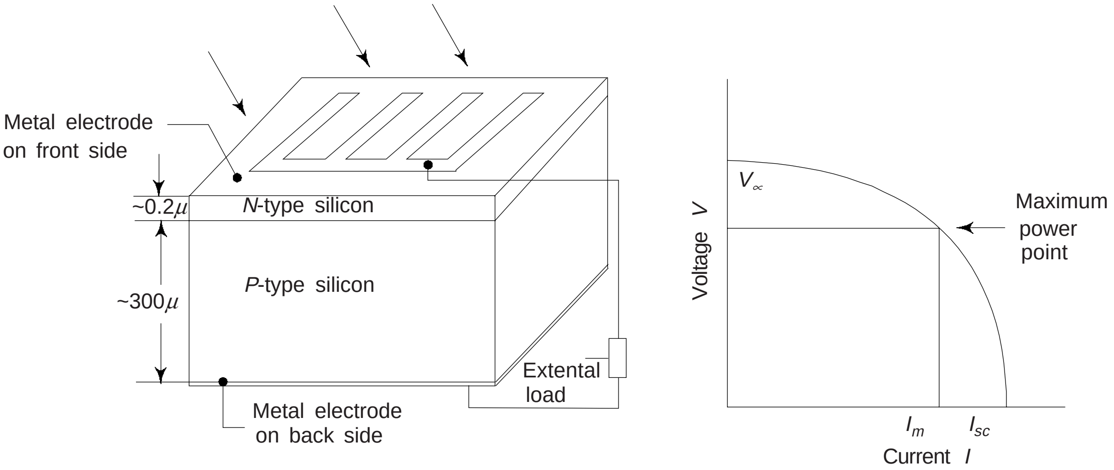
4.3 Solar Cells

Solar cells, which convert the sunlight directly into electricity, at present furnish the most important long duration power supply for satellites and space vehicles. These cells have also been successfully employed in small-scale terrestrial applications. Today, the solar cell is considered a major device for obtaining energy from the sun. Since it can convert sunlight directly to electrically with high conversion efficiency, and can provide nearly permanent power at low operating cost, and is virtually free of pollution. Recently, research and pollution development of low-cost, flat panel solar cells, thin film devices, concentrator systems, and many innovative concepts have increased. We can expect that in near future, the costs of small solar power modular units and solar power plants will be economically feasible for large-scale use of solar energy.

The first solar cell were developed by Chapin et al. in 1954 using a diffused silicon *P-N* junction. Subsequently, the cadmium-sulphide solar cell was developed by Raynolds et al. in 1954. To date, solar cells have been made of many other semiconductors, using various device configurations and employing single-crystal, polycrystal, and amorphous thin film structures.

Photo-voltaic Effect and Solar Cells

The photo-voltaic effect can be observed in nature in a variety of materials, but the materials that have shown the best performance in sunlight are the semiconductors. In photovoltaic conversion, the solar radiation falls on devices called solar cells which convert the sunlight directly into electricity. The principal advantages associated with solar cells are that they have no moving parts, require little maintenance, and work quite satisfactorily with beam or diffuse radiation.



(a) (b)

Fig. 15.26 (a) Schematic representation of a silicon solar cell, (b) A typical current-voltage characteristic for an ideal solar cell

A typical schematic representative of a conventional silicon solar cell is shown in Fig. 15.26(a). Silicon cells are thin wafers about 300 μm in thickness and 3 to 4 cm in diameter sliced from a single crystal of an type or *P-type* doped silicon. A shallow *P-N* junction is formed at one end by diffusion. Metal electrodes made from a Ti-Ag solder are attached to the front and back side of the solar cell. On the front and back side of the cell are attached metal electrodes made from a Ti-Ag solder. On the front side, the electrode is in the form of a metal grid with fingers which permit the sunlight to go through, while on the backside, the electrode completely covers the surface. An antireflection consists of SiO having a thickness of about 0.1 μm is also put on the top surface of the cell.

A well-made Si cell can have about 10 percent efficiency for solar energy conversion, providing approximately 100 W/m^2 of electric power under full illumination. This is modest amount of power per unit solar cell area, considering the effort in fabricating a large area of Si cells. One approach to obtaining more power per cell is to focus considerable light onto the cell using mirrors. Although Si cells lose efficiency at the resulting high temperatures, Ga As and related compounds can be used at 100°C or higher. In such solar concentrator system more effort and expense can be put into the solar cell fabrication. Since fewer cells are required. For example, a Ga As—Al Ga As hetrofunction cell provides good conversion efficiency and operates at the elevated temperatures common in solar concentrator systems.

Applications: Some of the important applications of photodiodes are *(i)* Photo detection (both visible and invisible), *(ii)* Demodulation, *(iii)* Logic circuits, *(iv)* Switching, *(v)* Optical communication system.

4.4 Light-emitting Diode (LED)

LED is a solid state (*P-N* junction diode) light source which has replaced incandascent lamps in many day to day applications.

LED is just not an ordinary *P-N* junction diode where silicon is used. Here we use compounds having elements like gallium, arsenic and phosphorus which are semi-transparent unlike silicon which is opaque. (Gallium-arsenide gives infrared radiations and gallium-arsenide-phosphide gives visible light either red or yellow.)

The advantages of an LED are:

(i) Low voltage of operation
(ii) Long life (more than 20 yrs)
(iii) Fast ON-OFF switching (10^{-9} sec)

The other uses of LED are:

1. For indicating power ON/OFF (Power level indicators)
2. Optical switching applications
3. Solid state video displays
4. Optical communication—Energy coupling circuits.

All natural colours are composed from three primary colours namely red, green and blue. LED in red and green have been available-but the blue one has been missing. Blue light emitters have several potential applications. They can be used for full colour displays of large area. They can be used in traffic lights as replacement for ordinary bulbs resulting in huge power savings as well as cost. Use of blue lasers can also result in higher density storage of information in optical CD-ROMs. Blue light emitters have been fabricated using several materials including ZnO, ZnSe and SiC. These attempts have been reasonably successful. However maximum success in last few years has been achieved with GaN. This material has proved to be very useful as it is also capable of operating at high power density, high temperature and unfriendly environment GaN having a band gap of 3.4 eV can give continuously varying band gap by combining with AIN to get upto 6.2 eV and with In N to get a band gap down to 1.9 eV. The high thermal conductivity and superior stability of this material makes it ideal for several applications over other competing materials.

4.5 (a) Liquid Crystal Display (LCD)

Just like LED we have another type of display that uses seven segment called Liquid Crystal Display also known as *electroluminescent display*. It consists of a thin layer of normally transparent liquid crystal material between two electrodes.

When an electric field is applied, the liquid crystal material between the two electrodes becomes turbulent, reflecting and scattering ambient light. It provides excellent brightness under high ambient light conditions and requires only 50 μW of power per segment, there by total power for one complete display of 7 segment is 350 μW. This power is much less than that of an LED display, but the life expectancy is not as high as that for LED which is 10,000 hours minimum.

These displays are used in Watches, Pocket Calculators, Pocket televisions and portable instrument displays.

4.6 Thermistors and Barretters

We have read that the electrical conductivity of a semiconductor changes significantly with temperature and has a negative temperature coefficient of resistivity. This property is utilised in a device called *thermistor*, whose resistance is temperature sensitive, usually decreasing with temperature. Conventional wire-wound metallic resistors have positive temperature coefficients of resistivity. Commercial thermistors are usually made of sintered mixtures of Mn_2O_3, NiO_2 and Co_2O_3. A thermistor consists of a semiconductor bead of approximately 0.04 cm in diameter. Two thin wires are attached to the bead to provide for the two terminals. Diameter of the wire is approximately 0.25 μm.

Thermistors finds use in control systems operated by temperature changes, in the measurement of microwave power, in thermometry, and as a thermal relay. In electronic circuits, themistors have been used to compensate for the change in resistance with temperature of ordinary components where variation of component values cannot be tolerated.

The I-V characteristic of a thermistor has a negative slope. Devices that exhibit, in some region, a negative slope in their I-V characteristics are useful for making oscillators, amplifiers and switching circuits. However, themistors are not suitable in these applications because their response characteristics are too slow. There are certain bulk semiconducting compounds which have negative resistance characteristics over a limited range of operating parameters utilizing mechanism unrelated to the temperature sensitivity of the resistivity. These materials have been used to obtain devices based on *Gunn effect*.

A heavily doped semiconductor shows metallic properties. It has a positive temperature coefficient of resistivity owing to the decrease of the carrier mobility with increasing temperature. Such a device is called *sensistor*. Thermistors also find extensive use as sensing elements in microwave power measuring equipments and as temperature sensors of electronic thermometers. Thermistors are capable of yielding power information over the power range of 10^{-5} mW to 20 mW with a typical burn out level of 400 mW.

A *barretter* has a positive temperature coefficient of resistance, consists of an approximately mounted piece of "Wollaston" platinum wire having diameter of approximately 1.25 μm. Barretters are capable of yielding power information over the range of 10^{-5} mW to 20 mW. A typical burnout level is 20 mW. For low level *rf* power application below 10^3 MHz, 0.001 A, "Littlefuse" may be used as a barretter.

4.7 Gunn Effect and Gunn Diode

Ridley-Watkins and independently Hilsum predicted that semiconductor materials under certain conditions can offer differential negative resistance. This differential negative resistance is a bulk effect and due to transfer of electrons from one valley to another in the conduction band. Gunn while experimenting on a sample of *N*-type GaAs and some other III-V compounds, found that the current through the sample increased linearly with voltage till a certain threshold voltage. Beyond the threshold voltage a number of current pulses appeared with a time interval proportional to the length of the sample. This threshold field is high (~ 400 V/mm). The oscillations lie in the microwave range and set in the negative differential conductance (NDC) region where the current decreases with increase in electric field *E* (Fig. 15.27). This behaviour is a consequence of the band structure of these materials. In the region of velocity field curve, where *v* decreases with increasing *E*, the differential mobility (*dv*/*dE*) of the electrons becomes negative. The reason is following.

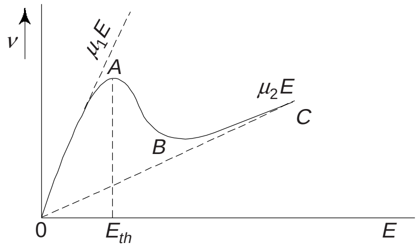


Fig. 15.27 A possible characteristic of electron drift velocity vs. field for a semiconductor exhibiting the transferred electron mechanism

n-GaAs has a direct energy band gap. The band gap is 1.4 eV. The free electrons in *n*-GaAs normally occupy the lowest energy states in the conduction band. The effective mass of electrons in this situation is and their mobility μ_1 is high (≈ 0.02 m^2/v-s). So as a result of the transfer of the electrons current begins to decrease with increase in field because $T = n_1 e \mu_1 E + n_2 e \mu_2 E$, where n_1 and n_2 are the concentration of electrons having the mobiltiy μ_1 and μ_2 respectively. The average drift velocity of an electron is

$$V = \left(\frac{n_1 \mu_1 + n_2 \mu_2}{n_1 + n_2}\right) E \qquad (23)$$

($n_1 + n_2$) being the total electron concentration.

If the transition of the electrons from the high-mobility state to the low-mobility state occurs rapidly over a range of field *E*, one observes that *v* diminishes with field *E* beyond a certain threshold field $E_{\text{th,}}$ shown by region *AB* in Fig. 15.27, when all the electrons move to the lob-mobility state, the drift velocity is $v = \mu_2 E$. One finds that *v* continues to increase slowly with *E*. This builds up the successive oscillations.

Owing to the occurrence of the NDM in the velocity-field curve, Gunn diodes are used as sources of microwaves especially where high power is not a requirement. Gunn diodes are also used as local oscillators for mixers in microwave receivers over the frequency range 1 to 100 GHz. A Gunn device can carry just before the threshold voltage one of two possible currents, depending on the presence or absence of a domain. This can, therefore, be used as a high-speed binary logic.

Gunn diode is a versatile semiconductor device. These are commercially available for pulsed operation yielding a power output of 5 W in the frequency range 5.0 to 12.0 GHz.

4.8 Impatt, Trapatt and Qwitt Diodes

Impatt (Impact Ionization Avalanche Transit Time) diodes employ impact-ionization and transit-time properties of semiconductor structures to produce negative resistance at microwave frequencies. TRAPATT (Trapped Plasma Avalanche Triggered Transit) diode is an IMPATT related device. These are quite new devices and used as oscillators in the microwave region. These devices work well in the breakdown region. Although IMPATT operation can be obtained in simpler structures, the Read diode is best suited for illustration of basic principles.

The Read diode (Fig. 15.28) consists essentially of two regions: *(i)* the n^+-p region at which avalanche multiplication occurs and *(ii)* the *(i)* (essentially intrinsic) region through which generated holes must drift in moving to the p^+ contact. One can built similar devices in the p^+-n-i-n^+ configuration, in which electrons resulting from avalanche multiplication drift through the *i*-region taking advantage of the higher mobility of electrons compared to holes.

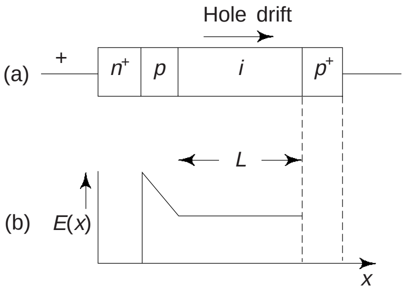


Fig. 15. 28 The Read diode (a) The basic configure (b) Electric field distribution in the device under reverse bias

The device operates in a negative conductance mode when the a.c. component of current is negative over a portion of cycle during which the a.c. voltage is positive, and *vice-versa.* The negative conductance occurs because of two processes, causing the current to lag behind the voltage in time: *(i)* a delay due to avalanche process and *(ii)* a further delay due to the transit time of the carriers across the drift region. When these two delay times combine to produce a net phase-shift between 90° and 270°, the diode resistance is negative at the corresponding frequency. Consequently, negative conductance occurs and the device can be used for oscillation and amplification.

The structures that have been successfully employed for IMPATT devices are the PN, PIN, P^+ NN^+, P^+ NIN^+ and N^+ PIP^+ (called read diodes). These devices are usually constructed from silicon and gallium arsenide. At present, IMPATT diode can generate the highest CW (Continuous Wave) power output at millimeter-wave frequencies, i.e. microwave power above 30 GHz. A major drawback of the IMPATT diode for very high frequency operation is the fact that the avalanche process, which depends on random impact ionization events, is inherently noisy.

In TRAPATT diode, the TRAPATT mode of operation, i.e. the operating frequency is substantially lower than the transmit time frequency and the efficiency is considerably higher. Under large-signal conditions the periodic avalanching of the diode begins at the high-field side and sweeps rapidly across the diode, leaving it substantially filled by a highly conducting plasma of holes and electrons whose space-charge depresses the voltage to very low values. Since the plasma cannot rapidly escape, this mode is called TRAPATT mode. This diode has been tried experimentally in pulsed transmitter of phased array radar systems.

A variety of approaches have been investigated to find alternative methods for injecting carriers into the drift region without relying on the avalanche mechanism. A particularly interesting device is QWITT (Quantum Well Infection Transit Time) diode, which employes resonating tunneling through a quantum well to inject electrons into the drift region. The device consists of a single GaAs quantum well between two $Al_x\,Ga_{1-x}$. As barriers, in series with a drift region of undoped GaAs. This structure is then placed between two n^+-GaAs regions to form two contacts (Fig. 15.29). In this device, one can achieve maximum resonant tunneling of electrons through the well if the dc bias is properly adjusted.

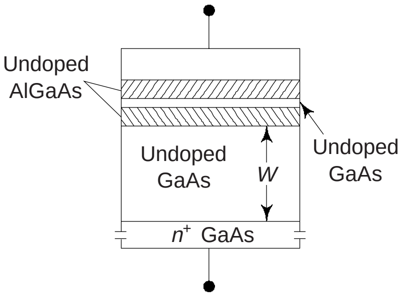


Fig. 15.29 Structure of a QWITT diode

QWITT diode is a low-noise injection mechanism with superior high frequency characteristics. The length of the transit time as well as the shape of the current pulse can be optimized to obtain the best power-frequency performance from the QWITT diode. This diode should extend the normal frequency limit associated with transit-time devices, while providing higher output power than simple quantum-well RTD oscillations.

4.9 Pin Diodes

This is a *P-N* junction with greatly improved switching times. Obviously this is a *PN* junction with a doping profile tailored so that an intrinsic layer the "*I*-region" is sandwitched between a player and an *N*-layer as shown in Fig. 15.30.

In practice, however, the idealized *I* (intrinsic)-region is approximated by either a high resistivity *P* layer (referred to as π-layer) or a high resistivity *N*-layer (referred to as v layer). The resistivity of *I* layer is typically 10^3 Ω-m.

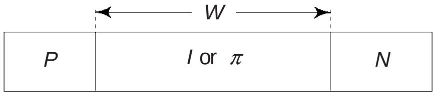


Fig. 15.30 ***PIN*** **diodes**

When no external voltage is applied between the terminals of PIN diode, the concentration gradient across the function cause diffusion of carriers. The width of the depletion region in the *I*-layer is relatively large. Now, if a reverse bias is applied and increased gradually, the depletion region becomes thicker. At *swept out voltage* (a particular value of the reverse bias), all the free carriers are swept out of the intrinsic layer. With further increase of reverse bias, the depletion region widens in the highly doped semiconductor regions. PIN diode offers a very high resistance under reverse bias. We must note that the breakdown voltage of the diode under reverse bias is also very large.

The PIN diode when used as a switch operates between the ON and OFF states. In ON state the diode is forward biased when minority carriers are infected into the high resistivity region between the highly doped regions. The increase in the densities of carriers reduces the resistance and the diode impedance is low. This causes a current flow. The forward resistance of the diode varies with the forward bias. In the OFF state the high resistivity region is completely swept out and the diode impedance is very high as the capacitance is very low.

The PIN diode has found wide applications in microwave circuits. It can be used as a microwave switch with essentially constant depletion layer capacitance and high power handling capability. When PIN diodes are used as microwave switches and when they are biased in the OFF condition the bias is usually beyond the swept-out voltage (usually –2 V). PIN diodes can also be used as a *variolosser* (variable attenuator) by controlling the device resistance which varies approximately linearly with the forward current. PIN diodes can also modulate signals upto the GHz range. We must note that the forward characteristics of a thyristor in its ON state closely resemble those of PIN diode.

5. THE TRANSISTOR

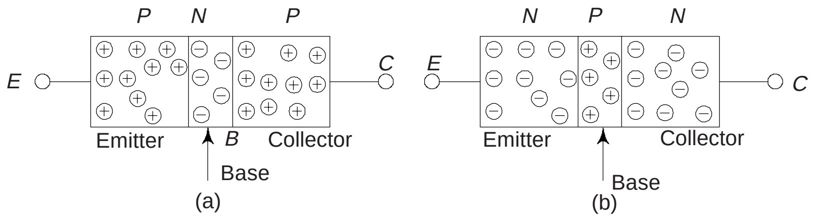
Transistors are three terminal (solid-state devices and are extremely important semiconductor devices in today's microelectronic circuitry.) In a way they have revolutionised the entire field of electronics. They are capable of two primary types of function. First, transistors can perform the same operation as their vacuum tube precursor, the triode valve; i.e., they can amplify an electrical signal (i.e., as amplifier). In addition they serve as switching devices in computers for the processing and storage of information. The major types of transistors are: the *junction* (or bimodal) *transistor* and the *metal-oxide semiconductor field-effect transistor* (MOSFET).

5.1 Junction Transistor

This is composed of two *P-N* junctions arranged back to back in either the *N-P-N* or *P-N-P* configuration. The charge carriers in a junction transistor have positive as well as negative polarity and so the nomenclature *Bipolar Junction Transistor* is used.

In actual practice a junction transistor consists of a silicon (or germanium) crystal, in which a very thin layer of *N*-type Si (Ge) is sandwitched between two layers of *P*-type Si (Ge) and thereby we get *PNP* junction transistor. Alternatively, it may consist of a very thin layer of *P*-type between two layers of *N*-type material and we get a *NPN* transistor. The semiconductor sandwitched is extremely thin, and small and is hermatically sealed against moisture inside a metal or a plastic case. The size is quite small.

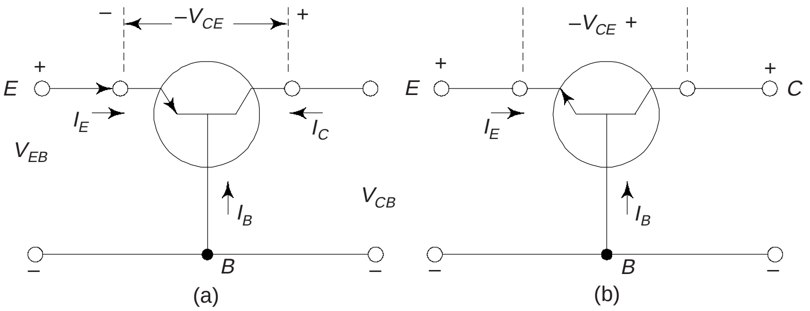
The three terminals taken from each section of a semiconductor are called *(i) Emitter (ii) Base* and *(iii) Collector* (Fig. 15.30). The middle section, i.e. base is very thin in comparison with the other two. Metal leads *E*, *B* and *C* come out of the package for connection to the emitter, the base and the collector, respectively. In a junction transistor both the majority and the minority carriers are involved.



(a) (b)

Fig. 15.30(a) Bipolar junction transistor

In a normal transistor operation, emitter-base junction is forward-biased and the collector-base junction is reverse-biased. The circuit symbols for both the types of transistors are shown in Fig. 15.31. The arrow on the emitter shows the direction of the current when the emitter-base junction is *forward-biased*. Obviously, the current *enters* the transistor through the emitter terminal for a *PNP* transistor and leaves the transistor through the emitter terminal for a *NPN* transistor. In both the cases, the emitter (*E*), base (*B*) and collector (*C*) currents, I_E, I_B and I_C, respectively, are taken *positive* when the currents go into the transistor. The symbols V_{EB}, V_{CB}, and V_{CE} represent respectively the emitter-base, collector-base and collector-emitter voltages. These are assumed positive when the terminal marked by the first subscript is positive with respect to the terminal marked by the second subscript.



(a) (b)

Fig. 15.31 Current symbols with reference to the current direction and voltage polarities for: (a) *PNP* transistor (b) *NPN* transistor

As the emitter junction is usually forward-biased, I_E is negative in the case of an *NPN* transistor and positive for a *PNP* transistor. Since the collector junction is reverse biased, the voltage V_{CB} is negative for a *PNP* transistor and positive for an *NPN* transistor. We may note that signs are opposite for the two type of transistors.

Transistors are generally fabricated with the help of the following four basic techniques: *(i)* Grown technique *(ii)* Alloy or fused technique *(iii)* Diffusion technique and *(iv)* Epitaxial technique.

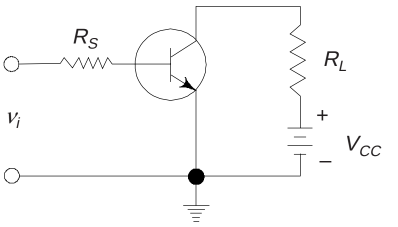


Fig. 15.32 Transistor as switching circuit

In many circuits such as those required in high speed counters and computer, a transistor is made to act as a *switch*. It is 'ON' when it is in saturation, and 'OFF' when it is in cutoff condition. By applying pulse voltages, the transistor is caused to switch over from one state to the other.

5.2 Field Effect Transistor (FET)

This is a semiconductor device with the output current controlled by an electric field. Since the current in FET is carried predominantly by one type of carriers, it is known as *unipolar transistor*. Obviously, FET is different from the bipolar transistor (BJT) which involves two types of carriers, i.e. both electrons and holes.

FET has several forms, e.g. *Junction Field Effect Transistor* (JFET), Metal Oxide Semiconductor Field-Effect (MOSFET) Transistor, etc.

The different classes of FETs are characterized by a high input impedance. These devices are used in controlled switching between conducting and non conducting states in digital circuits. FETs are also thermally stable. The main disadvantage of the FET is its relatively small gain-bandwith product in comparison with conventional transistor.

A schematic diagram of a FET is shown in Fig. 15.33. The JFET consists of a uniformly doped semiconductor bar *N*-type (or *P*-type) of Si or GaAs. The bar has Ohmic contacts at the two ends and semiconductor junctions on its two sides. If the semiconductor bar is *N*-type, the JFET is called an *N-channel JFET* and if the bar is *P*-type, the device is termed a *P-channel JFET*. Two sides of the semiconductor bar are heavily doped with impurities opposite to that of the bar, i.e. *P*-type impurities for a *N*-type bar and vice versa. For a *N*-type JFET, we can denote the three regions by P^+, N and P^+, where P^+ denotes a very heavily doped *P* region. By applying a voltage V_{DD} between the two ends of a semiconductor bar, a current is allowed to flow along the length of the bar. The central *N*-region through which the majority carriers (electrons) flow from the source (*S*) to the collector or drain (*D*) is called the *channel*. The P^+ regions form the *gate*. At each of the *PN* junctions there is a depletion layer extending into the *N*-region. These layers penetrate more and more as the V_{GG} (reverse bias) is increased, thus controlling the flow of electrons in the channel, i.e., the drain current I_D. The characteristic curves ($I_D - V_{DD}$ curves) are shown in Fig. 15.34. When I_D is increased from zero at constant V_{GS}, I_D increases linearly at first confirming to

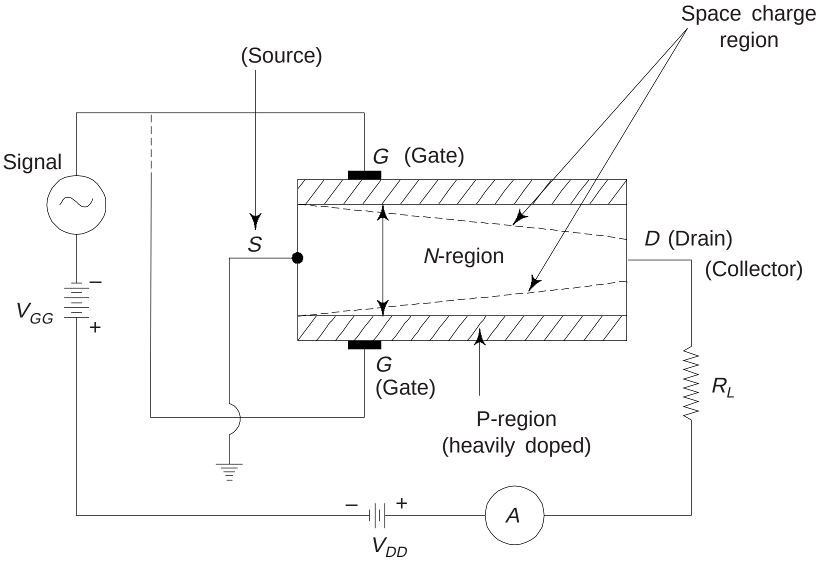


Fig. 15. 33 Basic structure of an N-channel JFET

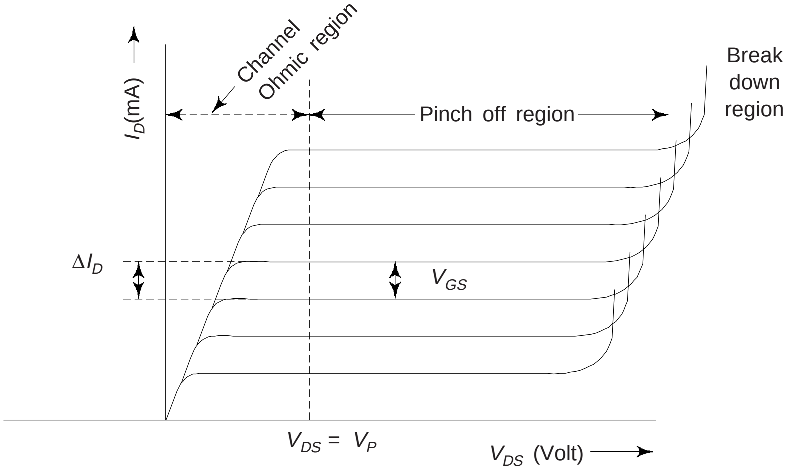


Fig. 15.34 $I_D - V_{DS}$ **static characteristic of an *N*-channel JFET**

the Ohm's law. With further increase in V_{DS}, the current attains saturation and this is called the *pinch-off* or the *saturation region*. Beyond a certain value of V_{DS}, I_D increases very rapidly indicating the break-down region. The name field effect is used for the device because the transverse *field* produced by the gate gives the *effect* of controlling the drain current.

JFETs are less temperature sensitive and are not affected much by radioactive radiations. The input resistance of JFET is very high. One can achieve much lower switching time in switching devices.

5.3 Metal Oxide Semiconductor Field Effect Transistor (MOSFET)

Like JFET, the MOSFET is also a field effect transistor, whose drain current (I_d) is controlled by the voltage on the gate. MOSFET and JFET differ physically as well as in operation. The MOSFET is also referred to as the IGFET (Insulated-Gate Field-Effect Transistor), MOST (Metal-Oxide-Semiconductor field effect Transistor) and MISFET (Metal-Insulator-Semiconductor-Field-Effect Transistor). The MOSFET is commercially more important than the JFET since MOS devices are suitable for large scale integration. MOSFETs are of two types: *(i)* Depletion type and *(ii)* Enhancement type. The MOSFET is an important power device.

MOSFETs can be of *N*- channel and *P*-channel types. MOSFETs have been constructed with various semiconductors, e.g. silicon and gallium arsenide, and with different insulators like SiO_2 and Al_2O_3. The Si-SiO_2 combination is the most common system.

One variety of MOSFET (depletion-mode *P*-type) is shown in Fig. 15.35. This consists of two small islands of *P*-type semiconductor that are created within a substrate of *N*-type Si, the islands are joined by narrow *P*-type channel. (A depletion mode *N*-type MOSFET is also possible, where in the *N*-and *P*-regions of Fig. 15.35 are reversed). Appropriate metal connections (source and drain) are made to these islands; an insulating layer of silicon dioxide is formed by the surface oxidation of the silicon. A final connector (gate) is then fashioned into the surface of this insulating layer.

The conductivity of the channel is varied by the presence of an electric field imposed on the gate. For example, imposition of a positive electric field on the gate will drive charge carriers (in this case holes) out of the channel, thereby reducing the electrical conductivity. Thus, a small alteration in the electric field at the gate will produce a relatively large variation in current between the source and the drain. In some respects, then, the operation of a MOSFET is very similar to that described for the junction transistor. The primary difference is that the gate current is exceedingly small in comparison to the base current of a junction transistor. Obviously, MOSFETs are, therefore used where the signal sources to be amplified cannot sustain an appreciable current.

Another important difference between MOSFETs and junction transistors is that, although majority carriers dominate in the function of MOSFETs (i.e., holes for the depletion-mode-*P*-type MOSFET (Fig. 15.35)), minority carriers do play a role with junction transistors (i.e., injected holes in the *N*-type base region).

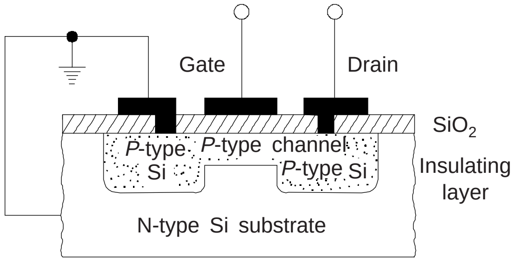


Fig. 15.35 Schematic cross-section view of a depletion-mode *P*-type MOSFET transistor

The MOSFETs can be used in any of the circuits wherever JFET is used, with the bulk (substrate) connected to source. We may note that handling of MOSFET is not so easy. It requires special precautions because of very fine layer of SiO_2 between gate and channel. It is very susceptible to high voltages and can get punctured. Even static electric charge can puncture it. This is why MOSFETs are protected by a shorting ring that is wrapped around all four terminals which must remain until the device is mounted and soldered properly.

6. SEMICONDUCTORS IN COMPUTERS

In addition to their ability to amplify an imposed electrical signal, semiconductor diodes and transistors may also act as *switching* devices, a feature utilized for arithmetic and logical operations, and also for storage of information in computers. Computer numbers and functions are expressed in terms of a binary code, i.e., numbers written to the *base 2*. Within this scheme or framework, numbers are represented by a series of two states (sometimes designated 0 and 1). Now, diodes and transistors within a digital circuit operate as switches that also have two states—ON and OFF, or conducting and non-conducting; ‘OFF’ corresponds to one binary state, and ‘ON’ to the other state. Obviously, a single number may be represented by a collection of circuit elements containing transistors that are appropriately switched.

7. MICROELECTRONIC CIRCUITS

We are all familiar with the impact of modern digital computers, communication systems, calculators, watches, etc. on society. One of the main cause of this great revolution is the advent of *integrated circuits* (ICs), which became possible because of tremendous progress in semiconductor technology in recent years. The operation of these systems and many other systems, is based on the principle of digital techniques and these systems are called as *digital systems*. Developments in ICs technology have made it possible to fabricate complex digital circuits, such as *microprocessors*, memory units etc. On tiny chips of silicon. The wonderchip—the microprocessor has been the most fantastic development of recent years. Inexpensive microelectronic circuits are mass produced by using some very igneous fabrication techniques. The process begins with the growth of relatively large cylindrical single crystals of high purity silicon from which thin circular wafers are cut. Many microelectronic or ICs, sometimes referred as ‘chips’, are prepared on a single wafer. A chip is rectangular, typically on the order of 6 mm (1/4 in) on a side and contains thousands of circuit elements: diodes, transistors, resistors, and capacitors. Presently, microprocessor chips containing 500 million transistors are being produced, and this number doubles about every 18 months.

Microelectronic circuits consist of many layers that lie within or are stacked on top of the silicon wafer in a precisely detailed pattern. Using a photolithographic techniques, for each layer, very small elements are masked in accordance with a microscopic pattern. Circuit elements are constructed in accordance with a microscopic pattern. Circuit elements are constructed by the selective introduction of specific materials (by diffusion or ion implantation) into unmasked regions to create localized *N*-type, *P*-type, high resistivity, or conductive areas. This procedure is repeated layer by layer until the total integrated circuit has been fabricated, as illustrated in the MOSFET scheme.

8. MICROELECTROMECHANICAL SYSTEMS (MEMS)

Almost all applications in industry, defence, medicine and other fields require sensing and control of various parameters. In recent years, taking advantage of the silicon IC technology, a new field has emerged which attempts to combine *sensor*, *actuator* and the *control* circuit on as one integrated unit. In this sense it emulates a biological system (Fig. 15.36). These are known as *smart sensors*, *microsystems technology* (MST) or *microelectromechanical systems* (MEMS). It has been possible to fashion miniature mechanical devices such as gears, motors, springs etc. Their combination with sensing and actuating functions has given engineers and scientists the tools to build microsystems that could not be imagined earlier. Some examples of the *smart structures* and systems are given in Fig. 15.37.

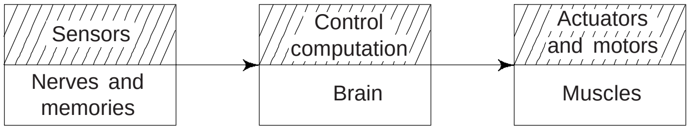


Fig. 15.36 **Integrated sensor-actuator systems with controller are analogous to biological systems**

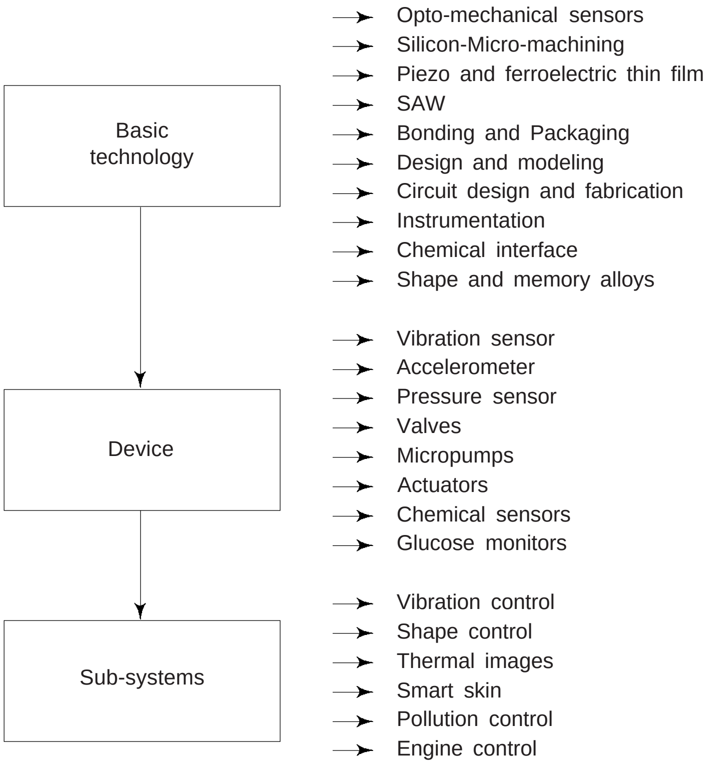


Fig. 15.37 **Few examples of smart structures and systems and their applications**

MEMS devices are small in size, light weight, low cost, reliable, with large batch fabrication technology. They generally consist of sensors that gather environmental information such as pressure, temperature, acceleration, etc., integrated electronics to process the data collected and actuators to influence and control the environment in the desired manner.

The MEMS technology involves a large number of materials. Silicon forms the backbone of these systems due to its excellent mechanical properties as well as mature micro-fabrication technology including lithography, etching and bonding. Other materials having piezoelectric, piezoresistive, ferroelectric and other properties are widely used for sensing and actuating functions in conjunction with silicon. The field of MEMs is expected to touch all aspects of our lives during this decade with revolutions in aviation, automobiles, pollution, control and industrial processes.

9. QUANTUM DOTS (QDS)

Rapid progress in the fabrication of semiconductor structures has resulted into the reduction of three dimensional systems to two-dimensional, one-dimensional, and finally to zero dimensional systems. These reduced dimensional systems are used in future applications like improved semiconductor lasers and microelectronics. *Quantum dots represent the ultimate reduction in the dimensionality of semiconductor devices*. These are three dimensional semiconductor structures only nanometers in size confining electrons and holes. QDs operate at the level of a single electron which is certainly the ultimate limit for an electronic device and are used as the gain material in lasers. QDs are use in quantum dot lasers, QD memory devices, QD photo-detectors and even quantum cryptography. The emission wavelength of a quantum dot is a function of its size. Obviously, by making QDs of different sizes, we can create light of different colours.

10. SPINTRONICS

A revolutionary new class of semiconductor electronics based on the spin degree of freedom could be created. The performance of conventional devices is limited in speed and dissipation whereas spintronics devices are capable of much higher speed at very low power. Spintronics transistor may work at a faster speed, are also smaller in size and will consume less power.

The study of electron spin in materials is called spintronics. Spintronics is based on the direction of spin and spin-coupling. Spins can be oriented in one direction or the other, called spin-up or spin-down. When electrons spins are aligned in one direction, these create a net magnetic moment as seen in magnetic materials like iron and cobalt. Magnetism is an intrinsic physical property associated with the spins of electrons in a material. The electron spin may exist not only in the up or down state but also in infinitely many intermediate states because of its quantum nature depending on the energy of the system. This property may lead to highly parallel computation which could make a *quantum computer* work much faster for certain types of calculations. In quantum mechanics, an electron can be in both spin-up and spin-down states, at the same time. The mixed state could form the base of a computer, built around not binary bits but the *quantum bits* or *qubit*. It is any combination of a 1 or a 0. The simplest device using spin-dependent effect is a sandwich with two magnetic layers surrounding a non-magnetic metal or insulator. If the two magnetic layers are different, then the magnetization direction of one can be rotated with respect to the other. This leads to the utilization of these structures as sensor elements and for memory elements. Scientists and engineers are now trying to use the property of the electron-like spin rather than charge to develop new generation of microelectronic devices which may be more versatile and robust than silicon chips and circuit elements. Spins appear to be remarkably robust and move relatively easily between semiconductors. In case of electron transport from one material to another, the spins do not lose its orientation or scatter from impurities or structural defects.

Explain 1 *Find the conductivity and resistivity of a pure silicon crystal at temperature 300°K. The density of electron hole pair per cc at 300°K for a pure silicon crystal is* 1.072×10^{10} *and the mobility of electron* $\mu_n = 1350$ *cm*^{2}*/volt-sec and hole mobility* $\mu_h = 480$ *cm*^{2}*/volt-sec.*

Solution Conductivity of pure silicon crystal is given by

$$\sigma = n_i e\,(\mu_e + \mu_h)\quad n_i = 1.072 \times 10^{10}$$

$$\sigma_i = 1.072 \times 10^{10} \times 1.6 \times 10^{-19}\,(1350 + 480)$$

$$= 3.14 \times 10^{-6} \text{ mho/cm}$$

$$\mu_h = 1350 \text{ vm}^2\text{/Volt-sec}$$

$$\mu_h = 480 \text{ cm}^2\text{/Volt-sec}$$

$$e = 1.6 \times 10^{-19} \text{ Coulomb}$$

Resistivity of silicon crystal is given by

$$\rho_i = \frac{1}{\sigma_i} = \frac{1}{3.14 \times 10^{-6}} = 3.18 \times 10^5 \text{ Ohm-cm}$$

$$= 3.18 \times 10^{-3} \text{ Ohm-m}$$

Example 2 *A silicon wafer is doped with phosphorus of concentration* 10^{13} *atoms/cm*^{3}*. If all the donor atoms are active, what is its resistivity at room temperature? The electron mobility is 1200 cm*^{2}*/Volt-sec charge on the electron is* 1.6×10^{-19}*Coulomb.*

Solution $\sigma = \mu e n$ $\mu = 1200$ cm/volt-sec

$\sigma = 1200 \times 1.6 \times 10^{-19} \times 10^{13}$ $e = 1.6 \times 10^{-19}$ C

$= 19.2 \times 110^{-4}$mho/cm $n = 10^{-13} = N_p$

Resistivity

$$\rho = \frac{1}{\sigma} = \frac{1}{19.2 \times 10^{-4}} = 5.2 \times 10^2 \text{ Ohm-cm}$$

Example 3 *Find the resistance of an intrinsic germanium rod 1 cm long, 1 mm wide and 1 mm thick at temperature of 300°K. For germanium* $n_i = 2.5 \times 10^{13}$*,* $\mu_n = 3900$ *cm*^{2}*/Volt-sec at 300°K.*

Solution $\sigma = n_i e\,(\mu_e + \mu_h)$

$$= 2.5 \times 10^{13} \times 1.6 \times 10^{-19}\,(0.39 + 0.19)$$

$$= 2.32 \text{ mho/m}$$

or $\rho = \frac{1}{232}$ Ohm × m

Now Resistance $R = \rho_\text{i} \left[\frac{\text{length}}{\text{area of cross section}}\right]$

$$= 2.32 \times \frac{10^{-6}}{10^{-2}}$$

$\therefore$ $R = \frac{1}{2.32 \times 10^{-4}}$ ohm = 4.31 kΩ

Example 4 *A sample of Germanium is made of P material by adding acceptor atoms at a rate of one atom per* 4×10^8 *Germanium atoms. The acceptor density is assumed to be zero and* $n_i = 2.5 \times 10^{19}$ *per m*^{3} *at 300°K. There are* 4.4×10^{28} *Germanium atoms/m*^{3}*. The acceptor density is found to be* 1.1×10^{20} *atoms/m*^{3}*.*

Solution $n_i^2 = np = n_p\, N_a$

$\therefore$ $n_p \simeq \frac{n_i^2}{N_a} = \frac{6.25 \times 10^{38}}{1.1 \times 10^{20}} = 5.6 \times 10^{18}$

$\therefore$ $\frac{n_p}{n_i} = \frac{5.6 \times 10^{18}}{2.5 \times 10^{19}} = 0.22$

Example 5 *Figure 15.38 shows the plot of log of resistivity versus reciprocal of temperature for two different semiconductors A and B. Assume that mobility is proportional to* $T^{-3/2}$*, find (a) which material has wider band gap? (b) which material will require light of shorter wavelength for generation of a electron-hole pair?*

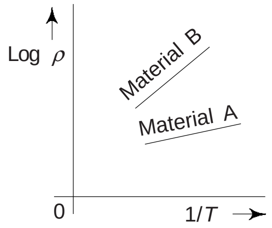


Fig. 15.38

Solution

(a) Resistivity,
$$\rho = \frac{1}{n_i(\mu_n + \mu_p)e}$$
$$\mu_n = \mu_{no} \times T^{-3/2}$$
$$\mu_p = \mu_{po} \times T^{-3/2}$$

∴
$$\rho = \frac{1}{\sqrt{A_0}\, T^{3/2}\, e^{-E_{go}/2kT}} \frac{1}{\left[\mu_{no} T^{-3/2} + \mu_{po} T^{-3/2}\right] e}$$
$$= \frac{1}{\sqrt{A_0}\,(\mu_{no} + \mu_{po})\, e \exp(-E_{go}/2kT)}$$

∴
$$\log \rho = -\log\left[\sqrt{A_0}\,(\mu_{no} + \mu_{po})\, e\right] + \frac{E_{go}}{2kT}$$
$$= C + \frac{m}{T}$$

[where C and D are constants

$C = -\log[\sqrt{A_0}\,(\mu_{no} + \mu_{po})e$

$m = E_{go/2k}]$

This is a equation of straight line whose slope, $m = E_{go}/2k$. From Fig. 15.38, we note that the graph of material B has higher slope and hence material B has wider band gap.

(b) We know that an electron-hole pair will be created if the energy of the incident photon is equal to or higher than the band gap. The critical wavelength (λ_c) is given by

$$\lambda_c = \frac{hc}{E_g e}$$
$$= \frac{6.6 \times 10^{-34} \times 3 \times 10^{8}}{E_g \times 1.6 \times 10^{-19}} = \frac{1.24 \times 10^{-6}}{E_g} \text{ m}$$

Obviously, the material B with higher value of E_g will require a shorter wavelength of light electron-hole pair creation.

SUGGESTED READINGS

1. S.L. Kakani and K.C. Bhandari, *Electronics: Theory and Applications* (3rd Ed; 2004;), New Age Int. Pub., New Delhi – 2.
2. S.L. Kakani and C. Hemrajani, *Solid State Physics* (4th Ed; 2004;) Sultan Chand, New Delhi – 2.
3. Pierret, R.F. and K. Harutunian, *Semiconductor Device Fundamentals*, Addition-Wesley Longman, Boston (1996).
4. Hummel, R.E., *Electronic Properties of Materials*, (3rd Ed; 2000;), Springer-Verlag, Newyork.
5. S.P. Keller, (Ed.), '*Hand Book on Semiconductors'*, Vol. 3, Materials, Preparation and Properties, North Holland, Amsterdam (1980).

REVIEW QUESTIONS

1. What are semiconductors? How they differ from metals and insulators? What are their characteristic properties? [BE]
2. Why an increase in temperature of semiconductor increases its conductivity? [BE]
3. Explain the terms: *(i)* Intrinsic semiconductors and *(ii)* Extrinsic semiconductors. [AMIE]
4. What is impurity conduction in semiconductors? Explain how the presence of a small impurity in a semiconductor modifies its conduction properties.
5. Define the terms: *(i)* Electron-hole pair, *(ii) Donor* and *(iii)* Acceptor. Give examples of each with suitable materials. [AMIE]
6. Describe *N*-type and *P*-type of semiconductors. [AMIE]
7. Describe a *P*-*N* junction and transistor with neat sketches. [BE]
8. Explain the process of recombination of electrons and holes in a semiconductor. If the electron-hole pair recombine, how can their concentration remain constant at a particular temperature?
9. What are extrinsic semiconductors? How can they be formed?
10. At a high temperature an extrinsic semiconductors behaves like an intrinsic one, why?
11. Is it possible for compound semiconductors to exhibit intrinsic behaviour? Explain.
12. For each of the following pairs of semiconductors decide which will have the smaller band gap energy, E_g, and then give reason: *(a)* Carbon (C), and Ge, *(b)* AlP and InSb *(c)* GaAs and ZnSe, *(d)* ZnSe and CdTe and *(e)* CdS and NaCl.
13. *(a)* Explain why no hole is generated by the electron excitation involving a donor impurity atom.
    (b) Explain why no free electron is generated by the electron excitation involving an acceptor impurity atom.
14. Will each of the following elements act as a donor or an acceptor when added to the indicated semiconducting material? Assume that the impurity elements are substitutional.

Impurity	*Semiconductor*
N	Si
B	Ge
Zn	GaAs
S	InSb
In	CdS
As	ZnTe

PROBLEMS

1. For intrinsic gallium arsenide, the room temperature electrical conducivity is $10^{-6}(\Omega\text{-m})^{-1}$, the electron and hole mobilities are, respectively, 0.85 and 0.04 m^2/V-s. Compute the intrinsic concentration n_i at room temperature. [*Ans.* 7.0×10^{12} m^{-3}]
2. Calculate the electrical conductivity of intrinsic silicon at 150°C (423 K) [*Ans.* 0.52 $(\Omega\text{-m})^{-1}$]
3. The high-purity silicon is added $10^{23}/m^3$ arsenic atoms: *(a)* Is this material *N*-type or *P*-type? *(b)* Calculate the room temperature electrical conductivity of this material. *(c)* Compute the conductivity at 100°C (373 K). [*Ans.* (a) *N*-type (b) 1120 $(\Omega\text{-m})^{-1}$ (c) 640 $(\Omega\text{-m})^{-1}$]
4. The following electrical characteristics have been found for both intrinsic and *P*-type extrinsic indium phosphide (InP) at room temperature:

	$\sigma(\Omega\text{-}m)^{-1}$	$N(m^{-3})$	$P\ (m^{-3})$
Intrinsic	2.5×10^{-6}	3.0×10^{13}	3.0×10^{13}
Extrinsic (*N*-type)	3.6×10^{-5}	4.5×10^{14}	2.0×10^{12}

Calculate electron and hole mobilities. [*Ans.* $\mu_e = 0.50$ m^2/V-s, $\mu_h = 0.02$ m^{2}V-s]

5. Mobilities of electrons and holes in a sample of intrinsic germanium at room temperature are 3600 cm^2/V-s and 1700 cm^2/V-s respectively. If the electron and hole densities are each equal to 2.5×10^{13} per cc, calculate its conductivity. [*Ans.* 2.12 mho/m]

OBJECTIVE QUESTIONS

1. *N*-type germanium is obtained on doping intrinsic germanium by
(a) Phosphorous *(b)* Aluminium
(c) Boron *(d)* Gold

2. Depletion region is a zone which contains
(a) holes only *(b)* electrons only
(c) both electrons and holes *(d)* neither electrons nor holes

3. In a semiconductor diode arrow represents
(a) *N* type material *(b)* *P* type material
(c) both *P* and *N* type materials *(d)* none of the above

4. Zener diode is used for
(a) rectification *(b)* amplification
(c) stabilization *(d)* none of the above

5. For a tunnel diode a decrease in current causes
(a) voltage constancy *(b)* decrease in voltage
(c) increase in voltage *(d)* none of the above

6. *PN* junction is formed when *P* type semiconductor and *N* type semiconductor are joined
(a) together *(b)* physically
(c) to get homogeneous material chemically
(d) in such a manner that electrons and holes diffuse to give depletion layer

7. The depletion region of a junction diode is formed
(a) when forward bias is applied to it *(b)* when the temperature of the junction is reduced
(c) under reverse bias *(d)* during the manufacturing process

8. The width of the depletion layer of a junction
(a) is independent of applied voltage *(b)* is increased under reverse bias
(c) decreases with light doping *(d)* increases with heavy doping

9. The LED or the light emitting diode
(a) is made from one of the two basic semiconducting materials, silicon or germanium.
(b) is made from the semiconducting compound gallium arsenide phosphide.
(c) emits light when forward biased.
(d) emits light when reverse biased.

10. The *p*-side of a junction diode is earthed and the *n*-side is given a potential of –2 V. The diode will
(a) not conduct *(b)* conduct partially
(c) break down (*d*) conduct

11. For detecting light intensity we use a/an
(a) photodiode in reverse bias *(b)* photodiode in forward bias
(c) LED is a reverse bias *(d)* ED in forward bias

12. When a *P-N* junction diode is forward biased, the flow of current across the function is mainly due to
(a) diffusion of charges *(b)* drift of charges
(c) depends on the nature of the material *(d)* both drift and diffusion of charges

13. A *P-N* junction diode cannot be used
(a) as a rectifier *(b)* for increasing the amplitude of an ac signal
(c) for getting light radiation *(d)* for converting light energy into electrical energy

14. A strong electric field across a *P-N* junction that causes covalent bonds to break apart is called
(a) reverse breakdown *(b)* avalanche breakdown
(c) lever breakdown *(d)* low voltage breakdown

15. A light emitting diode produces light when
(a) forward biased *(b)* reverse biases
(c) unbiased *(d)* none of the above

16. A solar cell is an example of
(a) photo emissive cell *(b)* photo radiation cell
(c) photo voltaic cell *(d)* photo conductive cell

17. When holes leave the P-material to fill electrons in the N-material to fill electrons in the N-material the process is called
(a) diffusion *(b)* depletion
(c) avalanche breakdown *(d)* zener breakdown

18. A varacter diode is optimised for
(a) high output current *(b)* high output
(c) its variable inductance *(d)* its variable

19. A diode which has zero breakdown voltage is known as
(a) tunnel diode *(b)* Zener diode
(c) Schottky diode *(d)* backward diode

20. A Zener diode is used as
(a) a coupler *(b)* a rectifier
(c) an amplifier *(d)* a voltage regular

Answers

1. *(a)* **2.** *(d)* **3.** *(b)* **4.** *(c)* **5.** *(c)* **6.** *(d)* **7.** *(d)* **8.** *(b)* **9.** *(b)*

SHORT QUESTION ANSWERS

1. On the basis of band structure how solids are classified?

Ans. Solids are classified as metals, semiconductors and insulators according to their band structure. Metals have partially filled bands or overlapping bands. Both semiconductors and insulators have a finite forbidden gap, the semiconductors have smaller gaps in the range of 1-3 eV.

2. On what factors the ability of semiconductor depends to conduct electricity?

Ans. *(i)* Thermal excitation and *(ii)* presence of impurities.

3. On what factors the conductivity of a semiconductor depends?

Ans. *(i)* Number of charge carriers *(ii)* mobility of charge carriers and *(iii)* the electronic charge.

4. On the basis of electrical behaviour, how semiconductors are classified?

Ans. Intrinsic or extrinsic. For intrinsic behaviour, the electrical properties are inherent to the pure material and electron and hole concentrations are equal. Electrical behaviour of extrinsic of an extrinsic semiconductor is dictated by impurities.

5. What are the main uses of transistors?

Ans. Transistors are used for amplification of electrical signals, as well as for switching devices in computer circuitries.

6. Why GaAs is an important semiconducting material?

Ans. This compound has high signal speed and low power consumption as compared to silicon. Its laser function makes it ideally suited for use in *satellite communications* and supercomputers.

7. What are the functions of donor and acceptor impurities?

Ans. Donor impurities introduce excess electrons; acceptor impurities introduce excess holes.

16

Superconductivity and Superconducting Materials

1. INTRODUCTION

When cooled to sufficiently low temperatures, a large number of metals and alloys can conduct electric current without resistance. Obviously, these specific materials undergo a phase tansition to a new superconducting state characterized by the complete loss of d.c. resistance below a well defined critical temperature, T_C. Thus zero resistivity ($\rho = 0$), i.e. infinite conductivity is observed in a superconductor at all temperatures below a critical temperature ($\rho = 0$ for all $T < T_C$). However, if we pass a current higher than the critical current density J_c, superconductivity disappears. This limits the maximum current that the material can sustain and is an important problem for applications of superconducting material. For elements, the transition temperature, T_C, lies generally below 10 K. Figure 16.1 shows resistance versus temperature for a low-temperature superconductor. At the transition temperature T_C the resistance drops abruptly to an unmeasurably small value. The transition from normal to the superconducting phase is often sharp and occurs within 10^{-2} to 10^{-4} K. This behaviour is remarkably different from the steadily decreasing resistance of non superconducting metals (Fig. 16.2) and suggests the existence of a physically different superconducting state. In pure metals the zero resistance state can be reached within a temperature range of 1 mK. In the case of impure metals the transition to the superconducting state may be considerably broadened. A transition width of ≈ 0.05 K was observed for impure tin. The resistivity of a superconductor to direct current is zero as far as it can be measured. The estimates of the resistivity in the superconducting phase place it at less than 4×10^{-25} Ω-m, which is essentially zero for all practical purposes. A striking way to demonstrate zero resistivity is to induce a current around a close ring of a superconducting metal. Experiments have been performed in which a 'persistent current' has run for over two and a half year without any measurable decay. The time dependence of the current I in the loop is given by $I(t) = I_0 e^{-t/\tau}$, where I_0 is the initial value of the current and t is the time which has elapsed since the supercurrent has been induced. The ratio of resistance R and self inductance L of the superconducting loop determines the time constant τ for the decay of the current. Above the critical temperature T_C the metal is in the normal state and resistance is proportional to T^5. In many metals the exponent is between 2 and 6, considerably different from the value of 5 predicted by Bloch theory.

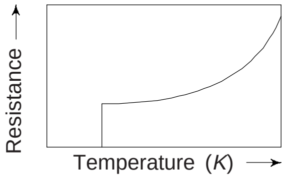


Fig. 16.1 Resistance versus temperature for a low temperature superconductor

The critical temperature, T_C varies from superconductor to superconductor but lies between less than 1 K and approximately 20 K for metals and metal alloys. Until 1986 the maximum T_C was observed in an alloy of niobium, aluminium and germanium. Recently it has been demonstrated that some complex cuprate

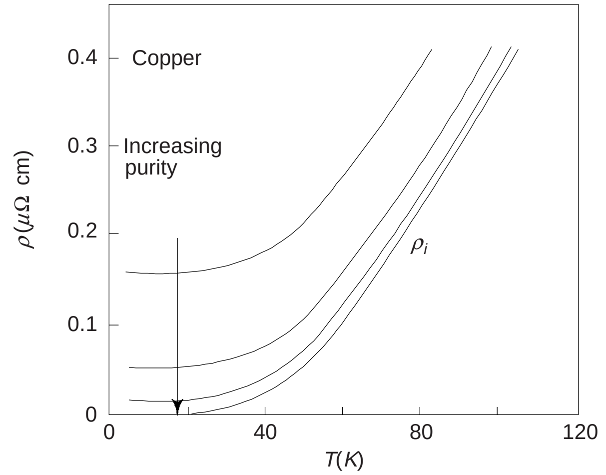


Fig. 16.2 Resistivity of copper for residual resistivity ratio (RRR) values of 10, 30 and 100. At very low temperatures the intrinsic resistivity $\rho(T)$ due to electron-phonon interaction approaches zero. RRR is defined as RRR = $\rho(273\ \text{K})/\rho(4\text{K})$

oxide ceramics have critical temperatures in excess of 100 K. Today, the highest known T_C is 133 K for mercury based cuprate oxide, $HgBa_2Ca_2Cu_3O_{8+\delta}$. When this compound is subjected to high pressure ~30 GPa, the onset of T_C increases to ~164 K. The dramatic evolution of critical temperatures that have been observed since its discovery in 1911 by K. Onnes are illustrated in Fig. 16.3 where the maximum value of T_C is plotted versus date. The superconductors with T_C < 25 K are called *conventional* or low T_C superconductors, whereas cuprate oxides and some other recently discovered sunderconductors with T_C > 25 K are termed as high temperature superconductors (HTSC).

In addition to resistanceless current transport, the superconducting state is characterized by perfect diamagnetism, i.e. $B = 0$ inside the superconductor. The magnetic inductance becomes zero inside the superconductor when it is cooled below T_C in a weak external magnetic field: the magnetic flux is expelled from the interior of the superconductor (Fig. 16.4). This effect is called the *Meissner-Ochsenfeld effect* after

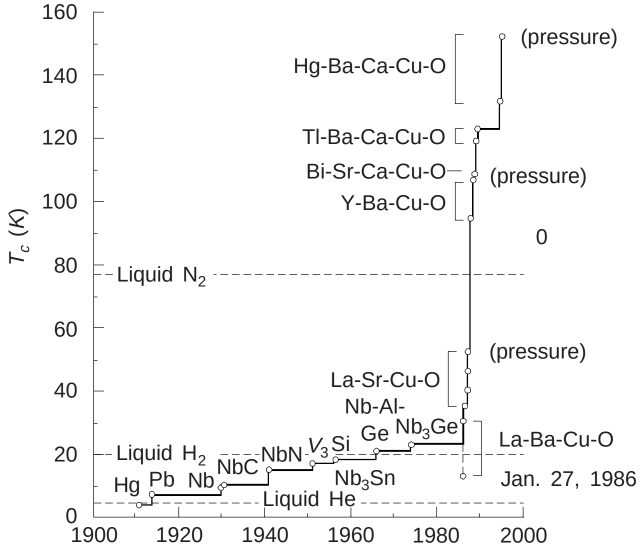


Fig. 16.3 The evolution of superconductivity critical temperature since its discovery

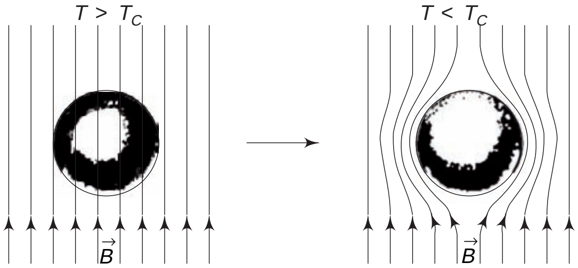


Fig. 16.4 **Explusion of a weak, external magnetic field from the interior of the superconducting material**

its discoverers and it is the ultimate practical test in any new material. We must note that there always exists some critical field, H_c, above which superconductivity disappears. Superconductivity disappears and the material returns to the normal state if one applies an external magnetic field of strength greater than H_c.

On the basis of magnetic response, superconducting materials may be divided into two classes designated as type-I an type-II. Type-I materials, while in the superconducting state, are completely diamagnetic; i.e., all of an applied magnetic field will be excluded from the body of material (Meissner effect). Several metallic elements including aluminium, lead, tin, and mercury belong to type-I group.

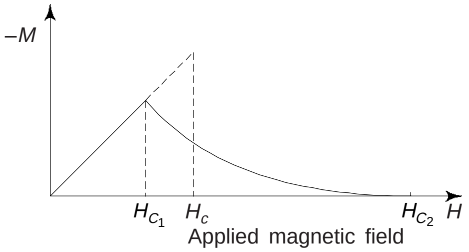


Fig. 16.4(a) **Variation of magnetization as a function of the magnetic field for (i) type-I Superconductor (dotted line) (ii) type-II ideal superconductor (solid line)**

Type-II superconductors are completely diamagnetic at low applied fields, and field exclusion is total. However, the transition from the superconducting state to the normal state is gradual and occurs between lower critical and upper critical fields, designated H_{c_1} and H_{c_2}, respectively [Fig. 16.4(a)]. The magnetic flux lines begin to penetrate into the body of the material at H_{c_1}, and with increasing applied magnetic field, this penetration continues; at H_{c_2}, field penetration is complete. For fields between H_{c_1} and H_{c_2}, the material exists in what is termed a mixed state-both normal and superconducting regions are present.

Type-II superconductors are preferred over type-I for most practical applications by virtue of their higher critical temperatures and critical magnetic fields.

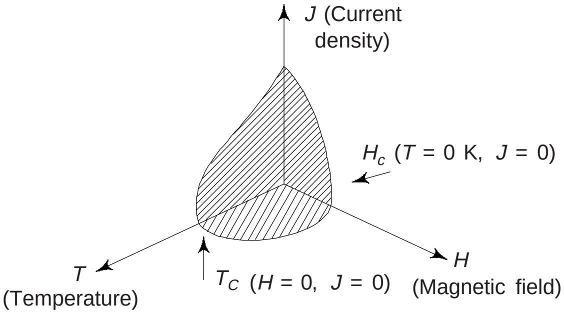


Fig. 16.5 **Critical temperature (T_c), current density (J_c) and magnetic field boundary separating superconducting and normal conducting states**

The three material parameters, T_C, H_c and J_c are of extreme importance in the practical applications of superconductivity. Figure 16.5 shows schematically the boundary in temperature magnetic field-current density space separating normal and superconducting states. The position of this boundary will, of course, depend on the material. For temperature, magnetic field, and current density values lying between the origin and this boundary, the material will be superconductive; outside the boundary, conduction is normal.

2. SUPERCONDUCTING MATERIALS

(i) Superconducting Elements: Metallic elements are mostly superconductors. Their T_C are typically of the order of a few Kelvin. Among metals, niobium exhibits the highest critical temperature of the pure elements, T_C = 9.2 K. Noble metals, copper, silver and gold, and alkaline metals, sodium and potassium, of which are excellent conductors of electricity at ambient temperatures, are not superconductors down to very low temperatures (if at all). Magnetic metallic elements do not exhibit superconductivity.

The best known semiconductors, Si and Ge, become superconductors under a pressure of ~2 K bar with T_C = 7 and 5.3 K respectively. Other elements that become superconductors under pressure include P, As, Se, Y, Sb, Te, Ba, Bi, Ce and U.

(ii) Binary Alloys and Compounds: In most alloys and compounds the critical temperatures are usually somewhat higher. Nb compounds, like Nb_3Sn, Nb_3Ge and in particular, Nb-Ti, are of technological interest. While the maximum current density one can pass through the standard water cooled copper wire at 300 K is about 2000 A-cm^2, one can pass very high current densities of upto 10^4 A-cm^2 in high magnetic fields of 10 Tesla at 4.2 K through a wire made of Nb-Ti alloy without destroying superconductivity. This enables the construction of powerful supermagnets which provide a basis for a range of large scale applications like energy storage or levitation of trains, etc.

Transition metals combined with other elements often, produce binary alloys or compounds with T_C's higher than those of starting elements. The intermetallic compounds and ordinary compounds usually exhibit the highest T_C's.

Intermetallic Compounds

Among the intermetallic superconductors, the most favourable group is the one based on the A_3B compound. In the cubic A-15 structure six binary compounds have T_C over 17 K. The highest known T_C prior to 1986, close to 23 K is obtained in Nb_3Ge stabilized by traces of oxygen or aluminium; it exhibits the upper critical field of 38 T. The A-15 structure exists in about 70 binary compounds. H_{c_2} (T) values for A-15 superconducting compounds is very high. H_{c_2}(O), the upper critical field values as the temperature approaches 0 K for A-15 compounds are also very high, e.g. 44 tesla for a composition Nb_{79} $(Al_{73}Ge_{27})_{21}$, 32 T for Nb_3Al, 39 T for Nb_3Ge, 23 T for Nb_3 Sn, 21 T for V_3Ga and 25 T for V_3Si.

(ii) Chevrel Phases: In 1971, Chevrel et al. discovered a series of ternary molybdenum chalcogenides having the general formula $M_xMO_6X_8$, where M represents any of a large number (nearly 40) of metallic elements and rare earths (RE) throughout the periodic table; X has values between 1 and 4, depending on the M element; and X is a chalcogens: sulphur (S), selenium (Se) or tellurium (Te). The highest T_c = (15 K) in the series is obtained in $PbMO_6S_8$ with unusual high H_{c_2} value (= 60 T = 600 kilogauss). This was the highest value of H_{c2} prior to the discovery of HTSC cuprates. The large values of H_{c_2} as compared with Nb_3Sn and NbTi make this material interesting for making superconducting wires. Critical currents (J_c) as high as ~3×10^{-5} A/cm^2 have been reported at 4.2 K and this provides an impetus for making wires out of these very brittle materials. It has been reported that antiferromagnetism of the rare earth can coexist with superconductivity like in Gd, Tb, Dy, Tr compounds where T_C is 1.4, 1.65, 2.1 and 1.85 K and T_N (Neel Temperature) is 0.84, 0.9, 0.4 and 0.15 K respectively. The compound $HoMo_6S_8$ exhibit 'reentrant' superconductivity phenomenon. The material is superconducting only between two critical temperatures, T_{C_1} = 2 K and T_{C_2} = 0.65 K. Below 0.65 K the material is ferromagnetic.

(iii) Tetragonal Rare Earth Rhodium Borides: This series has general formula $RERh_4B_4$. The compounds for RE = Y, Er, Tm and Lu exhibit superconductivity. The compound $ErRh_4B_4$ is a typical reentrant ferromagnetic superconductor (T_{C_1} = 8.67 K, T_{C_2} = 0.775 warming and 0.710 cooling and Curie temperature 1.2 K).

(iv) Organic Superconductors: These are a novel group of materials. The first organic superconductor $[TMTSF]_2\,PF_6$, where TMTSF denotes tetramethyltetraselenafulvalene, had a T_C of 1 K. Subsequent devel-

opments in the field have led to higher T_C organic materials. These systems were characterized by their nearly one-dimensional properties and by low carrier concentration. Later, a new series of organic materials with a two dimensional character was discovered: $(BEDT\text{-}TTF)_2$ X, where BEDT-TTF denotes bis-ethylenedithio-tetrathiafulvalene. The κ-modification of the X = Cu $(NCS)_2$ has T_C = 10.4 K and κ-(BEDT-$TTF)_2$Cu $[N(CN)_2]$ Br has T_C = 11.2 K, which is probably the highest known T_C so far observed in these materials.

(v) Rare-Earth Transition Metal Borocarbides: Superconductivity was reported recently in a series of compounds with the formula RNi_2B_2C with a maximum T_C of 16.5 K for R = Lu. These materials display both superconductivity and magnetic order and effects associated with them. Recently, intermetallic compounds YPd_5B_3C (T_C = 23 K) and $ThPd_3B_3C$ (T_C = 21 K) have been discovered.

(vi) Fullerene Superconductors: Fullerene are stable, cagelike molecules that constitute the third form of pure carbon; the other two are diamond and graphite. The archetype fullerene is C_{60}. When C_{60} is doped with alkaline metals, superconductivity is observed with T_C = 18 K, 30 K, 33 K and 40 K for K_3C_{60}, Rb_3C_{60}, $RbCs_2C_{60}$ and Cs_3C_{60} respectively. Hole doped C_{60}(for $C_{60}/CHBr_3$ with 3 to 3.5 holes per C_{60} molecule) was recently reported to be superconducting with a T_C as high as 117 K, although the nature of experiment meant that the supercurrents were confined to the surface of the C_{60} crystal, rather than probing the bulk.

(vii) Heavy Fermion Superconductors: A small number of compounds which include one Ce compound $CeCu_2Si_2$ and few uranium compounds: UBe_{13}, UPt_3, URu_2Si_2, UNi_2Al_3, UPd_2Al_3 exhibit superconductivity with $T_C < 1$ K. These compounds are characterized by enormous volumes of linear coefficient of the electronic specific heat ($C_e = \gamma T$) which can be as high as ~1 J mol^{-1} K^{-2}, and a corresponding large electron mass ($m^* \sim 10^2\text{-}10^3$ m_e, where m_e is free electron mass). Most of these systems exhibit the coexistence of superconductivity and antiferromagnetism.

(viii) Itinerant Electron Ferromagnetic Superconductors: The intermetallic compound Y_9Co_7 has been shown to exhibit an interesting interplay between very weak ferromagnetism and some form of superconductivity. The system shows ferromagnetism below 6-8 K and at a lower temperature (~3 K) superconductivity sets in. Recently, superconductivity in UGe_2 is reported below 1 K on the border of weak ferromagnetism.

(ix) Quantum Spin Ladder Materials: These materials consist of ladders made of AFM chains of S = 1/2 spins coupled by interchain AFM bonds. $SrCu_2O_3$ and $LaCuO_{2\cdot5}$ 2-leg ladder materials, whereas $Sr_2Cu_2O_5$ is a 3-leg ladder material. Superconductivity has apparently been discovered in the ladder material $Sr_{0.4}$ $Ca_{13.6}$ $Cu_{24}O_{41.84}$ under pressure with $T_C \approx 12$ K at 3 GPa.

(x) Magnesium Diboride (MgB_2): Very recently, superconductivity is reported at 39 K in the simple binary ceramic compound MgB_2. This is probably highest T_C yet determined for a non-copper oxide bulk superconductor.

(xi) Sr_2RuO_4: This superconducting compound has the same structure as the La_{2-x} M_xCuO_4 (M = Ba, Sr, Ca, Na) HTSC cuprates. While the T_C of $Sr_2RuO_4$is only ~1 K, this compound is of considerable interest because it is the only layered perovskite superconductor without copper. Several of these materials and their critical temperatures are listed in Table 16.1.

Table 16.1 Critical temperatures and magnetic fluxes for selected superconducting materials

Material	*Critical temperature T_C (K)*	*Critical magnetic flux density B_C (tesla)*
Elements		
Tungsten	0.02	0.0001
Titanium	0.40	0.0056

(*Contd.*)

Table 16.1 (*Contd.*)

Material	*Critical temperature T_C (K)*	*Critical magnetic flux density B_C (tesla)*
Aluminium	1.18	0.0105
Tin	3.72	0.0305
Mercury (α)	4.15	0.0411
Lead	7.19	0.0803
Niobium	9.5	0.2
Alloys and Compounds		
Nb-Ti alloy	10.2	12
Nb-Zr alloy	10.8	11
V_3Al	9.6	—
V_3Ga	15.4	23
V_3Si	17.1	23
V_3Ge	7	—
V_3Sn	4.3	—
Nb_3Al	18.9	33
Nb_3Ga	20.3	34
Nb_3Si	18.0	—
Nb_3Ge	23	38
Nb_3Sn	18.3	24
$SnMo_6S_8$	12	34
$PbMo_6S_8$	15	60
$LaMo_6S_8$	7	45
$SnMo_6Se_8$	4.8	—
$PbMo_6Se_8$	3.6	3.8
$LaMo_6Se_8$	11	5
Quasi one-dimensional organic superconductors $(TMTSF)_2X$		
X =	T_C	
ClO_4	1.2	
PF_6	1.2 (at 9 P_c/K bar)	
ReO_4	1.4 (at 9.5 P_c/K bar)	
Quasi bi-dimensional organic superconductor $(BEDT\text{-}TTF)_2X$		
X =	T_C	B_c
I_3 β_L	1.2	
I_3 β_H	8.1	
IBr_2 β	2.5	
AuI_2 β	4.2	
Cu $(NCS)_2$ κ	10	20
Rare-earth transition-metal borocarbide compounds		
$LuNi_2B_2C$	16.5	
YPd_5B_3C	23.0	
$ThPd_3B_3C$	21	17
Alkali-metal-doped C_{60} compounds		
K_3C_{60}	18	
Rb_3C_{60}	29	

(xii) High Temperature Cuprate Superconductors: Until 1986, most of the superconducting compounds were metals and alloys. However, some oxide superconductors were known for decades, but their transition temperatures were rather low. This was mainly due to a low number of carriers in the metallic state. Two

known exceptions were $LiTi_2O_4$ and $BaPbBiO_3$ with T_C ~13 K. This was quite unusual as their densities of carriers were also very small. The major breakthrough came in 1986 when J.G. Bednorz and K.A. Muller discovered superconductivity with a T_C (onset) of 35 K a new record in a mixed phase copper-oxide ceramic containing $La_{2-x}Ba_xCuO_4$. The end of 1986 and the beginning of 1987 was marked by synthesis of rare-earth metal oxides with the discovery of Yttrium barium copper oxide, $YBa_2Cu_3O_7$, which has a critical temperature of about 92 K. This was a significant breakthrough as it meant that for the first time the world has witnessed the existence of a superconductor with a T_C above that of liquid nitrogen (boiling point 77 K). Nitrogen is much more abundant than helium, much less expensive, and liquid nitrogen cryogenic systems using helium refrigeration. One application which could benefit from nitrogen cooling is the development of hybrid microelectronic technology (semiconductor-superconductor devices) both gallium arsenide and silicon can be tailored to perform better at liquid nitrogen temperatures.

In early 1988, Bismuth (Bi) and Thallium (Tl) based cuprates, $Bi_2Sr_2Ca_2Cu_3O_{10}$ (T_C = 110 K) and $Tl_2Ba_2Ca_2Cu_3O_{10}$ (T_C = 125 K) were discovered. These new HTSC cuprates may have some advantages over ceramic superconductors containing rare earths. Since the critical current density increases as T/T_C decreases, a T_C far above the opening temperature of liquid nitrogen (77 K) is advantageous. Moreover, the new materials are more stable than the rare earth cuprate superconductors; they do not lose oxygen or react with water.

The maximum value of T_C has now increased to 133 K for mercury based cuprate, $HgBa_2CA_2O_{8+\delta}$. When this compound is subjected to high pressure ~30 GPa, the onset of T_C increases to ~164 K (more than half way to room temperature). While $HgBa_2Ca_2O_{8+\delta}$ cannot be used in applications at such high pressures, this striking result suggests that values of T_C in the neighbourhood of 160 K, or even higher, are attainable in cuprate oxides. More than 100 different cuprate materials many of which are superconducting have been discovered since 1986. Several of the more important are listed in Table 16.2, along with the maximum values of T_C observed in each class of materials.

Table 16.2

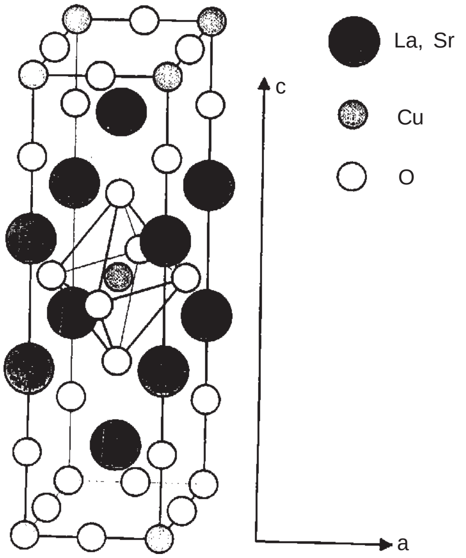
(i) Some important classes of HTSC cuprates along with the maximum value of T_C observed in each class.
(ii) Examples of the abbreviated names (nicknames) used to denote supeconducting cuprate materials.

(i) Material	*Maximum T_C (K)*
$La_{2-x}MCuO_4$; M = Ba, Sr, Ca, Na	~40
$Ln_{2-x}M_xCuO_4$; Ln = Pr, Nd, Sm, Eu; M = Ce, Th	~25
$YBa_2Cu_3O_{7-\delta}$	92
$LnBa_2Cu_3O_{7-\delta}$	~95
$RBa_2Cu_4O_8$	~80
$Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4}$ (n = 1, 2, 3, 4)	(n = 3) 110
$TlBa_2Ca_{n-1}Cu_nO_{2n+3}$ (n = 1, 2, 3, 4)	(n = 4) 122
$Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$ (n = 1, 2, 3, 4)	(n = 3) 122
$HgBa_2Ca_{n-1}Cu_nO_{2n+2}$(n = 1, 2, 3, 4)	(n = 3) 133

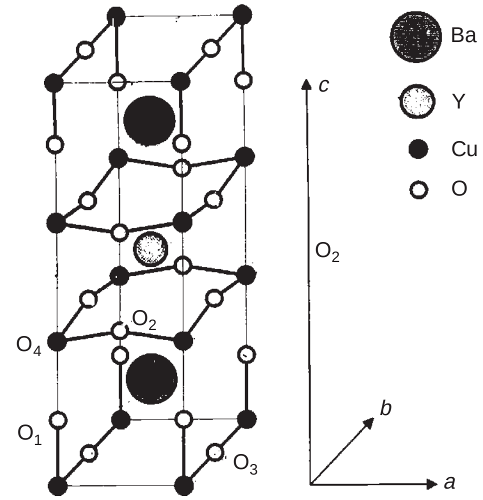
(ii) Material	*Nickname*	*T_C (K)*
$YBa_2Cu_3O_{7-\delta}$	YBCO; YBCO-123; Y-123	92
$Bi_2Sr_2Ca_2Cu_3O_{10}$	BSCCO; BSCCO-2223, Bi-2223	110
$Tl_2Ba_2Ca_2Cu_3O_{10}$	TBCCO; TBCCO-2223; Tl-2223	122
$HgBa_2Ca_2Cu_3O_8$	HBCCO; HBCCO-1223; Hg-1223	133

3. HTSC CUPRATE MATERIALS CHARACTERISTICS

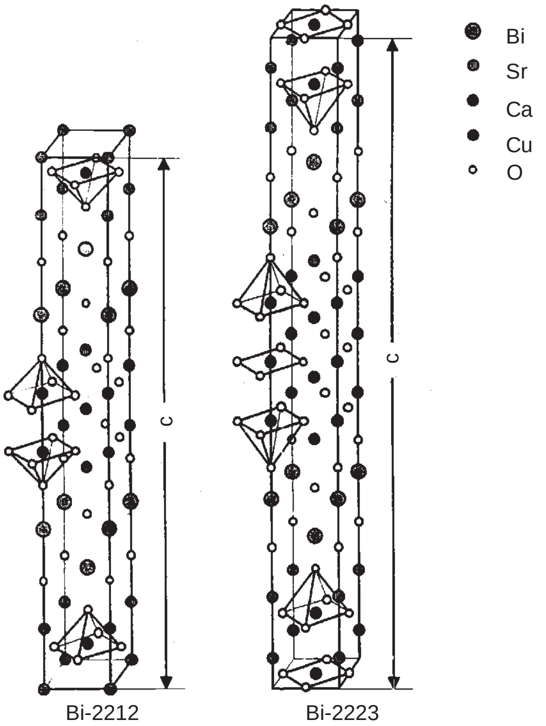
(i) Highly anisotropic, layered structures: Except for one material (isotropic, cubic $Ba_{1-x}K_xBiO_3$), all HTSC oxides are layered perovskites. One of the important characteristics of all cuprates is the presence of CuO_2 planes in their layered perovskite-like crystal structures (Fig. 16.6). The mobile charge carriers



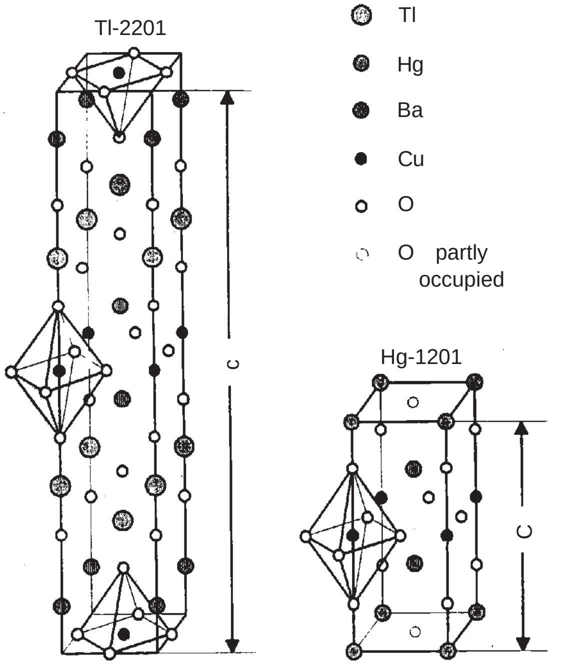
(a) Lattice structure of $La_{2-x}Sr_xCuO_4$



(b) Orthorhombic unit cell of $YBa_2Cu_3O_{7-\delta}$. A special feature of the Y-123 structure are CuO chains in *b* direction.



(c) Pseudo-tetragonal unit cells for the Bi-2212 and Bi-2223 structures



(d) Tetragonal unit cells of $Tl_2Ba_2CuO_{6+\delta}$ and $HgBa_2CuO_{4+\delta}$. The distance between neighbouring CuO_2 planes is much smaller in the Hg-1201 than in the Tl-2201 compound.

Fig. 16.6 Crystal structures of HTSC cuprate super conductors (a) Lattice structure of $La_{2-x}Sr_xCuO_4$ (b) Orthogonal unit cell of $YBa_2Cu_3O_{7-\delta}$ (c) Pseudo-tetragonal unit cells for Bi-2212 and Bi-2223 structures (d) Tetragonal unit cells of $Tl_2Ba_2CuO_{6+\delta}$ and $HgBa_2CuO_{4+\delta}$

which can be electrons but are usually holes, are believed to reside primarily within the CuO_2 planes. Obviously, the presence of CuO_2 layers dominate most properties of these systems.

(ii) Metallic Oxides: While most oxides are insulating materials, HTSC cuprates exhibit metallic behaviour. Many of these cuprates can be doped with charge carriers and rendered superconducting by substitution of appropriate elements into an insulating parent compound. For example, substitution of divalent Sr for trivalent La in the antiferromagnetic insulator La_2CuO_4 dopes the CuO_2 planes with mobile holes and produces superconductivity in $La_{2-x}Sr_xCuO_4$ with a maximum T_C of ~40 K at $x \approx 0.17$. Similarly, substitution of tetravalent Ce for trivalent Nd in the antiferromagnetic insulating compound Nd_2CuO_4 apparently dopes the CuO_2 planes with electrons, resulting in electron doped superconductivity in $Nd_{2-x}Ce_xCuO_{4-y}$ with a maximum T_C of ~25 K at $x \approx 0.15$ for $y \approx 0.02$. The temperature T versus x phase diagrams for both these systems are shown in Fig. 16.7.

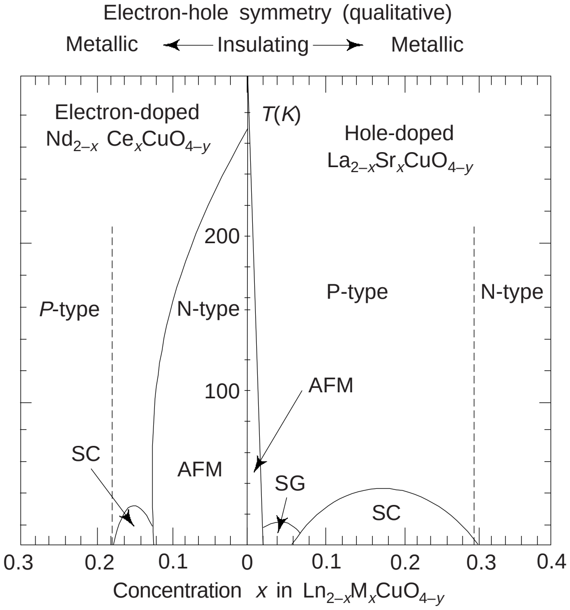


Fig. 16.7 Temperature-dopant concentration ($T-x$) phase diagram delineating the regions of superconductivity and antiferromagetic-ordering of the Cu^{2+} ions for the hole-doped $La_{2-x}Sr_xCuO_4$ and electron-doped $Nd_{2-x}Ce_xCuO_{4-y}$ systems. AFM → Antiferromagnetism SG → spin-glass phase, and SC → superconducting phase

(iii) Ceramic Materials: The original materials, $La_{2-x}Sr_xCuO_4$ and $YBa_2Cu_3O_7$, were synthesized by their discoverers as ceramic pellets. One mixes the correct ratio of constituent oxides, grinds and sinters them, make a pallet, and following a calcining procedure (at ~950°C) cools it down in oxygen. Typical ceramics, HTSC oxides also contain grains, grain boundaries, twins, voids and other imperfections. Even some of best thin films may consist of grains a few microns in diameter; all these are mostly detrimental to high critical current densities that are required for applications.

It is worthwhile to mention that even the best single crystals of HTSC cuprates often contain various defect and imperfections like oxygen vacancies, twins, impurities . . . . These imperfections are not only very relevant to their basic thermodynamic (meta) stability. We may note that various imperfections found in HTSC crystals are intrinsic to these materials.

We may note that materials science of HTSC cuprates is a non-trivial pursuit and that the understanding of phase diagrams, crystal chemistry, preparation and stability of these oxides is still very much in progress. The advancement of our understanding of these materials and appearance of applications depend very much on the advancements in material research.

4. CHARACTERISTIC PROPERTIES OF SUPERCONDUCTORS

(i) Zero Resistivity, i.e. Infinite Conductivity ($\rho = 0$ for all $T < T_C$): The d.c. (zero frequency) electrical resistance of a superconductor at all temperatures below a critical temperature T_C is practically zero (Fig. 16.8). In the first approximation, the transition is not accompanied by any change in structure of property of the crystal lattice and has been interpreted as an electronic phase transition.

The transition from the normal to the superconducting state occurs sharply in pure metals (Fig. 16.1) but not so in some impure, deformed and HTSC oxides. Bi, Tl, and Hg-based cuprate superconductors are chemically complex materials, in which there may exist several superconducting phases in one specimen. A two-step transition reflects the presence of atleast two superconducting phases. Figure 16.8 shows resistance versus temperature for a single and multi-phase high temperature superconductor. The transition width ΔT_C for single-phase high temperature superconductors is typically ≈1 K. In epitaxial $YBa_2Cu_3O_7$ films ΔT_C –values as small as 0.3 K has been achieved.

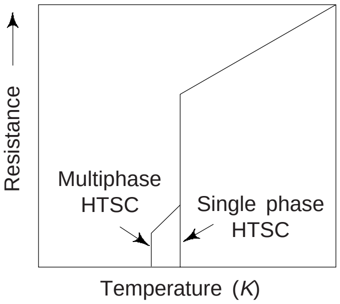


Fig. 16.8 Resistance versus temperature curves of a single – and a multi-phase high temperature superconductor

If we assume the usual Ohm's law ($V = RI$) describing the superconducting state

$$\vec{E} = \rho \vec{J} \qquad (1)$$

where $\vec{E}$ represents the electric field, ρ the resistivity and $\vec{J}$ the electric current density in the sample, then zero resistivity implies zero electric field.

So, if we take the Maxwell equation

$$\text{curl } \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

We have

$$\frac{\partial \vec{B}}{\partial t} = 0 \qquad (2)$$

For $\rho = 0$, $\vec{E} = 0$, i.e. for a superconducting state.

(ii) Meissner-Ochsenfeld Effect (B = O inside the superconductor): The magnetic inductance becomes zero inside the superconductor when it is cooled in a weak external field (Fig. 16.4). The effect is called the Meissner-Ochsenfeld effect. From Eq. (2), we see that the magnetic induction in the interior of the sample has to be constant as a function of time. The final state of the sample would have been different if it were cooled under an applied external field or if the field were applied after the sample has been cooled below T_C. In the former case the field would have remained within the sample, while in the latter it would have been zero. For the specimen to be in the same thermodynamic state, independent of the precise sequence that one uses in cooling or in applying the field, the superconducting metal always expels the field from its interior, and has

$$\vec{B} = 0$$

The superconducting state of a metal exists only in a particular range of temperature and field strength. The condition for the superconducting state to exist in the metal is that some combination of temperature and field strength should be less than a critical value. Superconductivity of the metal will disappear if the temperature of the specimen is raised above its T_C, or if a sufficiently strong magnetic is employed. There always exists some critical field H_c, above which superconductivity disappears. This field is temperature dependent (Fig. 16.9) and the empirical relation which describes well this dependence is

$$H_c(T) = H_0[1 - (T/T_C)^2] \qquad (3)$$

where H_0 is the value of H_c at absolute zero.

We have already mentioned in section 1 that the superconducting materials that completely expel magnetic flux until they become completely normal are called *type-I superconductors*. For a *type-II superconductor* there are two critical fields: the lower H_{c_1} and the upper H_{c_2} (Fig. 16.4(a)). So in applied field smaller than H_{c_1}, the type-II superconductor behaves just like a type-I superconductor below H_C. Above H_{c_1} the flux partially penetrates into the material until the upper critical field, H_{c_2}, is reached. Above H_{c_2} the material returns to the normal state. Between H_{c_1} and H_{c_2} the superconductor is said to be in a *mixed state*. For all applied fields $H_{c1} < H < H_{c_2}$, magnetic flux partially penetrates the superconducting specimen in the form of tiny microscope filaments called *vortices*. The diameter of vortex in conventional superconductors is typically 100 nm.

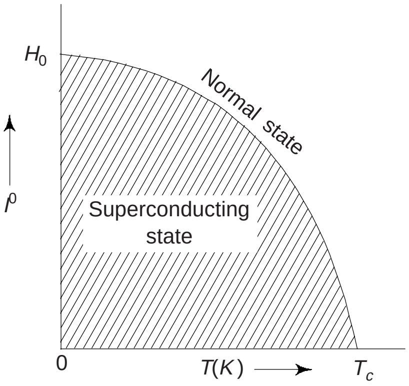


Fig. 16.9 The critical magnetic field at which superconductivity of a metal disappears

Type-II superconducting materials can withstand strong applied magnetic fields without returning to the normal state due to partial flux penetration. Superconductivity can and does persist in the mixed state upto the upper critical field, H_{c_2}, which is sometimes as high as 60 Tesla (Chevrel phases) or even ~ 150 Tesla in HTSC cuprates. At fields higher than H_{c_2} the superconductor returns to the normal state.

Practically all technologically interesting materials like Nb compounds, Chevrel phases or HTSC cuprates are all type-II superconductors. The reason for this is that the creation of vortices keeps the magnetic energy smaller than the condensation energy, so the overall free energy of the mixed superconducting state remains more favourable than the normal state even upto high magnetic fields. Since the supercurrent can flow in the mixed state through the superconducting regions between vortices, type-II superconductors allow one to construct wires needed for high field magnets.

(iii) Thermal Properties of Superconductors

(a) Entropy Entropy of all superconductors decreases considerably upon cooling below T_C. The entropy is a measure of the degree of disorder of a given system and hence this decrease in a superconductor signifies that the superconducting state is more ordered than the normal state. The fraction of electrons that is thermally excited in the normal state becomes ordered in the superconducting state. The entropy variation is relatively weak, of the order of 10^{-4} k_B/atom ($k_B \rightarrow$ Boltzmann constant).

(b) Heat capacity or Specific Heat The specific heat C_n in a normal metal consists of two contributions, C_n^e, from the electrons in the conduction band and C_n^l, from the lattice. Thus

$$C_n = C_n^e + C_n^l = \gamma T + \beta T^3 \tag{4}$$

The first term in Eq. (4) is linearly proportional to T while the second term is proportional to T^3. The specific heat of the superconductor (C_{es}). Changes at T_C in a characteristic way (Fig. 16.10). In zero magnetic field there appears a discontinuity at T_C.

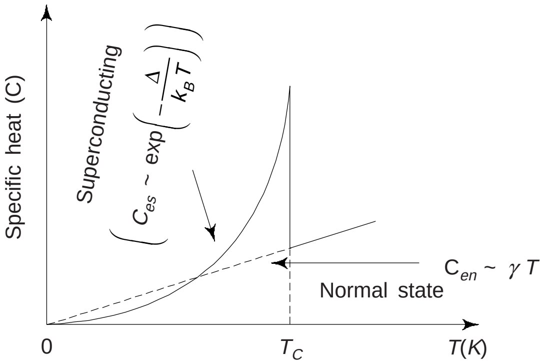


Fig. 16.10 Schematic representation of the specific heat of a metal in the normal and the superconducting phase. There is a characteristic jump of the transition point

At temperatures well below T_C the heat capacity fits an exponential form

$$C_{es}\ (T\ <\ T_C) = A \exp\ (-\Delta/K_B T) \tag{5}$$

where $E_g = 2\Delta$ is a constant for a given material, called the *energy gap*. Such an exponential temperature dependence is a hallmark of a system with a gap Δ in the spectrum of allowed energy states, separating the excited states from the ground states by energy Δ. Although there are also materials with gapless superconductivity, most materials of interest do have an energy gap. The relation between Δ and T_C is given by Bardeen-Copper-Schrieffer (BCS) theory of superconductivity.

$$2\Delta = 3.5\ K_B T_C \tag{6}$$

Heat capacity measurements provided the first indications of such a gap in superconductors, and one of the key features of BCS theory is its prediction of such a gap in superconductors.

Estimates shows that the gap in conventional superconductors with $T_C < 20$ K is of the order of 1 meV, while in HTSC oxides with $T_C \sim 100$ K, $\Delta \sim$ several (1-10) meV. We may note that in a superconductor 2 Δ corresponds to the energy needed to break a copper pair, whereas in semiconductors the gap in the energy spectrum corresponds to the energy difference between the valence and the conduction band (~1 eV).

(c) Thermal Conductivity The thermal conductivity of superconductors undergoes a continuous change between the two phases and usually lower in a superconducting phase and at very low temperatures approaches zero. This suggests that the electronic contribution drops, the superconducting electrons possibly plays no part in heat transfer. The thermal conductivity of tin ($T_C = 3.73$ K) at 2 K is 16 W $\text{cm}^{-1}\ \text{K}^{-1}$ for the superconducting phase and 34 W $\text{cm}^{-1}\text{K}^{-1}$ for the normal phase.

(d) Thermoelectric Properties A combined thermal and electric effect of interest and practical importance is the Peltier effect, which is the basis of operation of thermo couples used for temperature measurement. If the two junction regions of a loop made of two different metals are maintained at different temperature, an electrical current is driven around the loop. A greatly weakened form of this current occurs in the superconducting state, because some 'normal' electrons remain at any non-zero temperature. However, the presence of superconductivity 'shorts out' the usual thermoelectric voltage.

(iv) Acoustic Attenuation: When sound wave propagates through a metal, the microscopic electric fields due to the displacements of the ions can impart energy to the electrons near the Fermi level, thereby removing energy from the wave. One can express this by the attenuation coefficient α, of acoustic waves. The ratio of α for superconducting and normal state is given by

$$\frac{\alpha_s}{\alpha_n} = \frac{2}{1 + \exp\ (\Delta / k_B T)} \tag{7}$$

At low temperatures, one finds

$$\frac{\alpha_s}{\alpha_n} = 2\ [\exp(-\Delta/K_B T)] \tag{8}$$

(v) High Frequency Electromagnetic Properties: For all frequencies much higher than the frequency corresponding to the energy gap, $E_g = hv$, where v is the frequency in Hz, the electromagnetic response in the superconducting state is identical to the response of the normal state. Remembering that 1 eV $\sim 10^{14}$ Hz, one can easily understand that the change in the frequency response occurs at $v \sim 10^{11}$ and $\sim 10^{12}$ Hz in the conventional and HTSC oxides respectively.

High frequency electromagnetic properties of superconductors differ from zero frequency (or low frequency) behaviour. In the radio-frequeney ($<10^8$ Hz) and microwave frequency range ($\sim 10^8 - 10^{11}$ Hz) the resistance of the superconductor to current flow is not zero. However, the resistance and the accompanying energy loss are still rather small.

In the optical region of the spectrum (~10^{15} Hz), one finds no difference in electromagnetic response between the normal and the superconducting state so one does not see any change in the appearance of the sample as it undergoes the transition. In the range ~10^{11} to 10^{12} Hz (depending on the material) there is a sharp increase in the absorption of electromagnetic radiation by the superconductor. This is due to the existence of energy gap (E_g) in the electronic energy spectrum.

We may note that the sharp rise in the absorption occurs at the energy ($E = hv$) of a single photon which is just sufficient to produce an excitation (by depairing weakly coupled Cooper pairs).

HTSC oxides exhibit light excitations on the T_C values. These materials are ceramic in nature, black in colour and *p*-type (hole-type) in carriers. Though these materials are black in colour, they do not behave like black body. Their optical absorption spectra are continuous but consist of discrete bonds corresponding to different energy levels. These are responsible for additional *photoinduced* charge carriers and giving rise to change in T_C. One may call it as photodoping and superconductors as photoinduced superconductors.

(vi) The Energy Gap: We have seen that the heat capacity in the superconducting state well below T_C varies with temperature in an exponential manner in accordance with relation (5) with $\Delta = bk_B T_C$, where b is a constant. Near the transition temperature, the half width ($\Delta = E_g/2$) of the energy gap is approximately

$$\Delta(T) = 3.2\ k_B T_C\ (1 - T/T_C)^2 \qquad (9)$$

Such a temperature dependence is characteristic of a system that has an energy gap in its spectrum of allowed energy states. One can determine the energy gap of a superconductor by knowing the absorption of electromagnetic waves. One can detect this gap by photo absorption, quantum tunneling and other experiments. A plot of the temperature dependence of energy gap parameter $\Delta(T)$ is shown in Fig. 16.11. We note that the energy gap is zero at $T \rightarrow T_C$ and reaches a maximum value $\Delta_0(T)$ as the temperature lowered towards 0 K.

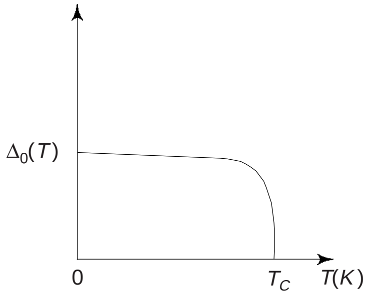


Fig. 16.11 The energy gap $\Delta(T)$ versus temperature. $\Delta(T)$ vanishers with infinite slope as $T \rightarrow T_c$ leading to the second order phase transition

(vii) Isotope Effect: It has been observed that the critical temperature (T_C) of superconductors varies with the ionic or isotopic mass. The relation valid for some simple metals is given by

$$T_C \propto M^{-\alpha}$$

or

$$M^{\alpha} T_C = \text{constant} \qquad (10)$$

where M is the atomic mass of the isotope and α is roughly 0.5. For example, for mercury, T_C varies from 4.185 K to 4.146 K as the isotopic mass M varies from 199.5 to 203.4. Obviously, the existence of the isotope effect indicated that, although superconductivity is an electronic phenomenon, it nevertheless depends in an important way on the vibrations of the crystal lattice in which the electrons move. Fortunately, not until after the development of the BCS theory was it discovered that the situation is more complicated than it had appeared. For some superconductors the exponent of M, i.e. α is not –1/2, but near zero, e.g. Ru and Zr, and for at least one it is positive.

The isotope effect, as studied by substitution of ^{18}O for ^{16}O in HTSC cuprates is very weak. The substitution of ^{18}O for ^{16}O in Eq. (10) gives $\alpha \approx 0.02$ in YBCO and $\alpha \approx 0.15$ in LBCO. This has prompted the exploration of non-phonon electronic coupling mechanisms responsible for superconductivity in these cuprate systems. However, a weak isotope effect is not conclusive.

(viii) Mechanical Effects: When a superconducting material is mechanically stressed, it is found experimentally that both T_C and H_c are slightly altered. One can easily see that many of the mechanical properties of the superconducting and normal states are thermodynamically related to the free energies of these states, and the critical field strength depends on the difference in the free energies of the two states. There is an

extremely small change in volume when a normal material becomes superconducting, and the thermal expansion coefficient and bulk modulus of elasticity must also be slightly different in the superconducting and normal states. However, the effects are extremely small.

(ix) Absence of Effects: We have seen that most of the electronic properties of a superconductor are profoundly affected by the transition to the superconducting state, while many other properties are changed very little if all. These include the mechanical and elastic properties, tensile strength, sound velocity and density among others.

(x) Characteristic Phenomenological Parameters

(a) Penetration Depth (λ) While studying Meissner effect, we mentioned that the superconductor expels a (weak) magnetic field B from its interior, i.e. $\vec{B} = 0$ in the interior of a superconductor. The finer experiments reveal that the field B penetrates into the superconductor within a very thin surface layer. Consider the boundary of a semi-infinite slab. When the external field is applied parallel to the boundary, the applied field does not suddendly drop to zero at the surface of the superconductor, but decays exponentially according to the relation

$$H(x) = H(0) \exp(-x/\lambda) \tag{11}$$

where $H(0)$ is the value of the magnetic field at the surface and λ is a characteristic length known as the penetration depth; λ is the distance for H to fall from $H(0)$ to $H(0)/e$. In most of the superconductors, λ is of the order of 500 Å. λ depends on the material and on the temperature, the latter variation being given approximately by

$$\lambda = \lambda_0 \left[1 - \left(\frac{T}{T_c}\right)^4\right]^{-1} \tag{12}$$

where λ_0 is the penetration depth at zero temperature for the particular material and is typically of order 500 Å.

If a superconducting film or filament is thinner than λ, its properties are significantly different from those of the bulk material. In particular, the value of H_c increases as thickness decreases, and the special property of type II superconductor arise from this.

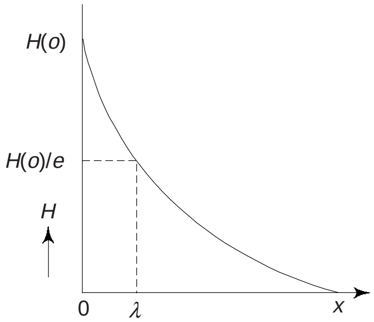


Fig. 16.12 **Decay of the magnetic field in the interior of a superconductor**

(b) The Coherence Length (ξ) It is a measure of the distance over which the gap parameter (Δ), can vary, for instance in a spatially varying magnetic field or near a superconductor-normal metal boundary. It is also referred as the distance between two electrons of the cooper pair within the highly coherent superconducting state. The intrinsic or BCS coherence length ξ_0 is defined as

$$\xi_0 = \frac{\hbar v_F}{\pi \Delta} \tag{13}$$

where v_F is the Fermi velocity and Δ is the energy gap.

Using order of magnitude values for v_F and Δ, one obtains $\xi = 16000$ Å in pure A1, $\xi = 380$ Å in pure Nb, but only about 10 Å in the new HTSC oxides. Table 16.3 provides values of λ and ξ for few selected superconductors.

When a small impurity is added to a metal λ increases very rapidly while ξ decreases.

(c) Ginzburg-Landau Parameter (κ) The ratio of two characteristic lengths, λ and ξ is called the Ginzburg-Landau ratio κ:

Table 16.3 T_c, H_{c_2}, λ and ξ for few selected superconductors

Material	T_C *(k)*	H_{c_2} *(Tesla)*	λ *(Å)*	ξ *(Å)*
Al	1.1	0.02 (Also exhibit type-I superconductivity)	500	16000
Nb	9.2	0.2	400	380
Nb-Ti	9.5	14	600	450
Nb_3Sn	18.3	24	800	35
Rb_3C_{60}	29.3	~50	1600	~20
$La_{1-x}Sr_xCuO_4$	38	~65	2500	~15
$Y\,Ba_2Cu_3O_7$	92	~120	4000	~10

$$\kappa = \frac{\lambda}{\xi} \tag{14}$$

Close to T_C, κ is independent of temperature and it allows one to distinguish between type-I and type II superconductors:

If $\kappa < 0.7$, material is type-I superconductor, and if $\kappa > 0.7$, the material is type-II superconductor. The exact critical value of κ which separates type-I from type-II behaviours is $1/\sqrt{2}$ (~0.7).

In the latter case the magnetic flux does penetrate the sample in the form of the cylindrical tubes called vortices. The vortices have a radius λ and destroy superconductivity locally within a cylinder of radius ξ. It is energetically favourable for type-II superconductors to let the flux penetrate partially in the form of vortices.

(xi) Flux Quantization: This refers to the fact that the magnetic flux threading a superconducting loop cannot have an arbitrary value; it has to be a multiple of $\phi_0 = h/2e$. The flux quantum is defined as

$$\phi = \frac{h}{2e} = 2.0678 \times 10^{-15} \text{ Weber} \tag{15}$$

Superconducting devices can measure this tiny variation of magnetic flux which is exceedingly important in metrology and advanced instrumentation.

5. JOSEPHSON EFFECTS

In 1962 B.D. Josephson predicted theoretically that Cooper pair tunneling through a very thin insulating layer (~2 nanometer) is possible. His theory predicted in addition to the Giaever current, the existence of supercurrent, arising from tunneling of the bound electron pairs. Such a junction is called *weak link*. The effects of pair tunnelling inlucde:

(i) D.C. Josephson Effect: A d.c. current flows across the junction resulting from cooper pair tunneling in the absence of any electric or magnetic field, i.e. at zero voltage across the insulating layer. One obtains the following expression for current through the contact for two identical superconductors as

$$I = I_c \sin \delta$$

where I_c is the critical current for the Josephson junction and δ is the phase difference of the two Cooper wave functions. Theoretical results for voltage drop across the junction yield

$$V_t = \frac{\pi\Delta}{2e} \tag{16}$$

For metals most commonly used as electrodes, lead and niobium, V_t is of the order of ~2.5 mV.

(ii) A.C. Josephson Effect: Josephson also predicted that if a constant non zero voltage V were maintained across the tunnel barrier, an alternating supercurrent would flow through the barrier in addition to the dc current produced by the tunneling of unpaired electrons. This effect has been utilized in a precision

determination of $\hbar/e$. Further, an r_f voltage applied with the dc voltage can then cause a dc current across the junction. The frequency of the ac Josephson current given by the Josephson relation

$$\omega = \frac{2eV}{\hbar} \tag{17}$$

where V is the dc voltage applied across the junction.

Josephson Junctions

These are sophisticated sandwitch structures of superconducting films (usually of Nb) separated by extremely thin (~10 Å) insulating oxide layers. In a suitable submicroscopic circuit they act as the fastest switching elements available today.

Josephson junction have also the lowest power consumption at their operating temperature, so they might provide the basis for the architecture of some of the fastest computers of the future. They are already used in the fastest commercially available oscilloscopes that operate at 10 GHz.

6. PROPERTIES OF HTSC OXIDES

All HTSC oxides possess the main properties of elemental or conventional superconductors, including zero resistivity, Meissner effect, flux quantization, Josephson effect, etc. In addition HTSC oxides are characterized by *(i)* highly anisotropic layered structures *(ii)* short zero temperature coherence length (ξ =10 Å), *(iii)* transport properties in the normal state significantly different from those of normal metals, with strongly anisotropic resistivity, anomalous magnetoresistance, pseudogap, etc. This is why that HTSC oxides are termed as *unconventional superconductors.* We briefly summarize the experimental results regarding HTSC oxides:

(i) All the HTSC oxides have layered crystal structure and contain two dimensioanl CuO_2 network.

(ii) The coordination of copper in Cu-O sheets in HTSC oxides is square planar.

(iii) Cu-O bonds are highly covalent with nearly 1.9 Å as the bond distance.

(iv) The normal state transport properties of HTSC oxides are highly anisotropic. The electrical resistivity ρ_{ab}(T) in the *ab*-plane of many hole-doped cuprate superconductors near optimal doping has a linear temperature dependence between T_C and high temperatures ~1000 K, with an extrapolated residual resistivity $\rho_{ab}(0)$ that is very small; i.e. $\rho_{ab}(\mathrm{T}) \approx \rho_{ab}(0) + cT$, with $\rho_{ab}(\mathrm{T}) \approx 0$ and the value of c similar within different classes of cuprate materials. However, the c-axis resistivity (ρ_c) shows a semiconducting ($\rho_c \propto T^{-1}$) dependence for less oxygenated samples and metallic ($\rho_c \propto T$) for well oxygenated samples. At room temperature, ρ_c/ρ_{ab}~ 30. The Hall coefficient, R_H is inversely proportional to T and the cotangent of the Hall angle $\theta_H = R_H/\rho$ varies as T^2, i.e. cot $(\theta_H) \equiv \sigma_{xx}/\sigma_{xy} = AT^2 + B$. The sign of R_H is positive, indicating that the charge carriers are holes. R_H for H parallel to the *ab*-plane is much smaller than for H parallel to the c-axis.

(v) Perhaps the most remarkable aspect of the normal state is the pseudogap in the charge and spin-excitation spectra of under-doped cuprates. The pseudogap has been inferred from features in various transport, magnetic, and thermal measurements including $\rho_{ab}(T)$, $R_H(T)$, and thermoelectric power, $S(T)$, etc.

(vi) Large positive Seeback coefficient, $S(T)$ is reported for $La_{2-x}Sr_xCuO_4$ and small positive value for $YBa_2Cu_3O_{7-\delta}$. Anistropy in $S(T)$ has been reported with the in-plane, $S_{ab}(T)$, showing temperature independence and $S_c(T)$ showing a linear variation with temperature.

(vii) Large anisotropy becomes evident in the appreciable difference of coherence length for the direction along the c-axis and direction perpendicular to it. The largest anisotropy is reported in the bismuth superconductor where the coherence length is ~2 Å along the c-axis, and ~40 Å in the transverse direction. For $YBa_2Cu_3O_7$ we have ξ_{ab} ~15 Å and ξ_c ~4.0 Å.

(viii) In conventional superconductors the coherence length (ξ) is ~ 10^{-4} cm. In HTSC oxides it is thousands of times smaller and equal to ~10-40 Å, which is comparable to the size of the unit cell. We may note that ξ_c is roughly equal to the interlayer distance and shorter than the corresponding unit

cell length. These remarkably short coherence lengths dominate all material related properties and cause a rather complex mixed state. Short coherence length also implies that HTSC oxides are type-II superconductors with very high H_{c_2}.

(ix) The dependence of T_C on the concentration of charge carriers has a non-monotonic character. The maximum value of T_C is attained at a relatively small density of charge carriers equal to ~10^{21} cm^{-3}. In conventional superconductors the T_C rises monotonically with rising concentration.

(x) Estimates of upper critical field $H_{c_2}^{11}$ (0) and $H_{c_2}^{\perp}$ (0) for $YBa_2Cu_3O_7$ give ~120 T and ~510 T respectively. Clearly, the upper critical fields are also anisotropic.

(xi) Energy gap measurements in the one-particle spectrum of a new superconductors by the tunneling and infrared radiation absorption and reflection methods give different values. The gap measured by tunneling method is larger. The values of $2\ \Delta\ (0)/K_BT_C$ is smaller for tunneling perpendicular to the plane and is in the range of 3.8 to 4 while parallel to the plane is in the range 5.8 to 6.0.

(xii) In HTSC oxides the magnitude of isotope effect is much lower than that predicted by BCS theory based on the electron-phonon interaction.

(xiii) The specific heat of HTSC oxides can be expressed by

$$C = AT^{-2} + \gamma T + \beta T^3 \tag{18}$$

The first term is due to Schottky-like anomaly and arises from magnetic impurities. The second term is a linear contribution similar to that found for free electrons in a metal and the third term is the contribution from the lattice by a Debye theory of solids. We note that there is a absence of the exponential dependence of the electronic specific heat expected on the BCS theory. This is a measure departure from the properties of the conventional superconductors. It is now believed that the linear term in (18) arise from the presence of impurities.

T_C, λ, ξ, H_{c_2} and H_{c_1} for few selected HTSC cuprates are given in Table 16.4.

Table 16.4 Critical temperature (T_C), penetration depth (λ_i), coherence length (ξ_i) and the critical fields B_{c_1} and B_{c_2} of few HTSC cuprates (i = ab or c)

HTSC cuprates	$T_C(K)$	$\lambda_{ab}(Å)$	$\lambda_c(Å)$	$\xi_{ab}(Å)$	B_{c_2}/T	B_{c_1}/T
LBCO	38	800	4500	35	80	15
YBCO	94	1500	7000	15	150	40
BSCCO	110	2500	9000	13	260	32

7. THERMODYNAMICS OF A SUPERCONDUCTOR

The transition between the normal and superconducting state is thermodynamically reversible. This change occurs because the Gibbs free energy in the superconducting state, G_S, is lower than its value G_N in the normal state. Gibbs free energy (G) is defined as

$$G(T,p) = U - TS + pV \tag{19}$$

Where U, T, S, p and V are internal energy, temperature, entropy, pressure and volume of the system respectively.

Let us consider that a superconductor is placed in a magnetic field, H_a. In analogy with Eq. (19), the Gibbs free energy in this case can be expressed by adding the magnetic contribution to the energy density of a magnetized specimen. If susceptibility, χ = is independent of H, the magnetic energy is

$$U_M = -\int_0^{H_0} \mu_0 H dM = -\int_0^{H_0} \mu_0 H\chi dH = -\frac{1}{2}\mu_0\, xH_a^2 \tag{20}$$

Now, neglecting the term pV in Eq. (19) for a magnetic solid and adding the magnetic energy term of Eq. (20), one obtains

$$G(T,\ H_a) = U - TS\ - \frac{1}{2}\ \mu_0 \chi H_a^2 \tag{21}$$

In the superconducting state, $\chi = -1$, and hence

$$G_S\,(T,\ H_a) = G_S\,(T,\ 0) +\ \frac{1}{2}\ \mu_0\, H_a^2 \tag{22}$$

In the normal state χ is very small and we may assume $\chi = 0$. Thus

$$G_N\ (T,\ H_a) = G_N(T,\ 0) \tag{23}$$

Normal and superconducting phases are at equilibrium at $H_a = H_c$. This means

$$G_S(T,\ H_c) = G_N(T,\ H_c) = \ G_N(T,\ 0)$$

$$= G_S(T,\ 0) +\ \frac{1}{2}\ \mu_0\, H_c^2\,(0) \tag{24}$$

or $$G_N(T,\ 0) -\ G_S(T,\ 0) =\ \frac{1}{2}\ \mu_0\, H_c^2 \tag{24a}$$

From Eq.s (22), (23) and (24), it is obvious that for $H_a > H_c$, $G_S > G_N$ while for $H_a < H_c$, $G_S < G_N$, and we note that the superconducting state is stable for $H_a < H_c$. Figure 16.13 shows free energy vs applied field. We can see that the transition from the superconducting state to the normal state for $H_a > H_c$ occurs because G_S exceeds G_N at this magnetic field.

From Fig. 16.13, we note that the normal and superconducting phases are in equilibrium along the critical field. For all the points along this curve, Eq. (24) is satisfied and hence for these points,

$$dG_N\ (T,\ H_C) = dG_S(T,\ H_c)$$

or $$\left(\frac{\partial G_N}{dT}\right)_{H_c}\ dT + \left(\frac{\partial G_N}{\partial H_c}\right)_T\ dH_c\ =\ \left(\frac{\partial G_N}{\partial T}\right)_{H_c}\ dT + \left(\frac{\partial G_s}{\partial H_c}\right)_T\ dH_c \tag{25}$$

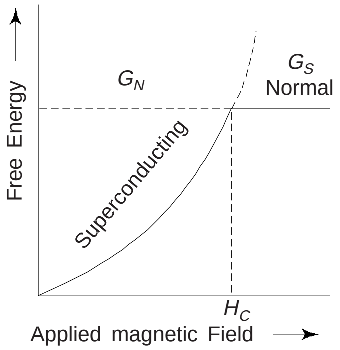


Fig. 16.13 Free energy vs. applied field

We have $G = U - TS + \frac{1}{2}\ \mu_0 H_a^2$. When H_a is constant,

$$dG = dU - TdS - SdT \text{ and } dU\ = TdS = C_v\, dT.$$

Thus

$$S = -\left(\frac{\partial G}{\partial T}\right)_{H_c} \tag{26}$$

and $$C_v^e\ = T\ \left(\frac{\partial S}{\partial T}\right)_{H_c} \tag{27}$$

The relation between entropy in the normal state, S_N, and that in the superconductivity state, S_S is as

$$S_N - S_S = -\frac{1}{2}\ \mu_0 \frac{d}{dT}\ \ H_c^2$$

or $$S_N = S_S - \mu_0\ \left(H_c\ \frac{dH_c}{dT}\right) \tag{28}$$

From Eq. (3)

$$\frac{dH_c\,(T)}{dT} = -2H_c(0)\;\frac{T}{T_c^2} < 0 \tag{29}$$

Sine H_c (dH_c/dT) is always negative and hence S_S is lower than S_N. Obviously, the superconducting state is more ordered than the normal state.

Using (27) and (28), one obtains the difference in specific heats C_N and C_S in the normal and superconducting states respectively as

$$C_N - C_S = -T\mu_0\left[\left(\frac{dH_c}{dT}\right)^2 + H_c\,\frac{d^2\,H_c}{dT^2}\right]$$

$$= \frac{2\,H_c^2\,(0)\,\mu_0}{T_c}\left[\frac{T}{T_c} - 3\left(\frac{T}{T_c}\right)^3\right] \tag{30}$$

Specific heat in the superconducting state has no linear term, as such we may identify

$$C_N = \gamma T = \frac{2\,H_c^2\,(0)}{T_c^2}\;\mu_0 T \tag{34}$$

with

$$\gamma = \frac{2\,H_c^2\,(0)}{T_c^2}\,\mu_0 \tag{35}$$

Thus

$$C_S = \frac{6\,H_c^2\,(0)\,\mu_0}{T_c}\left(\frac{T}{T_c}\right)^3 = 3\gamma T_C\left(\frac{T}{T_c}\right)^3 \tag{36}$$

Relation (36) explains the experimental result on the specific heat of the conventional superconductors.

8. THEORY OF SUPERCONDUCTIVITY

One of the simplest idea which was put forward in 1934 for describing superconductivity was the *two-fluid model*. Some properties can be understood with the simple assumption that some electrons behave in the normal way as nearly free electrons, while other exhibit anomalous behaviour. Developing this idea *F. and H. London* were able to describe the electrodynamics of what is now called type-II superconductors. London equation which replaces Ohm's law in a superconductor is as

$$\text{Curl}\;\vec{J}(\vec{r}) + \frac{\mu_0\,n_s\,e^2}{m}\,h(\vec{r}) = 0 \tag{37}$$

where $\vec{J}$ is a superconducting current density defined as

$$\vec{J} = n_s\,e\,\vec{v}_s \tag{38}$$

n_s is the fraction of the total density n of electrons which behaves in an abnormal way and $\vec{v}_s$ is their velocity.

In the gauge div $\vec{A} = 0$, we have

$$\vec{J}(\vec{r}) = -\frac{1}{\mu_0\,\lambda^2}\,\vec{A}(\vec{r}) \tag{39}$$

where $\vec{A}$ is vector potential and $\lambda = \sqrt{\dfrac{m}{\mu_0\,n_s\,e}}$ is the London penetration depth.

The London theory is not applicable to situations in which the number of superconducting electrons, n_s, varies; it does not link n_s with the applied voltage or current.

Ginzburg and Landau (GL) in 1950, using the Landau's theory of second order phase transitions, proposed a highly innovative phenomenological theory of superconductivity: *The Ginzburg-Landau Theory*. They first derive two equations which can be used to calculate both the distribution of fields and the variation of the number of superconducting electrons. This theory describes a superconductor close to T_C. The order parameter $\psi(\vec{r}) = |\psi| \exp i\phi(\vec{r})$ is a complex scalar and $\psi(\vec{r})$ is small. They introduce two unknown parameters α and β which can be estimated from experimental measurements. The characteristic quantities of the materials: λ, ξ or H_c, α and β can be eliminated and replaced by two of the measured characteristic quantities (λ, ξ or H_c). While both the characteristic lengths λ and ξ diverge at T_C, their ratio $\kappa = \lambda/\xi$, which is called the GL parameter, does not depend on temperature when close to T_C. If $\kappa < 1/\sqrt{2}$, one is dealing with type-I superconductor and if $\kappa > 1/\sqrt{2}$ one has a type-II superconductors. As flux expulsion is incomplete in type II superconductor, the mixed state exists between H_{c_1} and H_{c_2}. One of the greatest success fo GL theory was the prediction of the existence of type-II superconductors. An interesting relation can be obtained for the H_{c_1} H_{c_2} as

$$H_{c_1} H_{c_2} = H_c^2 \log \kappa \tag{40}$$

GL theory, although originally phenomenological, proved to be exact and very powerful.

Microscopic BCS Theory

The microscopic theory of superconductivity was formulated by Bardeen, Cooper and Schrieffer (BCS theory) in 1957, almost five decades after the discovery of the phenomenon in 1911 by Kammerlingh Onnes. It is an elegant but mathematically complex theory. We therefore present a brief discussion of the results of this theory.

The first clue to the basic interaction between electrons which gives rise to superconductivity was provided by the isotope effect. Two different isotopes of the same metal exhibit different T_C's (Eq. 10). Why is the mass of an atom involved in a purely electronic property? Obviously, the motion of ions has something to do with superconductivity.

The second clue was found by Leon Cooper. He showed that a normal metal (with standard metallic properties) could not be formed if there was a small attraction between electrons. In such a case, two electrons would form pairs however small the attractive interaction. And, if electrons did form pairs, completely different properties for the whole ensemble of electrons would be observed.

Thus the central idea is that motion of ions can lead to attractive interaction between electrons. How can this happen? When an electron moves among the positive ions of lattice in a solid, it interacts with the lattice vibrations (called *phonons*) and attracts them. However, the motion of ions is slow so the electron advances a great deal while the ions somewhat converge towards each other. They build a region of positive charge, which before relaxing attracts another electron. So, due to this slow response of ions, there appears an effective attractive interaction between electrons, called the *electron-phonon interaction* (Fig. 16.14). Of course the repulsive Coulomb interaction is also present and there is a delicate 'competition' between the Coulomb interaction and electron-phonon interaction which makes some metals superconducting and other non-superconducting.

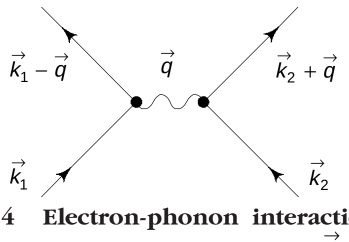


Fig. 16.14 **Electron-phonon interaction. An electron of wave vector $\vec{k}_1$ emits a virtual phonon q which is absorbed by electron $\vec{k}_2$. $\vec{k}_1$ is thus scattered as $\vec{k}_1 + \vec{q}$ and $\vec{k}_2$ as $\vec{k}_2 + \vec{q}$. If this phonon energy exceeds electronic energy, the electron-phonon interaction is attractive**

Cooper in 1956 showed that two electrons with an attractive interaction can bind together to form a 'bound pair' (often called a *Cooper pair)* if they are in the presence of a high density fluid of other electrons, no matter how weak the interaction is. The two partners of Cooper pair have opposite momenta ($\vec{K}$) and spin angular momenta, i.e. Cooper pair composed of particle pair $\vec{K}\uparrow$ and $-\vec{K}\downarrow$. This pairing is known as *s*-wave pairing. There are other forms of particle pairing possible, e.g. *d* wave, *d* + *s* wave, *p* wave etc. with the BCS theory. It is now established that superconducting gap in HTSC oxides is of $d_{x^2-y^2}$ symmetry. These electrons due to their charge also repel each other with an Coulomb interaction, V_c but BCS assume that in the superconducting phase electron-phonon mediated interaction V_{ph} dominates over V_C and net interaction V is positive, i.e.,

$$V = -V_{ph} + V_C < 0 \tag{41}$$

The main assumptions of BCS theory are

(i) The superconducting ground state can be expressed in terms of Cooper pairs so that the states ($\vec{K}\uparrow$, $-\vec{K}\downarrow$) are simultaneously occupied or empty.

(ii) Various interactions in the normal and superconducting states are identical and only the effective screened Coulomb interaction has to be considered.

(iii) The effective interaction is zero, except when two electrons of wave vectors $\vec{K}$ and $\vec{K}'$ have energies close to the Fermi energy. Then the attractive interaction is taken as a constant $-V$. If ξ_k is the energy measured from the Fermi energy, or more precisely from the chemical potential μ, then:

$$\xi_k = \varepsilon_k - \mu \tag{42}$$

In order for the electrons to attract each other, the energies of both the electrons have to satisfy the criterion

$$|\xi_k| \text{ and } |\xi_k| < k_B\theta_D \tag{43}$$

where θ_D is the Debye temperature.

BCS in 1957 showed how to construct a wave function in which all of the electrons (at least, all of important ones) are paired. Once this wave function is adjusted to minimize the free energy, it can be used as the basis for a complete microscopic theory of superconductivity.

At zero temperature, electrons in a superconductor are paired. In the superconducting state excitations are obtained by breaking up pairs which costs a minimum energy 2 Δ (T). The excitation spectrum is given by

$$E_k = \sqrt{\xi_k^2 + \Delta^2} \tag{44}$$

where Δ is a fundamental quantity introduced by BCS is called the gap or the superconducting order parameter. Δ depends on the temperature and obeys the relation, known as the *self-consistent BCS equation.*

$$\frac{1}{VN(E_F)} = \int_0^{\hbar\omega_D} d\xi(\xi^2+\Delta^2)^{-1/2} \tan h\left[\frac{1}{2k_B T}(\xi^2+\Delta^2)^{1/2}\right] \tag{45}$$

where $\hbar\,\omega_D$ is the characteristic energy of the attractive potential and $N(E_F)$ is the density of states at Fermi level.

When Δ (T) = 0, i.e., for $T = T_C$ the energy of the normal and the superconducting states are equal. An important BCS relation which relates Δ (0) and T_C is

$$2\Delta(0) = 3.52\ \mathrm{k_B T_C} \tag{46}$$

Close to T_C we have an approximate relation

$$\frac{\Delta(T)}{\Delta(0)} = 1.74\left(1-\frac{T}{T_C}\right)^{1/2} \tag{47}$$

The critical temperature (T_C) of simple conventional superconductors is given by the well-known BCS formula

$$k_B T_C = 1.13\ \hbar\omega_D \exp\left[\frac{-1}{\lambda_{ep}}\right] \tag{48}$$

λ_{ep} is the dimensionless electron-phonon coupling parameter. Its value for conventional superconductors is very close to ~ 0.3. ω_D is Debye or characteristic frequency, varies from one metal to another but only over a small range of value. Instead of ω_D, one can use the Debye temperature θ_D ($\because k_B\theta_D \approx \hbar\,\omega_D$), θ_D ranges from 100 K to 500 K. Such a range of θ_D (and λ_{ep} ~ 0.3) implies a maximum 'BCS' value of T_C ~25 K.

The extension of BCS weak coupling theory to strong coupling was presented by Eliashberg in 1960. MacMillan in 1968 obtained an approximate solution of Eliashberg's equations, good for $\lambda \le 2$ .BCS theory alone cannot explain the anomalous behaviour of HTSC oxides in normal and superconducting states. Several theoretical proposals have been advanced, e.g. Anderson's Resonance Valence Bond (RUB) theory; *t-J* model, interlayer tunneling model, Spin-Fermion model, Varma's Marginal Fermi Liquid Model, SO(2) and SO(5) symmetry models, Boson-Fermion model. None of these models provide satisfactory explanation of the anomalous features of HTSC oxides. However, the efforts to understand the HTSC oxides pairing mechanism, with enhancing future prospects for new HTSC materials and novel applications are in progress.

9. QUANTUM TUNNELING

This is a quantum mechanical process which permits electrons to penetrate from one side to the other through an extremely thin potential barrier to electron flow. Tunneling had been considered to be a possible electron transport mechanism between metal electrodes separated by either a narrow vacuum or a thin insulating film usually made of metal oxides. In 1960, Giaever demonstrates for the first time that, if one or both of the metals were in a superconducting state, the current-voltage curve in such metal tunnel junctions revealed many details of that state. Giaever's technique was sensitive enough to measure the most important feature of the BCS theory-the energy gap which forms when the electrons condense into correlated bound pairs, called as Cooper pairs.

A thin film of a superconductor, is deposited on a smooth glass surface by evaporation. The surface of the layer is coated by an insulator layer of thickness 10 nm. In case of aluminium, this can be achieved by oxidizing the layer first deposited. Let us suppose that a second metal coating (normal metal) is made on the top of the insulating layer. Electrodes be connected to the metal states on either side of the insulating layer and then a gradually increasing potential is applied. This will raise the Fermi level on the right side. Until a certain critical voltage V_c is attained, there will be no current in circuit. The current begins to grow afterwards as shown in Fig. 16.16. One can explain this on the basis of quantum tunneling. We know that quantum mechanically an electron on one side of the potential barrier has a finite probability of tunneling through it, if there is an allowed state of equal energy is available on the other side of the potential barrier. The density of states function in the energy space for a sandwich consisting of a superconductor, an insulator and a normal metal (all at absolute zero) is shown in Fig. 16.15. We know that in a normal conductor the electrons fill up to the Fermi level.

The Fermi level in the superconductor is in the middle of the band gap. When the sandwich is formed the Fermi levels are aligned (Fig. 16.15). Let us suppose that a voltage is applied across the barrier so as to raise the Fermi level on the right side of the junction. Until the Fermi level E_{F_2} on the right side of the junction raised upto the level E_1' electrons cannot move from right to the left as there are no available energy levels on the left side; since in that portion the band gap exist. Obviously, once the Fermi level E_{F_2} rises above the level of E_1', electrons can easily tunnel through the barrier from the right to the left since there are energy states available for the electrons on the left band side. Thus the current increases. However, there will be no current until the voltage becomes available equal to the band gap ($E_1' - E_2$). This has been confirmed experimentally.

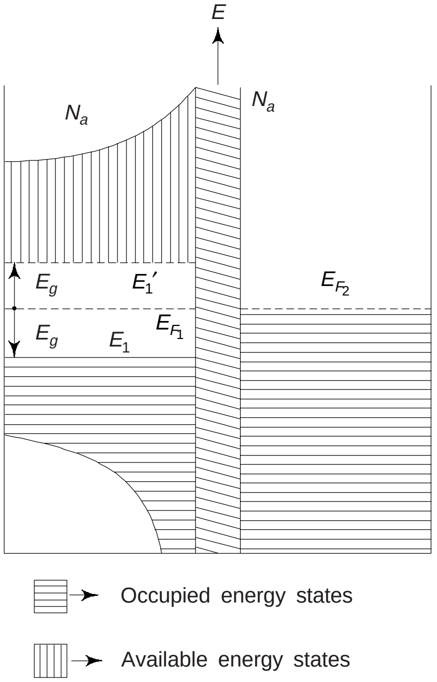


Fig. 16.15 Superconductor-oxide normal junction at 0 K

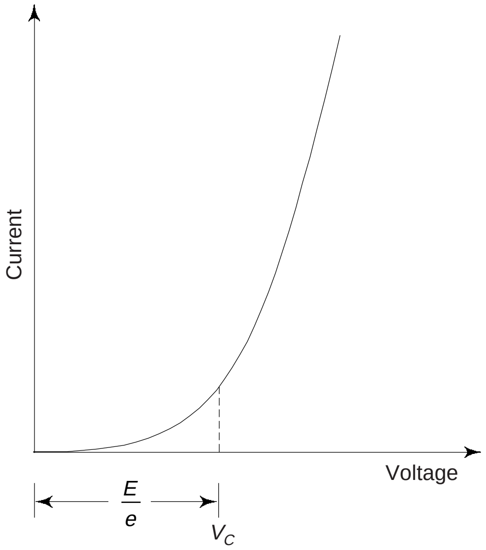


Fig. 16.16 Current (*I*) – Voltage (*V*) characteristic curve for Giaever tunneling

The tunneling phenomena has been exploited in many fields. For example, small area tunnel-junctions are used for mixing and synthesis of frequencies ranging from dc to the infrared region of the spectrum. This leads to absolute measurement of frequency in the infrared and provided the most accurate determination of the speed of light.

To study the non equilibrium superconducting properties, two tunnel junctions, one on the top of the other sharing the middle electrode, are used. Tunnel diodes are also used as a spectroscopic tool to study the phonon and plasmon spectra of the metals and the vibrational spectra of complex organic molecules introduced inside the insulating barriers.

10. APPLICATIONS OF SUPERCONDUCTIVITY

Thoughts concerning its technological applications are as old as superconductivity itself. Onnes himself tried to explore the potential of zero resistivity state for electrical applications. However, when he tried to build electromagnetic coils made with wires of pure superconducting metals, like lead or indium, he was unable to do so. These metallic superconductors easily lose their superconducting properties in weak external magnetic fields ($B < 0.1$ T). Furthermore, they were able to carry only weak electric currents. Soon it was realized that superconductors of interest to strong current applications, which can sustain very high magnetic fields, are usually fairly complex (Type- II) materials, like Nb compounds, Chevrel phases, or cuprate oxides.

Superconducting magnets of every desired size and geometry have now been planned and already manufactured. For equal power and uppermost field, the superconducting magnet is much smaller than its normal conunterpart made out of copper wires. Superconducting magnets are employed not just in scientific research, in high energy and solid state physics, as was the case in 1960's. They are already in planning for controlled thermonuclear fusion, for superconducting motors with output of several kilowatts, for energy storage facilities and for the magnetic suspension of trains (Levitation trains). Powerful superconducting magnets, *supermagnets*, are also used in medicine for diagnostics, for e.g. in nuclear spin tomographs. Superconducting magnets generate homogeneous, low noise and extremely high magnetic fields. A number of manufacturers now sell 20-T superconducting magnets with bore size in the range of 50 nm. Efforts are in process to design construct solenoidal magnets having fields approaching 25 T using new superconducting materials, including HTSC oxides.

Superconducting magnets seem more complicated than electromagnetic, especially because of their requirement of low temperature to keep the magnetic solenoid in its superconducting state.

Superconducting magnet technology allows users to produce extremely high magnetic fields without the kilowatt or megawatt power supplies needed for electromagnets. Once superconducting magnets are energized, users can easily disconnect them from their power source and they will remain energized which significantly reduces electricity costs.

However, the consideration of superconductivity for large scale applications is not limited to superconducting magnets. A great deal of effort is going into using superconductivity for the generation and transmission of electric power. Superconducting cables and storage elements will offer crucial advantages.

The application of superconductors in *measurement technology* is not revolutionary; here it opens up for us the opportunity to increase the sensitivity of many determinations by orders of magnitude over what could be achieved with normal conducting circuitary. Superconducting circuit and memory elements could bring decisive improvements to large computers, under certain conditions.

Thin film applications in electronics, together with superconducting magnets, are usually considered the most important area of superconductivity-based technology. Most of these applications are based on Josephson effect that enables the construction of the fastest nanoscopic switches, Josephson junctions, and related device structures, SQUIDs (Superconducting Quantum interference devices). SQUIDs explore subtle quantum interference effects: an analysis of superconducting loop shows that the magnetic flux that can thread the loop is quantized in quantum units of flux which have a value of 2×10^{-15} Weber. SQUIDs are suitably processed superconducting loops which detect minute changes in magnetic flux, i.e. they are high sensitivity magnetic flux detectors that can be used in the finest precision instruments at the forefront of meterology.

The HTSC oxides provide us with superconductors which can even be employed at the temperature of liquid nitrogen (T_B = 77 K) or even higher temperature. While the widespread use of HTSC cuprate superconductors in technology has not yet been realized, steady and significant progress has been made towards this objective during the past 15 years. Recent developments indicate that HTSC cuprates will begin to have a significant impact on technology during the next 5 to 10 years. The applications in superconducting electronics that are likely to be realized on this time scale include SQUIDs, NMR (nuclear magnetic resonance) coils, wireless communications subsystems, MRI cols and NMR microscopes, and ditigal instruments. In the area of superconducting wires and tapes, applications that appear to be feasible within this same time period include: power transmission lines, motors and generators, transformers, current limiters, magnetic energy storage, magnetic separation, research magnetic systems and current leads.

An example of recent progress in the area of superconducting wires and tapes is the development of flexible superconducting ribbons consisting of deposits of YBCO on textured substrates which have critical-current densities $J_c \sim 10^6$ A/cm^2 in fields up to 8 T at 64 K, a temperature that can be achieved by pumping on liquid nitrogen. The performance of these prototype conductors in strong magnetic fields already surpasses that of NbTi and Nb_3Sn, which are currently used in commercial superconducting wires at liquid helium temperatures in a comparable field range.

It is certain that HTSC oxides have opened new opportunities for technological applications of superconductivity. It is also certain, however, that a great deal of development will be necessary before we have in our hands suitable conductors for use in say magnet construction or for microcircuitry. HTSC oxides will have some impact in electronics related applications: initially in passive microwave devices and SQUIDs. The mixtures of HTSC oxides with silver may offer somewhat improved properties and might provide the basis for future wires and cables. It is economic questions that will play a crucial role here.

A brief account of potential applications of superconductivity in vital areas is given in Table 16.5.

Table 16.5 Potential applications of superconductivity in vital areas

Field	*Applications*
• Magnets	• High field Magnets • Magnetic Levitation • Magnetic Shielding • NMR (Magnetic diagnostics and spectroscopy) • Large Machines (colliders, Magnetic fusion confinement, r.f. cavities) • Ore refining
• Energy or Power-related	• Electric power storage and transmission • Magnetic energy storage • Power production by magnetic fusion and magnetohydrodynamics (MHD)
• Transportation	• High speed maglev trains • ship-drive system
• Electronic and other small devices	• SQUIDs • Bolometer • Electromagnetic shielding • Josephson devices (square-law detector, parameteric amplifier, mixer)
Computer and information processing	• Optoelectronics • Voltage standard • Active superconducting elements (FETs) • Malched filters • Semiconductor-superconductor hybrids (A-D converters)

The requirements for critical current density (J_c) in a given magnetic field for few selected applications are presented in Table 16.6.

Table 16.6 Requirement for J_c in a given magnetic field for several applications of superconductors

Application	*H (Tesla)*	J_c *(A/cm²)*
Interconnects	0.1	5×10^6
ac transmission lines	0.2	10^5
dc transmission lines	0.2	2×10^4
SQUIDs	0.1	2×10^2
Motors and generators	~4	$\sim 10^4$
Fault current limiters	>5	$>10^5$
Power generation	5-7	2×10^4
Power storage	5-10	10^3-10^4
MRI scanners	2	10^4
Mineral separation	2-5	10^4

We may not that most high current applications of superconductors require $J_c > 10^5$ A/cm^2 and $T_{\text{use}} \sim T_c/2$ and for weak current and most applications in electronics,

$$T_{\text{use}} \sim \frac{2}{3}\ T_c$$

T_c, upper critical field (H_{c_2}), λ and ξ of some technologically important superconductors are given in Table 16.7.

Table 16.7

Material	T_c (*k*)	*Penetration depth* λ *(nm)*	*Coherence length* ξ *(nm)*	H_{c_2} *(T)*
Nb	9.2	40	38	0.2
Nb-Ti	9.2	60	40	14
NbN	16	250	4	16
Nb_3Sn	18	80	3	24
$YBa_2Cu_3O_7$	92	150/600	1.5/0.4	150/40
$Bi_2Sr_2Ca_2Cu_3O_{10}$	110	200/1000	1.4/0.2	250/30

Example 1 *The critical temperature of mercury is 4.2 K.*

(a) *Calculate the energy gap in electron volts at T = 0?*

(b) *Calculate the wavelength of a photon whose energy is just sufficient to break up Cooper pairs in mercury at T = 0. In what region of the electromagnetic spectrum are such photons found?*

Solution

(a) The Cooper pair binding energy, or gap energy, is

$$E_g = 3\ kT_c$$
$$= 3 \times 1.4 \times 10^{-23}\ \text{J/°K} \times 4.2\ \text{K}$$
$$= 1.8 \times 10^{-22}\ \text{J} \simeq 1.1 \times 10^{-3}\ \text{eV}$$

(b) $E_g = h\nu = hc/\lambda$

$$\therefore \qquad \lambda = \frac{hc}{E_g} = \frac{6.6 \times 10^{-34}\ \text{J-s} \times 3 \times 10^{8}\ \text{m/s}}{1.8 \times 10^{-22}\ \text{J}} = 1.1 \times 10^{-3}\ \text{m}$$

Obviously, these photons are in the very short wavelength part of the microwave region.

Example 2 *Does the metal look like a superconductor to electromagnetic waves having wavelengths shorter than that found in example 1(b)?*

Solution No, since the energy content of shorter wavelength photons is sufficiently high to break up the Cooper pairs, or excite the conduction electrons through the energy gap into the non-superconducting states above the gap.

SUGGESTED READINGS

1. S.L. Kakani, Superconductivity: Current Topics, Arihant Pub. (opp. Rajasthan Univ.) Jaipur-4
2. S.L. Kakani, Superconducitivity: Key Problems, Arihant Pub, Jaipur-4
3. K.P. Sinha and S.L. Kakani, High Temperature Superconductivity, Nova Science Pub., New York, USA (1995).
4. K.P. Sinha and S.L. Kakani, Magnetic Superconductors, Nova Science, Newyork (1989)
5. S.L. Kakani, Superconductivity: Revisited, New Age Int. Pub., New Delhi-2 (In Press)
6. S.L. Kakani, Solid State Physics, (4th ed.), Sultan Chand, New Delhi-2
7. S.T. Ruggiero and D.A. Rudman (eds.), Superconducting Devices, Academic Press, London (1990)

REVIEW QUESTIONS

1. What is superconductivity? Mention some important property changes that occur in materials when they undergo phase change from normal to the superconducting state.

2. What are type-I and type-II superconducting materials? Give three examples of each why type-II materials are preferred for applications of superconductivity. [BE]

3. Define critical temperature, critical field and critical current density for a superconductor. [AMIE]
4. Explain the behaviour of a superconductor in a magnetic field. What is Meissner effect? How does the critical magnetic field varies with temperature? [M.Sc. (M.S.)]
5. What is isotope effect in superconductors? How it provided a clue to the microscopic BCS theory of superconductivity? [B.E.]
6. Describe the effects of *(i)* magnetic field *(ii)* frequency and *(iii)* isotopes on superconductors.
7. What are the characteristic lengths of a superconductor? How penetration depth varies with temperature? [B.E.]
8. Give a brief account of thermodynamics of a superconductor.
9. What is BCS theory of superconductivity? Write its salient features.
10. Write a short note on potential applications of superconductors. [B.E.]
11. What are d.c. and a.c. Josephson effects? Explain their importance.
12. What do you understand by quantum tunneling? Explain its significance. [B.E.]
13. What are HTSC oxides? How they are different from conventional superconductors? What is their future prospects from applications of superconductivity point of view?

SHORT QUESTIONS

1. Why do superconducting currents flow on the surface of a superconductor?
2. Why is the electric field zero inside a superconductor?
3. Does perfect conductivity require that the interior magnetic field of a body be zero? What does it require of the interior magnetic field?
4. How would you measure the critical field of a superconductor as a function of temperature?
5. The critical external magnetic field at absolute zero varies with the material as $M^{-1/2}$. Explain.
6. Can you say whether lead or aluminium has the higher T_c from the fact that at room temperature the electrical conductivity of aluminium is much larger than that of lead?
7. A superconducting film can be used as a high sensitivity bolometer (as instrument for measurement of heat radiation). Explain.
8. To what extent can the two electrons in a Cooper pair be thought of as moving as if they were bound to opposite ends of spring? What property of the system constitutes the spring?
9. What is the distinction between the energy states of an electron in a superconductor and the energy states of the superconductor itself?
10. Superconductors whose Cooper pairs are small enough to allow the existence of magnetic field carrying channels also have relatively high T_cs. What is the reason for this convenient behaviour of type-II superconductors?

PROBLEMS

1. *(a)* Show from Maxwell's equations that resistivity $\rho = 0$ (a perfect conductor) implies that $\vec{B} = 0$ inside the material. (b) Show, from MaxWell's equations, that $\vec{B} = 0$ inside a material (a superconductor) implies that the resistivity of the material is $\rho = 0$.
2. Show that Meissner effect implies perfect conductivity, but vice versa is not true.
3. The resistivites of Cu, Ag and Au at room temperature in units of 10^{-8} Ω–m are 1.6, 1.5 and 2.4 while those of Ti, Zr and Hf are 89, 45 and 32 respectively. Explain, why the former do not show superconductivity, where as the latter become superconductors?
4. The critical temperature (T_c) for Hg with isotopic mass 199.5 is 4.185 K. Calculate its critical temperature when its isotope mass changes to 203.4. [*Ans.* 4.139 K]
5. Show that the frequency of the electromagnetic waves radiated by a Josephson junction having a voltage of 650 μV across its terminals is 3.15×10^{11} Hz.

OBJECTIVE QUESTIONS

1. The temperature at which a metal becomes superconductor is called
(1) Curie Temperature (2) Debye Temperature
(3) Critical Temperature (4) Neel Temperature (3)

2. The transition temperatures of most superconducting elements falls within the range
(1) Zero to 143 K (2) 0 to 10 K
(3) Zero to 23 K (4) 0 to 50 K (2)

3. The relation between T_C and H_c for a superconductor is

(1) $H_c = H_0 \left[1-\left(\frac{T}{T_C}\right)^2\right]$ (2) $H_c = H_0 \left[1+T_C^2\right]$

(3) $H_c = H_0 [(T - T_c)^2]$ (4) $H_c = \dfrac{H_0}{(1-T/T_C)^2}$ (1)

4. A type-I superconducting material when placed in a magnetic field will
(1) Expel all the magnetic lines of forces passing through it
(2) Attract the magnetic field toward its centre
(3) Not influence the magnetic field
(4) None of the above (1)

5. A superconducting material on being subjected to critical magnetic field (H_c)
(1) Changes to normal state (2) Remain unaffected
(3) Exhibit both normal as well as superconducting state properties
(4) None of the above (1)

6. The width of the energy gap of a superconductor is maximum at
(1) T_C K (2) 0 K (3) (T_C/2) K (4) (T_C/3) K (2)

7. In superconductivity state
(1) Entropy decreases and thermal conductivity increases
(2) Entropy increases and thermal conductivity decreases
(3) Entropy and thermal conductivity both decreases
(4) Entropy and thermal conductivity both increases (2)

SHORT QUESTION ANSWERS

1. What are the two distinctive hallmarks of superconductivity?

Ans. *(i)* Zero resistivity, i.e. infinite conductivity below some critical temperature T_C and
(ii) Meissner effect, i.e., $\vec{B} = 0$ inside the superconductor.

2. When a superconducting state ceases to exist?

Ans. When temperature, magnetic field, or current density exceeds the critical value.

3. What are type-I and type-II superconductors?

Ans. Materials that always expel the flux before becoming normal are called type-I superconductors. All pure superconducting metals are type-I except for niobium and vanadium. For type-II, there are two critical fields: a lower one H_{c_1} and the upper one H_{c_2}. The flux is completely expelled only up to the field H_{c_1}. Between H_{c_1} and H_{c_2}, the superconductor is said to in the *mixed* state. The Meissner effect is only partial. Above H_{c_2} the material returns to the normal state.

4. What are the characteristics of HTSC oxides are?

Ans. These are ionic metals with highly anisotropic, layered structures. These are quasi two dimensional insulators with an unusal normal state. Superconductivity with T_C = 134 K has been achieved. These are extreme-type-II superconductors with very short coherence lengths (ξ ~ 10 Å) and large penetration depths (λ ~3000 Å). These allow inexpensive liquid nitrogen to be used as a coolant and therefore have a far reaching technological implications.

5. Mention some of the most important applications of superconductors.

Ans. The most important applications of superconductors are either in high current wires for powerful supermagnets or in thin film structures like Josephson Junctions for superconducting electronics. SQUIDs have been used to develop some of the most sensitive measurement instruments at the forefront of present technology.

17

Organic Materials: Polymers and Elastomers

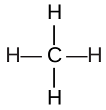
1. INTRODUCTION

The materials, which are derivatives of carbon, chemically combined with hydrogen, oxygen or any other non-metallic substance, and their structures, in most cases, fairly complex are known as organic materials. Organic materials may be natural or synthetic, i.e. prepared or manufactured artificially. The naturally available organic materials include wood, natural rubber, coal, petroleum and food products, etc. The synthetic organic materials include, plastics, lubricants, rubber, soap oils, synthetic fibres, etc. Obviously, the range of organic compounds is very extensive as the thousands of hydrocarbon compounds and their derivatives are in existence. Plastics and synthetic rubbers termed polymers and wood and wood products are common engineering materials of industrial importance.

The organic materials are marked by their non-crystalline nature. However, the non-crystalline solid structures of organic materials is not completely random as they are composed of ordered atoms in several sub-units. The structures of all non-crystalline materials are such that the sub-unit arrangements are tangled very easily and completely in liquid state itself. However, on solidification it is impossible for them to get untangled. Such materials result into three dimensional network or long chain molecules due to the entanglement. These non-crystalline organic materials may be elements or compounds and are usually referred to as amorphous solids. Glass is a common example of these materials. SiO_2 (silica) is another example of amorphous materials. The most important amorphous materials are *polymers.*

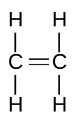
2. POLYMERS

As with metals and ceramics, the properties of polymers are intricately related to the structural elements of the material. Most polymers are organic in origin and are based on hydrocarbons, i.e. they are composed of hydrogen and carbon. Moreover, the intramolecular bonds are covalent. Each carbon has four electrons that may participate in covalent bonding, whereas every hydrogen atom has only one bonding electron. A single covalent bond exist when each of the two bonding atoms contribute one electron, as shown in Fig. 17.1 for methane. Double and triple bonds exist between two carbon atoms involve the sharing of two and three pairs of electrons, respectively. Figure 17.2 shows the structural formula for ethylene (C_2H_4). We note that in ethylene, the two carbon atoms are doubly bonded together, and each is also singly bonded to two hydrogen atoms. In Fig. 17.2, – and = denote single and double covalent bonds respectively. Acetylene (C_2H_2) is an example of *triple bond* (Fig. 17.3).



$$\begin{array}{c} \mathrm{H} \\ | \\ \mathrm{H—C—H} \\ | \\ \mathrm{H} \end{array}$$

Fig. 17.1 Methane



$$\begin{array}{ccc} \mathrm{H} & & \mathrm{H} \\ | & & | \\ \mathrm{C} & = & \mathrm{C} \\ | & & | \\ \mathrm{H} & & \mathrm{H} \end{array}$$

Fig. 17.2 Ethylene

H—C≡C—H

Fig. 17.3 Acetylene

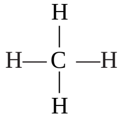
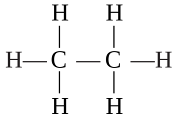
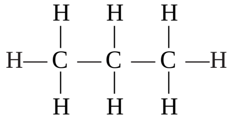
Organic molecules that have double and triple bonds are termed *unsaturated*, i.e., in these compounds, each carbon atom is not bonded to the maximum (four other atoms; as such, it is possible for another atom or group of atoms to become attached to the orginal molecule. On the otherhand, for a *saturated* hydrocarbon, all bonds are single ones; (and saturated), i.e., no new atoms may be joined without the removal of others that are already bonded. Though the saturated organic compounds are three dimensional structures, yet they are represented by two dimensional structural formula for convenience.

Most polymers are organic (carbon-based) materials that contain molecules composed of various combinations of hydrogen, oxygen, nitrogen and carbon. These four elements are among the most common found in organic polymers. Carbon forms the "spine" of the polymer chain, and the other constituents attach themselves to the carbon. These polymer chains become entangled and form irregular coils which give them added strength. A portion of this entanglement is natural and can be further induced by additives and controlled processes.

Most polymers are based on hydrocarbons, where the elements of carbon and hydrogen form predictable combinations based on the relationship, C_nH_{2n+2}. These petrochemical intermediates are chemicals manufactured from *paraffins* in petroleum and natural gas, which are further processed into polymer products.

The chain like paraffin molecules include methane (CH_4), ethane (C_2H_6), propane (C_3H_8), and butane (C_4H_{10}). Table 17.1 provides the compositions and molecular structures for some of paraffin molecules. The covalent bonds in each molecule are strong, but only weak hydrogen and Van der Waals bonds exist between molecules, and therefore these hydrocarbons have relatively low melting and boiling points. However, boiling temperature rise with increasing molecular weight. These intermediates are the basis for almost all rubber and polymer products. Intermediates are also produced from coal. The most important of these intermediates is ethylene. These are called *olefin intermediates* and include acetylene, propylene, butylene, isobutylene, and butadiene. Most of these are used in the production of rubbers. Nylon is produced from butadiene; polyvinyl chloride is produced from acetylene.

Table 17.1 Compositions and molecular structures for some of the paraffin compounds based on the relation C_nH_{2n+2}

Name	*Composition (formula)*	*Structure*	*Boiling point (°C)*
Methane	CH_4	 $H-\overset{H}{\underset{H}{C}}-H$	–164
Ethane	C_2H_6	 $H-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-H$	–88.6
Propane	C_3H_8	 $H-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-H$	–42.1
Butane	C_4H_{10}		–0.5
Pentane	C_5H_{12}		36.1
Hexane	C_6H_{14}		69.0

Hydrocarbon compounds with the same composition may have different atomic arrangement; a phenomenon called *isomerism*. For e.g., butane have two isomers. The structure of normal butane is shown in Fig. 17.4.

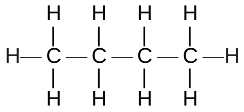

$H-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-H$

Fig. 17.4 Normal structure of butane

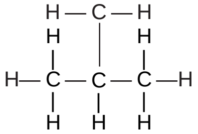
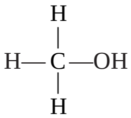
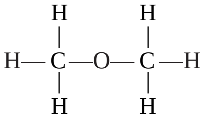
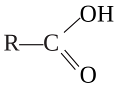
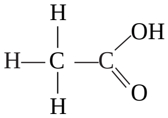
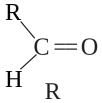
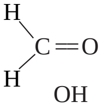
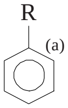
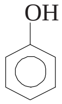


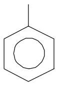
Fig. 17.5 Structure of isobutane

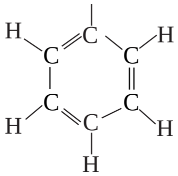
Whereas the structure of *isobutane* is shown in Fig. 17.5. We may note that some of the physical properties of hydrocarbons will depend on the isomeric state; for e.g., the boiling temperatures of normal butane and isobutane are –0.5 and –12.3°C, respectively.

There are several other organic groups, many of which are involved in polymer structures as mentioned above. Table 17.2 provides some common hydrocarbon groups, where *R* and *R′* represent organic radicals—groups of atoms that remain as a single unit and maintain their identity during chemical reactions. CH_5, C_2H_5 and C_6H_5 (methyl, ethyl and phenyl) groups are examples of singly bonded hydrocarbons.

Table 17.2 Few common hydrocarbon groups

	Family	*Characteristic unit*		*Representative compound*
(i)	Alcohols	R—OH	 $H_3C{-}OH$	Methyl Alcohol
(ii)	Ethers	R—O—R′	 $H_3C{-}O{-}CH_3$	Dimethyl ether
(iii)	Acids			Acetic acid
(iv)	Aldehydes			Formaldehyde
(v)	Aromatic hydrocarbons			Phenol

(a) The structure  represent a phenyl group:



The *aromatic intermediates* include benezene, toluene, and xylene. The organic compounds, which are derived from benezene (C_6H_6) by replacing the hydrogen atom with other groups of atoms are called as aromatic organic compounds. The structure of benezene is in the form of a ring in which the carbon atoms are arranged in a regular hexagon with the hydrogen atoms attached to each carbon atom outside the ring (Fig. 17.6). The single line between two carbon atoms indicates single covalent bond and double lines indicates double covalent bond in the structure. Obviously, the difference between the two structures (a) and (b) (Fig. 17.6) is in the position of double bond in the ring. In general, benezene structure or ring is drawn as a simple hexagon, the carbon and hydrogen atoms are not shown in the structure. Any group substituted

(a) (b)

Fig. 17.6 Benezene strurcture

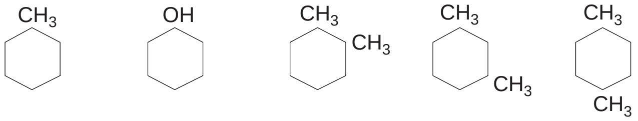


Fig. 17.7 Structures of few aromatic compounds

for hydrogen atoms are shown adjacent to the appropriate corner of the hexagon. Figure 17.7 shows some aromatic compounds on this pattern.

Napthenic intermediates is another category of intermediates in which cyclohexane is the most important, being used in the production of nylon. Aromatic, napthenic, and olefin intermediates are used to produce insecticides, detergents, rocket fuels, films, pharmaceuticals, explosives, alcohols, and other such products. Cellulosic plastics are produced from wood rather than petroleum, coal, or gas.

General characteristics of polymers can be summarized as follows:

(i) Polymers have long chain structures. The individual molecule of a polymer is very large, i.e., it may consists of thousands of similar small molecules, all bonded together covalently.

(ii) All polymers have one thing common, i.e., carbon, which further bonds with hydrogen, nitrogen, halogens or other organic or inorganic substances.

(iii) Although, polymer's structure may be crystalline in simple materials but generally they are non-crystalline solids at room temperatures. No doubt, polymers pass through a viscous stage during formation.

(iv) Polymers have light weight and they can be easily fabricated and shaped. They are poor conductors of electricity and their thermal conductivity is also low. Moreover, the polymers are resistant to chemical attack and decay.

Polymers are widely used in different fields. Polymers are used in indefinite number of forms, e.g. as telephone sets, paints, radio and television cabinets, coatings, adhesive and other countless objects. Few notable developments in polymer materials are:

(i) polymer materials far stronger than steel according to weight to weight ratio.

(ii) polymer materials suitable for repairing damaged kidney or heart.

(iii) materials act as *adhesive* so strong that the use of nails in wood and leather can be avoided.

(iv) *lining for spacecraft* to withstand high temperature, especially during re-entering into earth's atmosphere. We know that there is increase in temperature due to friction when the spacecraft re-enters the earth's atomsphere.

(v) *silicon film* does not permit water to enter but allows dissolved oxygen, in water, to pass through it. Using polymers, under water tents have been prepared which allows a man to live and work.

3. BROAD CLASSIFICATIONS

Usually polymers are broadly classified in three categories: *(i) Plastics (ii) Fibres* and *(iii) Elastomers.* Polymers are also classified according to their *composition, formability, geometrical structure* and *molecular structure.*

Thermoplastic resins are generally referred to as plastics and have the property of increasing plasticity, i.e., ability to deform plastically with rise in temperature. They have the long chain structure. On the otherhand, *thermosetting* resins have the three dimensional network of primary bonds. On heating they do not get soft. When heated, they become harder due to completion of any left over polymerization reaction. Due to their reaction with atmospheric air, they get decomposed at high temperatures. The process of decomposition in respect of thermosetting plastic is called *degradation*. The molecular chain structure of thermoplaistic and thermosetting plastic is shown in Fig. 17.8. Major characterstics of plastics are:

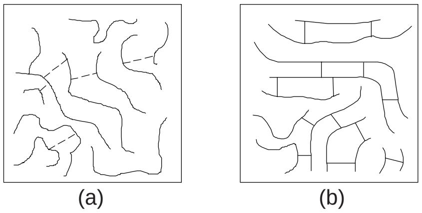


Fig. 17.8 (a) Thermoplastic and (b) Thermosets plastics

(i) Low softening temperature *(ii)* Non-crystalline structure *(iii)* can be readily formed into desired complex shapes *(iv)* With some exceptions resistant to chemical reaction *(v)* poor conductor of electricity and low thermal conductivity and *(vi)* visco-elastic behaviour.

Natural resin is generally used for naturally formed materials such as saps and extracts. Natural resins are obtained either directly through the excretions of certain trees or from the insects that feed on the sad juices. Natural resins are used in such products as paints, varnishes, enamels, soaps, inks, glues, and as plasticizers of other polymeric materials. These include shellacs, rosins and copal resins.

Shellac is a thermoplastic material used as a base in a wide variety of chemical solvents, as a hard glass coating, and as a bonding agent. Its use has declined in favour of less brittle materials, however, it is still used in the manufacture of abrasive inks, and lacquers. Rosin is a distillate of pine tree sap; turpentine is produced as a by-product. Rosins are used in paints, varnishes, soap, paper and ink. Copal resins are used in paints, varnishes, and lacquers. When copal resins are mixed with celluloid, they form a hard film that is moisture and abrasion resistant.

Fibres: These consists of long chain molecules with all the chains aligned to more or less logitudional direction of the fibre. The alignment shows highly directional behaviour with strength along the length being more than ten times that across the length.

Elastomers: These are also long chain molecules, but exhibit the unique property of high elasticity.

4. BASIC CONCEPTS OF POLYMER SCIENCE

(i) Polymer: It is a large molecule built up by the repetition of small, simple chemical units. In some cases the repetition is linear, much as a chain is built up from its links. In other cases chains are branched or interconnected to form three dimensional networks. The repeat unit of polymer is usually equivalent or nearly equivalent to the *monomer,* or starting material from which the polymer is formed. The *repeat unit* of poly (vinyl chloride) is—CH_2CHCl—; its *monomer* is vinyl chloride, $CH_2 = CHCl$ (Table 17.3).

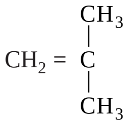
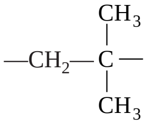
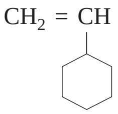
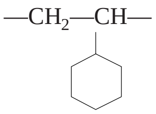
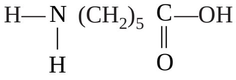
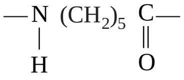
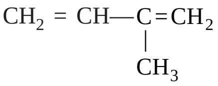
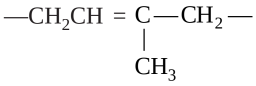
The length of the polymer chain is specified by the number of repeat units in the chain. This is called the *degree of polymerization* (DP). The DP is also defined as the number of repetitions of mers in a polymer chain.

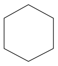
Mathematically,

$$\text{DP} = \frac{\text{Molecular weight of a polymer}}{\text{Molecular weight of a single monomer}}$$

The *molecular weight* of the polymer is the product of the molecular weight of the repeat unit and the DP. As an example, poly (vinyl chloride), a polymer of DP 1000 has a molecular weight of 63 × 1000 = 63,000. Most high polymers useful for plastics, rubbers, or fibres have molecular weights between 10,000 and 1,000,000. The molecular weight increases with the size of the molecule. The molecules having low DP are called *oligomers*, the oligo meaning few. Usually, the term polymer is used for macromolecule, i.e. a large size molecule.

Table 17.3 Some linear high polymers, their monomers and their repeat units

Polymer	*Monomer*	*Repeat unit*
Polyethylene	$CH_2 = CH_2$	$—CH_2CH_2—$
Poly (vinyl chloride)	$CH_2 = CHCl$	$—CH_2CHCl$
Polyisobutylene	 $CH_2 = \overset{CH_3}{\underset{CH_3}{C}}$	 $—CH_2—\overset{CH_3}{\underset{CH_3}{C}}—$
Polystyrene[1]		
Poly caprolactan (6-nylon)	 $H—\underset{H}{N}\ (CH_2)_5\ \underset{\overset{\|}{O}}{C}—OH$	 $—\underset{H}{N}\ (CH_2)_5\ \underset{\overset{\|}{O}}{C}—$
Polyisoprene (natural rubber)	 $CH_2 = CH—\underset{CH_3}{C}=CH_2$	 $—CH_2CH = \underset{CH_3}{C}—CH_2—$

1.  represents benezene ring, double bonds being omitted

When DP attains a *critical value* or the length of a polymer chain attains a sufficient length, the properties of polymer starts to appear. With the increase in DP, the properties of a polymer become more and more prominents, i.e. polymer size increases. Although, the molecular weight for most of the polymers ranges between 8,000 to 150,000 but higher molecular weights in the range of 10^8 are also in existence. For example, C_2H_4 (ethylene) is a gas, its oligomers having DP in the range 3 to 20 are liquids. With the increase in DP their viscosity goes on increasing. The polymers of ethylene (C_2H_4) with DP in the range 25 to 40 are like gases and with DP in the range 50 to 60, they are like waxes. When DP increases beyond 350, ethylene polymers become resins. The effect of DP on softening temperature is shown in Fig. 17.9. From figure it is obvious that with the increase in DP, the softening temperature also increases, but when DP increases beyond 1000, the rise in softening temperature is small.

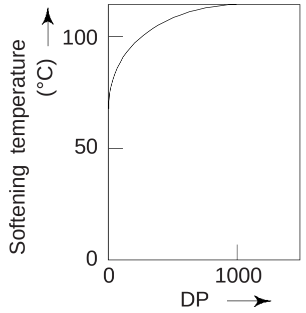


Fig. 17.9

The tensile strength increases with the molecular weight or DP. The viscosity also increases tensile strength rapidly in the initial stages but viscosity increases when molecular weight is large. Impact strength also increases with the increase in molecular weight (Fig. 17.10). Polymers with higher molecular weights possess superior mechanical properties but the fabrication becomes difficult due to higher viscosities. One has to make a compromise between properties and processability for commercial range.

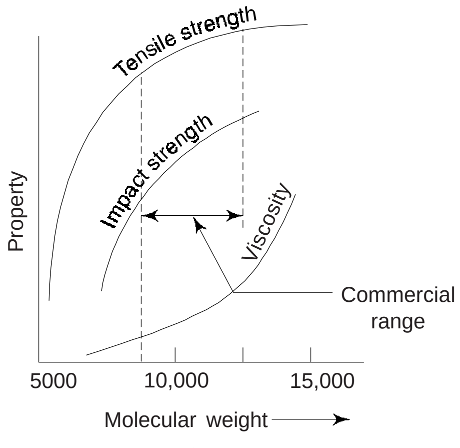


Fig. 17.10

(ii) Polymerization Process: Polymerization, or the joining of large unit molecules called monomers, utilizes the valence of the partially filled outer shell of the carbon atom (carbon has a valence of 4) to join smaller units together to form larger chains of molecules. Oxygen, sulphur, silicon, or nitrogen can be used to replace the carbon atom.

Two conditions must be met for polymerization to occur: *(i)* a molecule must have at least two locations that have unsatisfied bonds, which will easily join with other molecules. This requirement means starting with a molecule that has a double bond, such as carbon. Because the carbon molecule has a double bond, each bond is a pair of shared electrons. If one of the bonds between the carbon atoms opens up, a single bond exists, leaving the other two electrons to share with other atoms. If another carbon atom passes by that has opened up its double bond, the two can join to form a chain. This procedure continues, producing a polymer chain, and is called polymerization. The process continues as long as the second condition is met. *(ii)* The second condition necessary for polymerization is that after each reaction, at least two open locations must remain. Many chains form within polymers. These chains form threads that entwine themselves around each other for strength.

Polymers can be strengthened by *cross-linking.* Cross linking occurs when the double bonds between atoms within a chain are broken and these atoms or molecules form, or link up, with neighbouring atoms. This linkage provides additional strength to the chain and reduces the *slippage* that occurs between molecules. Slippage occurs when the polymer threads slide past each other when subjected to a load.

The properties of polymers also depend on the structure as well as the composition of the molecule. Two molecules with the same composition may form two different-configurations having different properties, such as propyl (1-propanol) and isopropyl (2-propanol) alcohol. These variations are called *isomers.*

Polymerization takes place through *addition polymerization, copolymerization,* or *condensation polymerization*. In the polymerization process, a large unit molecule, the *monomer,* is added to another monomer to form a large chain, the polymer (referring to many parts), which has a number of repeated units, mers. Mers are the smallest units recognizable in the chain. The degree of polymerization is the number of repeating units that have identical structures within the chain formed by the polymer. Addition polymerization involves only one type of mer. Figure 17.11 shows polymerization by addition.

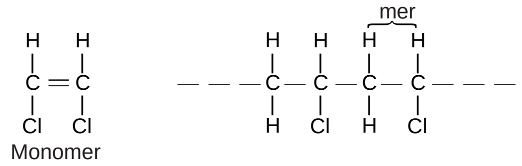


Fig. 17.11 Addition polymerization

In copolymerization, more than one molecule makes up the mer. Acrylonitride-butadiene-styrene (ABS) is an example of a copolymer. Figure 17.12 shows the copolymerization process for ABS polymers.

Condensation polymerization involves the chemical reaction of two or more chemicals to form a new molecule. This chemical reaction produces a condensate or nonpolymerizable by product, usually water. A

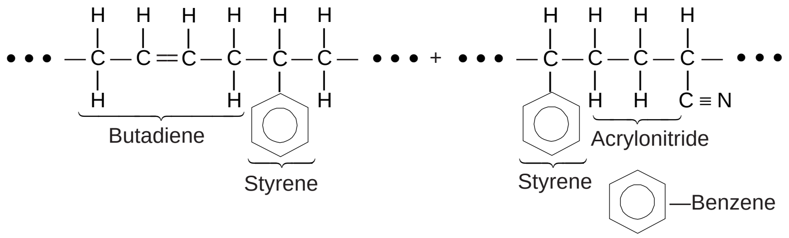


Fig. 17.12 Copolymerization

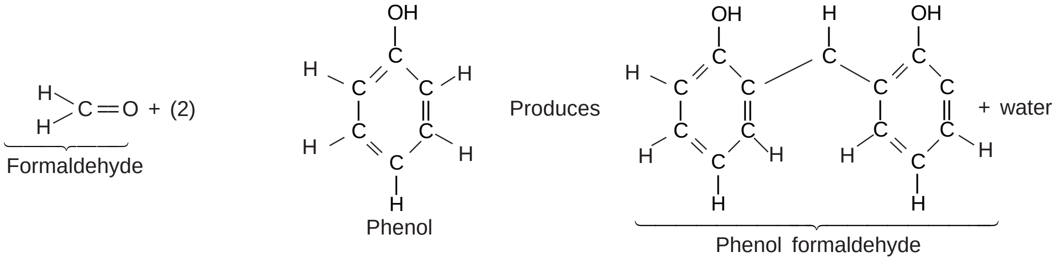


Fig. 17.13 Condensation polymerization

catalyst is often required to start and maintain the reaction. It can also be used to control the reaction rate. The process of condensation polymerization is shown in Fig. 17.13.

With some exceptions, polymers made in chain reactions often contain only carbon atoms in the main chain (*homochain polymers*), whereas polymers made in step reactions may have other atoms, originating in the monomer functional groups, as part of the chain (heterochain polymers).

(iii) Molecular weight and its Distribution: In both chain and stepwise polymerization, the length of a chain is determined by purely random events. In step reactions, the chain length is determined by the local availability of reactive groups at the ends of the growing chains. In radical polymerization, chain length is determined by the time during which the chain grows before it diffuses into the vicinity of a second free radical and the two react. In either case, the polymeric product contains molecules having many different chain lengths.

Extremely, large molecular weights (sometimes 'molecular mass', molar mass', and 'relative molar mass' are also used) are reported in polymers with very long chains. During the polymerization process in which these large macromolecules are synthesized from smaller molecules, not all polymer chains will grow to the same length; this results in a distribution of chain lengths or molecular weights. Normally, an average molecular weight is specified, which may be determined by the measurement of various physical properties, e.g. viscosity and osmotic pressure.

Average molecular weight can be defined through several ways. One can obtain the *number average molecular weight* $\overline{M}_n$ by dividing the chains into a series of size ranges and thus determining the number of fraction of chains within each size range. One can express this number average molecular weight as

$$\overline{M}_n = \sum_i x_i M_i \tag{1}$$

Here M_i represents the mean (middle) molecular weight of size range i, and x_i is the fraction of the total number of chains within the corresponding size range.

Another average, a *weight average* molecular weight $\overline{M}_w$ is based on the weight fraction of molecules within the various size ranges is defined as

$$\overline{M}_w = \sum_i w_i M_i \tag{2}$$

Here w_i is the weight fraction concentration of the molecules having molecular weight M_i. A typical molecular weight distribution along with these molecular weight averages for a polymer is shown in Fig. 17.14.

For typical polymers the number average lies near the peak of the weight-distribution curve or the most probable molecular weight. Both the averages represented by relations (1) and (2) are used in the interpretation of the properties of poly-

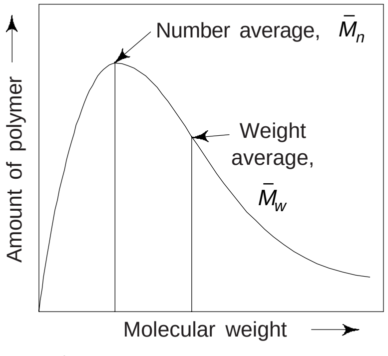


Fig.17.14 Distribution of molecular weights in a typical polymer

mers. The ration $\overline{M}_w / \overline{M}_n$ indicates the degree of distribution of the molecular weights. When all the molecules are identical, the said ratio is unity. However, in practice this ratio lies generally between 1.5 and 3 but in some cases it may exceed 30.

Another way of expressing average chain size of a polymer is the *degree of polymerization n*, which represents the average number of mer units in a chain. Both number-average (n_n) and weight-average (n_w) degrees are possible. These can be expressed as

$$n_n = \frac{\overline{M}_n}{\bar{m}} \tag{3a}$$

and

$$n_w = \frac{\overline{M}_w}{\bar{m}} \tag{3b}$$

Here $\overline{M}_n$ and $\overline{M}_w$ represents the number-average and weight-average molecular weights as defined above, while $\bar{m}$ is the mer molecular weight. One can determine $\bar{m}$ for a copolymer (having two or more different mer units) from

$$\bar{m} = \sum f_j \, m_j \tag{4}$$

where f_j and m_j are the chain fraction and molecular weight of mer j respectively.

(iv) Molecular Structure: Polymers consist of large number of molecular chains, each of which may bend, coil and kink (Fig. 17.15a). This leads to extensive interwining and entanglement of neighbouring chain molecules. These random coils and molecular entanglements are responsible for a number of important characteristics of polymers, to include the large elastic extensions displayed by the rubber materials.

There are some mechanical and thermal characteristics of polymers which are a function of the ability of chain regments to experience rotation in response to applied stresses or thermal vibrations. We may note that rotational flexibility is dependent on mer structure and chemistry, e.g., the region of a chain segment that has a double bond (C = C) is rotationally rigid. Moreover, introduction of a bulky or large side group of atoms restricts rotational movement, e.g., polystyrene molecules, which have a phenyl side group are more resistant to rotational motion than are polyethylene chains.

The physical characteristics of a polymer is found to depend not only its molecular weight and shape but also on differences in structures of molecular chains. In addition to various isomeric configurations, polymers have several molecular structures including linear, branched, crosslinked, and network. Modern polymer synthesis techniques permit one to have considerable control over various structural possibilities.

Linear Polymers: As mentioned earlier, these are the polymers in which the *mer* units are joined together end to end in single chains. These long chains are flexible and we may thought them as a mass of a spaghetti. Linear structure of polyethylene is shown in Fig. 17.16. Few common examples of polymers that form with linear structures are polyethylene, polyvinylchloride, polystyrene, polymethyl methacrylate, nylon, and the fluorocarbons. For these linear polymers, there may be extensive Vander Waals and hydrogen bonding between the chains. These linear polymers melt on heating. When the chains in a polymer are entangled with each other, mechanical strength of the polymer increases because of interweaving of molecules. Interestingly, bifunctional monomers produce linear polymers, e.g. ethylene produces polyethylene linear polymer (Fig. 17.16).

The polymer fibres used in textile industry are the linear polymers. These polymers have high molecular weight and fixed orientation of the molecules parallel to fibre axis.

Branched Polymers

In contrast to the linear chain molecules, some polymers have branched chains, often as a result of side reactions during polymerization (Fig. 17.17a). Branched polymers may be synthesized in which side branch chains are connected to the main ones as illustrated in Fig. 17.18. The chain packing efficiency is reduced

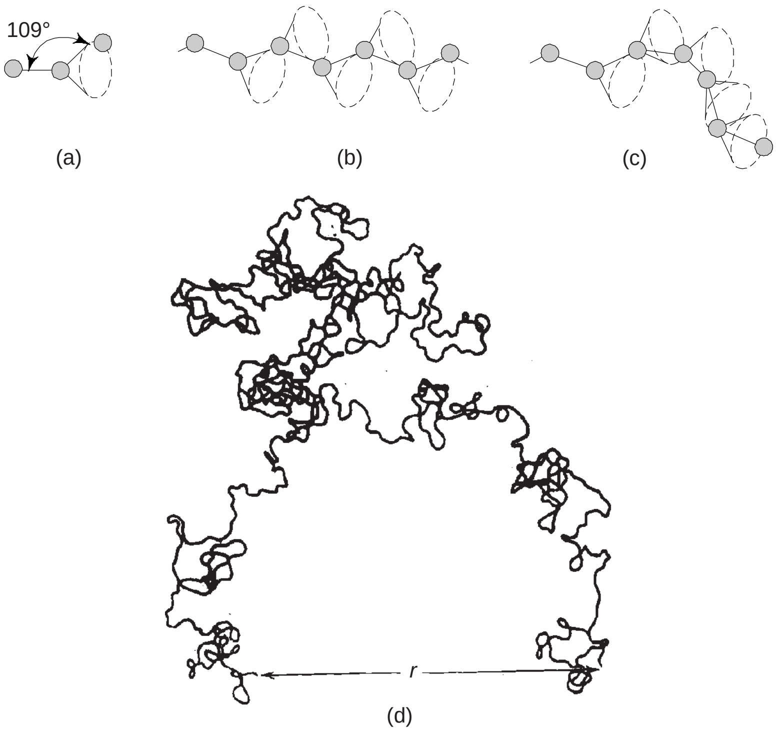


Fig. 17.15 How polymer chain shape is influenced by the positioning of backbone carbon atom (circle). (a) the rightmost atom may lie anywhere on the dashed circle and still subtend a 109° angle with the bond between the other two atoms (b) and (c) straight and twisted chain segments generation (d) single polymer chain molecule that has numerous random kinks and coils produced by chain body rotations

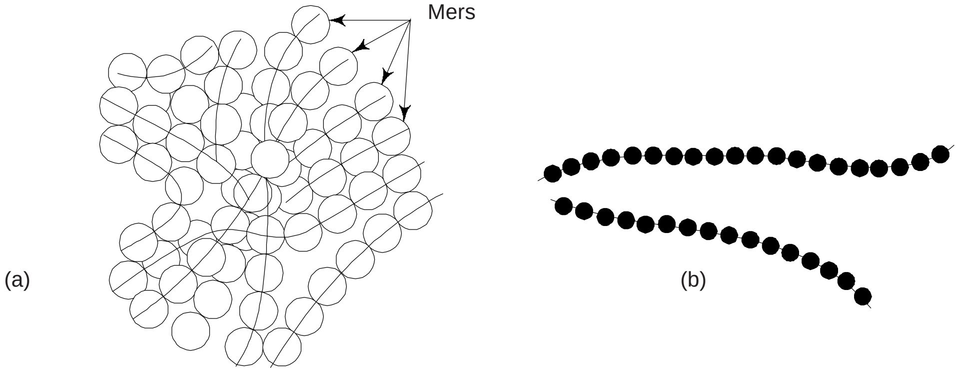
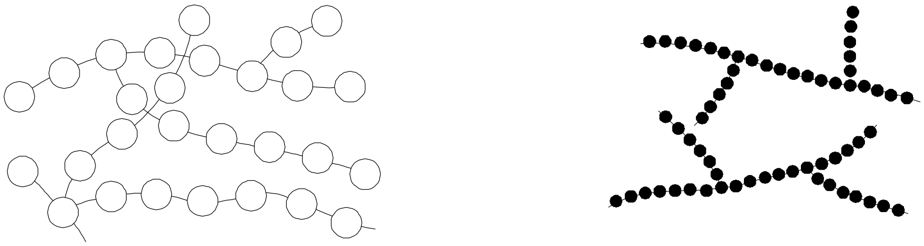


Fig. 17.16 (a) Linear structure of polyethylene polymer (b) Linear molecular structure

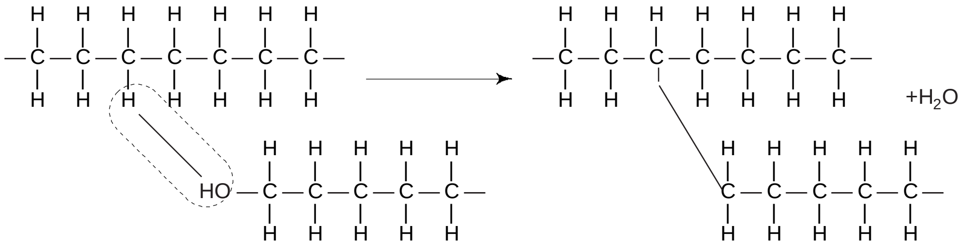
with the formation of side branches, which results in a lowering of polymer density. The polymers that form linear structures may also be branched.

Branched and linear polymers are thermoplastic and have gained a great industrial importance. By changing the temperature, the branched polymers can be hardened or softened.



(a) (b)

Fig. 17.17 (a) Branching (b) Alternative representation

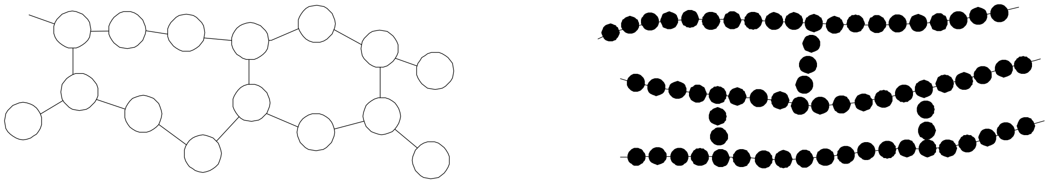


Polyethylene chains Branch linking of chain polyethyene

Fig. 17.18 Illustration of branching mechanism. This is achieved by removing a side atom from the main chain and attaching another C—C bond

Cross Linked Polymers

In these polymers, adjusted linear chains are joined one to another at various positions by covalent bonds (Fig. 17.19). Crosslinking is achieve either during synthesis or by a non reversible chemical reaction that is usually carried out at an elevated temperature. Usually, crosslinking is accomplished by additive atoms or molecules that are covalently bonded to the chains. The movement of adjacent chains is greatly restricted and the mechanical properties are affected to a great extent. Many of the rubber elastic materials are crosslinked. This is done to increase strength and reduce plasticity in natural rubber. This process is called *vulcanization.*



(a) (b)

Fig. 1719 (a) Crosslinked structure (b) Alternative representation

Vulcanization is the crosslinking process in elastomers, which is achieved by a non reversible chemical reaction, ordinarily carried out at an elevated temperature. In several vulcanizing reactions, sulphur compounds are added to the heated elastomer; chains of sulphur atoms bond with adjusted polymer backbone chains crosslink them, which is accomplished according the reaction given below (Fig. 17.20) in which two crosslinks shows consist of *m* and *n* sulphur atoms. Main chainsites of crosslink are carbon atoms that were doubly bonded prior to vulcanization but, after vulcanization, have become singly bonded.

Prior to vulcanization, i.e. unvulcanized rubber is soft and tacky and also has poor resistance to abrasion. Modulus of elasticity, tensile strength, and resistance to degradation by oxidation all are enhanced by this technique. The sulphur content controls the number of anchoring points. The magnitude of the modulus of

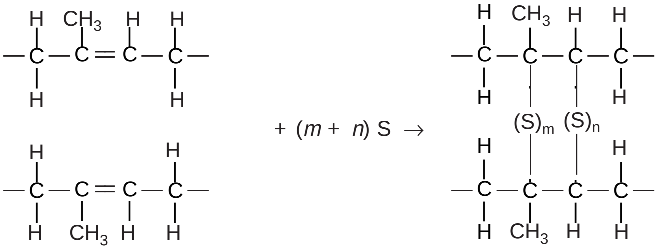


Fig. 17.20 Vulcanization

elasticity is directly proportional to the density of the crosslinks. Figure 17.21 shows the stress-strain curves for both vulcanized and unvulcanized natural rubber.

In order to produce a rubber that is capable of large extensions without rupture of the primary chain bonds, there must be relatively few crosslinks, and these must be widely separated. One obtains the useful rubber when about 1 to 5 parts (by weight) of sulphur are added to 100 parts of rubber. However, increasing further the sulphur content hardnes the rubber and also reduces its extensibility.

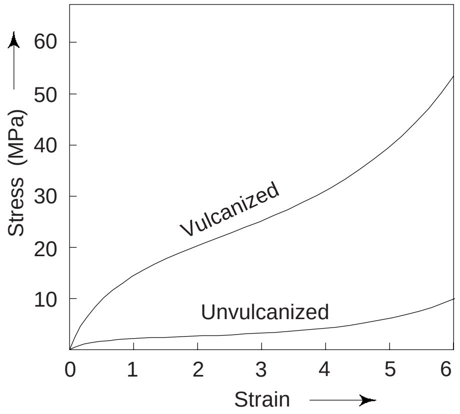


Fig. 17.21 Stress-strain curves to 600% elongation for unvulcanized and vulcanized natural rubber

Ageing of Polymers

We may note that cross-linking is sometimes harmful too. Under the catalytic action of sunlight, the rubber gets cross-linked with oxygen of air over a long period of time. This is termed as ageing of polymers (Fig. 17.22). The crossed linked rubber loses its flexibility and elasticity.

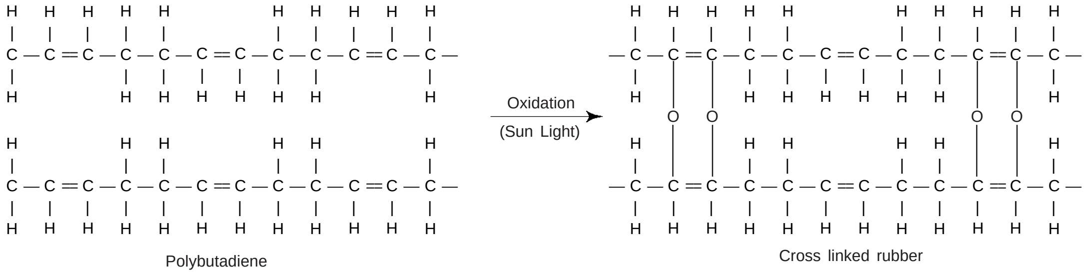


Fig. 17.22 Ageing of polymers

To retain the structure, the amount of cross-linkages should be as small as sufficient. Excessive cross linkages makes the internal structure so stiff that the rubbery state will be destroyed.

Network Polymers

Trifunctional mer units, exhibiting three covalent bonds, form three-dimensional networks (Fig. 17.23). These are termed *network* polymers. We may note that a polymer which is highly crosslinked may be classified as a network polymer. These types of polymeric materials have distinc-

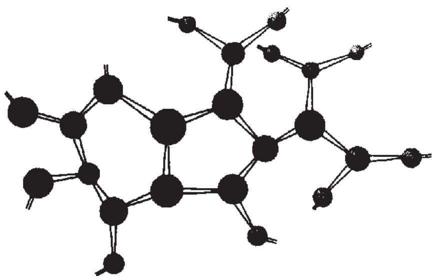


Fig. 17.23 Network (three-dimensional) polymer's molecular structure

tive mechanical and thermal properties. Few examples of materials belonging to this group are epoxies and phenol-formaldehyde.

(v) The Texture of Polymers: One can divide the geometrical arrangements of atoms in a polymer chain into two classes:

Configurations: Arrangements fixed by chemical bonding in the molecule, such as cis and trans isomers, or *d* and *l* forms, are described as configurations. This cannot be altered unless chemical bonds are broken and reformed.

Conformations: Arrangements arising from rotation about single bonds, which include the mainfold forms that the polymer chain may have in solution. These are described as *conformations.*

5. MOLECULAR CONFIGURATIONS

The polymers, which have one side atom or group of atom bonded to the main chain, the regularity and symmetry of the side group arrangement can significantly influence the properties. Let us consider the mer unit (Fig. 17.24) in which *R* represents an atom or the side group other than hydrogen, e.g. Cl, CH_3. When the *R* side groups of successive mer units bond to alternate carbon atom, one arrangement is possible as shown in Fig. 17.24.

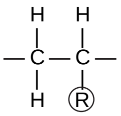


Fig. 17.24

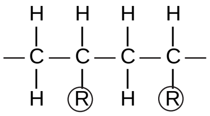


Fig. 17.25

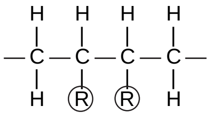


Fig. 17.26

This is named as head-to-tail configuration. When *R* group bond to adjacent chain atoms, the head-to-head configuration occurs (Fig. 17.26). The head-to-tail configuration predominates in most polymers. A polar repulsion occurs between *R* groups for the head-to-head configuration.

In polymer molecules, isomerism is also found, where in different atomic configurations are possible for the same composition. Stereoisomerism and geometrical isomerism are two subclasses of isomerism.

Stereoisomerism

This refers to the situation in which atoms are linked together in the same order (head to tail) but differ in their spatial arrangement. In an *isotactic configuration*, all the *R* groups are situated on the same side of the chain (Fig. 17.27). One stereoisomer is shown.

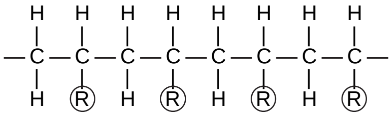


Fig. 17.27

The *R* groups alternate sides of the chain in a *syndiotactic configuration* (Fig. 17.28).

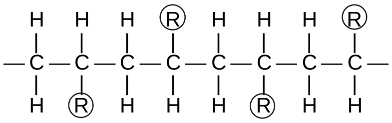


Fig. 17.28

The term *atactic configuration* is used for random positioning (Fig. 17.29).

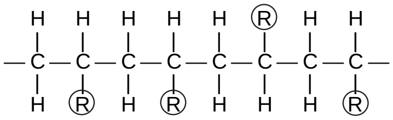


Fig. 17.29

Geometrical Isomerism

Geometrical isomers or other important chain configurations are possible within mer units having a double bond between chain carbon atoms. Bonded to each of the carbon atoms participating in the double bond is a single-bonded atom or radical, which may be situated on one side of the chain or its opposite. Let us consider the isoprene mer having the structure (Fig. 17.30) in which the CH_3 and the H are positioned on the same side of the chain. This is

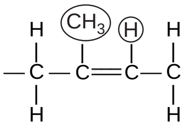


Fig. 17.30

called *cis* structure and the resulting polymer, cis-polyisoprene, is natural rubber. The alternate isomer, the *trans*-structure (Fig. 17.31), we can see that CH_3 and H reside on opposite chain sides. As a result of this configurational alteration, *trans*-polyisoprene, some times called gutta percha, has properties that are distinctly different from natural rubber. We may note that conversion of trans to cis, or vice versa, is not possible by a simple chain bond rotation as the chain double bond is extremely rigid.

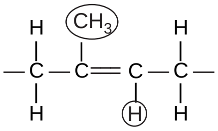


Fig. 17.31

6. THERMOPLASTIC AND THERMOSETTING POLYMERS

Polymers are seperated into these two general classifications:

(i) Thermoplastic Polymers: These polymers often exhibit plastic, ductile properties. They can be formed at elevated temperatures, cooled and remelted and reformed into different shapes without changing the properties of polymer. However, the heat used to melt and remelt the thermoplastic must be carefully controlled or material will decompose. Most linear polymers and those having some branched structures with flexible chains are thermoplastic. These materials are normally fabricated by the simultaneous application of heat and pressure. The properties of these materials are determined by the bonding method between polymer chains; in thermoplastic materials these bonds are weak, secondary bonds, as in Vander Waals forces. Through the application of heat and pressure, these bonds can be weakened, and the materials can be reshaped. Once the heat and pressure are removed, thermoplastic materials reharden in the new shape. Common thermoplastic polymers include acrylic, nylon (polyamide), cellulose, polystyrene, polyethylene, fluorocarbons, and vinyl. These are used for plastic walls and floor tiles, flouroscent light, plastic lenses, etc.

(ii) Thermosetting Polymers: These have strong primary bonds, often formed by condensation polymerization. These have strong, primary bonds throughout, and their structure resembles one large molecule. Their properties are the result of chemical changes undergone during processing, under heat or through the application of a catalyst. These polymers become permanently hard when heat is applied and do not softened or reshaped upon subsequent heating, due to the loss of part of the molecule (the by product of the reaction) during the curing process. Once *cured,* if further heat is applied to a thermosetting material, it will char, burn or decompose. During curing, the thermosetting polymers will become rigid and insoluble as the polymer chains becomes tangled and cross-linked. Cross-linking is usually extensive, in that 10% to 50% of the chain mer units are crossed linked. Thermoset polymers are generally harder and stronger than thermoplastics and have better dimensional stability. Common thermosetting polymers include phenolics, aminos, polyesters, epoxies and alkyds. Most of the cross-linked and phenolic and polyster resins are thermosetting.

7. COPOLYMERS

There is a continuous search for new polymer materials that can be easily and economically synthesized and fabricated, with improved properties or better property combinations that are offered by homopolymers.

One can prepare polymers with more than one kind of mer in the chain, i.e. differnet types of mers react to form a chain. Such resulting polymers are named as *copolymers.* The technique used for copolymers is named as *copolymerization*. There are different types of copolymers which can be produced depending upon the type of monomers and the method of synthesis.

Copolymers are of following types:

(i) Random Copolymer: This type has random distribution of various mers in its chain. If M and R represents two different mers, then the random copolymer can be depicted as shown in Fig. 17.32. Obviously, two different units are randomly dispersed along the chain.

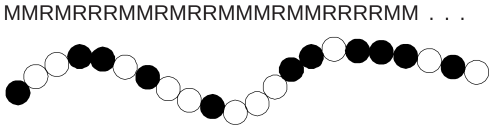


Fig. 17.32 Random copolymer (with two different type of mers)

(ii) Alternating Copolymer: As the name suggests, the two mer units alternate chain positions as shown in Fig. 17.33.

(iii) Block Copolymer: Identical mers are clustered in blocks along the chain (Fig. 17.34).

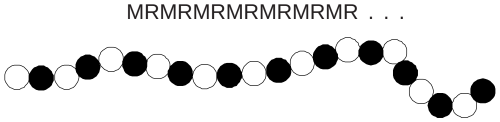


Fig. 17.33 Alternative copolymer (mers are arranged alternatively)

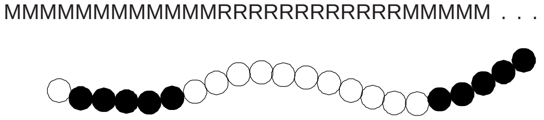


Fig. 17.34 Block copolymer (mers arranged in blocks)

(iv) Grafted Copolymer: It is the copolymer in which homopolymer side branches of one type may be grafted to homopolymer main chains that are composed of a different mer (Fig. 17.35).

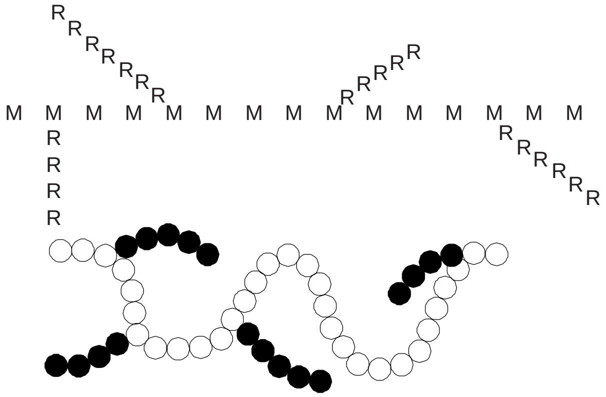


Fig. 17.35 Grafted copolymer (one types of mers are attached to one or more sides of the main chain formed by other type of mers)

Copolymers have better physical and mechanical properties. Copolymerization is just similar to the alloying process in metallurgy. Similar to alloys, one can synthesize various copolymers with desired combination of properties. Synthetic rubbers are often copolymers. Styrene-butadiene rubber (SBR) is a common random polymer from which automobile tyres are prepared. Nitride rubber (NBR) is another random copolymer composed of acrylonitrile and butadene. NBR is highly elastic and also resistant to swelling in organic solvents. NBR is used for making gasoline hoses.

8. POLYMER CRYSTALLINITY

The complexity of the molecular structure of the polymers results in a large variety of supermolecular structures and can extend from a well ordered crystalline structure to a completely amorphous structure. One can think of *polymer crystallinity* as the packing of molecular chains so as to produce an ordered atomic array. We know that crystal structure may be specified in terms of unit cells, which are often quite complex. The unit cell for polyethylene and its relationship to the molecular structure is shown in Fig. 17.36. This unit cell for polyethylene has orthorhombic geometry. However, the chain molecules also extend beyond the unit cell shown in Fig. 17.36.

Normally, molecular substances having small molecules, e.g. water and methane are either totally crystalline (as solids) or totally amorphous (as liquids). Polymer molecules, as a consequence of their size and often complexity are often only partially crystalline (or semicrystalline), having crystalline regions dispersed within the remaining amorphous material.

The density of a crystalline polymer is found to be greater than an amorphous one of the same material and molecular weight, since the chains are more closely packed together for the crystalline structure. One may determine the degree of crystallinity by weight from accurate density measurements, according to

$$\% \text{ crystallinity} = \frac{\rho_c \, (\rho_s - \rho_a)}{\rho_s \, (\rho_c - \rho_a)} \times 100 \qquad (5)$$

where ρ_s is the density of a specimen for which we have to determine the percent crystallinity, ρ_a is the density of the totally amorphous polymer, and ρ_c is the density of the perfectly crystalline polymer. One will have to determine the values of ρ_a and ρ_c by other experimental methods.

For a polymer, the degree of crystallinity depends on the rate of cooling during solidification and also on the chain configuration. The molecular chemistry also influence the ability of a polymer to crystallize. Crystallization is also not favoured in polymers that are composed of chemically complex mer structures, e.g. polyisoprene.

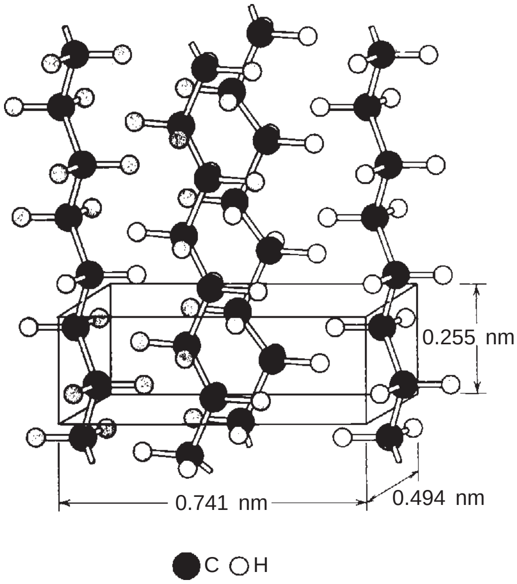


Fig. 17.36 **Molecular chains in a unit cell for polyethylene**

Crystallization is easily accomplished for linear polymers as there as virtually no restrictions to prevent chain alignment. However, branched polymers are never high crystalline as the side branches interfere with crystallization. Most network and crosslinked polymers are almost totally amorphous. However, a few crosslinked polymers are partially crystalline. There is a greater tendency for the development of non-crystallinity for coploymers due to irregular and random arrangement for the mer.

Obviously, the physical properties of polymeric materials to some extent are influenced by the degree of crystallinity. We may note that crystalline polymers are usually stronger and more resistant to dissolution and softening by heat.

9. DEFECTS IN POLYMERS

The point defect concept in polymers is different than in metals and ceramics due to the chain-like macromolecules and the nature of the crystalline state for polymers. However, point defects similar to those found in metals have been observed in crystalline regions of polymeric materials; these include vacancies and interstitial atoms and ions. Chain ends are also considered to be defects in as much as they are chemically dissimilar to normal chain units; vacancies are also associated with the chain ends. One may also incorporate impurity atoms/ions or group of atoms/ions in the molecular structure as interstitials. They may also be associated with main chains or as short side branches.

The surfaces of chain-folded layers are considered to be interfacial defects. Similarly boundaries between two adjacent crystalline regions also considered to be defects.

10. MECHANICAL PROPERTIES OF POLYMERS

It is observed that the mechanical strength of polymers increases with the degree of crystallinity (Fig. 17.37). This shows that the mechanical strength of polymer materials is a sensitive function of the process variables in the synthesis of these materials and therefore for the reproducibility and reliability of the products they should be properly controlled.

The properties of polymers, besides the degree of crystallinity, also depend on several parameters, e.g. nature of monomer unit, molecular weight, cross-linking, etc. This is why, it is very difficult to correlate the mechanical properties of the polymers with molecular structure directly. A simple mechanical model which can help to understand many of the properties of polymers is as follows:

When subjected to external stresses, many polymers undergo not only elastic deformation but also *viscous flow*. Viscoelastic mechanical behaviour, being intermediate between totally elastic and totally viscous, is displayed by a number of polymer materials. It is characterized by the *relaxation modulus*, a time-dependent modulus of elasticity,

$$E_r(t) = \frac{\sigma(t)}{\varepsilon_0} \tag{6}$$

where $\sigma(t)$ is the measured time dependent stress and ε_0 is the strain level, which is maintained constant. The magnitude of relaxation modulus is very sensitive to temperature and critical to in-service temperature range for elastomers in this temperature dependence.

One can define the viscoelastic modulus E_{ve} as

$$E_{ve} = \frac{\sigma}{(e_{\text{ela}} + e_{\text{vis}})} \tag{7}$$

The subscripts in Eq. (7) indicate the contribution from elastic and viscous flow deformations. Figure 17.37 show a plot of viscoelastic modulus as a function of temperature for different types of polymers.

For $E_{ve} > E_{ve}^3$, the properties of polymers are glassy and for $E_{ve}^3 > E > E_{ve}^2$ leather like and for $E_{ve}^2 > E > E_{ve}^1$ rubber like and for $E < E_{ve}^1$ viscous.

The polymers behaves as a glassy material for $T < T_g$. For $T > T_g$, one enters into the viscoelastic region in which the behaviour is leather like and then rubber like (Fig. 17.38). The modulus of elasticity in this region is low and is a function of time depending on the viscosity of the medium. We may note that in a crystalline polymer the rubber like region is not present.

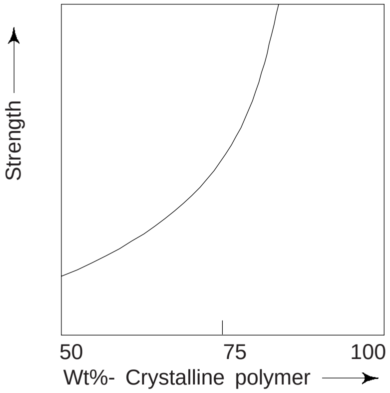


Fig. 17.37 **The increase of the strength of the polyethylene with crystallinity of the polymer**

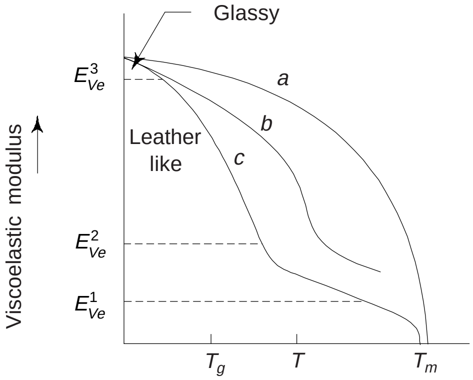


Fig. 17.38 **Viscoelastic modulus as a function of temperature: (a) crystalline, (b) crosslinked, and (c) amorphous polymers**

The mechanical properties are considerably affected by crystallization. Crystallization inhibit viscous flow. Although the composition of natural rubber and gutta percha is same but they have different molecular structure. Gutta percha being crystalline is hard and brittle, whereas rubber is soft and can easily be elongated. During the process of deformation, crystallization is also produced. Rubber can be elongated by 500-600% and also the strain is reversible. During such a stretching, crystallization takes place. However, with further extensions, when maximum crystallization is attained, rubber achieves high mechanical strength.

There are some textile fibres which show a linear relationship between stress and strain, i.e. stress $\propto$ strain, for moderate stresses. However, beyond a certain point, the elongation increases rapidly. We may note that polycrystalline structure does not change on the application of stress, e.g., in the case of nylon, such a relationship is almost entirely linear.

Concluding, we can say that in polymers, molecular structure which is controlled by synthesis parameters influences the properties to a great extent.

Viscoelastic Creep

When the stress level is maintained constant, many polymetric materials are found to be susceptible to time-dependent deformation. Such deformation is termed as *viscoelastic creep*. This type of deformation may be significant even at room temperature and under modest stresses that lie below the yield strength of the material.

Fracture of Polymers

Fracture strength of polymeric materials are low relative to metals and ceramics. Both brittle and ductile fracture modes are possible, and some thermoplastic materials experience a ductile-to-brittle transition with a lowering of temperature, an increase in strain rate, and/or an alteration of specimen thickness or geometry. In some glassy thermoplastics, the crack-formation process may be preceded by crazing; crazing can lead to an increase in ductility and toughness of the material.

11. MECHANISMS OF DEFORMATION

In order to be able to manage the mechanical characteristics of polymeric materials, an understanding of deformation mechanisms of these materials is important. In this regard, deformation mechanisms for two different types of polymers *semi-crystalline* and *elastomeric* are of importance. The deformation mechanisms are:

(i) Elastic Deformation Mechanism: The elastic deformation depends upon the amount of bond-straightening (Fig. 39a) and bond lengthening (Fig. 39b). The mechanism of elastic deformation in semicrystalline polymers in response to tensile stresses is the elongation of the chain molecules from their stable conformations, in the direction of the applied stress, by the bending and stretching of the strong chain covalent bonds. There may also be some slight displacement of adjacent molecules, which is resisted by relatively weak secondary or Vander Waals bonds. Semicrystalline materials may be considered as composite materials since they are composed of both crystalline and amorphous regions. Obviously, one may take elastic modulus as some combination of the moduli of crystalline and amorphous phases.

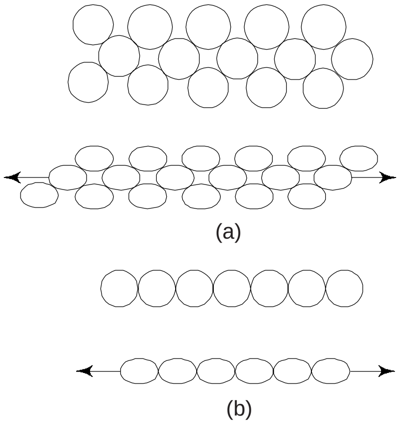


Fig. 17.39 Elastic deformation: (a) bond straightening (b) bond lengthening

(ii) Plastic Deformation Mechanism: Plastic deformation occurs due to slip between adjacent molecules of polymer material. Since the molecules are bonded by weak attractive forces and hence the slip occurs very easily. Plastic deformation is more prominent where alignment of molecules is linear. Plastic deformation occurs due to the slippage at the weaker points between the molecules and not due to the breaking of intermolecular bonds. The linear polymers with cross-linking show high degree of plastic deformation, whereas network polymers do not show high deformation and they are thus brittle.

The mechanism of plastic deformation can be best described by the interactions between lamellar and intervening amorphous regions in response to an applied tensile load. This process occurs in several stages.

12. CRYSTALLIZATION, MELTING AND GLASS TRANSITION PHENOMENA IN POLYMERS

With respect to the design and processing of polymeric materials, three phenomena, namely, crystallization, melting and glass transition are important. As mentioned earlier, crystallization is the process by which upon cooling, an ordered (i.e., crystalline) solid phase is produced from a liquid melt having a highly random molecular structure. when a polymer is heated, the occurence of melting transformation is the reverse process. The glass-transition phenomenon occurs with amorphous or non crystallizable polymers that, when cooled from a liquid melt, become rigid solids yet retain the disordered molecular structure, a characteristic of liquid state. The melting and glass transition temperatures of a polymer, i.e. T_m and T_g respectively, are important for knowing the temperature range over which a particular polymer may be utilized and processed. The magnitude of T_m and T_g increase with increasing chain stiffness; stiffness is enhanced by the presence of chain double bonds and side groups that are either bulky or polar. Molecular weight and degree of branching also affect T_m and T_g.

We may note that alterations of physical and mechanical properties attend crystallization, melting, and glass transition. Interestingly, for semicrystalline polymers, crystalline regions will experience melting (and crystallization), while non crystalline areas pass through the glass transition.

The time dependence of crystallization is the same as for many solid-state transformations. The fraction crystallized y is a function of time t according to the Avrami equation,

$$y = 1 - \exp\,[-kt^n] \tag{8}$$

where k and n are time-independent constants, whose values depend on the crystallizing system. Rate of crystallization is equal to the reciprocal of time required for crystallizaiton to proceed to 50% completion. This rate depends on crystallization temperature and also on molecular weight.

The melting of a polymer crystal on heating, at the *melting temperature*, T_m corresponds to the transformation of a solid material, having an ordered structure of aligned molecular chains, to a viscous liquid in which the structure is highly random. There are many features distinctive to the melting of polymers that are not normally observed with metals and ceramics, e.g. consequences of the polymer molecular structure and lamellar crystalline morphology.

The glass transition at temperature T_g occurs in amorphous (or glassy) and semicrystalline polymers, and is due to a reduction in motion of large segments of molecular chains with decreasing temperature. Upon cooling, this transition correspond to the gradual transformation from a liquid to a rubbery material, and finally to a rigid solid. However, this sequence of events occurs in reverse order when a rigid glass a temperature below T_g is heated. Moreover, abrupt changes in other physical properties also observed with glass transition; e.g. stiffness, heat capacity and coefficient of thermal expansion.

13. POLYMER TYPES

As stated in section 3, polymers are broadly classified in three categories:

(i) Plastics: These are the organic materials that can be easily shaped or moulded by mechanical or chemical action, with or without the application of heat. Obviously, a plastic is a permanently deformable or mouldable material. Probably, the largest number of different polymeric materials fall under this category. Polyethylene, polypropylene, polyvinyl chloride, polystyrene, and fluorocarbons, epoxies, phenolics and polyesters may be put under this category. Plastics have a wide variety of combination of properties. There are some plastics which are very rigid and brittle, whereas others are flexible, exhibiting both elastic and plastic deformations when stressed, and some time experiencing sufficient deformation prior to fracture.

Polymers belonging to this class of materials may have any degree of crystallinity, and all molecular structures and configurations (i.e., linear, branched, isotactic, etc.) are possible. Plastic materials may further be classified as thermoplastic or thermosetting. Table 17.4 summarizes the characterstics and typical applications for a number of plastics. Plastic have following properties:

(i) Plastics are light weight materials that is their densities and specific gravities are low.
(ii) They have high resistance to chemical attack, i.e. corrosion.
(iii) They have very good surface finish when obtained from the dies.
(iv) They have high thermal and electrical insulation.
(v) As compared to metals their strength is low and also low elastic modulus value.
(vi) As compared to metals, plastic have low softening and thermal degradation temperature.
(vii) Plastics are poor to fair dimensional stability, particularly in moist conditions.

Table 17.4 Characteristics and typical applications of some plastic materials

Plastic material (Type) (1)	*Trade names (2)*	*Characteristics (3)*	*Typical applications (4)*
		Thermoplastics	
Acrylonitrile-butadiene-styrene (ABS)	Abson Cycolac Kralastic Lustran Novodur Tybrene	Outstanding strength and toughness, resistant to heat distortion; good electrical properties; flammable and soluble in some organic solvents	Refrigerator linings, lawn and garden equipment, toys, high-way safety devices
Acrylics (poly-methyl methacrylate)	Acrylite Diakon Lucite Plexiglas	Outstanding light transmission and resistance to weathering; only fair mechanical properties	Lenses, transparent aircraft enclosures, drafting equipment, outdoor signs
Fluorocarbons (PTFE or TFE)	Teflon Fluon Halar Halon Hostaflon TF	Chemically inert in almost all environments, excellent electrical properties; low coefficient of friction; may be used to 260°C (500°F); relatively weak and poor cold-flow properties	Anticorrosive seals, chemical pipes and valves, bearings, antiadhesive coatings, high-temperature electronic parts
Polyamides (nylons)	Nylon Durethan Herox Nomex Ultramid Zytel	Good mechanical strength, abrasion resistance, and toughness; low coefficient of friction; absorbs water and some other liquids	Bearings, gears, cams, bushings, handles, and jacketing for wires and cables
Polycarbonates	Baylon Iupilon Lexan Makrolon Merlon Nuclon	Dimensionally stable; low water absorption; transparent; very good impact resistance and ductility; chemical resistance not outstanding	Safety helmets, lenses, light globes, base for photographic film
Polyethylene	Alathon Alkathene Ethron Fortiflex Hi-fax Petrothene Rigidex Zendel	Chemically resistant and electrically insulating; tough and relatively low coefficient of friction; low strength and poor resistance to weathering	Flexible bottles, toys, tumblers, battery parts, ice trays, film wrapping materials
Polypropylene	Bexphane Herculon Meraklon Moplen Poly-pro Pro-fax Propathene	Resistant to heat distortion; excellent electrical properties and fatigue strength; chemically inert; relatively inexpensive; poor resistance to UV light	Sterilizable bottles, packaging film, TV cabinets, luggage

(*Contd.*)

Table 17.4 (*Contd.*)

Plastic material (Type) (1)	*Trade names (2)*	*Characteristics (3)*	*Typical applications (4)*
Polystyrene	Carinex Celatron Hostyren Lustrex Styron Vestyron	Excellent electrical properties and optical clarity; good thermal and dimensional stability; relatively inexpensive	Wall tile, battery cases, toys, indoor lighting panels, appliance housings
Vinyls	Darvic Exon Geon Pee Vee Cee Pliovic Saran Tygon	Good low-cost, general-purpose materials; ordinarily rigid, but may be made flexible with plasticizers; often copolymerized; susceptible to heat distortion	Floor coverings, pipe, electrical wire insulation, garden hose, phonograph records
Polyester (PET or PETE)	Celanar Crastin Dacron Hylar Melinex Mylar Terylem	One of the toughest of plastic films; excellent fatigue and tear strength, and resistance to humidity, acids, greases, oils, and solvents	Magnetic recording tapes, clothing, automotive tire cords, beverage containers
		Thermosetting Polymers	
Epoxies	Araldite Epikote Epon Epi-rez Lekutherm Nepoxide	Excellent combination of mechanical properties and corrosion resistance; dimensionally stable; good adhesion; relatively inexpensive; good electrical properties	Electrical moldings, sinks, adhesives, protective coatings, used with fiberglass laminates
Phenolics	Bakelite Amberol Arofene Durite Resinox	Excellent thermal stability to over 150°C (300°F); may be compounded with a large number of resins, fillers, etc.; inexpensive	Motor housings, telephones, auto distributors, electrical fixtures
Polyesters	Aropol Baygal Derakane Laguval Laminac Selectron	Excellent electrical properties and low cost; can be formulated for room- or high-temperature use; often fiber reinforced	Helnets, fiberglass boats, auto body components, chairs, fans

(viii) Plastics are available in different colours as well as transparent form.
(ix) With respect to metals, plastics have high coefficient of thermal expansion.
(x) These are unsuitable for high temperature applications.
(xi) Plastics can be easily fabricated.

There are many plastics with outstanding properties. Polystyrene and polymethyl methacrylate are especially useful for applications in which optical transparency is critical, however, it is imperative that the material be highly amorphous or, if semicrystalline, have very small crystallites. The fluorocarbons have a very low coefficient of friction and moreover they are extremely resistant to attack by a host of chemicals, even at relatively high temperatures. They find wide range of uses as coatings. On nonstick cookware, in bearings and bushings, and for high temperature electronic components.

(ii) Elastomers: These materials have properties similar to polymers. These materials may be repeated stretched or elongated and will return to their original condition upon release of the force producing the elongation. They, exhibit elastic behaviour, as compared to polymers, which exhibit greater plastic properties. Included in this category are rubber and rubber like materials. A distinction is made between rubber and elastomers. Rubber must withstand a 200% elongation and rapidly return to its original dimensions. This property of a material to recover from elastic deformation is termed as *resilience.* Table 17.5 summarizes properties and applications of few common elastomers.

There are several criteria which have to be met for a polymer to be elastomeric: *(i)* It must be easily crystallize. We may note that elastomeric materials are amorphous, having molecular chains that are naturally coiled and kinked in the unstressed state, *(ii)* For the coiled chains, chain bond rotations must be relatively free so that they may readily respond to an applied force, *(iii)* To experience relatively large

Table 17.5 Important characteristics and typical applications of few common elastomers

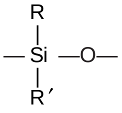
Material (Chemical Type)	*Trade/ common name*	*Elongation (%)*	*Useful temperature range [°C]*	*Characteristics*	*Applications (Typical)*
Natural polyisoprene	Natural rubber (NR)	500-760	–60 to 120	Excellent physical properties; good resistance to cutting, gouging, and abrasion; low heat, ozone, and oil resistance; good electrical properties	Pneumatic tyres and tubes; heels and soles; gaskets
Styrene-butadiene copolymer	GRS, Buna S (SBR)	450-500	–60 to 120	Good physical properties; excellent abrasion resistance; not oil, ozone, or weather resistant; electrical properties good, but not outstanding	Same as natural rubber
Acrylonitrile-butadiene copolymer	Buna A, Nitrile (NBR)	400-600	–50 to 150	Excellent resistance to vegetable, animal, and petroleum oils; poor low-temperature properties; electrical properties not outstanding	Gasoline, chemical, and oil hose; seals and O-rings; heels and soles
Chloroprene	Neoprene (CR)	100-800	–50 to 105	Excellent ozone, heat, and weathering resistance; good oil resistance; excellent flame resistance; not as good in electrical applications as natural rubber	Wire and cable; chem. tank linings; belts, hoses, seals, and gaskets
Polysiloxane	Silicone (VMQ)	100-800	–115 to 315	Excellent resistance to high and low temperatures; low strength; excellent electrical properties	High-and low- temperature insulation; seals, diaphragms; tubing for food and medical uses

elastic deformations for elastomers, the onset of the plastic deformation must be delayed. This objective is achieved by restricting the motions of chains past one another by crosslinking. The crosslinks acts as anchor points between the chains and prevent chain slippage to occur. In many elastomers, crosslinking is carried out in a process called vulcanization, *(iv)* The elastomer must be above its glass transition temperature. The lowest temperatures at which rubber like behaviour persists for good number of common elastomers is between –50 and –90°C. An elastomer becomes brittle below its glass transition temperature.

The properties of common elastomers are typical and, of course, depend on the degree of vulcanization and on whether any reinforcement used. Natural rubber has an outstanding combination of desirable properties and still utilized to a large degree. However, the most important synthetic elastomer is SBR (styrene butadiene copolymer) which is used especially in automobile tyres, reinforced with carbon black. Acrylonitrile butadiene copolymer (NBR) is another common synthetic elastomer, which is highly resistant to degradation and swelling.

The mechanical properties of even vulcanized rubbers for many applications, e.g. automobile tyres, are not satisfactory in terms of tensile strength, abrasion and tear resistance, and stiffness. One may improve these characteristics by additives such as carbon black.

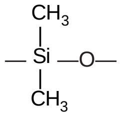
Nowadays, silicon rubbers are widely used for several applications. For these materials the backbone carbon chain is replaced by a chain that alternates Si and oxygen atoms (Fig. 17.40) where *R* and *R′* represent the side-bonded atoms such as hydrogen or group of atoms, e.g. CH_3. As an example, the mer structure of polydimethylsiloxane is (Fig. 17.41). No doubt, as elastoners these materials are crosslinked.



$$-\overset{R}{\underset{R'}{\mathrm{Si}}}-\mathrm{O}-$$

Fig. 17.40

At low temperatures, the silicone elastomers possess a high degree of flexibility (–90°C) and yet these are stable to temperatures as high as 250°C. Moreover, they are resistant to weathering and lubricating oils. A more attractive characteristic is that some silicone rubbers vulcanize at room temperature (RTV rubbers).



$$-\overset{CH_3}{\underset{CH_3}{\mathrm{Si}}}-\mathrm{O}-$$

Fig. 17.41

Thermoplastic elastomers, sometimes referred to as *elastoplastics,* result from the copolymerization of two or more monomers. One monomer is used to provide the hard, crystalline features, whereas the other monomer produces the soft, amorphous features. When combined, these form a thermoplastic material that exhibits properties similar to the harder, vulcanized elastomers. Thermoplastic elastomers can be either molded or extruded into finished shapes. Urethanes were the first thermoplastic elastomers to be produced. These products are used in such applications as gaskets, gears, and fuel lines. Other thermoplastic elastomers include copolyester, styrene, styrene-butadiene, and the olefins. Copolyester elastomers have more properties but are more expensive than urethanes. They are used as hydraulic hoses, couplings, and cable insulation. Styrene copolymers are less expensive than either urethanes or copolyesters. They have a lower tensile strength but exhibit greater elongation. SBR is used in medical products, tubing, packaging materials, adhesives, and sealants. Thermoplastic olefins (TPO) are used in tubing, seals, gaskets, and electrical products. The molecular structure for a thermoplastic elastomer is shown in Fig. 17.42.

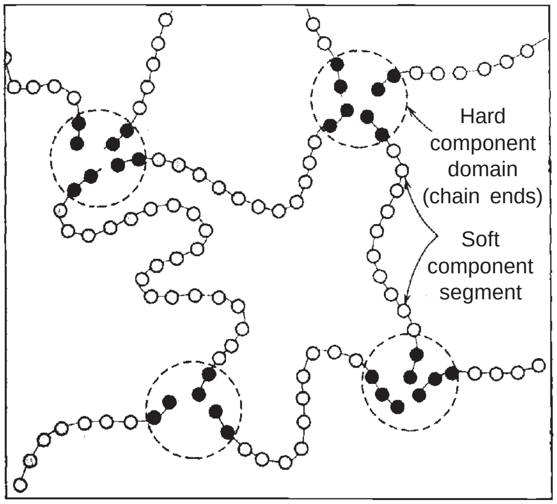


Fig. 17.42 Molecular structure for a thermoplastic elastomer. It consists of ‘soft’ (i.e. butadiene or isoprene) mer centre-chain segments and hard (i.e. styrene) domains (chain ends), which act as physical crosslinks at room temperature

(iii) Fibres: These consists of long molecular chains and all the chains are aligned in the direction of length of the fibre. The directional properties are improved by alignment, i.e. the strength of fibres are greatly enhanced in the direction parallel to the fibre length. These provide strength to the molecular units.

The fibre polymers are quite capable of being drawn into long filaments having at least a 100 : 1 length to diameter ratio. Some polymers used as fibres, such as nylon and cellulose acetate, serve equally well as plastics. Most commercial fibre polymers are utilized in the textile industry, being woven or knit into cloth or fabric. In addition, the aramid fibres are widely employed in *composite* materials. A fibre polymer, to be used as a textile material must have some restrictive physical and chemical properties. Fibres, while is use may be subjected to a variety of mechanical deformations, e.g. stretching, twisting, shearing and abrasion. Obviously, they must have a high tensile strength over a relatively wide temperature range and a high modulus of elasticity and also abrasion resistances. We may note that these properties are governed by the chemistry of polymer chains and also by fibre drawing process.

Fibres materials should have relatively high molecular weight. The structure and configuration of the chains should allow the production of a highly crystalline polymer as the tensile strength increases with degree of crystallinity. This translates into a requirement for linear and unbranched chains that are symmetrical and have regularly repeating mer units.

Synthetic fibres are produced by the following methods: *(i)* Melt spinning *(ii)* Dry spinning and wet spinning. To improve fibres properties, i.e., to increase/decrease resistance, water or fire proofs, etc., chemical treatment is done. The choice of method mainly depends upon the properties of polymers, i.e., heat stability, melting point and solubility in suitable solvents.

The thermal properties of fibre polymers are important from the point of view of washing and maintainin clothes. These properties include melting and glass transition temperatures. Also, fibre polymers must show chemical stability to a rather extensive variety of enviornments, including acids, bases, dry cleaning solvents, bleaches, and sunlight. Moreover, they must be relatively non flammable and amenable to drying.

Metallic and ceramic fibres have also similar qualities as that of polymeric fibres. However, metallic and ceramic fibres have very high strength and therefore they have many industrial applications. There are also natural fibres, e.g. cotton, silk, wool, cellulose, etc. Cellulose fibres are flexible and are strong in tension. Cotton is produced naturally as a plant fibre. However, the synthetic fibres are cheap, durable and have more dimensional stability and that is why they have almost replaced the natural fibres.

14. MISCELLANEOUS APPLICATIONS OF POLYMERS

In order to serve one or more functions, e.g. *(i)* to protect the item from the environment that may produce corrosive or deteriorative reactions, *(ii)* to improve the appearance of item, and *(iii)* to provide electrical insulation, coatings are frequently applied to the surface of materials. Many of the ingredients in coating materials are polymers and majority of which are organic in origin. One finds that these organic coatings fall into several different classifications, e.g. paint, varnish, enamel, lacquer, and shellac.

Adhesives: This is a substance which is used to join together surfaces of two solid materials (usually termed as “adherends”) to produce a joint with a high shear strength. The inherent strength of the adhesive may be much less than that of the adherend materials, nevertheless, a strong joint may be produced if the adhesive layer is thin and continuous. If a good joint is formed, the adherend material may fracture or rupture before the adhesive.

Polymeric materials belonging to the groups: thermoplastics, thermosetting resigns, elastomeric compounds, and natural adhesives (animal glue, casein, starch, and rosin) may serve adhesive functions. Polymers adhesives are widely used to join a large variety of material combinations; e.g. metal-metal, metal-plastic, metal-ceramic, etc. The service temperature limitation is a major primary drawback. The strength of organic polymers decreases rapidly with rise in temperature but they maintain their mechanical integrity only at relatively low temperatures.

Films: Polymeric materials have recently found widespread use in the form of *thin films.* Films, whose thicknesses lies between 0.025 and 0.125 mm are fabricated and widely used as bags for packaging food products and other items, as textile products, and a host of other uses. The main characteristics of the materials produced and used as films include low density, a high degree of flexibility, high tensile and tear strengths, resistance to attack by moisture and other chemicals, and low permeability to some gases, especially water vapour. The polymers which satisfy these criteria and are manufactured in film form include polyethylene, polypropylene, cellophane, and cellulose acetate, etc.

Foams: These are the plastic materials which contain a relatively high volume percentage of small pores. Thermoplastic as well as thermosetting materials are used as foams. These materials include polyurethane, rubber, polystyrene, and polyvinyl chloride. Foams are widely used as cushions in automobiles and furniture as well as in packaging and thermal insulations. The process of foaming is carried out by incorporating into the batch of material a blowing agent that upon heating, decomposes with the liberation of a gas. Gas bubbles are created throughout the now fluid-mass, which remains as pores upon cooling and yield to a sponge-like structure. One can obtain the same effect by bubbling an inert gas through a material while it is in molten state.

15. ADVANCED POLYMERIC MATERIALS

In the past few years a number of new polymer materials having unique and desirable combinations of properties have been developed. Many of these polymer materials have found niches in new technologies and also replaced several other materials. Some of these materials include: *ultrahigh molecular weight polyethylene, liquid crystal polymers,* and *thermoplastic elastomers.*

(i) Ultrahigh Molecular Weight Polyethylene (UHMWPE): Most linear polyethylenes have weight-average molecular weights in the range 100,000-200,000, commercially two higher-molecular-weight grades are widely used.

High-Molecular weight (HMW) polyethylene has $\overline{M}_w$ in the range 300,000-500,000. This can be processed by usual techniques and moreover, it has improved environmental stress crack resistance, impact and tensile strength, and long-term strength retention. HMWPE is used for pipe, film and large blow-molded containers where these properties are important.

UHMWPE has $\overline{M}_w$ in the range 3,000,000-6,000,000 g/mol, which an order magnitude greater than that of high density polyethylene. In fibre form, the trade name of UHMWPE is *spectra.* These materials exhibit following extraordinary characteristics: *(a)* exceptionally high impact-resistance, *(b)* outstanding resistance to wear and abrasion *(c)* very low-coefficient of friction *(d)* extremely good chemical resistance to normally encountered solvents *(e)* self-lubricating and nonstick surface *(f)* high sound damping and energy absorption characteristics *(g)* excellent low-temperature properties *(h)* excellent dielectric properties and electrically insulating.

However, the mechanical properties of this material diminish rapidly with decreasing temperature as it has relatively low melting temperature.

This material does not melt or flow in the normal thermoplastic manner because of its extremely high molecular weight. Therefore, most of its fabrication is based on modifications of the *compression-molding technique*.

This material finds numerous and diverse applications due to its unusual combination of properties, e.g. bullet proof vests, composite military helmets, ski-bottom surfaces, fishing line, bowling alley, golf ball cores, bearings, sprockets, gaskets, value seats, conveyor-belt parts and ice-skating rink surfaces, biomedical prostheses, blood filters, marking pen nibs, pump impellers, bushings, material handling equipment (for coal, cement, gravel, grain, etc.).

(ii) Liquid Crystal Polymers (LCPs): These polymer materials are a group of chemically complex and structurally distinct materials. These materials have unique properties and used in diverse applications.

These materials are composed of extended, rod-shaped, and rigid molecules. In view of their molecular arrangement, LCPs do not fall within any of conventional liquid, amorphous, crystalline, semicrystalline classifications, but these may be considered as a new state of matter, i.e., the *liquid crystalline state,* being neither crystalline nor liquid. In the melt (or liquid) condition, other polymer molecules are randomly oriented, whereas LCPs molecules can become aligned in highly ordered configurations. Based on orientation and positional orderings, there are three types of liquid crystals: *smectic*, *nematic* and *cholesteric*.

LCPs finds principal use in liquid crystal displays (LCDs), on digital watches, laptop computers, and also for other digital displays. Here cholesteric types of LCPs are used, which, at room temperature, are fluid liquids, transparent, and optically anisotropic. The displays consist of two sheets of glass between them is sandwitched the LCP material. The outer face of each glass is coated with a transparent and electrically conductive film and also, in addition, into this film on the side that is viewed or etched the character-forming number/letter elements. When a voltage is applied through the conductive films, i.e. between these two glass sheets, over one of these character-forming regions causes a disruption of the orientation of the LCP molecules in this region, a darkening of LCP material, and, in turn, the formation of a visible character.

There are some nematic type of LCPs, which are rigid solids at room temperature. The outstanding combination properties and processing characteristics of these LCPs have found uses in variety of commercial applications. These LCPs materials exhibit the following behaviours:

(a) They exhibit excellent thermal stability and may be used upto 230°C.
(b) They are stiff and strong. Their tensile moduli range between 10 and 24 GPa and tensile strengths between 125 to 255 MPa.
(c) They have high impact strengths, which are retained upon cooling to relatively low temperatures.
(d) They exhibit chemical interness to a wide variety of acids, bleaches, solvents, etc.
(e) They show inherent flame resistance and combustion products that are relatively nontoxic.

As regards the processing and fabrication characteristics of these materials, we have

(a) One may use all conventional processing techniques available for thermoplastic materials.
(b) These materials have extremely low shrinkage and warpage take place during molding.
(c) These materials have exceptional dimensional repeatability from part to part.
(d) The low melt viscosity of these materials permits molding of thin sections and/or complex shapes.
(e) The low heats of fusion of these materials result in rapid melting and subsequent cooling. This helps to shorten molding cycle times.
(f) The anisotropric finished part properties of these materials mean that molecular orientation effects are produced from melt flow during molding.

LCPs are widely used by electronics industry (interconnect devices, relay and capacitor housings, brackets, etc.), by the medical equipment industry (in components to be repeatedly sterilized), and in photocopiers and fibre optic components.

(iii) Thermoplastic Elastomers: See section 13.

16. POLYMER ADDITIVES

The properties of the polymers discussed so far are intrinsic ones, i.e., characteristic of or fundamental of or fundamental to the specific polymer. Many times, one find it necessary to modify the mechanical, chemical, and physical properties to a greater degree than is possible by the simple alteration of this fundamental molecular structure. To enhance or modify many of these properties, foreign substances called *additives* are knowingly introduces so that a polymer material may found more serviceable. Following are the typical additives:

(i) Fillers: Many plastics are virtually useless alone but are converted into highly serviceable products by combining them with particulate or fibrous solids. This improves their tensile and compressive strengths, abrasion resistance, toughness, dimensional and thermal stability, and other properties. Materials used as particulate fillers include silica flour and sand, wood flour (finely powered sawdust), glass, clay, talc,

limestone and also some synthetic polymers. The principal fillers used can be divided into two types: *particulate* and *fibrous.* Among the particulate fillers are silica products, including sand, quartz and diatomaceous earth; silicates, including clay, mica, talc, asbestors, etc. Fibrous fillers, some quite old (cellulosics) and some new "space-age" products (metallic fibres, whiskers) include cellulosic fibres, such as alpha-cellulose and cotton flock; synthetic fibres, etc. Particle sizes range all the way from 10 nm to macroscopic dimensions. In recent years the use of mineral fillers has increased markedly, with emphasis on calcium carbonate. As these inexpensive materials replace some volume of the more expensive polymer, the cost of final product is reduced.

(ii) Plasticizers: These are added to plastic materials to improve flow and, therefore, processability, and to reduce the brittleness of the product. Thus the addition of plasticizers to plastics improves their flexibility, ductility, and toughness of polymers. Moreover, their presence also produces reductions in hardness and stiffness. Usually, plasticizers are liquids having low vapour pressures and low molecular weights. The small plasticizer molecules occupy positions between the large polymer chains thereby effectively increasing the interchain distance with a reduction in the secondary intermolecular bonding. Plasticizers are commonly used in polymers to make them soft, flexible and tough materials from those of a hard, brittle, glass like solid. An example is the plasticization of poly (vinyl chloride) and vinylchloride acetate copolymers. One can also achieve similar changes in properties, of course, by alterating the molecular structure of polymer (e.g. by copolymerization, sometimes called *internal plasticization*). We may note that the plasticizer lowers the glass transition temperature, so that at ambient conditions one may use the polymers in applications requiring some degree of pilability and ductility. These applications are thin sheets or films, raincoats, tubing, and curtains.

(iii) Stabilizers: Under normal environmental conditions, some polymeric materials are subject to rapid deteriorian, usually in terms of mechanical integrity. Usually, this deterioration is a result of exposure to light, in particular ultraviolet (UV) radiation, and also oxidation. UV radiation interacts with and causes a severance of some of the covalent bonds along the molecular chain, which may also result in some crosslinking. Oxidation deterioration is a consequence of the chemical interaction between oxygen atoms and the polymer molecules. Additives that counteract these deteriorative processes are named as *stabilizers.* An important example is vinyl resins. Other applications of stabilizers include the use of carbon black to prevent photochemical degradation (by excluding light), as in polyethylene, and of ultraviolet light absorbers, such as hydrobenzophenones, to improve the light stability of both plastics and their colorants.

(iv) Colorants: Colorants for plastics include a wide variety of inorganic and organic materials. These impart a specific colour to a polymer. Colorants may be added in the form of dyes or pigments. A few are molecularly dispersed (oil-soluble dyes) or have small particle size and a refractive index near that of the plastic (*organic pigments* such as the phthalocyanines) and lead to transparent coloured products when incorporated into transparent plastics. Others, including *inorganic pigments,* impart opacity to the plastic. Common colorants for plastics include, among many others, titanium dioxide and barium sulphate (white), phthalocyanine blues and greens, ultramarine blues, chrome greens, quinacridone reds and magnets, molybdate orange, cadmium reds and yellows, iron oxide and chrome yellow, carbon black, flake aluminium for a silver metallic effect, and lead carbonate or mica for pearlescence.

The molecules in a dye dissolve and become part of the molecular structure of the polymer, whereas pigments are filler materials which do not dissolve and remain as a separate phase, i.e., usually they have a small particle size, are transparent, and have a refractive index just equal to that of the parent polymer. Other may impart *opacity* as well as *colour* to the polymer as mentioned above.

(v) Flame Retardants: The flammability is a major concern for polymeric materials, especially in the manufacturing of textile goods and children's toys. In pure form, most of the polymers are flammable; exceptions are those polymers which contain significant contents of chlorine and/or fluorine, such as polyvinyl chloride and polytetrafluoroethylene. One can enhance the flammability resistance of the remain-

ing combustible polymers by additives called *flame retardants.* These flame retardants may function by initiating a chemical reaction that causes a cooling of the combustion region and a cessation of burning. The most useful material imparting flame retardance to plastics is antimony trioxide.

17. MANUFACTURING PROCESSES INVOLVING POLYMERS

Polymers are produced from a variety of naturally occurring materials, including wood, air, water, petroleum, natural gas, and salt, etc. Chemists separate these naturally occurring materials into their basic elements and then reassemble these basic elements into different, distinct polymers using heat, pressure and chemical action. For example, polystyrene is produced from coal and petroleum or natural gas. The chemist begins by extracting benezene from coal and ethylene, gas from the natural gas or petroleum. The basic elements benzene and ethylene, are combined to form ethylbenezene. Through the application of heat and pressure, the ethylbenzene is formed and ground to produce polystyrene beads. Several polymers are available in powder, pellet, or premixed form.

Many polymers are also available as liquid casting system. Epoxies are an example of a liquid casting system. Another type of casting polymer is *plastisol.* A plastisol is a liquid vinyl polymer that solidifies when it contacts a heated surface. Plastisols are widely used in dipping processes for insulated coatings.

Polymers are available in expanded or expandable forms. Expanded polymers are low-density cellular polymers available in two varieties, open-and closed cell polymers. These expanded, or *foamed, polymers* come in two forms, rigid and flexible. An example of a rigid, foamed polymer is polystyrene, or styrofoam, which is widely used in packaging. One type of flexible foamed polymer is urethane, which is used in seat cushions and craft work. Foam insulation is an example of expandable polymers. Expandable polymers are used to fill voids and provide thermal and acoustic insulation.

Several methods are used to fashion products from polymers. The method used for the specific polymer depends on several factors: *(a)* whether the polymer material is thermoplastic or thermosetting; *(b)* if material is thermoplastic, the temperature at which if softens; *(c)* the atmospheric stability of the material being formed; and *(d)* the geometry and size of the finished product.

Several methods are used to fashion products from polymers. Some of the more common production methods are casting, blow moulding, compression moulding, transfer moulding, injection moulding, extrusion, lamination, vacuum forming, cold forming, filament winding, calendaring, and foaming. Each of these methods has advantages and disadvantages, depending on the type of material used, the product design, and the cost involved.

(i) Castings: This is the simplest of the manufacturing methods, because no fillers are used and no pressure is applied. However, a mould is required. There are four basic type of casting methods: *simple casting, dip casting, slush casting,* and *rotational casting.* In the simple casting process, a prototype, or model, of the material is made first. Dip casting requires a preheated mandrel, a dipping station, a curing oven, a cooling station, and a method to remove the finished casting. Slush casting uses a split method, which has a internal cavity formed in the shape of the finished part. Rotational casting or rotational moulding utilizes a predetermined amount of plastisol, which is placed inside a preheated split mould.

(ii) Moulding: This is the widely used common method for forming plastic polymers. There are several moulding techniques used, which include compression, transfer, blow, injection and extrusion moulding. For each of the said technique, a finely pelletized for granulized plastic is forced, at an elevated temperature and by pressure, to flow into, fill, and assume the shape of the mould cavity.

Compression and Transfer Moulding

In compression moulding, the moulding powder is compressed between two parts of heated metal mould (Fig. 17.43). Both mould pieces are heated, however, only one is movable. The mould is closed, and heat and pressure are applied, causing the plastic material to become viscous and conform to the mould shape. Prior to moulding, raw material may be mixed and cold pressed into a disc, which is called a preform.

Preheating of the preform reduces moulding time and pressure, extend the die lifetime, and produces more uniform finished piece.

This technique is suitable for large and small parts of thermosetting plastics as well as large and bulky parts of thermoplastics. However, it is used with thermoplastics, is more time consuming and expensive.

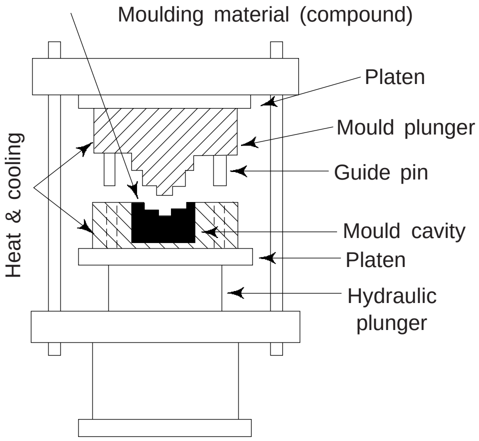


Fig. 17.43 Compression moulding apparatus

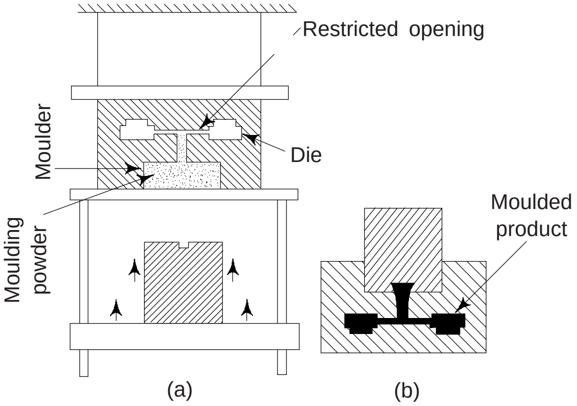


Fig. 17.44 Transfer moulding apparatus.
(a) Moulding powder in chamber
(b) Resin forced into mould

Transfer moulding is generally used for thermosetting plastics, specially for manufacturing small parts having small deep holes, cavities, metal inserts, etc., because compressive moulding is not suitable for these applications. In this technique, a variation of compression moulding, the solid ingradients are first melted in a heat transfer chamber. As the molten material is injected into the mould chamber, the pressure is distributed more uniformly over all surfaces (Fig. 17.44). This procedure is more suitable for thermosetting polymers and for pieces having complex geometries.

Injection Moulding

This is used to form thermoplastic materials. Commonly used thermoplastic polymers include acrylics, fluorocarbons, nylon, polyethylene, polystyrene, and vinyls. It is an extremely fast method of production. A schematic cross-section of the apparatus is shown in Fig. 17.45. Raw material is fed into a hopper, which is screw-fed into the pressure chamber. As the screw turns, the material is fed into the heating chamber, where it is melted. The melted material continues through a nozzle that is seated against the mould. The nozzle allows the molten material to flow into the cavity through a system of gates and runners. The die remains relatively cool; thus, the product solidifies almost upon filling the cavity. To prevent the premature cooling (and resultant solidifying) of the molten material, the material is forced through the barrel into the

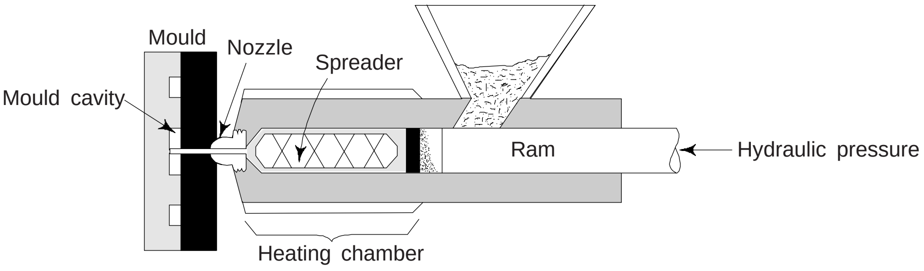


Fig. 17.45 Injection moulding apparatus

cavity under pressure. The mould is then opened and the part is ejected. The method is used to produce such items as toys, plastic models and parts, tile, food containers, battery cases, and radio cabinets.

Extrusion Moulding

This is simply injection moulding of a viscous thermoplastic through an open-ended die. This process is used to form materials that have long, uniform-cross-section. It is typically involved in forming thermoplastic polymers, such as acrylics, cellulosics, fluorocarbons, nylon, styrene, polyethylene and vinyls. In this process raw material is fed into a hopper, where it is gravity-fed into a screw chamber. In a screw chamber, a rotating screw forces the material into and through a preheated section. Here the material is compressed and then forced through a heated die onto a conveyor. As the material leaves the die and lands on the belt, it is cooled by an air or water spray to harden it sufficiently, preserving its shape during transportation (Fig. 17.46). Usually, the material is then cut to length or coiled. This process can be used to produce any material that has a uniform cross-section, such as pipe, tubing, rods, and sheets. It is also used to form pellets for injection moulding. Extrusion is very fast and allows for *continuous production*.

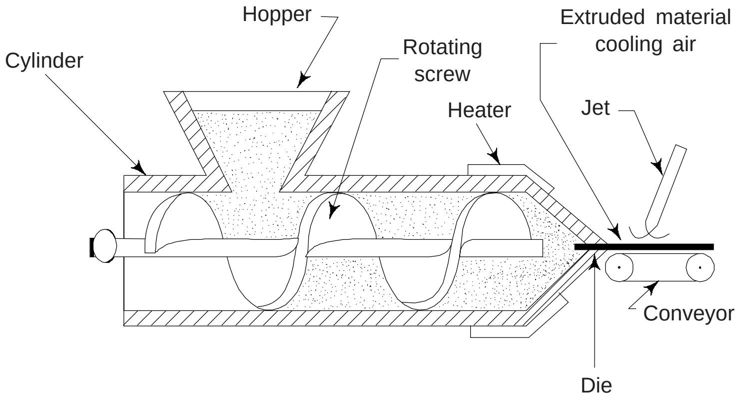


Fig. 17.46 Continuous extrusion moulding apparatus

Extrusion Blow Moulding

This process is used to form hollow shapes, such as bottles and containers. A thermoplastic tube called a *parison* is fed through the opening formed by the two halves of split mould. Once the necessary length of the parison has entered the mould, the two halves close, sealing one end of the parison, forcing the parison to expand, following the contour of the inner mould cavity. This mould is kept cool during the blow moulding process, so that the thermoplastic parison solidifies as it comes into contact with the mould. Once the material has cooled, the mould is opened and the part is removed. The two other types of blow moulding are *injection blow moulding* and *stretch blow moulding*. These differ in the way the raw material is introduced into the mould. The blow moulding process is shown in Fig. 17.47. In this process the temperature and viscosity of the parison have to be carefully regulated.

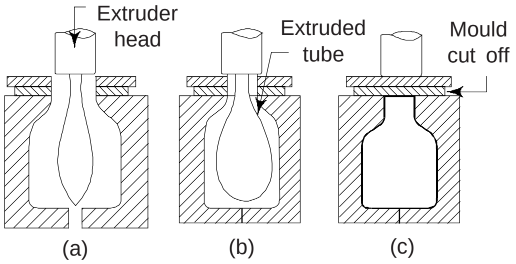


Fig. 17.47 Extrusion blow moulding (a) mould open (b) mould closed, tube expanding under pressure and (c) tube fully blows

18. REINFORCED POLYMERS

These are polymers whose properties have been altered through the use of reinforcement material. Glass, mineral and carbon fibres can be added to polymers to alter their properties. Most polymers are available as glass fibre-reinforced products. The glass fibres, whose diameters range from 0.0002 to 0.001 inch in diameter, are coated with a *resin* and a coupling agent. Property increase of 200% or better can be obtained through the glass fibre reinforcement.

The strength of these reinforced polymers depends on the type and quantity of *filler* used and the degree to which the resin has wetted the reinforcement material. One of the more common filler materials is glass fibre.

Carbon fibre-reinforced materials are more expensive than fibre-reinforced materials. Carbon fibre-reinforced materials have high tensile strength, stiffness, and greater related mechanical properties. They also produce lower coefficients of expansion, improved creep resistance, better wear resistance, greater toughness, and higher strength-to-weight ratios than glass fibre-reinforced materials.

Glass and carbon are the most common fibre reinforcements. Other materials used in fibre reinforcements include graphite, boron, and cotton.

Reinforcement materials are available in several forms, including mat, fibre bundles, chopper fibres, and cloth.

Most of the reinforced thermosetting polymers (90% to 95%) are polyesters and epoxies, with the majority being polyesters.

These reinforced polymers are manufactured through a variety of methods, including hand lay-up, spray-up, matched moulding, premixed moulding, and vacuum-or pressure bag casting.

These materials are replacing metals in a variety of applications. They offer similar or better strength, characteristics at lower weight and often, at lower cost. They are frequently easier to produce, are resistant to corrosion and chemicals, and offer a wide variety of manufacturing processes, from the very simple and inexpensive to the higher volume and more intricate processes.

19. BEHAVIOUR OF POLYMERS

(i) Electrical Behaviour: Polymer materials are very god insulators. They can be easily applied on the wires as polymerized liquid as well as partially polymerized liquid to provide a *uniform coating*. This coating gets polymerized around the wire. There are some plastics which allow the formation of film easily and therefore such plastics are used for wire insulation.

Polymers do not exhibit any magnetic properties. However, when polymers are combined with filler materials, e.g. the addition of finely powdered graphite, they exhibit magnetic properties and conduct electricity. Obviously, polymers can be made conductive by the inclusion of second conductive phase. One can achieve the conductivity in polymers through the following methods: *(a) mistreating by gamma ray irradiations*. This exposes the chain and bond regularities that contain donor and acceptor sites. Resistivities in the range 1 Ω-m to 10^{10} Ω-m can be achieved. (b) by producing a molecular chain which has radicals. These radicals provide donor or acceptor site. One can achieve resistivities as low as 0.12 Ω-m.

Those polymers which do not absorb water are good insulators, e.g. polyethylene, polytetrafluorethylene PTFE, polyvinyle chloride. Tuflon and perpex are rigid insulators.

The polymers which are sensitive to polarization of structure exhibit *dielectric* properties. The polymers which contain natural dipoles, the value of dielectric constant is more, e.g. the polymers containing polar groups, e.g. chloride, hydroxal or sulphur atom readily become electrically polarized when placed in a electric field, resulting in high dielectric constant.

The dipoles at low frequencies respond to the changes in electric field. With the rise in frequency, the response of polar groups decreases and only electronic polarization occurs. Figure 17.48 shows the behaviour of dielectric constant with frequency. The dielectric constant decreases with the increase of frequency and also with temperature.

The dielectric constant of plastics which contain plasticizers or fillers, increases artificially. This phenomenon is known as *interfacial polarization*

(ii) Mechanical Properties: See section 10.

(iii) Chemical Properties: Although the organic materials are covalently bonded but they can be ionized to some extent, e.g. acetic acid produces negative acetate radical and positive hydrogen ion, as follows:

$$CH_3COOH \rightarrow CH_3COO^- + H^+$$

Obviously, due to ionization of organic salts and when the reaction proceeds backward, a weak acid is generated.

The polymers which are capable of only limited ionization are termed as *polyelectrolytes*. These polyelectrolyte materials possess both covalent and ionic bonds. They are called ionic conductors of electricity. Since their diffusion rate is slow and hence conduction through them is lesser.

During chemical reaction, the chemical bonds break and recombine. However, for polymerization process, the chemical reactions are essential.

Most of the organic materials are flammable materials. Organic compounds react with oxygen and produce CO_2 and H_2O, i.e. complete combustion of hydrocarbon,

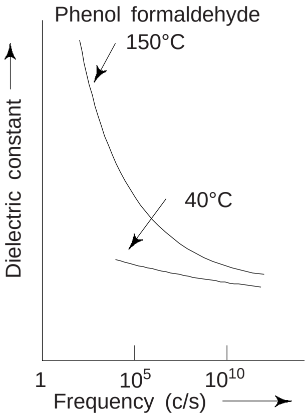


Fig. 17.48 Frequency versus dielectric constant

$$C_nH_m + \left(n + \frac{m}{4}\right) O_2 \rightarrow nCO_2 + \frac{m}{2} H_2O$$

During this process due to formation of CO_2 energy is released but energy is also required to dissociate the hydrocarbons into their constituent elements. Hydrocarbons dissociate into carbon and hydrogen and then again combine with oxygen to form CO_2 and H_2O. If hydrogen in the chain is replaced by some other element in the chain, the combustibility reduces.

The explosives have sufficient oxygen and therefore combustion in them is easy and quick. Such materials require negligible energy for their dissociation. The dissociation of cellulose nitrate and nitroglycerine are as follows:

$$\underset{\text{(Cellulose nitrate)}}{2C_6H_7(ONO_2)_3O_2 + 2O_2} \rightarrow \underset{\text{(Combustion products)}}{CO_2 + 11CO + 7H_2O + 6NO}$$

$$\underset{\text{(Nitroglycerine)}}{2C_3H_5(ONO_2)_3} \rightarrow \underset{\text{(Combustion products)}}{CO_2 + 5CO + 5H_2O + 6NO}$$

Nitroglycerine for its complete combustion does not require any external source of oxygen, whereas cellulose nitrate requires very little oxygen from air.

(iv) Hardness: Most plastic materials have low scratch resistance. One can improve the hardness of these materials by lamination with glass fibre.

(v) Thermal Properties: Due to depolymerisation, the polymerised thermosetting polymers lose their strength at high temperature. However, one can expose thermoplastic polymers to high temperatures.

Polymers become unsuitable at subzero temperature as they get embrittled at low temperature.

These materials show low thermal diffusivity, poor heat resistance, high flammability and poor ignition and burning capacity in comparison to metals and ceramics.

(vi) Optical Properties: Clear reflectivity, permanent colouring and lustre are some advantages of polymers. One can make them fully transparent to fully opaque as per requirement.

20. FABRICATION OF FIBRES AND FILMS

(i) Fibres: The process by which fibres are formed from bulk polymer material is called *spinning*. Most often, fibres are spun from the molten state in a process called melt spinning. First, the material to be spun is heated until it forms a relatively viscous liquid and then it is pumped down through a plate called a

spinnerette, which contains numerous small round holes. When the molten material passes through each of these orifices, a single fibre is formed, which solidifies almost immediately upon passing through the air.

The crystallinity of a spun fibre will depend on its rate of cooling during the process of spinning. The strength of fibres is improved by a *postforming process* called *drawing*. However, drawing is simply the mechanical elongation of a fibre in the direction of its axis. The molecular chains during this process become oriented in the direction of drawing, such that the tensile strength, modulus of elasticity, and toughness are improved to a certain extent. Although the mechanical strength of a drawn fibre is improved in this axial direction, strength is reduced in a transverse or radial direction. Moreover, since fibres are normally stressed only along the axis, this strength differential is not critical. We may note that the cross-section of drawn fibres is nearly circular, and the properties are uniform throughout the cross-section.

Films: Many films are simply extruded through a thin die slit and then this may be followed by a rolling operation that serves to reduce thickness and improve strength. Alternatively, film may be blown: continuous tubing is extruded through an annular die; then, by maintaining a very carefully controlled positive gas pressure inside the tube, one may reduce wall thickness continuously to produce a thin cylindrical film, which may be cut and laid flat. Some of the newer films are produced by coextrusion; that is, multilayers of more than one polymer type are extruded simultaneously.

21. WOOD

Wood is a natural, renewable, organic substance with a wide variety of uses. The main constituent of wood is cellulose. Of particular interest are trees that grow to sufficient height with diameters large enough to be useful in the production of lumber and wood products. Other plants, such as bushes, vines, and shrubs, are also wood, but these plants have little use in structural application.

Wood is widely available and easy to work, although it burns easily, has a variety of defects, and is subject to attack by insects and fungi. Therefore, special precautions must be taken to preserve it. When wood has been prepared for subsequent use in building and construction, it is called *lumber*.

Wood is generally divided into two easily recognizable areas, *heatwood* and *sapwood.* Very young trees are sapwood. Heatwood is usually called as *truewood*. Heatwood cells are dead and no longer contribute to the life processes of a tree.

Wood typically consists of tiny, hollow, fibrous cell that are composed of *cellulose* and *lignin*. The typical composition of wood is 60% cellulose, 28% lignin, and 12% other materials. Lignin acts primarily as a cement to hold the wood cells together, thus supporting the tree. Cellulose is made up primarily of carbohydrates produced by photosynthesis.

Trees are often classified as *hardwoods* or *softwoods*. This distinction is based on their cell structure and not on their mechanical properties. However, most hardwoods are harder and more durable than softwoods, but this relationship is not an absolute. In general, hardwoods are *deciduous*, or broad-leaves, trees, which drop their leaves in the fall. Softwoods are *conifers,* which bear needles instead of leaves and produce seed cones. The softest and hardest of all woods are found in the tropics; the lightest is balsa and the heaviest is iron wood, or lignum vitae. The principal distinction between the two types of wood is that hardwoods contain very large cells, known as *vessels*, or *pores*, whereas softwoods do not. In addition, hardwoods have wood *parenchyma* and ray cells. Wood parenchyma are thin, soft-walled cells that often surround the vessels and contribute to the grain pattern of the wood. The principal woods used are listed in Table 17.6.

Table 17.6 Principal types of wood

	Wood	*Uses*
(i)	*Softwoods*	
	Cedar	Clothes, closets, pencils
	Cypress	Tanks, silos, storage containers

(*Contd.*)

Table 17.6 (*Contd.*)

	Wood	*Uses*
	Douglas fir	Plywood
	Sitka spruce	Aircraft applications
	Southern pine	Railroad ties, boxes, trim
	Western white pine	Matches, boxes, crating
(ii)	*Hardwoods*	
	Ash	Tool handles, baseball bats
	Basswood	Boxes, crating
	Hickory	Tool handles
	Maple	Flooring, boxes, crating
	Oak	Posts, flooring, pallets, trim

Wood, for construction purposes, is primarily used as structural timber; light framing, siding; exterior finish; interior finish, sashes, doors, and frames; sheathing, roofing, and subflooring; and flooring. The principal woods used for each of these purposes are shown in Table 17.7.

Table 17.7 Woods used for construction

Use	*Wood*
Structural	• Yellow pine, douglas fir, hemlock, red wood, cedar
Light framing	• Spruce, hemlock, yellow pine, fir, cedar
Siding	• Yellow pine, cypress, red wood, spruce, cedar, hemlock, fir
Exterior finish	• White pine, cypress, red wood, spruce, cedar, hemlock, fir
Interior finish	• *Painted:* White and yellow pine, birch, gum, red wood, popular, spruce, cedar, fir *Natural or Stained:* Oak, white pine, birch, red wood, yellow pine, walnut, ash, cherry, gum, maple
Sashes, doors, frames	• White pine, fir, birch, red wood, yellow pine
Sheathing, roofing	• Fir, hemlock, pine, red wood, spruce
Flooring	• Yellow pine, maple, oak, hemlock, birch, beech, spruce, gum, fir

The physical properties of wood, including the grain pattern, physical characteristics, weight and moisture content, relate to the structure of the wood. Mechanical properties, such as compressive strength, flexural strength, and modulus of elasticity are associated with the ability of the wood to withstand applied forces.

Lumber grading is performed to allow the user to select the best quality product suited to the application. In almost all grades, for all types of lumber, grading is based on the number, character, and location of various features and defects that may lower the strength, durability or visual value of the product. Common features include knots, checks, splits, pitch pockets, and strains.

SUGGESTED READINGS

1. L.H. Sperling, '*Introduction to Physical Polymer Science*', Wiley, New York (2001).
2. A. Rudin, *'The Elements of Polymer Science and Engineering'*, 2nd Ed., Academic Press, San Diego (1998).
3. C.A. Harper (Ed.), *'Hand Book of Plastics, Elastomers and Composites'*, 3rd Ed., McGraw Hill, New York (1996).

REVIEW QUESTIONS

1. What is a polymer? Write their main characteristics? What are their uses?
2. What are meant by the terms *saturated* and *unsaturated* in relation to polymers?
3. Differentiate between *polymorphism* and *isomerism*.
4. Explain three types of polymerization.
5. What do you understand by polymerization? What is the difference between addition polymerization and condensation polymerization? [AMIE]
6. What is meant by polymerization? What are the two broad classifications of polymers? Discuss them. [AMIE]
7. Define the term 'polymerization'. Explain the different types of polymers with examples and applications. [AMIE]
8. What is polymerization? With the help of suitable examples, compare and contrast and processes of addition polymerization and condensation polymerization. [AMIE, M.Sc. (M.S.)]
9. What are the basic differences between the thermosetting and thermoplastic polymers? [B.E.]
10. Name three commonly used thermoplastic polymers and their applications. [Diploma]
11. Name three commonly used thermosetting polymers and their applications. [Diploma]
12. *(a)* Discuss briefly the different polymerization mechanisms.
(b) What are the general properties of polymeric materials? [AMIE]
13. *(a)* What is the difference between *configuration* and conformation in relation to polymer chains? [B.E.]
14. Make comparisons of thermoplastic and thermosetting polymers *(i)* on the basis of mechanical characteristics upon heating, and *(ii)* according to possible molecular structure. [AMIE, B.E.]
15. *(a)* What feature is necessary in a monomer for addition polymerization to be possible? Is this same feature necessity for condensation polymerization?
(b) Describe the difference between thermoplastics and thermosetting polymers in terms of applied stress, increased temperature and atomic structure.
(c) Describe briefly how each of the following modifies the strength of a polymer: *(i)* increased degree of polymerization *(ii)* increased branching *(iii)* increased cross-linking and *(iv)* increased crystallinity. [AMIE, M.Sc.(M.S.)]
16. What is the effect of the use of plasticizer or thermoplastic materials? [AMIE]
17. Explain the types of bonding that exist in polymers? [B.E.]
18. Explain the lamination process and the purpose for laminating plastics. [Diploma]
19. List and describe the six major production techniques that are used in the plastics industry. [M.Sc.]
20. *(a)* List the principal additives used in the manufacture of plastic articles. *(b)* What are the special properties of plastics that make them useful engineering materials? [AMIE]
21. What is the difference between compression moulding and transfer moulding of plastic parts? For what type of work is the latter used? [AMIE]
22. What are linear polymers? Explain the difference between addition and condensation polymerization. [AMIE]
23. Explain briefly why the tendency of a polymer to crystallize decreases with increasing molecular weight. [B.E.]
24. Describe the phenomenon of viscoelasticity. [B.E.]
25. For thermoplastic polymers, cite five factors that favour brittle fracture. [B.E.]
26. Which of the following would you expect to be elastomers and which thermosetting polymers at room temperatures? Justify your answer:
(i) linear and crystalline polyethylene *(ii)* phenolformaldehyde
(iii) heavily cross-linked polyisoprene having a glass-transition temperature of 50°C

(iv) lightly cross-linked polyisoprene having a glass-transition temperature of –60°C.

(v) linear and partially amorphous polyvinyl chloride. [B.E.]

27. What is the distinction between dye and pigment colorants? [B.E.]

28. Cite four factors that determine what fabrication technique is used to form polymeric materials. [B.E.]

29. Why must fibre materials that are melt spun and then drawn by thermoplastic? Mention two reasons: [B.E.]

30. Write short notes:
(a) Fillers for plastics *(b)* Plasticizers *(c)* Baekelite *(d)* Polyvinyl chloride *(e)* Plastic as a insulating material *(f)* Wood as a natural polymer *(g)* Polymerization *(h)* Difference between thermosetting and thermoplastic resins *(i)* Protective coatings *(j)* Natural rubber *(k)* Copolymerization *(l)* Vulcanization *(m)* Die and pigment colorants *(n)* Lamination processes and the purpose for laminating plastics *(o)* Rheology *(p)* Foam *(q)* Elastomer *(r)* Cis- and trans-forms of elastomers

31. In what ways do elastomers differs from polymers?

32. Why are elastomers and not polymers used in applications such as tyres?

PROBLEMS

1. The number average molecular weight of a polypropylene is 1,000,000 gm/mol. Compute the number average degree of polymerization. [*Ans.* 23,700], [B.E.]

2. For a linear polymer molecule, the total chain length L depends on the bond length between chain atoms d, the total number of bonds in the molecule N, and the angle between adjacent backbone chain atoms θ, as follows: $L = Nd \sin(\theta/2)$. Moreover, the average end to end distance for a serier of polymer molecule r is equal to, $r = d\sqrt{N}$. A linear polyethylene has a number-average molecular weight of 300,000 gm/mol. Determine the average values of L and r for this material.
[*Ans.* L = 2682 nm, r = 22.5 nm] [B.E., REC, M.Sc. (M.S.)]

3. The tensile strength and number-average molecular weight for two polymethyl methacrylate materials are given below

Tensile strength MPa	*Number average molecular weight (gm/mol)*
107	40,000
170	60,000

Find the tensile strength at a number average molecular weight of 30,000 gm/mol.
[*Ans.* 44 MPa] [B.E.]

OBJECTIVE QUESTIONS

1. The side group in polyvinyl chloride is
(1) H, Cl, Cl, Cl (2) H, H, H, Cl
(3) H, H, H, H (4) H, H, CH_3, Cl (2)

2. The factors that promote non crystallinity in polymers are
(1) branching and addition of plasticizers (2) copolymerization
(3) large random side groups (4) all the above (4)

3. The time average molecular weight is expressed as (symbols have usual meanings)
(1) $\overline{M}_n = \sum x_i M_i$ (2) $\overline{M}_n = \Sigma M_i/x_i$
(3) $\overline{M}_n = \sum \sqrt{x_i M_i}$ (4) None of the above (1)

4. A weight-average molecular weight $\overline{M}_w$ is calculated according to (symbols have usual meanings)
(1) $\overline{M}_w = \sum w_i M_i$ (2) $\overline{M}_w = \sum \sqrt{w_i M_i}$
(3) $\overline{M}_w = \sum M_i/w_i$ (4) $\overline{M}_w = \sum (W_i M_i)^2$ (1)

5. Relaxation modulus $E_r(t)$ for viscoelastic polymers is given by (symbols have usual meanings)
(1) $E_r(t) = \sqrt{\sigma(t)}/\varepsilon_0$ (2) $E_r(t) = \sigma(t)/\varepsilon_o$
(3) $E_r(t) = \varepsilon_0 \sqrt{\sigma(t)}$ (4) $E_r(t) = \sqrt{\varepsilon_0}\,\sigma(t)$ (2)

6. Apart from three *H* atoms, the fourth side group in the monomer of polyproplene is
(1) C_6H_6 (2) CH_3 (3) C_6H_5 (4) $COOCH_3$ (2)

7. If C-C bond length is 1.54 Å, the length of uncoiled polyethylene chain with 500 mers is
(1) 650 Å (2) 1258 Å (3) 1628 Å (4) 1836 Å (2)

8. The degree of polymerization in polyethylene is 50,000. The average molecular weight is
(1) 650,000 amu (2) 1,300,000 amu
(3) 1,400,000 amu (4) 1,720,000 amu (3)

9. The number of double bonds required in the monomer for cross-linking is
(1) 2 (2) 3 (3) 4 (4) 1 (1)

10. Plasticizers are added in plastics to
(1) reduce hardness (2) improve flexibility
(3) improve their strength (4) reduce glass transition temperature (t_g) (2)

11. The windows of aeroplane are made in
(1) PVC (2) PTFE (3) PMMA (4) None of the above (3)

12. Which one is not a characteristic of plastics?
(1) low density (2) machinability (3) high strength (4) large plastic deformation (3)

13. In teflon, i.e., PTFE monomer, the four side groups are
(1) H, H, H, CH_3 (2) F, F, F, F
(3) H, H, H, H (4) H, H, Cl, Cl (2)

14. The weight of sulphur required to achieve 5% cross-linking of 68 kg of polyisoprene is
(1) 0.8 kg (2) 1.6 kg (3) 3.2 kg (4) 5.6 kg (2)

15. An ebonite is
(1) copolymer (2) natural wood (3) natural rubber (4) none of the above (3)

16. Plastics are used in automobiles because of their
(1) low melting point (2) high strength
(3) capacity to be moulded with good finish
(4) none of the above (3)

17. Choose the correct statement
(1) Thermoplastics are either amorphous or crystalline but thermosets are amorphous
(2) Thermoplastics are crystalline
(3) Thermosetting and thermoplastic polymers are essentially amorphous
(4) Thermosetting plastics are crystalline (1)

SHORT QUESTIONS ANSWERS

1. What are polymers?

Ans. Most of the polymeric materials are composed of very large molecules-chains of carbon atoms, to which are side-bounded various atoms or radicals.

2. What do you understand by mers?

Ans. The macromolecules of polymers may be though of as being composed of mers, smaller structural entities, which are repeated along the chain.

3. When molecular entanglements occur?

Ans. When the chain assume twisted, coiled, and kinked shapes or contours.

4. What are the structures possible with regard to molecular structure of polymers?

Ans. With regard to molecular structure, linear, branched, cross-linked, and network structures are possible, in addition to isotactic, syndiotactic, and atactic, stereoisomers, and cis and trans geometrical isomers.

5. How polymers are classified with regard to their behaviour at elevated temperature?

Ans. Thermoplasts and thermosets. The thermoplasts comprise of plastics, fibres and elastomers.

6. What are thermoplastic polymers?

Ans. These materials have linear and branched structure and they soften when heated and harden when cooled.

7. What are thermoset polymers?

Ans. In contrast to thermoplast materials, thermosets, once having hardened, will not soften upon heating. Their structures are cross-linked and network.

8. When the crystallinity exist?

Ans. When the packing of molecular chains is such as to produce an ordered atomic arrangement, the condition of crystallinity is said to exist. In addition to being entirely amorphous, polymers may also exhibit varying degrees of crystallinity; i.e., crystalline regions are interdispered within amorphous areas. Crystallinity is facilitated for polymers that are chemically simple and that have regular and symmetrical chain structures. The crystallinity of long chain polymers is influenced by a number of factors, e.g. chain length, branching, presence of large random side groups and plasticizers.

9. How you will grow polymer single crystals?

Ans. One can grow this from dilute solutions as thin platelets and having chain-folded structures.

10. What are spherulites?

Ans. Many semicrystalline polymers form spherulites. Each spherulite consists of a collection of ribbon-like chain-folded lamellar crystallites that radiate outward from its centre.

11. On the basis of stress-strain behaviour, how you will classify polymers?

Ans. Brittle, plastic and highly elastic.

12. What do you understand by viscoelastic mechanical behaviour of polymers?

Ans. This type of behaviour, being intermediate between totally elastic and totally viscous, is displayed by a number of polymeric materials. It is characterized by the relaxation modulus, a time-dependent modulus of elasticity. The magnitude of the relaxation modulus is very sensitive to temperature; critical to the in-service temperature range for elastomers in this temperature dependence.

13. What is the importance of melting and glass transition temperatures for polymers?

Ans. These parameters are important relative to the temperature range over which a particular polymer may be utilized and processed. The magnitude of T_m and T_g increase with increasing chain stiffness; stiffness is enhanced by the presence of chain double bonds and side groups that are either bulky or polar. Molecular weight and degree of branching also affect T_m and T_g.

18

Composites

1. INTRODUCTION

Many of modern technologies require materials with unusual combinations of properties that cannot be met by the metals, conventional metallic alloys, ceramics, and polymeric materials, e.g. materials needed for aerospace, underwater, and transportation applications. For example engineers working in aircraft industry are looking and searching for structural materials that have low densities, are strong, stiff, and abrasion and impact resistant, and are not easily corroded. Obviously, this is a rather formidable combination of characteristics. Usually strong materials are relatively dense; also, increasing the strength or stiffness generally results in a decrease in impact strength. Therefore, in composites the whole is greater than the sum of the parts (synergy). Wood, celery, bamboo, and corn are all examples of nature's composites, where two materials combine to reinforce and bind together. We have already discussed in previous chapters, composites of sort which include multiphase metal alloys, ceramics and polymers. A composite is considered to be any multiphase material that exhibits a significant proportion of the properties of both constituent phases such that a better combination of properties is realized. This is termed as the *principle of combined action*. According to this principle, better combinations are fashioned by the judicious combination of two or more distinct materials.

All composites generally have one thing in common: a *matrix* or *binder* combined with a reinforcing material. Obviously, a composite consists of a matrix material, dispered within which is a dispersion of one or more phases of another material. If the fibres are directionally oriented and continuous, the material is termed an *advanced composite*. Reinforced concrete is a good example of composite material. When concrete is reinforced with steel rebar, the concrete becomes the matrix, which surrounds the reinforcing fibre, the rebar. Another example is reinforced fiber glass products such as fishing rods. Glass fibres are set in a thermosetting resin matrix. This produces a strong, lightweight, flexible fishing rod. Other fibres are produced from aramid (Kevlar and Nomex), boron, carbon, graphite, and ultrahigh-molecular-weight polyethylene (spectra). The matrix for these materials is typically a thermosetting epoxy resin. These materials provide some exceptional increase in mechanical properties, sometimes three to six times greater than steel. Another example of a composite material is pearlitic steels. The microstructure of this material consists of alternating layers of α-ferrite and cementite. The ferrite phase is soft and ductile, whereas cementite is hard and very brittle. The combined mechanical characteristics of this composite, i.e. pearlite (reasonably high ductility and strength) are superior to those of either of the constituent phases. Other common varieties of composites include combinations: fibre-resin, fibre-ceramic, carbon-metal, metal concrete, metal resin and wood plastic. Most of the contemporary advanced composites use glass, kevlar (an aramid), or one of the various types of graphite fibres. *Fibre composites* form an important subset of this class of engineering materials. Obviously, a composite is a multiphase material that is artificially made, as to one that occurs naturally. In addition, the constituent phases in a composite must be chemically dissimilar and separated by a distinct interface. This is why most metallic alloys and many ceramics do not fit this definition because their multiple phases are formed as a consequence of natural phenomena.

There are many reasons for making composites: the incorporation of fibres into brittle ceramics produces a composite of enhanced toughness. Fillers, such as the presence of aggregate in concrete, reduce the overall cost of the product, and additionally improve the compressive strength. The second phase may furthermore be a gas, as in the manufacture of foamed products of low density.

On the basis of strength and stiffness alone fibre reinforced composite materials may not be superior to metals of comparable strength, but when the specific modulus (i.e. modulus per unit weight) and specific strength are considered, then their use implies that the weight of components can be reduced. This is an important factor in all forms of transport, where reductions of weight result in greater energy savings.

In order to produce a new generation of extraordinary materials, scientists and engineers while designing composite materials, have ingeniously combined various metals, ceramics, and polymers.

One can classify composite materials as per simple scheme shown in Fig. 18.1. There are three main divisions: particle reinforced, fibre-reinforced, and structural composites. We note that there exist atleast two subdivisions for each main division.

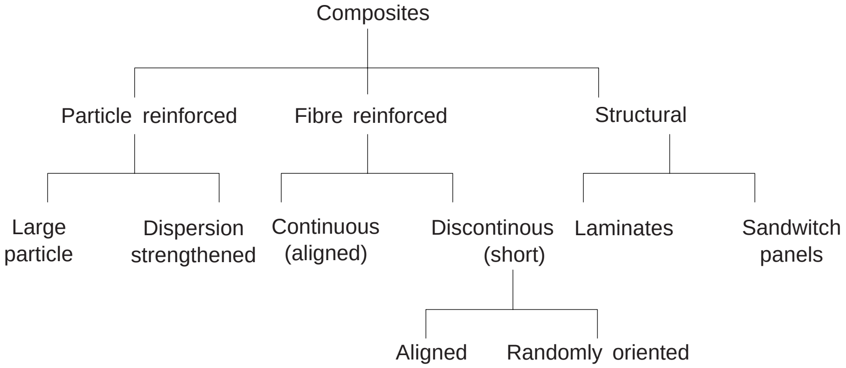


Fig. 18.1 A simple classification schemes for the various composite types

The dispersed phase for particle-reinforced composites is equiaxed (i.e., particle dimensions are approximately the same in all directions) and the dispersed phase for fibre-reinforced composites has the geometry of a fibre-reinforced composites has the geometry of a fibre. (i.e., a large-length-to-diameter ratio). Structural composites are combinations of composites and homogeneous materials.

Examples of these three groups include concrete, a mixture of cement and aggregate, which is a particulate composite; fibre glass, a mixture of glass fibres imbedded in a resin matrix, which is a fibre composite; and plywood, alternating layers of laminate veneers, which is a laminate composite. These are discussed in detail in section 3.

2. GENERAL CHARACTERISTICS

(i) Composite materials are superior to all other known structural materials in specific strength and stiffness, high temperature strength, fatigue strength and other properties. The desired combination of properties can be tailored in advance and realized in the manufacture of a particular material. Moreover, the material can be shaped in this process as close as possible to the form of final products or even structural units.

(ii) Composite materials are complex materials whose components differ strongly from each other in the properties, are mutually insoluble or only slightly soluble and divided by distinct boundaries.

(iii) The principle of manufacture of composites has been borrowed from nature. Trunks and stems of plants and bones of man and animals are examples of natural composites. In wood, cellulose fibres are bonded by plastic lignin, in bones, thin and strong fibres of phosphates are bonded by plastic collagen.

(iv) The properties of composites mainly depend on the physico-mechanical properties of their components and the strength of bonds between them. A characteristic feature of composite materials is that the merits of their components are fully utilized. Composite materials may acquire certain valuable properties not found in the components. For obtaining the optimal properties in composites, their components are chosen so as to have sharply different, but complementary properties.

(v) The base, or matrix, of composites may consist of metals or alloys (metallic composites), polymers, carbon and ceramic materials (non metallic composites).

(vi) The matrix is essentially the binding and shaping component in composites. Its properties determine to a large extent the process conditions for the manufacture of composite materials and the important operating characteristics, e.g., working temperature, fatigue strength, resistance to environmental effects, density, and specific strength. Some composites have a combined matrix which consists of alternating layers (two or more) of different composition.

(vii) Composites with combined matrix may be called *multi-matrix*, or *multi-layer* composites (Fig. 18.2(a)). Multi-matrix composites can be characterized by a wider spectrum of useful properties. For example, use of titanium as an addition to aluminium may increase the strength of a composite material in directions transverse to fibres. Aluminium layers in a matrix diminish the density of composite material.

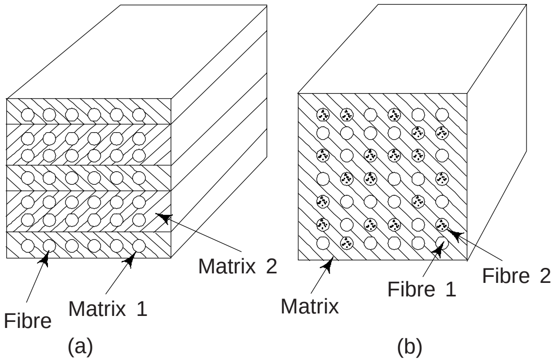


Fig. 18.2 Schemes of (a) multi-matrx and (b) polyfibre composite materials

(viii) Fillers, i.e., other components are uniformly distributed in a matrix. These plays the major part in strengthening of composites and thus they are called *strengtheners*. Fillers should possess high values of strength, hardness and elastic modulus. These characteristics should be substantially higher than those of the matrix. With an increase of the elastic modulus and ultimate strength of a filler, the corresponding properties of a composite material also increase, but do not reach the value of the filler. Fillers are alternatively called reinforcing components. This is a broader term than ‘strengthener’, it does not specify the particular strengthening role of filler which may used for improving other properties of a composite.

The properties of a composite material can also depend on the shape (geometry), dimensions, concentration and distribution of filler (reinforcement pattern).

As regards their shape, fillers are divided into three main groups *(i)* zero-dimensional *(ii)* one-dimensional, and *(iii)* two dimensional.

By the reinforcement pattern, composite materials are divided into three groups: with uniaxial, biaxial, and triaxial reinforcement.

(ix) Fillers of different shape may be used for obtaining a wider complex of properties or enhancing a particular property of a composite material. For example, the strength of bond between one-dimensional filler elements (glass or carbon fibres) and a polymer matrix can be increased by introducing a zero-dimensional filler (particles of asbestos, silicon carbide, etc.). The same purpose can be achieved by reinforcing a composite material with fillers of the same shape, but different composition. For example, the modulus of elasticity of composite materials with a polymer matrix reinforced by glass fibres can be increased by additional reinforcement with boron fibres.

Composite materials containing two-or more different fillers are termed *complex-reinforced composites*. (Fig. 18.2b).

3. PARTICLE-REINFORCED COMPOSITES

These can be further classified under two subgroups: *(i) large particle* and *(ii) dispersion- strengthened composites*. The distinction between these is based upon reinforcement or strengthening mechanism. The term 'large' indicate that particle-matrix interactions cannot be treated on the atomic or molecular level; rather continum mechanics is used. The particulate phase for most of these composites is harder and stiffer than the matrix. In the vicinity of each particle, these reinforcing particles tend to restrain movement of the matrix phase. Obviously, the matrix transfers some of the applied stress to the particles, which bear a fraction of the load. We may note that the degree of reinforcement or improvement of mechanical behaviour depends on strong bonding at the matrix particle interface.

Particles for dispersion-strengthened composites are normally much smaller (diameter between 0.01 and 0.1 μm). Particle-matrix interactions occur on the atomic or molecular level and lead to strengthening. We may note that the mechanism of strengthening is similar to that for precipitation. The matrix bears the major portion of an applied load, where as the small dispersed particle hinder or impede the motion of dislocations. Obviously, plastic deformation is restricted such that yield and tensile strengths, as well as hardness improve.

3.1 Large-particle Composites

Fillers added to some polymeric materials produce large-particle composites. The fillers modify or improve the properties of the material and/or replace some of the polymer volume with less expensive filler material.

Concrete is another familiar example of large-particle composite. Concrete is composed of cement (the matrix), and sand and gravel (the particulates).

We may note that particles can have quite a variety of geometries, but they should be approximately the same dimension in all directions (equiaxed). Particules should be small and evenly distributed throughout the matrix for effective reinforcement. Moreover, the volume fraction of the two phases influences the behaviour; mechanical properties are enhanced with increasing particulate content. The *rule of mixtures* equations reveal that the elastic modulus should fall between an upper bound given by

$$E_c(u) = E_m V_m + E_p V_p \tag{1}$$

and a lower bound, or limit,

$$E_c(t) = \frac{E_m E_p}{V_m E_p + V_p E_m} \tag{2}$$

Where E and V denote the elastic modulus and volume fraction, respectively and the subscript c, m, and p represent composite, matrix and particulate phases respectively.

Large-particle composites are utilized with all three material types, i.e. metals, polymers and ceramics. Examples of ceramic-metal composite are *cermets*. Cemented carbide, which is composed of extremely hard particles of a refractory carbide ceramic such as tungsten carbide (WC) or titanium carbide (TiC), embedded in a matrix of a metal such as cobalt or nickel is the most common cermet. These composites are widely used as cutting tools for hardened steels.

Interestingly, both elastomers and plastics are frequently reinforced with various particulate materials. The use of many of the modern rubbers would have been severely restricted without reinforcing particulate materials, e.g. as carbon black. Carbon black consists of very small and essentially spherical particles of carbon. Carbon black is produced by the combustion of natural or oil in an atmosphere that has only a limited air supply. When carbon black is added to vulcanized rubber, this material which is very cheap, enhances tensile strength, toughness, and tear and abrasion resistance. Automobile tire contain on the order of 15 to 30 volume percentage of carbon black.

In order that carbon black may provide significant reinforcement, the particle size must be evenly small, with diameters between 20 and 50 nm. Moreover, the particle must be evenly distributed throughout the rubber and must form a strong adhesive bond with the rubber matrix. Particle reinforcement using other materials, e.g. silica, is much less effective as this special interaction between the rubber molecules and particle surfaces does not exist.

Concrete

This is a common large-particle composite in which both matrix and disperse phases are ceramic materials. Broadly speaking, concrete implies a composite material consisting of an aggregate of the particles that are bound together in a solid body by some type of binding medium, i.e., a cement. The two most familiar concretes are those made with portland and asphaltic cements, where the aggregate is gravel and sand. Asphaltic cement is used primarily as a paving material on a wider scale, whereas portland cement concrete is used extensively as a structural building material.

Portland Cement Concrete

The ingredients for portland cement concrete are portland cement, a fine aggregate (sand), a coarse aggregate (gravel), and water. The aggregate particles act as a filler material to reduce the overall cost of concrete product as they are cheap, whereas cement is relatively costly. The ingredients have to be added in correct proportions so that one may achieve the optimum strength and workability of concrete mixture. One can achieve dense packing of the aggregate and good interfacial contact by having particles of two different sizes. The fine particles of sand should fill be void spaces between the gravel particles. Normally, these aggegrates comprise between 60% and 80% of the total volume. However, the amount of cement-water paste should be sufficient to coat all the sand and gravel particles, otherwise the cementitious bonds will be incomplete.

This is a major material for construction. One can pour portland cement concrete in place and it hardens at room temperature, and even when submerged in water. However, as a structural material, it is found to be relatively weak and extremely brittle; the tensile strength is approximately 10 to 15 times smaller than its compressive strength. Moreover, large concrete structure can experience considerable thermal expansion and also contraction with temperature fluctuations. Water also penetrates into external pores, which can cause severe cracking in cold weather (as a consequence of freeze-thaw cycles).

Reinforced Concrete

One can increase the strength of portland cement concrete by additional reinforcement. One can achieve this by means of steel rods, wires, bars (rebar), or mesh, which are embedded into the fresh and uncured concrete. Obviously, the reinforcement renders the hardened structure capable of supporting greater tensile, compressive and shear stresses. Considerable reinforcement is maintained even when cracks develop in the concrete.

The coefficient of thermal expansion for steel is nearly the same as that of concrete and thus it serves as a suitable reinforcement. Moreover, steel is not rapidly corroded in the cement environment and also a relatively strong adhesive bond is formed between it and the curved concrete. One can enhance this adhesion by the incorporation of contours into the surface of the steel member. This permits a greater degree of interlocking.

One can also reinforce portland cement concrete by mixing into the fresh concrete fibres of a high-modulus material, e.g. glass, nylon, steel and polyethylene. However, when exposed to cement environment some fibre materials experience rapid deterioration and therefore care must the exercised in using this type of reinforcement.

There is another reinforcement technique of strengthening concrete. This involves the introduction of residual compressive stresses into the structural member. The resulting material is called *prestressed concrete*. The characteristic of brittle ceramics, i.e. they are stronger in compression than in tension is utilized in this method. Obviously, to fracture a prestressed concrete member, the magnitude of the precompressive stress must be exceeded by an applied tensile stress.

In one prestressed technique, inside the empty moulds high-strength steel wires are positioned and stretched with a high tensile force, which is maintained constant. After placing the concrete and allowing them to harden, the tension is released. As the wires contract, they put the structure in a state of compression because the stress is transmitted to the concrete through the concrete-wire bond that is formed.

There is also another technique known as *posttensioning.* In this technique stresses are applied after the concrete hardens. Sheet metal or rubber tubes are situated inside and pass through the concrete forms, around which the concrete is cast. Steel wires are fed through the resulting holes after the cement has hardened and tension is applied to the wires through jacks attached and abutted to the faces of the structure. Also, a compressive stress is imposed on the concrete piece, this time by the jacks. To protect the wire from corrosion, the empty spaces inside the tubing are filled finally with a grout.

We must note that the concrete that is prestressed should be of a high quality and there must be a low shrinkage and low creep rate. Usually, prestressed concretes are prefabricated and used mostly for railway bridges and highway.

3.2 Dispersion-strengthened Composite

This type of composite contains small particulates or dispersions, which increase the strength of the composite by blocking the movement of dislocations. The *dispersoid* is typically a stable oxide of the original material. A common example is sintered aluminium powder (SAP). SAP has an aluminium matrix, which contains up to 14% aluminium oxide (Al_2O_3). This composite is produced with the *powder metallurgy process*, where the powders are mixed, compacted at high pressures, and sintered together. Sintering involves heating a material until the particles of the material fuse together. Only the edges of the particles are generally bonded together; the whole particle does not melt. In terms of sintered ceramics, the product is a strong, rigid, brittle product that exhibit good compressive strength, high melting points, and good heat resistance. Examples of dispersion-strengthened composites include Ag–CdO, used as an electrical contact material, Pb–PbO, used in battery plates, and Be–BeO, used in nuclear reactor and aerospace components. The high-temperature strength of nickel alloy may be enhanced significantly by the addition of about 3 volume percentage of thoria (ThO_2) as finely dispersed particles. This material is known as thoria-dispersed (or TD) nickel.

4. FIBRE-REINFORCES COMPOSITES

These are strong fibres imbedded in a softer matrix to produce products with high strength-to-weight ratios. The matrix material transmits the load to fibres, which absorb the stress. Under an applied stress, fibre-matrix bond ceases at the fibre ends, yielding a matrix deformation pattern (Fig. 18.3).

In order to have effective strengthening and stiffening of the composite material, some critical fibre length is esential. This critical length l_c is dependent on the fibre diameter d and its ultimate (or tensile) strength σ_f, and on the fibre-matrix bond strength (or the shear yield strength of the matrix, whichever is smaller) τ_c according to the following relation

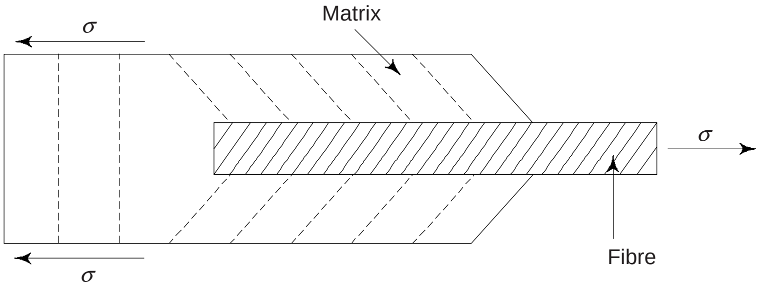


Fig. 18.3 The deformation pattern in the matrix surrounding a fibre, subjected to an applied tensile load

$$l_c = \frac{\sigma_f^* \, d}{2\,\tau_c} \qquad (3)$$

The critical length, l_c is on the order of 1 mm for a number of glass and carbon fibre-matrix combinations and ranges between 20 and 150 times the fibre diameter.

Obviously, the strength of these composites comes from the bonding between the reinforcement fibres and the matrix. The length-to-diameter, or *aspect*, ratio of the fibres used as reinforcement influences the properties of the composite. The higher the aspect ratio, the stronger the composite. Therefore, long, continuous fibres are better than short ones for composite construction. However, continuous fibres are more difficult to produce and place in the matrix. Shorter fibres are easier to place in the matrix but offer poor reinforcement. Some trade-off is made when shorter, discontinuous fibres are used with aspect ratios greater than a specified minimum value. The greater the number of fibres, the stronger the composite. This holds true upto about 80% of the volume of the composite, where the matrix can no longer completely surround the fibres.

Fibres for which $l >> l_c$ (normally $l > 15\ l_c$) are called *continuous*, whereas fibres which have lengths shorter than this are termed *discontinuous*. The matrix deforms around the discontinuous fibre having length less than l_c, such that there is virtually no stress transference and little reinforcement by the fibre. These are essentially the particulale composites as discussed earlier. In order to affect a significant improvement in strength of the composite, the fibres must be continuous. Mechanical properties of some reinforcing fibres are given in Table 18.1.

Table 18.1 Mechanical properties of some reinforcing fibres

Material	*Density,* ρ *(kg* $\times$ *m*$^{-3}$ $\times$ *10*^{3}*)*	*Young's Modulus, E (GPa)*	*Tensile strength* σ^* *(GPa)*	*Fibre radius, r (*μ*m)*
• E-glass fibres	2.56	76	1.4-2.5	10
• Carbon fibres (High modulus)	1.75	390	2.2	8.0
• Carbon fibres (High strength)	1.95	250	2.7	8.0
• Kevlar fibres	1.45	125	3.2	12
• Silicon carbide (Monofilament)	3.00	410	8.6	140
• Silicon carbide (Nicalon)	2.50	180	5.9	14
• Alumina (Saffil)	2.80	100	1.0	3

The arrangement or *orientation* of the fibres, relative to each other, the fibre *concentration*, and *distribution* all have a significant influence on the strength and other properties of fibre-reinforced composites. There are two possible extremes with respect to orientation: *(i)* a parallel alignment of the longitudinal axis of the fibres in the single direction, and *(ii)* a totally random alignment. Continuous fibres are normally aligned as shown in Fig. 18.4(a), discontinuous fibres may be aligned shown in Fig. 18.4(b) and randomly oriented, as or partially oriented as shown in Fig. 18.4(c). One can realize overall better composite properties when fibre distribution is uniform.

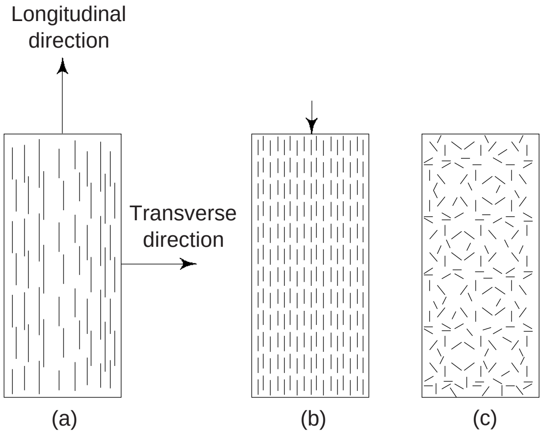
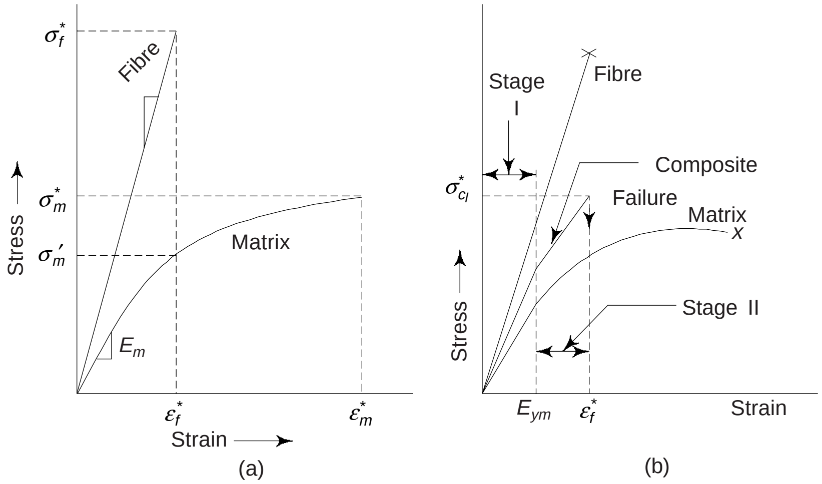


Fig. 18.4 **Representation of (a) continuous and aligned, (b) discontinuous and aligned, and (c) discontinuous and randomly oriented fibre-reinforced composites**

Continuous and Aligned Fibre Composites

Mechanical behaviour of this type of composite depend on: *(i)* stress-strain behaviours of fibre and matrix phases, *(ii)* the phase volume fractions, and *(iii)* the direction in which the stress or load is applied. Moreover, the properties of a composite having its fibres aligned are highly anisotropic, i.e., dependent on the direction in which they are measured. Let us first consider the situation shown in Fig. 18.5(a), wherein the stress is applied along the direction of alignment, i.e. the *longitudinal direction*.



(a) (b)

Fig. 18.5 **(a) Stress-strain curves for brittle fibre and ductile matrix materials. For both the materials stresses and strains are noted. (b) Stress-strain curve for an aligned fibre-reinforced composite, which is exposed to a uniaxial stress applied in the direction of alignment. We may note curve for fibre and matrix materials of part (a) are also superimposed**

Let us assume the stress versus strain behaviours for fibre and matrix phases shown in Fig. 18.5(a). Here we consider the fibre to be totally brittle and the matrix phase to be reasonably ductile. The fracture strengths in tension for fibre and matrix are indicated by σ_f^* and σ_m^* respectively, in Fig. 18.5(a). The corresponding fracture strains are represented by ε_f^* and ε_m^* respectively. It is assumed that $\varepsilon_m^* > \varepsilon_f^*$, which is normally the case.

A fibre-reinforced composite consisting of these fibre and matrix materials will exhibit the stress-strain behaviour shown in Fig. 18.5(b). In the initial stage (region I), both fibres and matrix deform elastically. Normally, this portion of the curve is linear. Typically, for a composite of this type, the matrix yields and deform plastically as shown in Fig. 18.5(b) (at ε_{ym} while the fibres continue to stretch elastically, in as much as the tensile strength of the fibres significantly higher than the yield strength of the matrix. This constitutes the stage II. This stage is ordinarily very linear but of diminished slope in comparison to stage I. Moreover, in passing from stage I to stage II, the proportion of the applied load that is borne by the fibres increases.

Corresponding to a strain $\sim\varepsilon_f^*$ (Fig. 18.5b), the onset of composite failure begins as the fibres start to fracture. For a couple of reasons, composite failure is not catastropic: all the fibres do not fracture at the same time, since there will always be considerable variations in the fracture strength of brittle fibre materials. Moreover, even after fibre failure, the matrix is still intact in as much as $\varepsilon_f^* < \varepsilon_f^*$ (Fig. 18.5a). Obviously, these fractured fibres, which are shorter than the original ones, are still embedded within the intact matrix, and consequently are capable of sustaining a diminished load as the matrix continues to plastically deform.

Elastic Behaviour: Longitudinal Loading

Let us study the elastic behaviour of a continuous and oriented fibrous composite that is loaded in the direction of fibre alignment direction (i.e., *longitudinal loading*). Assuming that the fibre-matrix interfacial bond is very good and the situation is that of *isostrain*, i.e. deformation of both matrix and fibres is same. Thus the total load sustained by the composite, F_c is given by

$$F_c = F_m + F_f \tag{4}$$

where F_m and F_m are loads carried by the matrix phase and the fibre phase respectively. Now, stress $F = \sigma A$, thus the expressions for F_c, F_m and F_f in terms of their respective stresses, i.e. σ_c, σ_m and σ_f and cross-sectional areas, i.e. A_c, A_m and A_f are possible. Equation (4) yields

$$\sigma_c A_c = \sigma_m A_m + \sigma_f A_f \tag{5}$$

or

$$\sigma_c = \sigma_m \frac{A_m}{A_c} + \sigma_f \frac{A_f}{A_c} \tag{6}$$

Now, $A_m/A_c = V_m$ and $A_f/A_c = V_f$, where V_m and V_f are volume fraction of the matrix and fibres respectively. Here we have assumed that composite, fibre and phase lengths are all equal. Truly speaking A_m/A_c and A_f/A_c are the area fractions of the matrix and fibre phases. Since phase lengths for both matrix and fibre are taken equal and hence we have taken A_m/A_c and A_f/A_c equivalent to V_m and V_f respectively.
Thus, Eq. (6) becomes

$$\sigma_c = \sigma_m V_m + \sigma_f V_f \tag{7}$$

In the light of assumption of isostrain state, we have

$$\varepsilon_c = \varepsilon_m = \varepsilon_f \tag{8}$$

Dividing each term in Eq. (7) by its respective strain, we obtain

$$\frac{\sigma_c}{\varepsilon_c} = \frac{\sigma_m}{\varepsilon_m} V_m + \frac{\sigma_f}{\varepsilon_f} V_f \tag{9}$$

If composite, matrix, and fibre deformations are all elastic, then $\sigma_c/\varepsilon_c = E_c$, $\sigma_m/\varepsilon_m = E_m$ and $\sigma_f/\varepsilon_f = E_f$, where E's being the moduli of elasticity for the respective phase. Substitution in (9), we obtain an expression for the modulus of elasticity of a continuous and aligned fibrous composite *in the direction of alignment*, E_{cl}, as

$$E_{cl} = E_m V_m + E_f V_f \tag{10a}$$

or

$$E_{cl} = E_m (1 - V_f) + E_f V_f \tag{10b}$$

Since the composite consists of only matrix and fibre phase, i.e., $V_m + V_f = 1$.

Obviously, E_{cl} is equal to the volume-fraction weighted average of the moduli of the elasticity of the fibre and matrix phases. We may note that other properties, including density, also have this dependence on volume fractions.

We can also show that for longitudinal loading, the ratio of the load carried by the fibres to that carried by the matrix is

$$\frac{F_f}{F_m} = \frac{E_f V_f}{E_m V_m} \tag{11}$$

Elastic Behaviour—Transverse Loading

A continuous and oriented fibre composite may be loaded in transverse direction (i.e., *transverse loading*). Obviously, load is applied at a 90° to the direction of fibre alignment. For this situation the stress σ to which the composite as well as both phases are exposed is the same. i.e.

$$\sigma_c = \sigma_m = \sigma_f = \sigma \tag{12}$$

This state is called as *isostress state*. Now, the strain or deformation of entire composite ε_c is

$$\varepsilon_c = \varepsilon_m V_m + \varepsilon_f V_f \tag{13}$$

or
$$\frac{\sigma}{E_{ct}} = \frac{\sigma}{E_m} V_m + \frac{\sigma}{E_f} V_f \qquad (\because\ \varepsilon = \sigma/E) \tag{14}$$

where E_{ct} is the modulus of elasticity in the transverse direction. Dividing (14) through σ, yields

$$\frac{1}{E_{ct}} = \frac{V_m}{E_m} + \frac{V_f}{E_f} \tag{15}$$

or
$$E_{ct} = \frac{E_m E_f}{V_m E_f + V_f E_m} = \frac{E_m E_f}{(1 - V_f)\, E_f + V_f E_m} \tag{16}$$

Longitudinal Tensile Strength

Let us study the strength characteristics of continuous and aligned fibre-reinforced composites that are loaded in the *longitudinal direction.* In this case, strength is usually taken as the maximum stress on the stress-strain curve (Fig. 18.5b). Usually, this point corresponds to fibre fracture, and indicates the onset of composite failure. Longitudinal tensile strength values for three common fibrous composites are given in Table 18.2. We may note that the failure of this type of composite material is relatively complex process and there are several different failure modes possible. Again, the mode that operates for a specific composite will depend on fibre and matrix properties, and the nature and strength of the fibre-matrix interfacial bond.

Table 18.2 Longitudinal and transverse tensile strengths (typical) for unidirectional fibre-reinforced composites. The fibre content in each is about 50 Vol%

Material	*Longitudinal tensile strength (MPa)*	*Transverse tensile strength (MPa)*
Glass-polyester	700	20
Carbon (high modulus)–epoxy	1000	35
Kevlar-epoxy	1200	20

For $\varepsilon_f^* < \varepsilon_m^*$ (Fig. 18.5a), we note that fibres will fail before the matrix. Once the fibres have fractured, the majority of the load that was borne by fibres is now gets transferred to the matrix. One obtains the expression for the longitudinal strength of the composite, σ_{cl}^* as

$$\sigma_{cl}^* = \sigma_m^* (1 - V_f) + \sigma_f^*\ V_f \tag{17}$$

where σ_m^* is the stress in the matrix at fibre failure (Fig. 18.4a) and σ_f^* is the fibre tensile strength.

Transverse Tensile Strength

The strength of continuous and unidirectional fibrous composites are highly anisotropic. Normally, such composites are designed to be loaded along the high strength, longitudinal direction. However, transverse tensile loads may also be present during in-service applications. Under these prevailing circumstances, premature failure may result in-as much as transverse strength is usually extremely low. Sometimes, it lies below the tensile strength of the matrix. Obviously, the reinforcing effect of the fibres is negative one. Transverse tensile strengths for few unidirectional composites are given in Table 18.2.

Properties of both the fibre and matrix, the fibre-matrix bond strength, and the presence of voids have significant influence on the transverse strength. In order to improve the transverse strength of these composites, one will have to modify the properties of matrix.

Discontinuous and Aligned Fibre Composites

Although reinforcement efficiency is lower for discontinuous than for continuous fibres, discontinuous and aligned fibre composites (Fig. 18.4b) are more in demand. Chopped glass fibres are used mostly, however, carbon and aramid discontinuous fibres are also used. One can produce these short fibre composites with moduli of elasticity and tensile strengths approaching 90% and 50%, respectively, of their continuous fibre counterparts.

The longitudinal strength (σ^*_{cd}) for a discontinuous and aligned fibre composite having a uniform distribution of fibres and in which $l > l_c$, is given by

$$\sigma^*_{cd} = \sigma^*_f \; V_f \left(1 - \frac{l_c}{2l}\right) + \sigma'_m \; (1 - V_f) \tag{18}$$

Here σ^*_f and σ'_m represent, respectively, the fracture strength of the fibre and the stress in the matrix when the composite fails (Fig. 18.5a).

For $l < l_c$, i.e. fibre length is less than critical, then the longitudinal strength $(\sigma^*_{cd'})$ is given by

$$\sigma^*_{cd'} = \frac{l\tau_c}{d} V_f + \sigma'_m \; (1 - V_f) \tag{19}$$

where d is the diameter of the fibre and τ_c is the smaller of the either the fibre matrix bond strength or the matrix shear yield strength.

Discontinuous and Randomly Oriented Fibre Composites

Normally, fibres are used when the fibre orientation is random, short and discontinuous. Reinforcement of this type is shown in Fig. 18.4(c). One can use a 'rule-of mixture' expression for the elastic modulus similar to Eq. 10(a), as follows:

$$E_{cd} = K \; E_f \; V_f + E_m \; V_m \tag{20}$$

Here K is a fibre efficiency parameter, which depends on V_f and the ratio E_f/E_m. However, its magnitude will be less than unity in the range 0.1 to 0.6. Thus, the modulus increases in some proportion to the volume fraction of fibre for random fibre reinforcement (as with oriented).

We have seen that aligned fibrous components are inherently anisotropic in that the maximum strength and reinforcement are achieved along the alignment (longitudinal) direction. Fibre reinforcement is virtually non-existent in the transverse direction and fractures usually occurs at relatively low tensile stresses. The composite strength lies between these extremes for other stress orientations.

Laminar Composites

When multidirectional stresses are imposed within a single plane, aligned layers that are fastened together one on top of another at different orientations are frequently utilized. These are called *laminar composites*. These are generally designed to provide high strength and low cost at a lighter weight. A familiar laminar composite is plywood, where the veneers are joined by adhesives, typically phenolic or amine resins. The individual odd number of piles are staked so that the grain in each layer runs perpendicular to that of the layers above and below it (Fig. 18.6). This technique offers plywood that is strong and yet cheaper. Some safety glass is a laminated structure, where an adhesive such as polyvinyl butyral is used between two outer layers of glass to keep the glass from flying when broken. Formica is another common laminate used for countertops. Laminates require two or more layers be bonded together. Laminations may also be constructed using fabric material such as cotton, paper, or woven glass fibres embedded in a plastic matrix. Obviously, a laminar composite has relatively high strength in a number of directions in the two-dimensional plane; however, the strength in any given direction is,

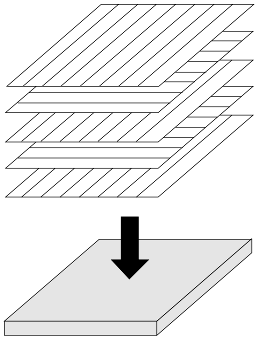


Fig. 18.6 Laminar composite. The stacking of successive oriented, fibre-reinforced layers

of course, lower than it would be if all the fibres were oriented in that direction. Modern ski is one example of a relatively complex *laminated structure.*

Applications involving totally multidirectional applied stresses generally use discontinuous fibres, which are randomly oriented in the matrix material. We may note that the reinforcement efficiency is found only 1/5th that of an aligned composite in the longitudinal direction; however, the mechanical characteristics are isotropic.

For a particular composite, consideration of orientation and fibre length will depend on the level and nature of the applied stress as well as fabrication cost. Short-fibre composites (both aligned and randomly oriented) production rates are rapid, and intricate shapes can be formed that are not possible with continuous fibre reinforcement. Moreover, fabrication costs are considerably lower than for continuous and aligned.

Sandwich Structures

These have thin layers of facing materials over a low density material, or combcore, such as a polymer foam or expanded metal structure. A familiar sandwich-structured composite is corrugated cardboard. The corrugated paper core is covered by two faces of thin paper. In structures of this type, the facing material serves to fix the inner core in place. The core provides the strength. Typical face materials include aluminium alloys, fibre-reinforced plastics, titanium, steel and plywood.

Structurally, the core serves two purposes: *(i)* it separates the faces and resists deformations perpendicular to face plane, and *(ii)* it provides a certain degree of sheer rigidity along planes that are perpendicular to the faces. Foamed polymers, synthetic rubbers, inorganic cements, balsa wood, etc. materials and structures are used for cores. Core has lower stiffness and lower strength.

Another popular core consists of a '*honeycomb*' structure, which finds wide use in industries such as the aircraft industry, where higher strength and lower weight are important factors. The honeycomb structure consists of thin foils that have been formed into interlocking hexagonal cells, with axis oriented perpendicular to the face panels. The material used may be similar to the face material. These structures are light weight, stiff, and strong and can be filled to provide sound and vibration damping. The honeycomb structure is shown in Fig. 18.7.

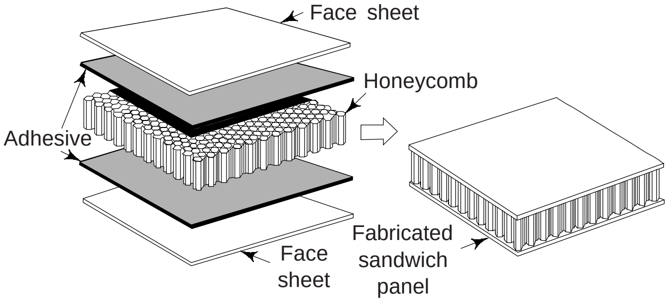


Fig. 18.7 Honeycomb structure

The Fibre Phase

An important characteristics of most materials, especially of brittle ones, is that a small diameter fibre is much stronger than the bulk material. We have read that the probability of the presence of a critical surface flaw that can lead to fracture diminishes with decreasing specimen volume. This feature of materials is used to advantage in the fibre-reinforced composites. Moreover, the materials used for reinforcing fibres have high tensile strengths.

Fibres can be grouped into three different classed based on their *diameter* and *character*: *whiskers, fibres* and *wires*. Whiskers are not utilized extensively as a reinforcement medium because they are extremely expensive and moreover, it is difficult and impractical to incorporate whiskers into a matrix. Graphite silicon carbide, silicon nitride, and aluminium oxide are few whisker materials. Some mechanical characteristics of wiskers, fibres and metallic wires are given in Table 18.3.

Table 18.3 Some mechanical characteristics of few fibre-reinforcement materials

Material	*Specific gravity*	*Tensile strength [GPa (10^6 psi)]*	*Specific strength (GPa)*	*Modulus of elasticity [GPa (10^6 psi)]*	*Specific modulus (GPa)*
		Whiskers			
Graphite	2.2	20 (3)	9.1	700 (100)	318
Silicon nitride	3.2	5–7 (0.75–1.0)	1.56–2.2	350–380 (50–55)	109–118
Aluminium oxide	4.0	10–20 (1–3)	2.5–5.0	700–1500 (100–220)	175–375
Silicon carbide	3.2	20 (3)	6.25	480 (70)	150
		Fibres			
Aluminium oxide	3.95	1.38 (0.2)	0.35	379 (55)	96
Aramid (Kevlar 49)	1.44	3.6–4.1 (0.525–0.600)	2.5–2.85	131 (19)	91
Carbon[a]	1.78–2.15	1.5–4.8 (0.22–0.70)	0.70–2.70	228–724 (32–100)	106–407
E-Glass	2.58	3.45 (0.5)	1.34	72.5 (10.5)	28.1
Boron	2.57	3.6 (0.52)	1.40	400 (60)	156
Silicon carbide	3.0	3.9 (0.57)	1.30	400 (60)	133
UHMWPE (Spectra 900)	0.97	2.6 (0.38)	2.68	117 (17)	121
		Metallic Wires			
High-strength steel	7.9	2.39 (0.35)	0.30	210 (30)	26.6
Molybdenum	10.2	2.2 (0.32)	0.22	324 (47)	31.8
Tungsten	19.3	2.89 (0.42)	0.15	407 (59)	21.1

[a] The term "carbon" instead of "graphite" is used to denote these fibres, since they are composed of crystalline graphite regions, and also of non crystalline material and areas of crystal misalignment.

Materials, which fall under the head *fibres* are either polycrystalline or amorphous and have small diameters. Fibrous materials are usually either polymers or ceramics, (e.g., the polymer aramids, glass, carbon, boron, aluminium oxide, and silicon carbide). Fibre-reinforced composites have been used for centuries, e.g. straw has been used in the making of mud bricks since the Pharaohs. Steel rebar has been used for many years in reinforced concrete structure. Fiberglass cloth and resin have been used to repair many automobiles fenders. These applications are among the more familiar.

Fine wires have relatively large diameters. Typical materials under this group include steel, molybdenum, and tungsten. Wires are widely used as a radical steel reinforcement in automobile tyres, in filament-wound rocket castings, and in wire-wound high pressure houses.

The Matrix Phase

The matrix phase of a fibrous composites may be a polymer, or ceramic and a metal. Metals and polymers are generally used as matrix materials because some ductility is desirable. To improve fracture toughness for ceramic-matrix composites, the reinforcing component is added.

The matrix phase serves several functions for fibre-reinforced composites: *(i)* it binds the fibres together and acts as a medium by which an externally applied stress is transmitted and distributed to the fibres. However, only a very small proportion of an applied load is sustained by the matrix phase. Also, the matrix material should be ductile. The elastic modulus of the fibre should be much higher than that of the matrix, *(ii)* Matrix has to protect the individual fibres from surface damage as a result of mechanical abrasion or chemical reactions with the environment. These interactions may cause surface flaws capable of forming cracks, which may lead to failure at low tensile stress levels. Lastly, the matrix separates the fibres and, by virtue of its relative softness and plasticity, prevents the propagation of brittle cracks from fibre to fibre, which may result in catastrophic failure. Obviously, matrix phase serves as a barrier to crack propagation. It is possible that some of the individual fibres fail, total composite fracture will not occur until large number of adjacent fibres, one having failed, form a cluster of critical size.

To minimize fibre pull-out, it is essential that adhesive bonding forces between fibre and matrix be high. In reality, in the choice of the matrix-fibre combination bonding strength is an important factor. The magnitude of this bond plays vital role in the ultimate strength of the composite. To maximize the stress transmittance from the weak matrix to the strong fibres, adequate bonding is essential.

Polymer-Matrix Composites (PMCs)

These materials consist of a polymer resin (here resin denote a high-molecular weight reinforcing plastic) as the matrix, with fibres as the reinforcement medium. In light of their room-temperature properties, ease of fabrication, and cost, these PMCs are used in the greatest diversity of composite applications, as well as in huge quantities. In accordance with reinforcement type (i.e., glass, carbon and aramid), various classifications of PMCs, along with their applications and the various polymer resins that are used, are as follows:

(a) Glass Fibre-Reinforced Polymer (GFRP) Composites Fiberglass is a composite consisting of glass fibres, which may be continuous or discontinuous and contained within a polymer matrix. These are commonly produced in '*E*' glass (*E* is for electrical), because it draws well and has good strength and stiffness. A typical composition (Wt%) would be $52SiO_2$, $17CaO$, $14Al_2O_3$, $10Ba_2O_3$ with some oxides of Mg, Na and K, and molten glass is gravity fed into a series of platinum bushings each of which has several hundred holes in its base. Fine glass filaments are drawn mechanically as the glass exudes from the holes, then wound on to drums at speeds of several thousand metres per minute.

The strength of the glass fibres is dependent upon the surface damage arising when they rub against each other during processing. The application of a size coating early at an early stage during manufacture minimises this degradation in properties, by reducing the propensity for forming these 'Griffith' cracks. The size consists of an emulsified polymer in water, and also has the effect of binding the fibres together for ease of further processing.

We are familiar with many fibre glass applications; e.g. automotive and marine bodies, plastic pipes, industrial floorings, and containers. In order to decrease vehicle weight and boost fuel efficiencies, the transport industries are utilizing to maximum extent of glass fibre-reinforced plastics. Many new applications are under investigation by the automative industry.

(b) Carbon Fibre-Reinforced Polymer (CFRP) Composites These consist of small crystallites of graphite (crystal structure see chapter 3). The atoms in the basal planes are held together by very strong covalent bonds, and there are weak Vander Waals forces between the layers. To obtain high modulus and high strength the layer planes of the graphite have to the aligned parallel to the axis of the fibre, and the modulus of carbon fibres depends on the degree of perfection of alignment of atom planes. This varies considerably with the particular manufacturing route adopted, of which there are three main possibilities:

(i) Starting with the polymer PAN (polyacrylonitride), which closely resembles polyethylene in molecular confirmation, it is converted into a fibre and then stretched to produce alignment of the molecular chains along the fibre axis. White still under tension, it is heated in oxygen to form cross-links between ladder-molecules and finally chemically reduced to give (at high temperatures) a graphic structure. The final graphitisation temperature determines whether the fibres have maximum stiffness but a relatively low strength (Type-I fibres), or whether they develop maximum strength (Type-II).

(ii) Alternatively, fibres may be produced by melt-spinning molten pitch. During the spinning process, the orifice causes the planar molecules to become aligned. It is then treated, whilst held under tension in order to maintain its preferred orientation, at temperatures upto 2000 C to form the requisite grains of graphite.

(iii) It is also possible to stretch either of the fibre types described above during the graphitization stage, giving further orientation of the layers parallel to the fibre axis.

Carbon fibres have the highest specific modulus and specific strength of all reinforcing fibre materials. Carbon fibres retain their high tensile modulus and high strength at elevated temperatures. However, high temperature oxidation may be a problem. Carbon fibres are not affected by moisture or a wide variety of solvents, acids and bases at room temperature. Carbon fibres exhibit a diversity of physical and mechanical characteristics, allowing composites incorporating these fibres to have specific engineering properties.

On the basis of tensile modulus, carbon fibres can have four classes: *standard*, *intermediate*, *high* and ultrahigh moduli. Fibre diameters generally range between 4 and 10 μm. Both *continuous* and *chopped* forms of these fibres are available. Moreover, carbon fibres are usually coated with a protective epoxy size that also improves adhesion with the polymer matrix.

Carbon-reinforced polymer composites are extensively used in sports and recreational equipment (fishing rods, golf clubs), filament-wound rocket motor cases, pressure vessel and aircraft structural components—both military and commercial, fixed wing and helicopters.

(c) Aramid Fibre-Reinforced Polymer Composites Aramid fibres are high-strength, high modulus materials, especially desirable for their outstanding strength-to-weight ratios, which are superior to metals. Chemically, this group of materials is called as poly paraphenylene terephthalamide. There are a number of aramid materials. Trade names of two of these most common materials are Kevlar and Nomex. Kevlar has several grades (viz, Kevlar 29, 49, and 149) that have different mechanical behaviours. Mer chemistry and mode of chain alignment for aramid (Kevlar) are shown in Fig. 18.8. These fibres have longitudinal tensile strengths and tensile moduli (Table 18.3) and they are higher than for other polymeric fibre materials, but they are relatively weak in compression. This material is also known for its toughness, impact resistance, and resistance to creep and fatigue failure. Although aramids are thermoplastics, they are nevertheless, resistant to combustion and stable to relatively high temperatures. The temperatures range over which aramids retain their high mechanical properties is between –200 and 200°C. Aramids are susceptible to degradation by strong acids and bases, but they are relatively inert in other solvents and chemicals.

Usually, aramid fibres are used in composites having polymer matrices. Common matrix materials are the epoxies and polyesters. Aramid fibres are relatively flexible and somewhat flexible. Aramid composites are used in ballistic products (bullet-proof vests), sporting goods, tyres, ropes, missle cases, pressure levels, and as a replacement for asbestos in automotive brake and cluch linings, and gaskets.

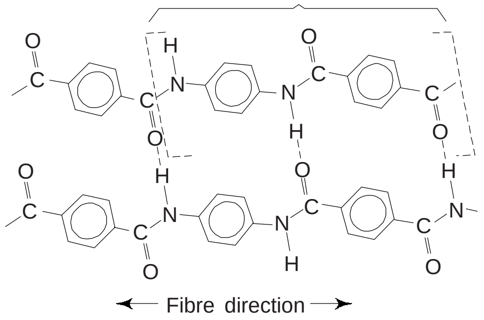


Fig. 18.8 Mer and chain structures for aramid (Kevlar) fibres

Other Fibre Reinforcement Materials

Other fibre materials are boron, silicon carbide, and aluminium oxide. However, these fibre materials are used to much lesser degrees. Boron fibre-reinforced polymer composites find use in military aircraft components, helicopter robot blades, and some sporting goods. Silicon carbide and alumina fibres are used in tennis rackets, rocket nose cones and circuit boards.

Polymer-Matrix Materials

Polyesters and vinyl esters are the most widely used and least expensive polymer resins. These matrix materials are basically used for fibre glass reinforced composites. For mutations of a large number of resin provide a wide range of properties for these materials. The epoxies are more expensive and in addition to wide ranging commercial applications, also find extensive use in PMCs for aerospace applications. In comparison to polyesters and vinyl resins, epoxies have better mechanical properties and resistance to moisture. Polymide resins are used for high temperature applications (~230°C). High-temperature thermoplastic resins may find wider future aerospace application. These include polyetheretherketon (PEEK), polyphenylene sulphide (PPS), and polyetherimide (PEI).

Metal-Matrix Composites (MMCs)

The matrix in these composites is a ductile metal. These composites can be used at high service temperatures than their base metal counterparts. The reinforcement in these materials may improve specific stiffness, specific strength, abrasion resistance, creep resistance, thermal conductivity, and dimensional stability. In comparison to polymer-matrix composites, these materials have higher operating temperatures, non flammability and greater resistance to degradation by organic fluids. In comparison to PMCs, MMCs are much more expensive and therefore, their use is somewhat restricted.

Alloys of aluminium, magnesium, titanium, and copper, and superalloys are used as matrix materials. The reinforcement may be in the form of particulates, both continuous and discontinuous fibres, and whiskers (concentrations range 10 to 60 Vol%) carbon, silicon carbide, boron, alumina, and refractory metals are continuous fibre materials, whereas silicon carbide whiskers, chopped fibres of alumina and carbon, and particulates of silicon carbide and alumina are discontinuous reinforcements sense, cermets fall within MMCs.

At elevated temperatures, some matrix-reinforcement combinations are highly reactive. This problem can be resolved either by using a protective surface coating to the reinforcement or by modifying the matrix alloy composition.

Recently, automobile industries have begun to use MMCs in their products, e.g. some engine components have been introduced consisting of an aluminium-alloy matrix that is reinforced with alumina and carbon fibres. This MMCs is light in weight and can resist wear and thermal distortion. MMCs are also used in driveshafts (that have higher rotational speeds and reduced vibrational noise levels), forged suspension and transmission components, and extruded stabilizer. MMCs are used by aerospace industry. Aluminium alloy metal-matrix composites, boron fibres are used as the reinforcement for the Space Shuttle Orbiter, and continuous graphite fibres for Hubble Telescope.

Using refractory metals such as tungsten, the high temperature creep and rupture properties of some of the superalloys (Ni-and Co based alloys) may be enhanced by fibre reinforcement.

Ceramic-Matrix Composites (CMCs)

The fracture toughnesses of ceramics have been improved significantly by the development of CMCs-particulates, fibres, or whiskers of one ceramic material that have been embedded into a matrix of another ceramic. The fracture toughnesses of CMCs materials lies between 6 and 20 MPa$\sqrt{m}$. Increasing fibre content improves strength and fracture toughness. There is a considerable reduction in the scatter of fracture strengths for whisker-reinforced ceramics relative to their unreinforced counterparts. These CMCs exhibit improved high-temperature creep behaviour and resistance to thermal shock.

SiC whisker-reinforced aluminas are being utilized as cutting tool inserts for machining hard metal alloys. We may note that tool lives for these materials are greater than for cemented carbides.

Carbon-Carbon Composites

Carbon-carbon composite, i.e., carbon fibre-reinforced carbon-matrix composite is one of the most advanced and promising engineering material. Obviusly, in this material, both reinforcement and matrix are carbon. These materials have high-tensile modulii and tensile strengths that are retained to temperatures in excess of 2000°C, resistance to creep, and relatively large fracture toughness values. These materials have low coefficients of thermal expansion and relatively high thermal conductivities. These characteristics of these materials, coupled with high strengths, give rise to a relatively low susceptibility to thermal shock. The major drawback of these materials is a propensity to high-temperature oxidation. These materials are relatively new and expensive, and therefore, are not in wide use.

These materials are employed in rocket motors, as friction, materials in aircraft and high-performance automobiles, for hot-pressing molds, in components for advanced turbine engines, and as ablative shields for re-entry vehicles.

Hybrid Composites

The composites obtained by using two or more different kinds of fibres in a single matrix are termed as *hybrid*. Hybrid composites have a better all-round combination of properties than composites containing only a single fibre type. Although a variety of fibre combinations and matrix materials are used, but in the most common system, both carbon and glass fibres are incorporated into a polymeric resin. The carbon fibres are expensive, but they are strong and relatively stiff and provide a low density reinforcement. Glass fibres lack the stiffness of carbon, but they are inexpensive. The glass-carbon hybrid may be produced at a lower cost than either of the comparable all-carbon or all-glass reinforced plastics, and also it is stronger and tougher.

The properties of hybrid composites are anisotropic. When these are stressed in tension, failure is usually non catastrophic, i.e., does not occur suddenly. The carbon fibres are the first to fail, at which time the load is transferred to the glass fibres. Upon failures of the glass fibres, the matrix phase will have to sustain the applied load. We may note that eventual composite failure concurs with that of the matrix phase.

Hybrid composites find applications in lightweight land, water and air transport structural components, lightweight orthopedic components and sporting goods.

5. FABRICATION

Fabrication of composite materials can be accomplished through hand lay-up, injection moulding, lamination, filament winding, or pultrusion. The oldest method of composite manufacture is *hand lay-up*. In it, the resin is poured into a mould or applied to the surface and then the fabric is rolled into the resin or applied to the surface. The hand lay-up technique is used in boat building, auto repair, and home projects, because it requires minimal equipment and the result cures at room temperature.

Pultrusion is used for the manufacture of components having continuous lengths and a constant cross-sectional shape, (i.e. rods, tubes, beams). With this technique, a continuous fibre *rovings or tows* (A roving or tow is a loose and untwisted bundle of continuous fibres that are drawn together as parallel strands), are first impregnated with a thermosetting resin. These are then pulled through a steel die that performs to the desired shape and also establishes the resin/fibre ratio. Then the stock passes through a curing die that is precision machined so as to impart the final shape. This die is also heated to initiate curing of the resin matrix. A pulling device draws the stock through the dies and also determines the production speed. By using centre mandrels or inserted hollow cores, tubes and hollow sections are made possible. Glass, carbon and aramid fibres are principal reinforcements and normally added in concentrations between 40 and 70 Vol%. Polyesters, vinyl esters, and epoxy resins are commonly used matrix materials.

Pultrusion is a continuous process that is easily automated. Production rates are relatively high which makes it very cost effective. Moreover, a wide variety of shapes are possible. Also, there is really no practical limit to the length of stock that may be manufactured.

Prepag fibres are woven fibres impregnated with resin prior to manufacture. The prepag fabric is laid in place, heated while it is being laid, and cooled by a blast of cold air to set it in place initially. The formed product is then cured at elevated temperatures and pressures.

Resin Transfer Moulding (RTM)

This is a two-piece matching cavity-mould system used for moulding composites (Fig. 18.9).

The reinforcement material is placed in the mould cavity. A chopped mat is often draped along the inside of the mould. The mould halves are clamped together and the resin is introduced into the cavity. RTM offers fast curing times and is labour-intensive than other methods.

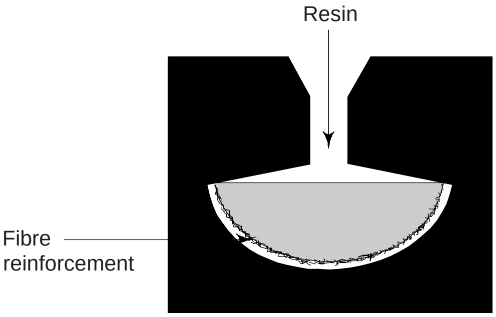


Fig. 18.9 Resin transfer moulding

Ultimate Reinforced Thermoset Reaction Injection (URTRI)

This process allows the use of cores in the moulding process. A core is cast from a high-temperature epoxy foam. The core is wrapped with layers of graphite fabric and is positioned in the mould cavity. Epoxy resin is pumped into the heated mould and cured for about 5 min. The mould is then opened, and the part is extracted. This process is shown in Fig. 18.10.

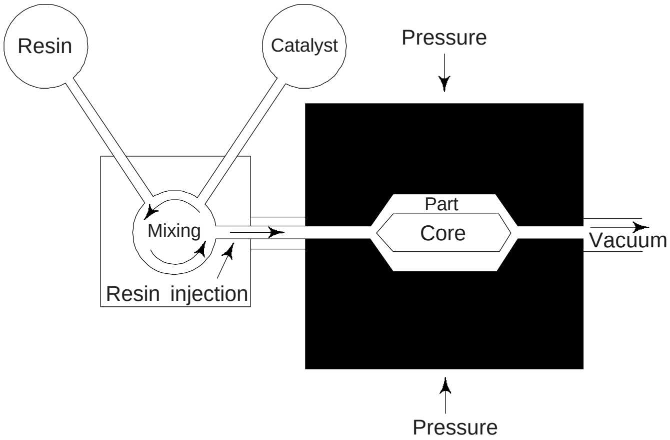


Fig. 18.10 Ultimately reinforced thermoset reaction injection process (URTRI)

The primary drawback of composites is the time it takes to manufacture them. Parts that are stamped out of metal every few seconds take several minutes to complete if made of composites. The increased time often increases the cost of manufacturing. However, with the rapid advances in technology, composite production has increased and applications have broadened, e.g. more and more automotive parts are being prepared from composites.

Example 1 *A continuous and aligned glass fibre-reinforced composite consists of 40 Vol% of glass fibres having a modulus of elasticity 69 GPa and 60 Vol% of a polyester resin that, when hardened, displays a modulus of 3.4 GPa. Calculate the modulus of elasticity of this composite in the longitudinal directions.* [B.E.]

Solution We have

$$\begin{aligned} E_{cl} &= E_m V_m + E_f V_f \\ &= (3.4 \text{ GPa}) \times 0.6 + (69 \text{ GPa}) \times (0.4) \\ &= 30 \text{ GPa} \\ &= 4.3 \times 10^6 \text{ psi} \end{aligned}$$

Example 2 *Calculate the elastic modulus of the composite material described in Example 1, but assume that the stress is applied perpendicular to the direction of the fibre alignment.*

Solution We have

$$\begin{aligned} E_{cl} &= \frac{E_m E_f}{V_m E_f + V_f E_m} \\ &= \frac{(3.4 \text{ GPa})(69 \text{ GPa})}{(0.6)(69 \text{ GPa}) + (0.4)(3.4 \text{ GPa})} \\ &= 5.5 \text{ GPa} = 0.81 \times 10^6 \text{ psi} \end{aligned}$$

SUGGESTED READINGS

1. K.K. Chawla, '*Composite Materials Science and Engineering*', 2nd ed., Springer-Verlag, New York (1998).
2. D. Hull and T.W. Clyne, '*An Introduction to Composite Materials*', 2nd ed., Cambridge Univ. Press, New York (1996).
3. S.T. Peters, '*Handbook of Composites*', 2nd ed. Kulwer Academic, Norwell, MA (1998).

REVIEW QUESTIONS

1. Define the term composite and explain what are advanced composites? Give few examples. [AMIE]

2. What factors influence the final properties of composites?

3. What is the difference between matrix and dispersed phases in a composite material?

4. Contrast the mechanical characteristics of matrix and dispersed phase for fibre reinforced composite materials? [B.E.]

5. What is difference between cement and concrete?

6. Mention three important limitations that restrict the use of concrete as a structural material.

7. What are the functions of the matrix phase for a polymer-matrix fibre-reinforced composite?

8. For a polymer-matrix-fibre-reinforced composite, compare the desired mechanical characteristics of matrix and fibre phases.

9. Why glass fibres are most commonly used for reinforcement?

10. What is the difference between carbon and graphite?

11. Why fibre glass reinforced composites are utilized extensively?

12. What is a hybride composite? Mention two important advantages of hybrid composites over normal fibre composites.

PROBLEMS

1. A tensile load of 200 N is applied to a composite material of 1 sq. mm cross-sectional area. The volume of parallel fibre is 40%. Determine the stress in fibres, if axis of load is *(i)* parallel to the fibres, and *(ii)* perpendicular to the fibres. Given that the ratio of Young's moduli of fibre material and matrix material is 7. [*Ans. (i)* 58.8 N/mm^2 *(ii)* 411.6 N/mm^2] [B.E.]

2. Calculate the longitudinal tensile strength of an aligned glass fibre-epoxy matrix composite in which the average fibre diameter and length are 0.015 mm and 2.0 mm, respectively, and the volume fraction of fibres is 0.25. Assume that *(i)* the fibre-matrix bond strength is 100 MPa *(ii)* the fracture strength of the fibres is 3500 MPa, and *(iii)* the matrix stress at composite failure is 5.5 MPa.

[*Ans.* σ_{cd}^{*} = 822 MPa] [B.E.]

OBJECTIVE QUESTIONS

1. The cermets are examples of
(1) ceramic-metal composites (2) carbon-carbon compasites
(3) metal-matrix composites (4) none of the above (1)

2. The critical fibre length (l_c) is related to the fibre disameter *d*, ultimate (or tensile) strength σ_f^* and fibre-matrix bond strength τ_c as

(1) $l_c = \dfrac{\sigma_f^* d}{\tau_c^2}$ (2) $l_c = \dfrac{\sigma_f^* d}{2\tau_c}$ (3) $l_c = \dfrac{\sqrt{\sigma_f^* d}}{\tau_c}$ (4) $l_c = \sqrt{\sigma_f^* \, d\tau_c}$ (2)

3. Whiskers are
(1) very thin steel wires (2) very thin copper wires
(3) very thin single crystals (4) none of the above (3)

4. The matrix phase of fibrous composites may be
(1) a metal only (2) a polymer only
(3) a ceramic only (4) a metal, polymer, or ceramic (4)

5. For metal-matrix composites the matrix is a
(1) ductile metal (2) polymer
(3) ceramic (4) none of the above (1)

6. A laminar composite is composed of
(1) one-dimensional sheets or panels (2) two-dimensional sheets or panels
(3) three-dimensional sheets or panels (4) none of the above (2)

7. Most advanced and promising composite is
(1) carbon-carbon (2) metal-matrix
(3) polymer-matrix (4) none of the above (1)

8. Trade names for two most common aramid materials are
(1) silicon carbide, Silicon nitride (2) e-glass, Aluminium oxide
(3) kevlar, nomex (4) high strength steel, Aluminium oxide (3)

SHORT QUESTION ANSWERS

1. What are ‘composites’?

Ans. These are artificially produced multiphase materials having a desirable combination of the best properties of the constituent phases. Usually, one phase (the matrix) is continuous and completely surrounds the other (the disperse phase).

2. What is ‘concrete’?

Ans. This is a type of large particle, composite consists of an aggregate of particles bonded together with cement.

3. What is ‘portland cement concrete’?

Ans. The aggregate consists of sand and gravel. The cementitious bond develops as a result of chemical reaction between the portland cement and water.

4. Out of the several composite types, for which the potential for reinforcement efficiency is greatest?

Ans. Fibre reinforced.

5. What types of fibres will be for short and discontinuous fibrous composites?

Ans. The fibres may be either aligned or randomly oriented.

6. Can polymer-matrix be reinforced with glass, carbon, and aramid fibres?

Ans. Yes.

7. Which are the more advanced composites?

Ans. *(i)* Carbon-Carbon and *(ii)* hybrids

8. Write two kinds of structural composites.

Ans. *(i)* Laminar composites and *(ii)* Sandwich panels.

19

Nanostructured Materials

1. INTRODUCTION

The discovery of fullerenes and carbon nanotubes has opened a new research area in physics, chemistry and material science. Nanostructured materials may be defined as those materials whose structural element clusters, crystallites or molecules—have dimensions in the 1 to 100 nm range. Clusters of atoms consisting of typically hundreds to thousands on nanometer (nm) scale are commonly called as nanoclusters. These small groups of atoms, in general, go by different names such as nanoparticles, nanocrystals, quantum dots and quantum boxes. Substantial work is being carried out in the domain of nanostructured materials and nanotubes in the past one decade. The explosion in both academic and industrial interest in these nanomaterials over the past decade arises from the remarkable variations in fundamental electrical, optical and magnetic properties that occur as one progresses from an 'infinitely extended' solid to a particle of material consisting of a countable number of atoms. Nanostructured materials has led to new basic science phenomena and several applications for the short, medium and long term. Examples are: nanoelectronic devices, quantum wires, electron field emitters for ultra-thin TV screens, nanoprobes, high resolution tips for scanning and atomic force microscopes, sensors, ultrahigh strength composites, gas storage, nanodevices, and parts of nanomachines, among others. Carbon-based nanomaterials and nanostructures including fullerenes and nanotubes play an increasingly pervasive role in nanoscale science and technology. Carbon nanotubes are currently being studied in an effort to understand their novel structural, electronic and mechanical properties to explore their immense potential for many applications in nanoelectronics, and as actuators and sensors.

Fullerenes and carbon nanotubes can be seen as curves pieces of graphite. Graphite is another polymorph of carbon.

Graphite

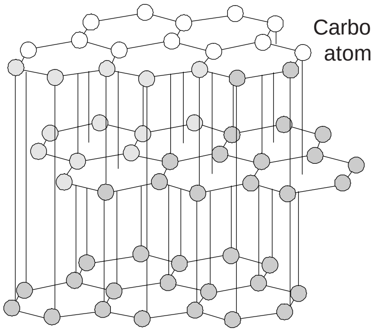


Fig. 19.1 The structure of Graphite

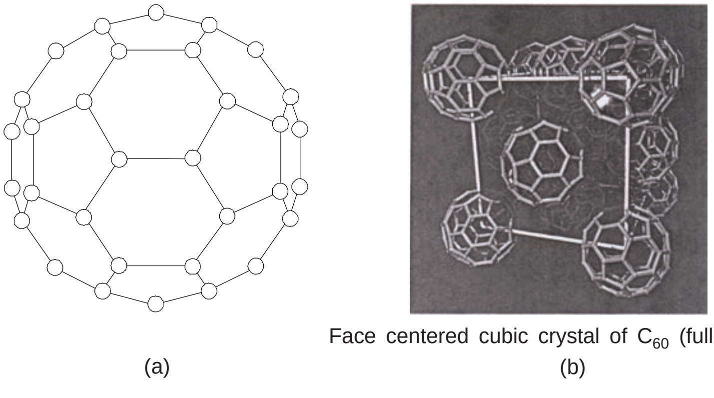
Graphite has a crystal structure distinctly different from that of diamond and is also more stable than diamond at ambient temperature and pressure (Fig. 19.1). Graphite is formed by flat hexagonal layers of carbon atoms separated 3.35Å. The C-C distance is 1.42Å and the structure belongs to P 63 mc space groups with lattice constants a = 2.46 Å and c = 6.71 Å. Graphite is highly anisotropic solid, structurally its interplaner spacing (3.35 Å) is quite large compared to the in plane interatomic spacing (1.42 Å). Physically its stiffness along the plane is quite large because of strong σ bonds and in the perpendicular direction it is weak because of the Vander Waal's force. The graphite structure is composed of layers of hexagonally arranged carbon atoms; within the layers, each carbon atom is bonded to three coplanar neighbour atoms by strong covalent bonds. Carbon when sp^2 hybridized contains three strong π-bonds which lie in a triangular plane with bond angle of 120° between them and one weakly bonding π-bond perpendicular to the plane. Each carbon atom in the plane is covalently bonded to three others in plane with strong σ bonds of the sp^2 hybridization, this network forms a continuous sheet of carbon. The fourth bonding electron participates in a weak Vander Waals type of bond between the layers. As a consequence of these weak interplanar bonds, interplanar cleavage is facile, which gives rise to the excellent lubricative properties of graphite. Also, the electrical conductivity is relatively high in crystallographic directions parallel to the hexagonal sheets. Graphite has metallic conductivity along the plane and semiconducting perpendicular to the plane. The anisotropy in conductivity is about 10^3.

Other desirable properties of graphite include: high strength and good chemical stability at elevated temperatures and in non oxidizing atmospheres, high thermal conductivity, low coefficient of thermal expansion and high resistance to thermal shock, high absorption of gases, and good machinability. Graphite is commonly used in heating elements for electric furnaces, as electrodes for ac welding, in metallurgical crucibles, in casting moulds for metal alloys and ceramics, for high temperature refractories and insulations, in rocket nozzles, in chemical reactor vessels, for electrical contacts, brushes and resistors, as electrodes in batteries, and in air purification devices.

There are other layered materials which also can acquire curvature to generate new nanomaterials with novel applications. The other types of nanostructured materials which has been considered for applications in opto-electronic devices and quantum-optic devices are nonosized powders of silicon, silicon-nitride (SiN), silicon carbide (SiC) and their thin films. Some of these nanomaterials, e.g. SiC and SiN are also used as advanced ceramics with controlled microstructures because their strength and toughness increase when the grain size diminishes.

Fullerenes

In the case of fullerenes, graphite acquires curvature due to the presence of pentagonal rings in the hexagonal graphite sheet. Fullerenes exists in discrete molecular form and consists of a hollow spherical cluster of sixty carbon atoms; a single molecule is denoted by C_{60}. Each molecule is composed of groups of carbon atoms that are bonded to one another to form both hexagon (six-carbon atom) and pentagon (five carbon atom) geometrical configurations. One such molecule (Fig. 19.2), is found to consist of 20 hexagons and 12 pentagons which are arrayed such that no two pentagons share a common side; the molecular surface thus exhibits the symmetry of soccer ball. The material composed of C_{60} molecules that contains sixty carbon atoms in a network of sp^2 bonding which forms a spherical structure, is known as *buckminsterfullerene*, from a number of perspectives, a molecule almost 'custom built for nanoscience. Discovered in 1985 by Kroto et al., C_{60} is approximately 1 nm in diameter (the cage diameter is 0.71 nm;


Face centered cubic crystal of C_{60} (fullerite.)

(a) (b)

Fig. 19.2(a) The structure of C_{60} molecule (b) Face centered cubic crystal of C_{60} fullerite

the intermolecular space in fullerite, the bult crystal is 1.004 nm readily synthesized in large (i.e. gram) quantities and of key significance—its physical and chemical properties may be tuned via the addition of a range of elemental and molecular species either to the fullerene lattice, to the fullerene cage or within the cage. The term *fullerene* is used to denote the class of materials that are composed of this type of molecule. Fullerenes are made up of hexagonal and pentagonal rings of carbon arranged exactly as of an Indian football.

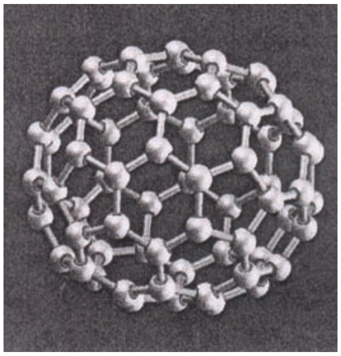


Fig. 19.3 Structure of fullerene C_{70}

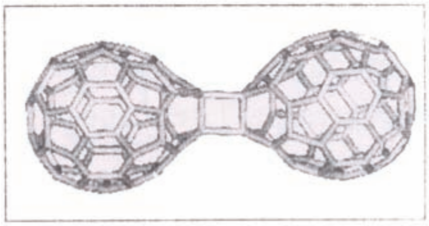


Fig. 19.4 Structure of C_{120} fullerene

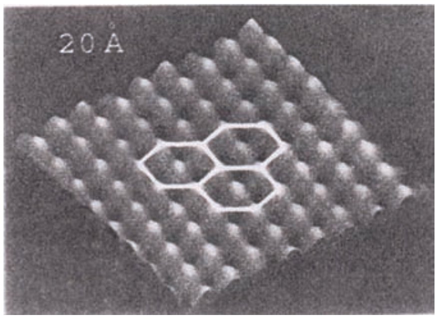


Fig. 19.5 AFM image of (HOPG) graphite

Based on fullerenes (C_{60}) presently we find the existence of C_{36}, C_{70} (Fig. 19.3), C_{80}, C_{100} and also few members of geometrically allowed fullerene family. These fullerene molecules differ in the number of hexagonal rings present. The shape of this molecule changes as the number of carbon atom differs. The C_{70} looks similar to a rugby ball and C_{120} (Fig. 19.4) takes a dumbbell shape.

Diamond and graphite are what may be termed *network solids*, in that all of the carbon atoms form primary bonds with adjacent atoms throughout the entirety of the solid. The common *synthetic graphite* is usually manufactured as composites. In this process petroleum coke is mixed into a paste with a coal tar pitch and then heated to around 120°C–1400°C. This step drives out all volatile material from the petroleum coke. Further heating this to 2500°C–3000°C causes an ordering of the carbon atoms to graphitize the mixture into a completely graphitic structure (Fig. 19.5).

Presently, there are methods to prepare C_{60} molecules in large quantities, so the study of this molecular form of carbon is advancing in all possible arenas. By way of contrast, the carbon atoms in the buckminsterfullerene bond together so as to form these spherical molecules. In the solid state, the C_{60} units form a crystalline structure and pack together in a face-centered cubic array.

As a pure crystalline solid, this material is electrically insulating. However, with proper impurity additions, it can be made highly conductive and semiconductive. Presently we know that C_{60} doped with alkali metals become superconducting with $T_c \sim 50$ K. Fullerene research has got immense scope in nanoscience and technology. They find use in catalytic chemistry, biomolecular recognition, as molecular sieves, nanoreactors, and also as inhibitors to the activity of HIV virus, etc.

Carbon Nanotubes

Another molecular forms of carbon has recently been discovered that has some unique and technologically promising properties. Ijiyama of NEC Japan in 1991 first announced the synthesis of carbon nanotubes. In his experiments on the arc discharge between graphite electrodes, alongwith the collected shoot particles there were some fine tube like structures. This on careful analysis revealed that they are carbon nanotubes (CNTs). Its structure consists of a single sheet of graphite, rolled into a tube, both ends of which are capped with C_{60} fullerene hemispheres. The tubes can be either open at their ends or capped at one or both ends with half a spheroidal fullerene. Figure 19.6 shows a schematic representation of a carbon nanotube. A nanotube is completely specified by what is referred to as a roll up vector, which identifies its helical nature and diameter. Depending on its roll-up vector, a nanotube can have either metallic Fig. 19.6(a) or semiconducting (Fig. 19.6b, c) properties.

[There are a vast number of ways in which a graphite sheet can be rolled up to form a seamless cylinder and hence a wide variety of nanotubes exist. Nanotubes are characterized by their *chiral* (or wrapping) vector, $\vec{c}$ such that $\vec{c} = n\vec{a}_1 + m\vec{a}_2$, where $\vec{a}_1$ and $\vec{a}_2$ are the basis vectors of the graphite (more correctly graphene) lattice and m, n are integers (Fig. 19.7(a)). The chiral vector spans the circumference of the tube formed by joining the dotted lines shown in Fig. 19.7(a). Those tubes with chiral vectors of the form (n, o) are termed zig-zag tubes (Fig. 19.7(b)) whereas $n = m$ a so called ‘arm chair’ tube results Fig. 19.7(c)). All other values of n and m produce a chiral tube (Fig. 19.7(d)).

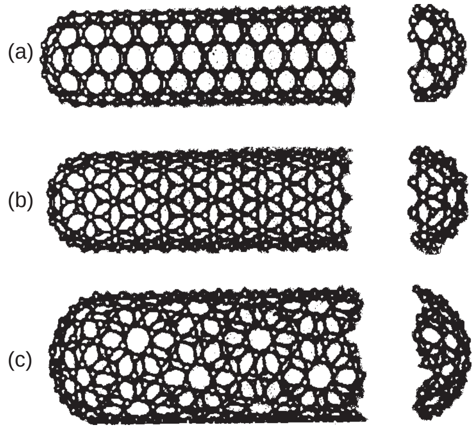


Fig. 19.6 The structure of carbon nanotubes (a) metallic (b) and (c) semiconducting

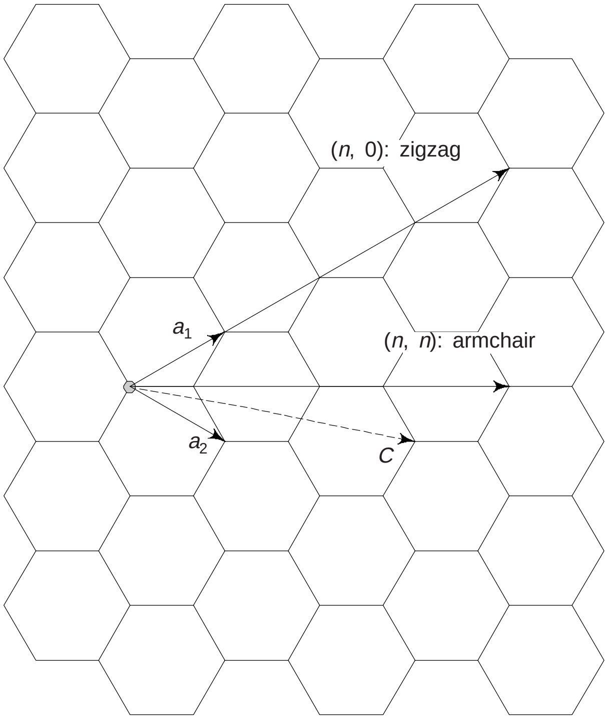


Fig. 19.7a The relationship between the graphite lattice basis vectors $\vec{a}_1$, $\vec{a}_2$ and the chiral vector $\vec{c} = n\vec{a}_1 + m\vec{a}_2$ used to characterize carbon nonotubes. Two limiting cases are shown: (n, 0) indices are associated with zig-zag tubes whereas (n, n) indices are associatd with armchair tubes. All other tubes are chiral

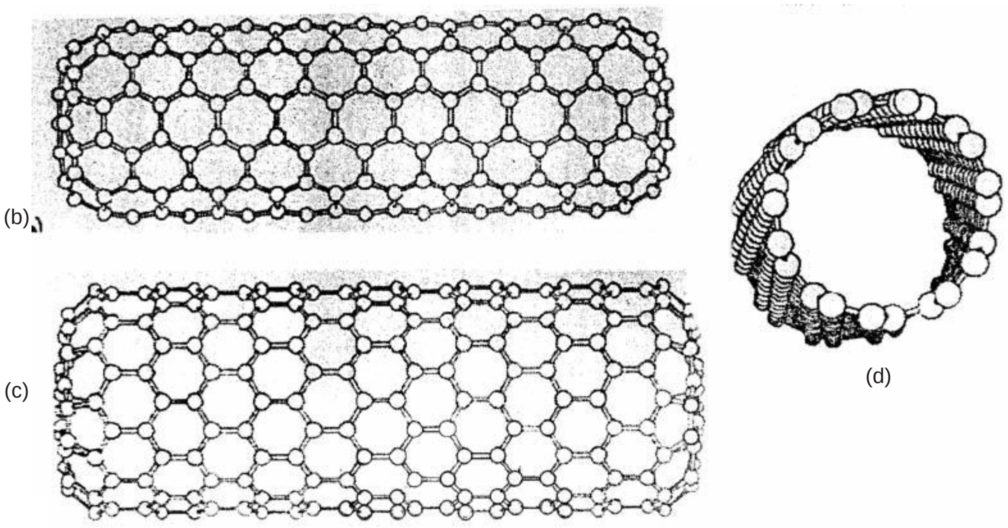


Fig. 19.5 (b), (c) and (d) Diagrams of three types of nanotubes: (b) armchair (c) zigzag and (d) chiral

Closely following the synthesis of nanotubes came the remarkable theoretical prediction that their electronic properties could be changed between metallic and semiconducting simply by varying the tube diameter or its helicity i.e. by changing the values of *n* and *m*. A key theoretical result is that armchair nanotubes are metallic, whereas for all other SWNTs, when n–m = 3 l (l an integer) the tubes are metallic; otherwise they are semiconducting.

These striking electronic properties can be understood within a tight-binding framework. In the direction along *c* (i.e. along circumference of the tube), periodic boundary conditions constrain the values of electron wavevector to those given by

$$\vec{c} \cdot \vec{k} = 2\pi q \quad (q = 1, \ldots\ n) \tag{1}$$

Taking the particular example of an armchair nanotube, this can be reduced to

$$k_x\, n\sqrt{3}\ a_o = 2\pi q \tag{2}$$

where $a_o = (\sqrt{3})\ 0.142$ nm is the graphite lattice constant and defines the chiral vector for the armchair tube (in this case $\vec{c} = n(\vec{a_1} + \vec{a_2})$).

The well-known tight binding dispersion relation for a 2 D graphite sheet is

$$E_{2D}\,(k_x, k_y) = +\gamma_o \left\{1 + 4\cos\left(\frac{\sqrt{3}\,k_x\ a_0}{2}\right)\cos\left(\frac{k_y\,a_0}{2}\right) + 4\cos^2\left(\frac{k_y\,a_0}{2}\right)\right\}^{1/2} \tag{3}$$

γ_o is the C-C overlap integral. Substituting into (11) the allowed values of k_x given by (10) yields

$$E_{10}^{A}\,(k_y) = \pm\gamma_o \left\{1 \pm 4\cos\left(\frac{q\pi}{n}\right)\cos\left(\frac{k_y\,a_0}{2}\right) + 4\cos^2\left(\frac{k_y\,a_0}{2}\right)\right\}^{1/2} \quad (-\pi < k_y a_o < \pi) \tag{4}$$

where the subscript *A* denotes that (4) is valid only for armchair tubes. It is assumed that the value of the overlap integral for graphite and nanotube are identical. The 1 D dispersion relations given by (4) are shown graphically in Fig. 19.7(e) for the case of (5, 5) armchair SWNT. There are six dispersion curves for both the valence and conduction bands. In each case, four bands are doubly degenerate (bold continuous lines) and two bands are non-degenerate (thin, dashed lines). Thus ten electronic levels comprise the valence

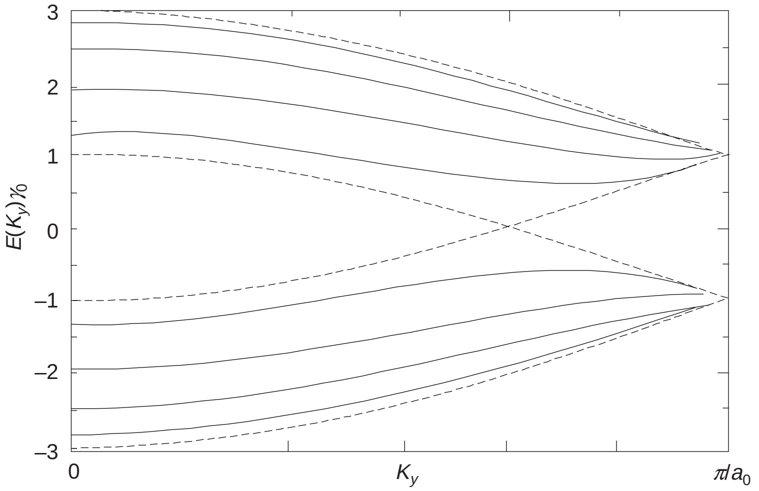


Fig. 19.7(e) One dimensional dispersion relation for an armchair (5, 5) nanotube

bands and ten comprise the conduction bands, as expected from the fact that there are ten hexagons around the circumference of a (5, 5) nanotube. There are two particularly noteworthy features of the dispersion relations shown in Fig. (7(b), (c) and (d)). First there is a very high degree of degeneracy at the Brillouin Zone boundary. One can readily understood this from the fact that at $k_y a_o = \pi$, Eq. (4) reduces to

$$E_{1D}^{A} = \pm\gamma_o$$

The second and more important feature of Fig. (19.7(b), (c) and (d)) in terms of electronic properties of armchair nanotubes is that the highest valence band and lowest conduction band are degenerate at $k_y = \pm 2\pi/(3a_o)$ and each cross the Fermi energy ($E = 0$) at that point. This is true of all armchair nanotubes-thus all armchair nanotubes be metallic.

We must note that Eqs. (2)–(4) represent taking a 1 D slice-in a direction given by

$$E(\theta) = -KV\cos^2\theta$$

where K is the magnetic anisotropy constant, V the volume of the particle, θ the angle between the magnetization vector and an easy direction of magnetization and $E(\theta)$ is the magnetic anisotropy energy for a uniaxial particle, through the 2 D graphite band structure. The crossing at $E = 0$ in Fig. 19.7(e) occurs because the corresponding 2 D graphite bands cross at the K points (i.e. the corners) of the 2 D Brillouin Zone. For armchair tubes, there is always an allowed $\vec{K}$-vector that goes through K point of the 2 D lattice. However, this alignment of the tube wavevector and K-points of the graphite Brillouin zone does not automatically occur for zigzag or chiral tubes and hence these can have semiconducting character. Only for n – m = 3 l is there an alignment of an allowed $\vec{k}$-vector and a K-point of the 2 D Brillouin Zone, giving rise to metallic behaviour.]

The *nano* in the nanotube prefix denotes that tube diameters are on the order of a nanometer (i.e., 100 nm or less). Each nanotube is a single molecule composed of millions of atoms; the length of this molecule is much greater (on the order of thousands of times greater) than its diameter.

According to number of layers, carbon nanotubes (CNTs) can be single-walled and multi-walled (MWNTs). In multi-walled nanotubes, more than one CNTs are coaxially arranged (Fig. 19.8). Iijima realized

Fig. 19.8 SEM image of CNTs

that graphite could be bent to form MWCNTs with different helicities or chiralities which refer to the way hexagonal rings are arranged with respect to the tube axis (Fig. 19.9).

It is found theoretically that the electronic properties of carbon nanotubes depended on the diameter and helicity: in particular, all of the so-called armchair type nanotubes are conductors (Fig. 19.9) and most zig-zag nanotubes are semiconductors (Fig. 19.9).

These nanotubes are extremely strong and stiff, and relatively ductile. According to indirect measurements, it has been found that carbon nanotubes are 100 times stronger than steel and six times lighter. For single walled nanotubes, tensile strength range between 50 and 200 GPa (approximately an order of magnitude greater than for carbon fibres); this is the strongest known material. Elastic modulus values are on the order of one tetrapascal [TPa: (1 TPa = 10^3 GPa)], with fracture strain between 5% and 20%. Furthermore, nanotubes have relatively low densities. On the basis of unique properties, the carbon nanotube has been termed the 'ultimate fibre' and is extremely promising as a reinforcement in composite materials.

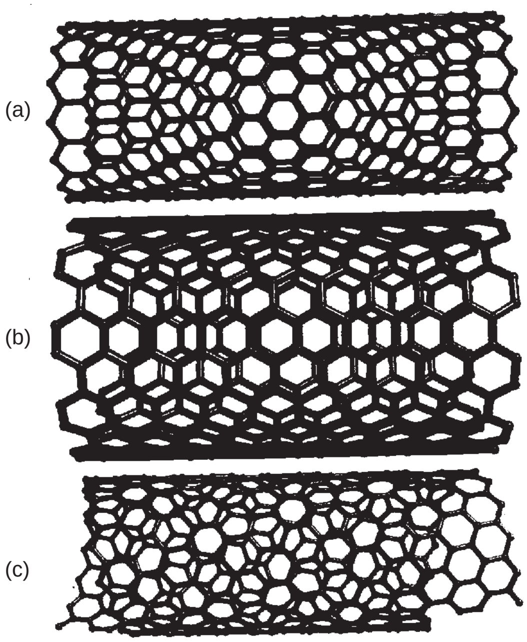


Fig. 19.9 **Structures of carbon nanotubes (a) Zig-zag nanotube; (b) Arm-chair nanotube; and (c) Chiral or helical nanotube**

Carbon nanotubes also have unique and structure-sensitive electrical characteristics. Depending on the orientation of the hexagonal units in the graphite plane, i.e., tube well, with the tube axis, the nanotube may behave electrically as either a metal or a semiconductor. Semiconducting and metallic nanotubes recently are put into use in field effect transistors (FET) and single electron transistors. The flat-panel and full-colour displays, i.e. TV and computer monitors, have been fabricated using carbon nanotubes as field emitters. These displays should be cheaper to produce and will have lower power requirements than CRT and liquid crystal displays. The mechanical strength of CNT is amazingly 600 times tougher than steel, which finds applications in micro electro mechanical systems (MEMS) and in aerospace. Progress has been made in using CNT's as field emission devices e.g. diodes, transistors, etc. for high resolution display systems. Construction of nanoprobes and electrodes in biological and other applications are also being aimed.

Nanowires

Fullerenes and CNT are hollow, so it should be possible to fill them up with different elements and compounds. Introducing a metal inside CNT might form a nanowire. Using electric arc discharge with a metal powder inside the anode produces CNT filled with metal carbides. A difficulty from the method mentioned above is that the tubes are filled partially; this might affect the conductivity properties. It is found that nanotubes could be filled more if we use small amounts of selenium, sulfur and germanium with the electric arc discharge method. Moreover, there are chemical techniques in which by nitric acid treatment, the tubes can be opened and then materials such as palladium, silver, gold, cobalt, iron, uranium and nickel oxides, molybdenum, tin, nyodinium, europium, and cadmium, can be inserted in their interior. It is important to know that enzymes and proteins have been put inside nanotubes. Capillarity effects have also been used to introduce elements inside CNT such as tin, lead, bismuth, cesium, sulfur and selenium. Nanowires have been used to prepare ultimate interconnects in integrated circuits.

Carbon Nanocones

Theoretical studies predicted the formation of graphitic cones. Subsequently, isolated graphitic cones were produced by carbon condensation on a graphite substrate and by pyrolysis of heavy oils. More recently, single walled aggregates of conical graphitic structures have been prepared by laser ablation of graphite targets. In addition, conical structures consisting of other layered materials such as BN have also been prepared by reacting boron oxide vapours with multi-walled CNTs.

Recently, it is reported that pyrolysis of palladium precursors, always produce conical nanofibers. An important feature of these new nanostructures is that they are held together by Vander Waals forces since the fibre is composed of an arrangement of stacked cones which can be opened (lamp-shade), or closed. It is believed than anocones may be good electron field emitters. Calculations of the electronic properties of nanocones reveal that there is a charge accumulation towards the tip and that there are localized states near the Fermi level. These features make them suitable as field emitters.

Haeckelites

The presence of defects such as pentagons and heptagons in fullerenes change the electronic properties. A new hypothetical type of graphene sheets has been proposed which admit pentagons, heptagons, and hexagons, observing that the number of heptagons and pentagons should be the same in order to compensate the negative curvature of the heptagons and the positive curvature of pentagons (Fig. 19.10). These arrangements have been named as haeckelites. They all exhibit a considerable signal at the Fermi level in the local density of states. Obviously, all these systems should possess a metallic behaviour. Thus, it is possible to roll up haeckelites to form nanotubes, which will be conductors independent of the diameter and the helical arrangement.

Another property of haeckelite tubes is that they keep the stiffness of classical graphite nanotubes composed with just hexagons; the Young's modulus of haeckelite tube is around 1.0 Tera Pa; of the same order as classical carbon nanotubes. Haeckelites also exhibit local rugosity due to the local bending produced by heptagons and pentagons. It might be possible, that electron irradiation experiments can generate heptagon-pentagon pairs (rugosity) in SWCNT.

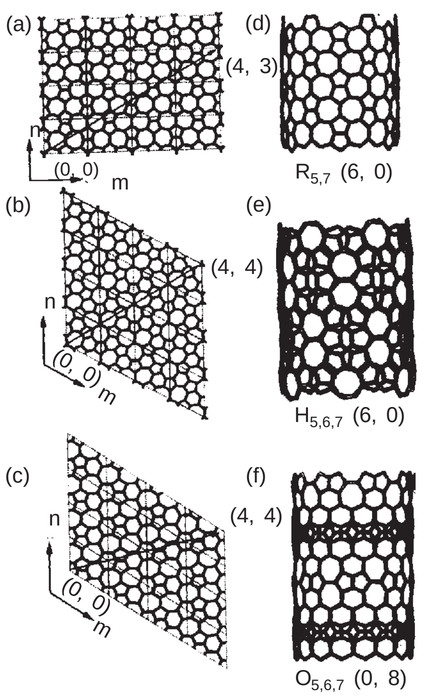


Fig. 19.10 Haeckelites (a) Rectangular haeckelite sheet (R); (b) Hexagonal haeckelite sheet (H); (c) Oblique haeckelite sheet (O); (d) to (f) are the corresponding haeckel tubes. All haeckelites are metallic (sheets and tubes)

2. PRODUCTION METHODS FOR CNTS

CNTs can be obtained by using four different methods: *(i)* arc-discharge *(ii)* pyrolysis *(iii)* laser vaporization and *(iv)* electrolysis.

The electric arc discharge set-up is shown in Fig. 19.11. This method is very similar to the one used for obtaining fullerenes with two main differences: first, the pressure are higher, around 500 torr (for fullerenes the pressure is around 100 torr), and second, the nanotubes are found deposited in the cathode and not in the soot. This method produces well ordered MWCNT with diameters from 2 nm upto 30 nm. The separation between graphite layers is around 3.4 Å, which is very close to that of graphite 3.35 Å and corresponds to turbostratric graphite. The length of these nanotubes can be upto 30 μ. The optimum conditions for generating nanotubes with this method consist of using currents

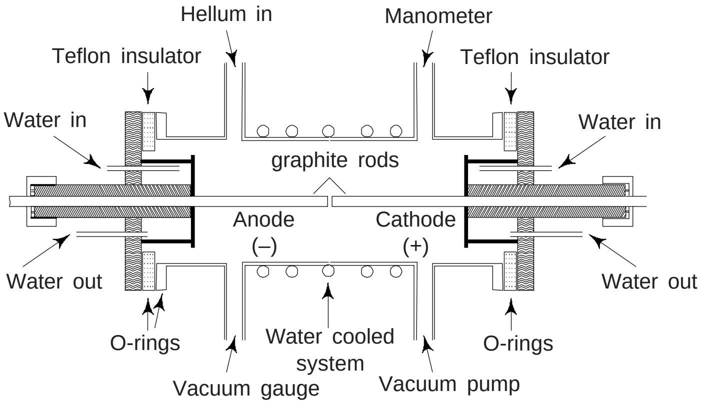


Fig. 19.11 Arc-discharge set-up

of 150 amps, voltages of 25 volts and graphite electrodes between 6 mm and 8 mm in diameter with a separation between them of 1 mm under helium atmosphere. The electric arc reaction is too violent, so it is very difficult to control the formation reaction.

The pyrolysis method (Fig. 19.12) consists of heating a hydrocarbon or a precursor which contains carbon under the presence of a catalyst such as nickel, cobalt or iron. Prior to the discovery of C_{60}, it was well known that carbon fibres could be obtained by pyrolytic methods using hydrocarbons. We must note that carbon fibres are not nanotubes since the fibres are bigger in diameter and exhibit a great amount of defects and impurities. Nevertheless, the application of carbon fibres is widespread, showing a strength comparable to steel.

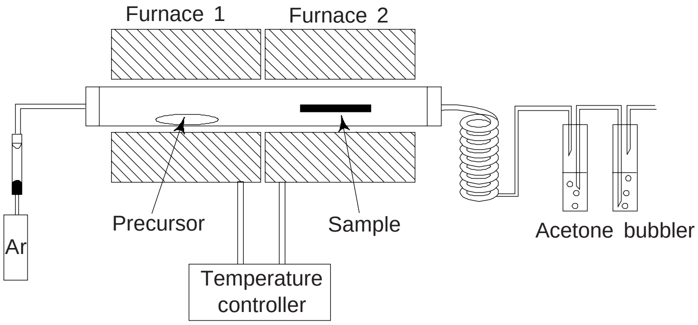


Fig. 19.12 Pyrolysis set-up

In 1994, Ajayan et al. found that carbon nanotubes could be aligned by embedding them in a polymer, thus forming a composite material. When the polymer was cut, they observed an aligned arrangement of nanotubes. In this context, a better alternative is to use laser ablation of cobalt thin films and then pyrolyse them with an organic precursor such as 2-amino 4, 6-dichloro-S-tryazine. The nanotubes obtained in this way are aligned in bundles which can grow upto 40 μ and diameters between 30 nm and 50 nm.

The laser vapourization method involves firing a high power laser towards a graphite target inside a furnace at 1200°C. The condensation of material generated by the laser is responsible for the nanotube formation. If we add nickel or cobalt to the graphite target, we obtain single-walled carbon nanotubes (SWCNT). The SWCNT show diameters around 14 Å forming a two-dimensional crystal with a lattice constant of 17Å. Arrangements of SWCNT can also be obtained by electric arc discharge with mixtures of carbon, nickel and ytrium.

The electrolysis method is based on using graphite electrodes inside a molten salt such as lithium chloride (700°C) under an argon atmosphere. Depending on conditions 20% to 30% consists of nanotubes. The depth of the cathode and the current (3-10 amps) play an important role in the formation of nanotubes.

Boron Nitride Curved Nanostructures (BNNTs)

Hexagonal boron nitride (h-BN) is another layered material with the structure very similar to graphite having hexagonal ring layers separated by 3.33 Å, in which every boron is connected to three nitrogen atoms and vice-versa; and B-N distance is 1.44 Å. Between the layers, every boron interacts with a nitrogen atom through a Vander Waal's force. Regarding the electronic properties, h-BN is an insulator with a band gap of 5.8 eV, whereas graphite is a semi-metal (band overlap = 0.04 eV). The first BN nanotube were prepared using the electric arc discharge with a high melting point metal (tantalum and tungsten for e.g.) electrodes and BN powder inside the anode. In 1996, Loiseau et al. reported the production of BN nanotubes by acting HfB_2 electrodes in an inert atmosphere. This technique lead to the formation of mainly single-and double-walled BN tubes.

It was reported that in BN nanotubes the tips were squared which means that instead of having pentagonal rings to close the structure there are squares of BN. This makes a sense because a pentagon will exhibit B-B or N-N bonds, which will not be very stable, changing also the proportions of B and N. Therefore, BN cages should possess squares. In terms of curvature, BN admits, with high stability, squares for the positive curvature and octagons for the negative curvature. From functional tight binding calculations, it is found that nitrogen excess cages with B and N are stable. Theoretical studies revealed that BNNTs are semiconductors with a band gap of 5.5 eV, smaller than in bulk BN. This property of BNNTs is independent of tube diameter, chirality, and number of walls. These BNNTs may also be extremely resistant to oxidation and should exhibit high Young's modulii. Therefore, the BNNTs may find important uses in ceramics.

BNNTs bundles can also be produced when carbon nanotubes are exposed to mixture of B_2O_3, V_2O_5, and M_0O_3 in a nitrogen atmosphere at elevated temperatures. There are alternative approaches to generate BNNTs and other BN nanostructures: *(a)* thermolysis of borazine ($B_3N_6H_6$) over nickel boride catalyst particles at 1000-1100°C, *(b)* thermal annealing of powdered rhombohedral boron and h-BN mixtures at 1200°C in lithium vapour, *(c)* plasma jets and *(d)* ball milling of h-BN, followed by thermal annealing at 1300°C.

At this stage, novel composites and nanoscale components need to be fabricated employing BNNTs, which appear to be extremely resistant to oxidation. Experimentally obtained values of Young's modulus of BNNTs (Y or $E = 1.22 \pm 0.24$ TPa) agrees well with the values obtained using theoretical approaches, and reveals that BN tubes are highly crystalline and probably, they may be the strongest insulating nanofibres.

Graphite and h-BN have opened new avenues for novel nanomaterials with different combinations of the type $B_xC_yN_z$. Pyrolysis experiments with melamine have been found very helpful in synthesizing C_xN_y nanofibres.

Boron doped carbon nanotubes have been synthesized by electric arc discharge with h-BN inside the graphite anode. During this experiment, it is observed that carbon nanotubes are highly graphitic and also very long (~ 20 μ). It was found that B was located at the tips of the nanotubes and that the zig-zag chirality was preferred. It is reported by ab-initio calculations that boron was acting as surfactant avoiding the zig-zag tubes closure. In addition, boron doped nanotubes exhibit conducting properties.

It is reported that nitrogen doped carbon nanotubes can also exhibit interesting features such as piridine like bonding in the graphitic mesh, which causes peak close to the Fermi level in the LDOS. All these nanotubes, arm chair and zig-zag are found to be conductors. Obviously, nitrogen is donating electrons to the nanotube system to produce conductivity.

3. KEY ISSUES IN NANOMANUFACTURING

Similar to large-scale fabrication, nanomanufacturing research issues revolve around precursor materials; fabrication processes and characterization techniques; instrumentation and equipment; theoretical modeling and control; and design and integration of structures into devices and systems.

4. MECHANICAL AND ELECTRONIC PROPERTIES OF CARBON NANOTUBES

Direct observation using high resolution transmission electron microscopy (HRTEM) reveals that CNTs are extraordinarily flexible. For example, they can be bent mechanically by an ultrasonic bath, by mechanical milling, and by using a polymeric resin. Theoretical calculations predict this flexibility. The first attempt to determine the stiffness of carbon nanotubes was when Treacy et al. used HRTEM to measure the amplitudes of the vibrating tubes, to calculate indirectly the Young's modulus. They found that MWCNTs exhibit Young's modulus of the order of 1.8 Tera Pa (1 Tera Pa = 1×10^{12} Pascals), which is higher than conventional carbon fibres (Ca 800 Giga Pa). Direct measurement using atomic force microscopy (AFM) reveal that the Young's modulus of MWCNTs is around 1.28 Tera Pa. This high Young's modulus makes the nanotubes a very strong structure. However, the challenge of building a super strong material out of nanotubes remains.

The first measurements of the electrical conductivity of CNTs were made using gold micro contacts attached to the CNTs by lithographic techniques finding that the resistance depends on the temperature and that in the range between 2 K and 300 K the tubes were semimetallic. Conductivity measurements on aligned MWCNTs show that the material behave as a nanoconductor. Also it has been confirmed that depending on the diameter and helicity, carbon nanotubes can be conductors or semiconductors. SWCNTs around 14 Å diameter exhibit quantum conductivity. We may note that the atomic arrangement, i.e., the geometry of the structure, determines the mechanical and electronic properties of CNTs. Now a days, nanocrystals have been prepared.

We may note that SWCNT can coalesce under high electron irradiation and high temperature conditions (around 800°C), Tight binding calculations suggest that vacancies, generated by knock on from the electrons, trigger the coalescence. Moreover, in order to have coalescence, the tubes should possess the same chirality, otherwise, tube polymerization might take place.

5. NANOSTRUCTURES IN MOTION

Micromechanics, micromachines and moving nanostructures represent a new "micro-revolution". However, the concepts related to micromechanics and micromachining of silicon are not new. A number of commercial products including accelerometers, pressure transducers, thermal print heads and ink jet nozzle arrays have been fabricated for internal industrial use and for the open market.

Since 1982, the field of micromechanics has expanded considerably, adding micrometers, movable linkages, gears, micro-turbines, electrical drives and microweezers to an ever growing suite of micromechanical devices.

6. NANOMATERIAL ADVANTAGE

Nanocrystalline materials are three dimensional solids composed of nanometersized grains, or crystallites. Because of their unique structure, which is characterized by ultrafine grains and a rather high density of crystal lattice defects, these materials have extra ordinary fundamental properties that could be exploited to make 'next-generation' *superstrong* materials.

Strength and ductility are the central mechanical properties of any material. They are determined by the physical nature of plastic deformation, which in conventional, coarse-grained metals is mainly carried by dislocations-line defects of regular crystal lattice-within individual grains. Study suggest, however, that the mechanism of plastic deformation in nanocrystalline materials may be different and so may lead to novel mechanical properties.

High strength, or hardness has already been observed in many nanomaterials. But in most cases such nanomaterials have very low ductility—they fail when their shape is changed. Some even become brittle when a force or deforming load is applied. Strength and ductility are usually opposing characteristics: the higher the strength, the lower the ductility, and vice versa. This correlation is associated with the nature of plasticity: the more difficult it is for dislocations to appear and to move, the stronger but brittle and less ductile is any crystalline material.

Very recently, Wang et al. [Nature 419, 912 (2002)] report their success in engineering an unusual nanocrystalline material that combines two useful properties that are often mutually exclusive-strength and ductility. They created a copper nanostructure by rolling the metal at temperature below 77 K and then heating it to around 450 K. The result was a 'bimodel' structure of micrometer-sized grains (at a volume fraction of around 25%) embedded in a matrix of nanocrystalline grains. The material showed extraordinary high ductility, but also retained its high strength. The reason for this behaviour seems to be that while the nanocrystalline grains provide strength, the embedded larger grains stabilize the tensile deformation of the material.

Processing nanomaterials to achieve both strength and ductility can enhance the material's resistance to fatigue. More than anything else, it is fatigue that at present limits the life-time, and hence the range of applications, of many advanced materials.

Conclusions

We have seen that different layered materials can acquire curvature to form several nanostructures with novel properties. Graphite is just one example of a planar structure that can form fullerenes, nanotubes (SWCNT or MWCNT), nanocones, etc. Boron nitride, tungsten and molybdenum disulphides can also be bent to form new nanostructures. Very recently Mickelson et al. [Nature 300, 467, (2003)] have created insulated C_{60} nanowire by packing C_{60} molecules into the interior of insulating boron nitride nanotubes (BNNTs). For small diameter BNNTs, the wire consists of a linear chain of C_{60} molecules. With increasing BNNT inner diameter, unusual C_{60} stacking configurations are obtained (including helical, hollow core, and incommensurate) that are unknown for bulk or thin film forms of C_{60}. C_{60} in BNNTs presents a model system for studying the properties of dimensionally constrained 'silo' crystal structures. For the linear-chain case, they have fused the C_{60} molecules to form a single walled carbon nanotube inside the insulating BNNT.

Computer calculations based and experimental evidence indicate that curvature, introducted through different kinds of 'defects', produces stable atomic arrangements. These 'defects', which can be pentagons, heptagons, rings with more atoms, or more complex architectures, change the electronic, magnetic, mechanical properties of the structure. We have seen that doping of curved nanostructures is a promising area of research. Nanotechnologists are seeking to build nanometer-scale electronic devices in which the functional unit is a single molecule or atom. Carbon nanotubes have proved particularly useful as molecular 'wires' in this quest, thanks to their long length of several micrometers. So it could be possible to wire up a short molecule, or even a single atom, to create a nanoscale transistor.

Carbon nanotubes or tips coated with carbon nanotubes are found to be good field emitters. They can operate at lower electric field giving larger currents. They are highly stable too. Obviously, they may be used in future as 'tips' in flat panel display technology, electron microscopes, scanning electron microscope (STM) or atomic force microscope (AFM). It is expected that the silicon technology may be replaced in future by carbon nanotechnology, i.e., by carbon nonotube based devices. Carbon nanotubes are also thought to be used as 'hydrogen storage' materials. Carbon nanotubes can be useful in weighting nanoparticles. From an applications point of view, the emphasis is on such methods which can produce high yields of nanotubes at low cost on a commercial scale.

SUGGESTED READINGS

1. M.S. Dresselhaus, et al. Science of Fullerenes and Carbon Nanotubes (San Diego, CA: Academic 1996).
2. P. Moriarty, Nanostructured Materials, Rep. Prog. Phys. 64, 297 (2001).

REVIEW QUESTIONS

1. What are nanostructured materials?
2. Explain with crystal structures the salient features of *(a)* graphite *(b)* fullerenes.
3. What are carbon nanotubes? Mention their few important technological applications.
4. Write a note on nanowires and nanocones.
5. What are the different methods for the production of CNTs. Briefly mention them.
6. What are the key issues in nanomanufacturing?
7. Mention some important mechanical and electronic properties of CNTs.

Appendix 1

Units, Conversion Factors, Physical Constants

UNITS

	Quantity	Symbol	*S.I. Units*		Other units
			Expressed as	*Full Form*	
1.	Activation energy	*Q*	J/mol	Joule per mole	cal/mole, cal/gm
2.	Atomic diameter		nm	Nanometre	Å
3.	Bond energy	—	KJ/mole	Kilo joule/mole	eV/atom; Kcal/mole
4.	Bond length	—	nm	Nanometre	Å
5.	Bragg angle	θ	rad	Radian	0 (degree)
6.	Calorific value	—	KJ/m^3	Kilo joule per cubic metre	
7.	Coercive field	H_c	A/m	Ampere per metre	Oersted
8.	Concentration	*C*	mol/m^3	Mole per cubic metre	No. of atoms/cm^3
9.	Contact angle	Φ	rad	Radian	0 (Degree)
10.	Current density	J	A/m^2	Ampere per square metre	amp/cm^2
11.	Dipole moment	—	C–m	Coulomb metre	debye
12.	Dislocation energy	*E*	J/m	Joule per metre	erg/cm; eV/plane
13.	Diffusion co-efficient	*D*	m^2/s	Metre square per second	—
14.	Diffusion constant	D_0	m^2/s	Metre square per second	—
15.	Dielectric field strength	*E*	V/m, N/c	Volt per metre	Newton per Coulomb
16.	Electrical conductivity	σ	Ohm^{-1}m^{-1}	Per Ohm per metre	mho/cm
17.	Electron affinity	—	KJ/mol	Kilojoule per mole	eV/atom, kcal/ mole
18.	Electron energy level		J	Joule	eV (electron volt), erg.
19.	Energy of surface imperfections	γ	J/m^2	Joule per square metre	erg/cm^2
20.	Enthalpy	*H*	J/mol	Joule per mole	cal/mol, cal/gm
21.	Entropy	*S*	J/mol/K	Joule per mole/Kelvin	cal/mole/°C cal/gm/°C
22.	External energy	—	J/mol	Joule per mole	cal/mole, cal/gm
23.	Flux	*J*	mol/m^2/s	Mole per square metre per second	No. of atoms/cm^2/sec
24.	Force	*N*	Newton	lbf.	
25.	Fracture stress	r_f	MN/m^2	Mega Newton per square metre	kgf/mm^2, PSi

(*Contd.*)

(*Contd.*)

	Quantity	Symbol	S.I. Units		Other units
			Expressed as	Full Form	
26.	Fracture toughness (Strain energy release rate)	Gc	J/m^2	Joules per square metre	lb-in/in^2
	—Stress intensity factor	K_c	MN/m$^{3/2}$	Mega newton per (metre)$^{3/2}$	KSi in$^{1/2}$
27.	Frequency of radiation	v	S^{-1}, Hz	per second; Hertz	
28.	Gibbs free energy	G	J/mol	Joule per mole	cal/mole, cal/gm
39.	Hysteresis loss		J/m^3	Joule per cubic metre	—
	Hall Petch constant	K	MNm^{-32}	Mega Newton/(metre)^{32}	—Kgf/m$^{3/2}$
30.	Interatomic force	F	N	Newton	Kgf, lb, dyne
31.	Interfacial energy	Y	J/m^2	Joule per square metre	erg/cm^2
32.	Internal energy	E	J/mol	Joule per mole	cal/mole, cal/gm
33.	Interplanar spacing	d	nm	Nanometre	Å
34.	Kinetic energy	E	J	Joule	erg, eV
35.	Lattice parameter	a	nm	Nanometre	Å
36.	Magnetic induction	B	W$_b$/m^2 or T	Weber per square metre of Tesla	Gauss
37.	Magnetic moment	m	Am^2	Ampere square metre	
38.	Magnetic permeability	μ	H/m	Henry per metre	
39.	Magnetic field strength	H	A.T/m	Ampere turns per metre	Oersted
40.	Mass	—	Kg	Kilogram	lb
41.	Momentum	—	Kg m/s	Kilogram metre per second	—
42.	Power		W	Watt	ft.lbf.
43.	Pressure	P	MP$_a$, MN/m^2	mega pascal; mega Newton/ square metre	atmosphere, psi Kg/cm^2, dyne/ cm^2, mm of Hg
44.	Resistivity	ρ	Ohm.m	Ohm metre	ohm.cm
45.	Shear modulus	u	GN/m^2	Giga Newton per square metre	Kgf/mm^2; PSi, dyne/cm^2
46.	Shear stress	r	N/m^2 P$_a$	Newton/m^2 pascal	Kgf/mm^2, PSi, dyne/cm^2
	Shear strain	γ	—	—	—
47.	Specific heat	C_v	J/mol/k	Joule per mole per Kelvin	cal/mol/°C; cal/gm/°c
48.	Specific heat	C_p	J/mol/k	Joule per mole per Kelvin	cal/gm/°C
49.	Specific volume	—	m^3/Kg	Cubic metre per kilogram	ft^3/lb, cm^3/gm
50.	Specific energy	—	J/Kg	Joule per kilogram	Btu/lb
51.	Strain energy	ε	J/m^3	Joule per cubic metre	erg/cm^3
52.	Surface energy	Y	J/m^2	Joule per square metre	erg/cm^2
53.	Surface tension	γ	N/M	Newton per metre	dyne/cm
54.	Temperature	T	K	Kelvin	°C, °F
55.	Temperature co-efficient of resistance	α_0	K^{-1}	per Kelvin	per °F
56.	Tensile strength	—	MN/m^2	Mega Newton per square metre	Kgf/mm^2, Psi, dyne/cm^2
57.	Tensile stress	σ	MN/m^2	Mega Newton per square metre	Psi, Kgf/mm^2
58.	Thermal energy	RT	J/mol	Joule per mole	cal/mole, cal/gm
59.	Thermal conductivity	—	W/m/K	Watt per metre per deg Kelvin	Kcal/m/s/°C
60.	Ultimate tensile strength	UTS	MN/m^2	Mega Newton per square metre	PSi, Kgf/mm^2

(*Contd.*)

(*Contd.*)

	Quantity	*Symbol*	*S.I. Units* *Expressed as*	*Full Form*	*Other units*
61.	Vibration frequency	ν or S^{-1}	S^{-1}	Per second	—
62.	Viscosity	η	Pas; N/m^{2}s	Pascal second Newton per square metre second	Poise
63.	Wavelength of radiation	λ	nm	nanometre	A
64.	Young's modulus	E	GPa,	Giga pascal;	PSi, Kgf/mm^2
65.	Yield strength	σ_y	N/m^2	Newton/m^2	Kgf/mm^2

Appendix 2

Conversion Factors

1. 1° = 0.01745 rad
2. 1 Å =10^{-10} m = 0.1 nm
3. 1 A.hr = 3.6 Kc
4. 1 ampere = 1 C/s
5. 1 atm = 0.101325 × 10^{6} N/m^{2} = 760 mm Hg = 10^{5} Pascal
6. 1 Btu = 1.05506 kJ
7. 1 Btu/lb = 2326 J/Kg
8. 1 Btu/ft^{3} = 37.2589 KJ/m^{3}
9. 1 bar = 10^{-1} MPa
10. 1 Calorie = 4.18 J
11. 1 Coulomb = 1 A-s
12. 1 Debye = 3.33 × 10^{-30} cm
13. 1 Dyne/Cm = 10^{-3} N/m
14. 1 dyne/cm^{2} = 0.1 N/m^{2}
15. 1 erg = 10^{-7} J
16. 1 erg/cm = 10^{-5} J/m
17. 1 eV = 1.602 × 10^{-19} J
18. 1 eV/entity = 96.49 KJ/mol (of entities)
19. 1 eV/particle = 96.49 KJ/mol
20. 1 farad = 1 C/V
21. 1 Gauss = 10^{-4} Wb/m^{2}
22. 1 gram-calorie = 4.185 J
23. 1 Henry = 1 V-s/A
24. 1 Hertz = 1 cycle/s
25. 1 HP = 0.7457 KW
26. 1 Kgf/mm^{2} = 9.81 MN/m^{2}
27. 1 KSi/in = 1.1 MN/m$^{3/2}$
28. 1 lb/Cu.ft = 16.01 kg/m^{3}
29. 1 lumen = 0.0016 W (at 0.55 μm)
30. 1 Newton = 1 kg.m/s^{2}
31. 1 Oersted = 79.6 A/m
32. 1 Poise = 0.1 PaS
33. 1 PSi = 6.89 KN/m^{2} 10^{5}
34. T°C = (T + 273.15) K
35. T°F = 5/9 (T + 459.67) K
36. 1 torr (mm of Hg) = 133.3 N/m^{2}

37. 1 TR = 3024 KCal
38. 1 Watt = 1 Joule/sec
39. TR = 3024 Kcal/hg
40. 1 Joule = 10^7 ergs
41. 1 K-cal = 4.18 kJ

Appendix 3

Physical Constants

	Constant	*Symbol*	*Value*
1.	Acceleration due to gravity	g	= 9.81 m/s^2
2.	Atomic mass unit	amu	= 1.660×10^{-24} g
3.	Avogadro's number	N	= 6.023×10^{-23}/g mol
4.	Bohr magneton (magnetic moment)	β	= 9.273×10^{-24} A m^2
5.	Boltzmann's constant	k	= 1.380×10^{-23} J/K = 8.620×10^{-5} eV/k
6.	Electron rest mass	m_e	= 9.109×10^{-31} Kg
7.	Electronic charge	e	= 1.602×10^{-19} C
8.	Faraday's constant	F	= 9.649×10^{4} C/mol
9.	Gas constant	R	= 8.314 J/mol/K
10.	Mass of proton	m_p	= 1.673×10^{-24} g
11.	Mass of electron	m_e	= 9.108×10^{-28} g
12.	Planck's constant	h	= 6.626×10^{-34} Js
13.	Permeability of free space	μ_0	= 1.257×10^{-6} H/m
14.	Velocity of light in free space	c	= 2.998×10^{8} m/s
15.	Permittivity of free space	ε_0	= 8.854×10^{-12} F/m
16.	Volume of 1 kg mole of ideal gas at N.T.P.	V	= 22.41 m^3
17.	Magnetic constant	μ_i	= 1.2566×10^{-8} H/cm
18.	Gravitation constants	G	= 6.670×10^{-1}
19.	Radius of electron	r_e	= 2.81777×10^{-15} m.

Appendix 4

Prefix Names, symbols and Multiplication factors

	Name	*Symbol*	*Factor by which unit is multiplied*
1.	1 Deci	d	$= 10^{-1}$
2.	1 exa	E	$= 10^{18}$
3.	1 Giga	G	$= 10^{9}$
4.	1 Kilo	K	$= 10^{3}$
5.	1 Mega	M	$= 10^{6}$
6.	1 micro	μ	$= 10^{-6}$
7.	1 milli	mm	$= 10^{-3}$
8.	1 micron	mie	= 0.001 in = 25 μm
9.	1 Nano m.	nm.	$= 10^{-9}$ m
10.	1 peta	P	$= 10^{15}$
11.	1 tera	T	$= 10^{12}$

Subject Index