Chapter 1. Introduction

1.1 Historical Perspective and Materials Science

1.1.1 Historical Perspective

Materials are so important in the development of human civilization that the historians have identified early periods of civilization by the name of most significantly used material, e.g.: Stone Age, Bronze Age. This is just an observation made to showcase the importance of materials and their impact on human civilization. It is obvious that materials have affected and controlling a broad range of human activities through thousands of decades.

From the historical point of view, it can be said that human civilization started with *Stone Age* where people used only natural materials, like stone, clay, skin, and wood for the purposes like to make weapons, instruments, shelter, etc. Thus the sites of deposits for better quality stones became early colonies of human civilization. However, the increasing need for better quality tools brought forth exploration that led to *Bronze Age*, followed by *Iron Age*. When people found copper and how to make it harder by alloying, the *Bronze Age* started about 3000 BC. The use of iron and steel, a stronger material that gave advantage in wars started at about 1200 BC. Iron was abundant and thus availability is not limited to the affluent. This commonness of the material affected every person in many aspects, gaining the name *democratic material*. The next big step in human civilization was the discovery of a cheap process to make steel around 1850 AD, which enabled the railroads and the building of the modern infrastructure of the industrial world. One of the most significant features of the *democratic material* is that number of users just exploded. Thus there has been a need for human and material resources for centuries, which still going strong. It’s being said and agreed that we are presently in *Space Age* marked by many technological developments towards development materials resulting in stronger and light materials like composites, electronic materials like semiconductors, materials for space voyage like high temperature ceramics, biomaterials, etc.

In summary, materials constitute foundation of technology. The history of human civilization evolved from the Stone Age to the Bronze Age, the Iron Age, the Steel Age, and to the Space Age (contemporaneous with the Electronic Age). Each age is marked by the advent of certain materials. The Iron Age brought tools and utensils. The Steel Age brought railroads, instruments, and the Industrial Revolution. The Space Age brought the materials for stronger and light structures (e.g., composite materials). The Electronic Age brought semiconductors, and thus many varieties of electronic gadgets.

1.1.2 Materials Science

As engineering materials constitute foundation of technology, it’s not only necessary but a must to understand how materials behave like they do and why they differ in properties. This is only possible with the atomistic understanding allowed by quantum mechanics that first explained atoms and then solids starting in the 1930s. The combination of physics, chemistry, and the focus on the relationship between the properties of a material and its microstructure is the domain of Materials Science. The development of this science allowed designing materials and provided a knowledge base for the engineering applications (Materials Engineering).

Important components of the subject Materials Science are structure, properties, processing, and performance. A schematic interrelation between these four components is shown in figure 1.1.

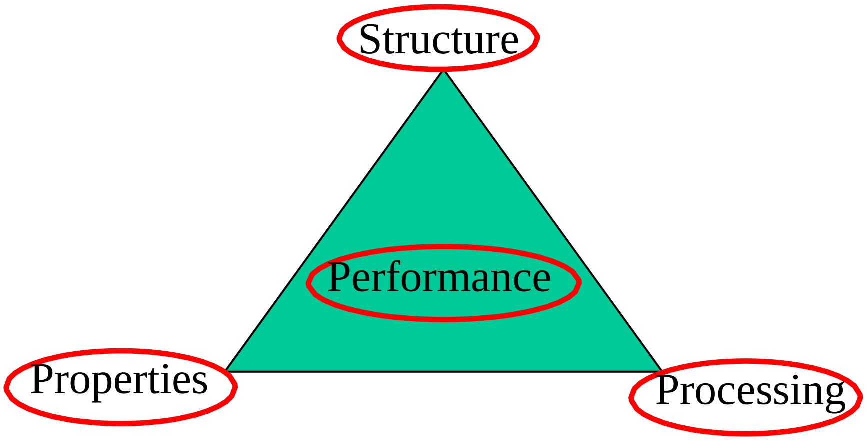


Figure 1.1: *Interrelation between four components of Materials Science.*

1.2 Why Study Materials Science and Engineering? and Classification of Materials?

1.2.1 Why Study Materials Science and Engineering?

All engineers need to know about materials. Even the most "immaterial", like software or system engineering depend on the development of new materials, which in turn alter the economics, like software-hardware trade-offs. Increasing applications of system engineering are in materials manufacturing (industrial engineering) and complex environmental systems.

Innovation in engineering often means the clever use of a new material for a specific application. For example: plastic containers in place of age-old metallic containers. It is well learnt lesion that engineering disasters are frequently caused by the misuse of materials. So it is vital that the professional engineer should know how to select materials which best fit the demands of the design - economic and aesthetic demands, as well as demands of strength and durability. Beforehand the designer must understand the properties of materials, and their limitations. Thus it is very important that every engineer must study and understand the concepts of Materials Science and Engineering. This enables the engineer

- To select a material for a given use based on considerations of cost and performance.
- To understand the limits of materials and the change of their properties with use.
- To be able to create a new material that will have some desirable properties.
- To be able to use the material for different application.

1.2.2 Classification of Materials

Like many other things, materials are classified in groups, so that our brain can handle the complexity. One can classify them based on many criteria, for example crystal structure (arrangement of atoms and bonds between them), or properties, or use. Metals, Ceramics, Polymers, Composites, Semiconductors, and Biomaterials constitute the main classes of present engineering materials.

Metals: These materials are characterized by high thermal and electrical conductivity; strong yet deformable under applied mechanical loads; opaque to light (shiny if polished). These characteristics are due to valence electrons that are detached from atoms, and spread in an *electron sea* that *glues* the ions together, i.e. atoms are bound together by metallic bonds and weaker van der Waalls forces. Pure metals are not good enough for many applications, especially structural applications. Thus metals are used in alloy form i.e. a metal mixed with another metal to improve the desired qualities. E.g.: aluminum, steel, brass, gold.

Ceramics: These are inorganic compounds, and usually made either of oxides, carbides, nitrides, or silicates of metals. Ceramics are typically partly crystalline and partly amorphous. Atoms (ions often) in ceramic materials behave mostly like either positive or negative ions, and are bound by very strong Coulomb forces between them. These materials are characterized by very high strength under compression, low ductility; usually insulators to heat and electricity. Examples: glass, porcelain, many minerals.

Polymers: Polymers in the form of thermo-plastics (nylon, polyethylene, polyvinyl chloride, rubber, etc.) consist of molecules that have covalent bonding within each molecule and van der Waals forces between them. Polymers in the form of thermo-sets (e.g., epoxy, phenolics, etc.) consist of a network of covalent bonds. They are based on H, C and other non-metallic elements. Polymers are amorphous, except for a minority of thermoplastics. Due to the kind of bonding, polymers are typically electrical and thermal insulators. However, conducting polymers can be obtained by doping, and conducting polymer-matrix composites can be obtained by the use of conducting fillers. They decompose at moderate temperatures (100 – 400 C), and are lightweight. Other properties vary greatly.

Composite materials: Composite materials are multiphase materials obtained by artificial combination of different materials to attain properties that the individual components cannot attain. An example is a lightweight brake disc obtained by embedding SiC particles in Al-alloy matrix. Another example is reinforced cement concrete, a structural composite obtained by combining cement (the matrix, i.e., the binder, obtained by a reaction known as hydration, between cement and water), sand (fine aggregate), gravel (coarse aggregate), and, thick steel fibers. However, there are some natural composites available in nature, for example – wood. In general, composites are classified according to their matrix materials. The main classes of composites are metal-matrix, polymer-matrix, and ceramic-matrix.

Semiconductors: Semiconductors are covalent in nature. Their atomic structure is characterized by the highest occupied energy band (the valence band, where the valence electrons reside energetically) full such that the energy gap between the top of the valence band and the bottom of the empty energy band (the conduction band) is small enough for some fraction of the valence electrons to be excited from the valence band to the conduction band by thermal, optical, or other forms of energy. Their electrical properties depend extremely strongly on minute proportions of contaminants. They are usually doped in order to enhance electrical conductivity. They are used in the form of single crystals without dislocations because grain boundaries and dislocations would degrade electrical behavior. They are opaque to visible light but transparent to the infrared. Examples: silicon (Si), germanium (Ge), and gallium arsenide (GaAs, a compound semiconductor).

Biomaterials: These are any type material that can be used for replacement of damaged or diseased human body parts. Primary requirement of these materials is that they must be biocompatible with body tissues, and must not produce toxic substances. Other important material factors are: ability to support forces; low friction, wear, density, and cost; reproducibility. Typical applications involve heart valves, hip joints, dental implants, intraocular lenses. Examples: Stainless steel, Co-28Cr-6Mo, Ti-6Al-4V, ultra high molecular weight poly-ethelene, high purity dense Al-oxide, etc.

1.3 Advanced Materials, Future Materials, and Modern Materials needs

1.3.1 Advanced Materials

These are materials used in *High-Tech* devices those operate based on relatively intricate and sophisticated principles (e.g. computers, air/space-crafts, electronic gadgets, etc.). These materials are either traditional materials with enhanced properties or newly developed materials with high-performance capabilities. Hence these are relatively expensive. Typical applications: integrated circuits, lasers, LCDs, fiber optics, thermal protection for space shuttle, etc. Examples: Metallic foams, inter-metallic compounds, multi-component alloys, magnetic alloys, special ceramics and high temperature materials, etc.

1.3.2 Future Materials

Group of new and state-of-the-art materials now being developed, and expected to have significant influence on present-day technologies, especially in the fields of medicine, manufacturing and defense. Smart/Intelligent material system consists some type of sensor (*detects an input*) and an actuator (*performs responsive and adaptive function*). Actuators may be called upon to change shape, position, natural frequency, mechanical characteristics in response to changes in temperature, electric/magnetic fields, moisture, pH, etc.

Four types of materials used as actuators: Shape memory alloys, Piezo-electric ceramics, Magnetostrictive materials, Electro-/Magneto-rheological fluids. Materials / Devices used as sensors: Optical fibers, Piezo-electric materials, Micro-electro-mechanical systems (MEMS), etc.

Typical applications: By incorporating sensors, actuators and chip processors into system, researchers are able to stimulate biological human-like behavior; Fibers for bridges, buildings, and wood utility poles; They also help in fast moving and accurate robot parts, high speed helicopter rotor blades; Actuators that control chatter in precision machine tools; Small microelectronic circuits in machines ranging from computers to photolithography prints; Health monitoring detecting the success or failure of a product.

1.3.3 Modern Materials needs

Though there has been tremendous progress over the decades in the field of materials science and engineering, innovation of new technologies, and need for better performances of existing technologies demands much more from the materials field. More over it is evident that new materials/technologies are needed to be environmental friendly. Some typical needs, thus, of modern materials needs are listed in the following:

- Engine efficiency increases at high temperatures: requires high temperature structural materials

- Use of nuclear energy requires solving problem with residues, or advances in nuclear waste processing.
- Hypersonic flight requires materials that are light, strong and resist high temperatures.
- Optical communications require optical fibers that absorb light negligibly.
- Civil construction – materials for unbreakable windows.
- Structures: materials that are strong like metals and resist corrosion like plastics.

References

1. M. F. Ashby and D. R. H. Jones, Engineering Materials 1, An introduction to Their Properties and Applications, second edition, Butterworth-Heinemann, Woburn, UK, 1996
2. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
3. V. Raghavan, Materials Science and Engineering, third edition, Prentice Hall of India Private Limited, New Delhi, 1990.

Chapter 2. Atomic Structure, Interatomic Bonding and Structure of Crystalline Solids

2.1 Atomic Structure and Atomic Bonding in Solids

2.1.1 Atomic Structure

Atoms are composed of electrons, protons, and neutrons. Electrons and protons are negative and positive charged particles respectively. The magnitude of each charged particle in an atom is 1.6×10^{-19} Coulombs.

The mass of the electron is negligible with respect to those of the proton and the neutron, which form the *nucleus* of the atom. The unit of mass is an atomic mass unit (*amu*) = 1.66×10^{-27} kg, and equals 1/12 the mass of a carbon atom. The Carbon nucleus has Z=6, and A=6, where Z is the number of protons, and A the number of neutrons. Neutrons and protons have very similar masses, roughly equal to 1 *amu* each. A neutral atom has the same number of electrons and protons, Z.

A *mol* is the amount of matter that has a mass in grams equal to the atomic mass in *amu* of the atoms. Thus, a mole of carbon has a mass of 12 *grams*. The number of atoms in a mole is called the Avogadro number, $N_{av} = 6.023 \times 10^{23}$. Note that N_{av} = 1 *gram*/1 *amu*.

Calculating n, the number of atoms per cm^3 of a material of density δ (g/cm^3):

$$n = N_{av} \frac{\delta}{M}$$

where M is the atomic mass in *amu* (*grams* per *mol*). Thus, for graphite (carbon) with a density $\delta = 1.8$ g/cm^3, M =12, we get 6×10^{23} atoms/mol × 1.8 g/cm^3 / 12 g/mol) = 9×10^{22} C atoms/cm^3.

For a molecular solid like ice, one uses the molecular mass, $M_{(H2O)} = 18$. With a density of 1 g/cm^3, one obtains n = 3.3×10^{22} H_2O molecules/cm^3. Note that since the water molecule contains 3 atoms, this is equivalent to 9.9×10^{22} atoms/cm^3.

Most solids have atomic densities around 6×10^{22} atoms/cm^3. The cube root of that number gives the number of atoms per centimeter, about 39 million. The mean distance between atoms is the inverse of that, *or* 0.25 nm. This is an important number that gives the scale of atomic structures in solids.

2.1.2 Atomic bonding in solids

In order to understand the why materials behave like they do and why they differ in properties, it is necessary that one should look at atomic level. The study primarily concentrates on two issues: what made the atoms to cluster together, and how atoms are arranged. As mentioned in earlier chapter, atoms are bound to each other by number of bonds. These inter-atomic bonds are primarily of two kinds: Primary bonds and Secondary bonds. Ionic, Covalent and Metallic bonds are relatively very strong, and grouped as primary bonds, whereas van der Waals and hydrogen bonds are relatively weak, and termed as secondary bonds. Metals and Ceramics are entirely held together by primary bonds - the ionic and covalent bonds in ceramics, and the metallic and covalent bonds in metals. Although much weaker than primary bonds, secondary bonds are still very important. They provide the links between polymer molecules in polyethylene (and other polymers) which make them solids. Without them, water would boil at -80˚C, and life as we know it on earth would not exist.

Ionic Bonding: This bond exists between two atoms when one of the atoms is negative (has an extra electron) and another is positive (has lost an electron). Then there is a strong, direct Coulomb attraction. Basically ionic bonds are non-directional in nature. An example is NaCl. In the molecule, there are more electrons around Cl, forming Cl^- and fewer electrons around Na, forming Na^+. Ionic bonds are the strongest bonds. In real solids, ionic bonding is usually exists along with covalent bonding.

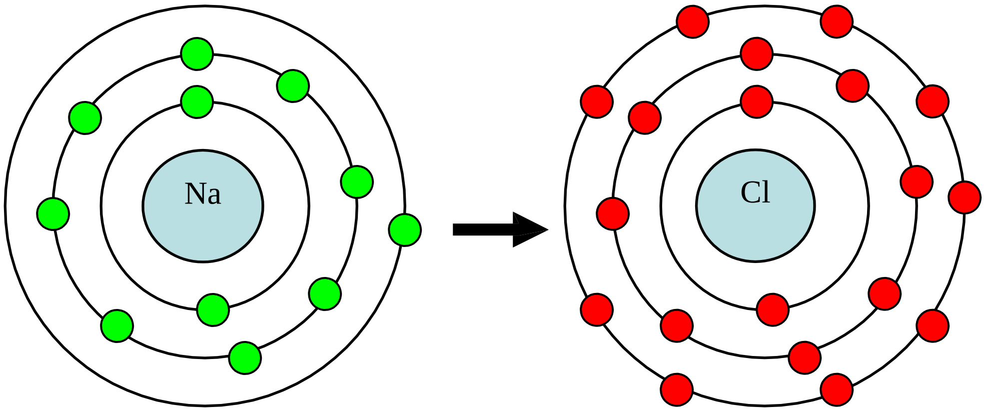


Fig.1 Schematic representation of ioning bonding. Here, Na is giving an electron to Cl to have stable structure

Covalent Bonding: In covalent bonding, electrons are shared between the atoms, to saturate the valency. The simplest example is the H_2 molecule, where the electrons spend more time in between the nuclei of two atoms than outside, thus producing bonding. Covalent bonds are stereo-specific i.e. each bond is between a specific pair of atoms, which share a pair of electrons (of opposite magnetic spins). Typically, covalent bonds are very strong, and directional in nature. The hardness of diamond is a result of the fact that each carbon atom is covalently bonded with four neighboring atoms, and each neighbor is bonded with an equal number of atoms to form a rigid three-dimensional structure.

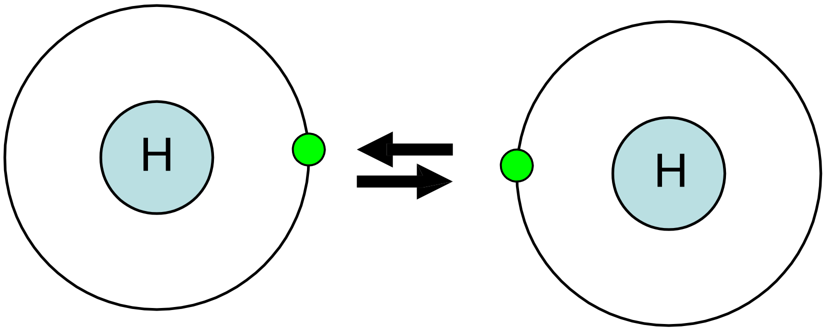


Figure 2. Schematic representation of covalent bond in Hydrogen molecule (sharing of electrons)

Metallic Bonding: Metals are characterized by high thermal and electrical conductivities. Thus, neither covalent nor ionic bondings are realized because both types of bonding localize the valence electrons and preclude conduction. However, strong bonding does occur in metals. The valence electrons of metals also are delocalized. Thus metallic bonding can be viewed as metal containing a periodic structure of positive ions surrounded by a sea of delocalized electrons. The attraction between the two provides the bond, which is non-directional.

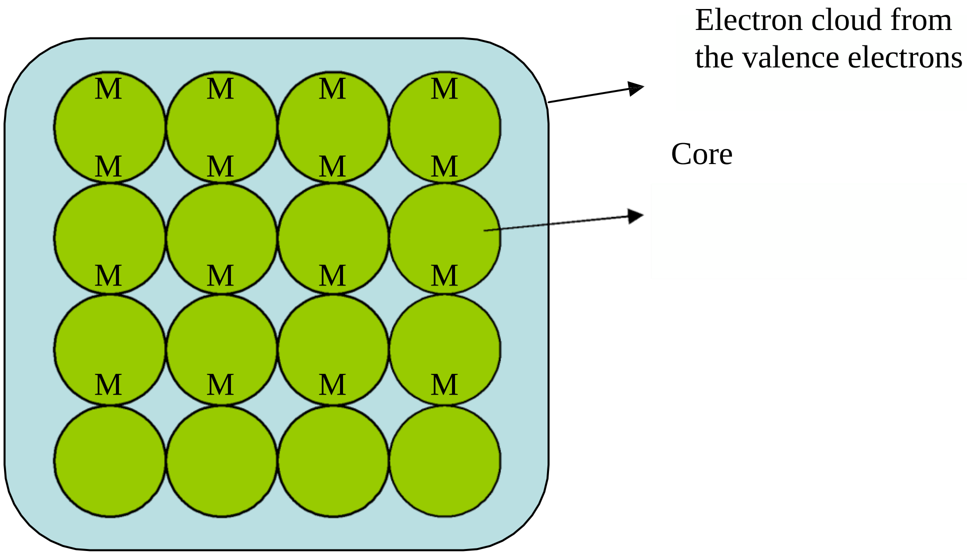


Figure 3. Metallic bonding

Fluctuating Induced Dipole Bonds: Since the electrons may be on one side of the atom or the other, a dipole is formed: the + nucleus at the center, and the electron outside. Since the electron moves, the dipole fluctuates. This fluctuation in atom A produces a fluctuating electric field that is felt by the electrons of an adjacent atom, B. Atom B then polarizes so that its outer electrons are on the side of the atom closest to the + side (or opposite to the – side) of the dipole in A.

Polar Molecule-Induced Dipole Bonds: Another type of secondary bond exists with asymmetric molecules, also called polar molecules because of positively and negatively charged regions. A permanent dipole moment arises from net positive and negative charges that are respectively associated with the hydrogen and chlorine ends of the HCl molecule, leading to bonding. The magnitude of this bond will be greater than for fluctuating induced dipoles.

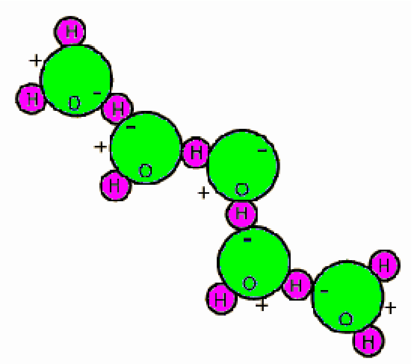


Figure 4. Dipole bond in water

These two kinds of bonds are also called van der Waals bonds. Third type of secondary bond is the hydrogen bond. It is categorized separately because it produces the strongest forces of attraction in this category.

Permanent Dipole Bonds / Hydrogen bonding: It occurs between molecules as covalently bonded hydrogen atoms – for example C-H, O-H, F-H – share single electron with other atom essentially resulting in positively charged proton that is not shielded any electrons. This highly positively charged end of the molecule is capable of strong attractive force with the negative end of an adjacent molecule. The properties of water are influenced significantly by the hydrogen bonds/bridges. The bridges are of sufficient strength, and as a consequence water has the highest melting point of any molecule of its size. Likewise, its heat of vaporization is very high.

2.2 Crystal Structures, Crystalline and Non-Crystalline materials

2.2.1 Crystal structures

All metals, a major fraction of ceramics, and certain polymers acquire crystalline form when solidify, i.e. in solid state atoms self-organize to form *crystals*. Crystals possess a long-range order of atomic arrangement through repeated periodicity at regular intervals in three dimensions of space. When the solid is not crystalline, it is called amorphous. Examples of crystalline solids are metals, diamond and other precious stones, ice, graphite. Examples of amorphous solids are glass, amorphous carbon (a-C), amorphous Si, most plastics.

There is very large number of different crystal structures all having long-range atomic order; these vary from relatively simple structures for metals to exceedingly complex structures for ceramics and some polymers. To discuss crystalline structures it is useful to consider atoms as being hard spheres, with well-defined radii. In this scheme, the shortest distance between two like atoms is one diameter. In this context, use of terms *lattice* and *unit cell* will be handy. *Lattice* is used to represent a three-dimensional periodic array of points coinciding with atom positions. *Unit cell* is smallest repeatable entity that can be used to completely represent a crystal structure. Thus it can be considered that a unit cell is the building block of the crystal structure and defines the crystal structure by virtue of its geometry and the atom positions within.

Important properties of the unit cells are

- The type of atoms and their radii R.
- Cell dimensions (Lattice spacing a, b and c) in terms of R and
- Angle between the axis α, β, γ
- a*, b*, c* - lattice distances in reciprocal lattice , α^*, β^*, γ^* - angle in reciprocal lattice
- n, number of atoms per unit cell. For an atom that is shared with m adjacent unit cells, we only count a fraction of the atom, $1/m$.
- *CN*, the coordination number, which is the number of closest neighbors to which an atom is bonded.
- *APF*, the atomic packing factor, which is the fraction of the volume of the cell actually occupied by the hard spheres. *APF* = Sum of atomic volumes/Volume of cell.

Some very common crystal structures and relevant properties are listed in table 2.1.

Table 2.1: Common crystal structures and their properties.

Unit Cell	n	*CN*	a/R	*APF*
Simple Cubic	1	6	$4/\sqrt{4}$	0.52

Body-Centered Cubic	2	8	4/√ 3	0.68
Face-Centered Cubic	4	12	4/√ 2	0.74
Hexagonal Close Packed	6	12		0.74

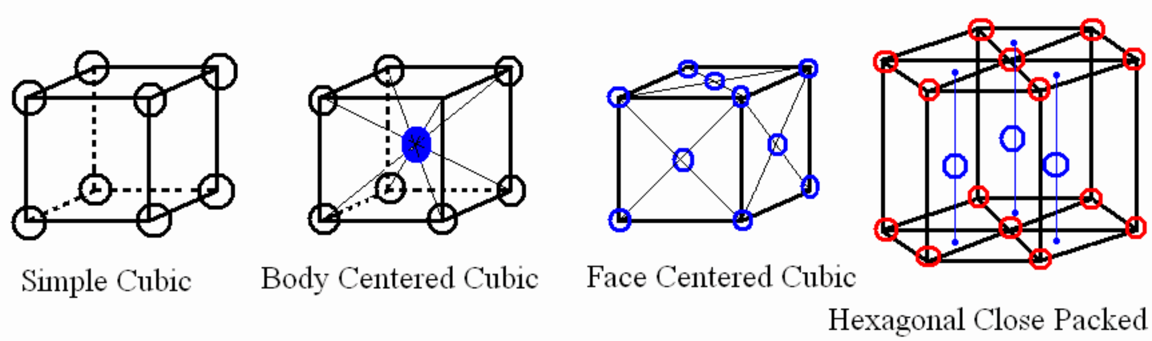


Figure 2.1: *Common metallic crystal structures.*

2.2.2 Crystalline and Non-crystalline materials

Single Crystals: Crystals can be *single crystals* where the whole solid is one crystal. Then it has a regular geometric structure with flat faces.

Polycrystalline Materials: A solid can be composed of many crystalline grains, not aligned with each other. It is called *polycrystalline*. The grains can be more or less aligned with respect to each other. Where they meet is called a *grain boundary*.

Non-Crystalline Solids: In amorphous solids, there is no long-range order. But amorphous does not mean random, since the distance between atoms cannot be smaller than the size of the hard spheres. Also, in many cases there is some form of short-range order. For instance, the tetragonal order of crystalline SiO_2 (quartz) is still apparent in amorphous SiO_2 (silica glass).

2.3 Miller Indices, Anisotropy, and Elastic behavior of composites

2.3.1 Miller indices:

It is understood that properties of materials depend on their crystal structure, and many of these properties are directional in nature. For example: elastic modulus of BCC iron is greater parallel to the body diagonal than it is to the cube edge. Thus it is necessary to characterize the crystal to identify specific directions and planes. Specific methods are employed to define crystal directions and crystal planes.

Methodology to define crystallographic directions in cubic crystal:

- a vector of convenient length is placed parallel to the required direction.
- the length of the vector projection on each of three axes are measured in unit cell dimensions.
- these three numbers are made to smallest integer values, known as indices, by multiplying or dividing by a common factor.
- the three indices are enclosed in square brackets, *[uvw]*. A family of directions is represented by <*uvw*>.

Methodology to define crystallographic planes in cubic crystal:

- determine the intercepts of the plane along the crystallographic axes, *in terms of unit cell dimensions*. If plane is passing through origin, there is a need to construct a plane parallel to original plane.
- take the reciprocals of these intercept numbers.
- clear fractions.
- reduce to set of smallest integers.
- The three indices are enclosed in parenthesis, *(hkl)*. A family of planes is represented by *{hkl}*.

For example, if the x-, y-, and z- intercepts of a plane are 2, 1, and 3. The Miller indices are calculated as:

- take reciprocals: 1/2, 1/1, 1/3.
- clear fractions (multiply by 6): 3, 6, 2.
- reduce to lowest terms (already there). => Miller indices of the plane are (362).

Figure 2.2 depicts Miller indices for number of directions and planes in a cubic crystal.

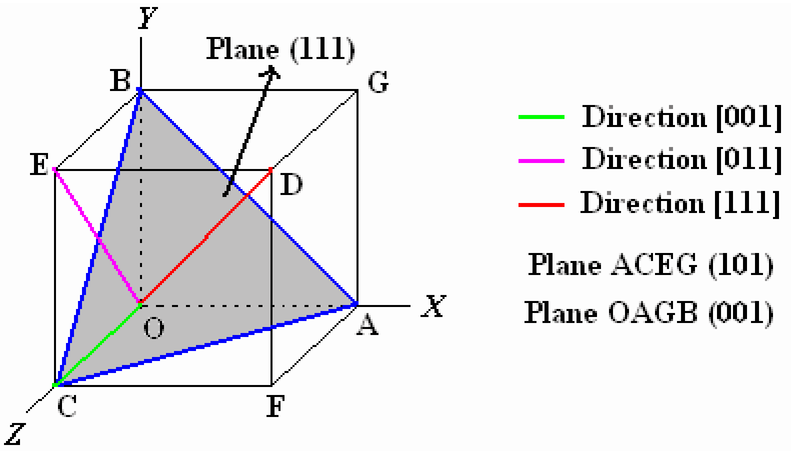


Figure 2.2: *Miller indices in a cubic crystal.*

Some useful conventions of Miller notation:

- If a plane is parallel to an axis, its intercept is at infinity and its Miller index will be zero.

- If a plane has negative intercept, the negative number is denoted by a bar above the number. *Never alter negative numbers.* For example, do not divide -1, -1, -1 by -1 to get 1,1,1. This implies symmetry that the crystal may not have!
- The crystal directions of a family are not necessarily parallel to each other. Similarly, not all planes of a family are parallel to each other.
- By changing signs of all indices of a direction, we obtain opposite direction. Similarly, by changing all signs of a plane, a plane at same distance in other side of the origin can be obtained.
- Multiplying or dividing a Miller index by constant has no effect on the orientation of the plane.
- The smaller the Miller index, more nearly parallel the plane to that axis, and vice versa.
- When the integers used in the Miller indices contain more than one digit, the indices must be separated by commas. E.g.: (3,10,13)
- By changing the signs of all the indices of (a) a direction, we obtain opposite direction, and (b) a plane, we obtain a plane located at the same distance on the other side of the origin.

More conventions applicable to cubic crystals only:

- [*uvw*] is normal to (*hkl*) if $u = h$, $v = k$, and $w = l$. E.g.: (111) ⊥ [111].
- Inter-planar distance between family of planes {*hkl*} is given by:

$$d_{\{hkl\}} = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

- [*uvw*] is parallel to (*hkl*) if $hu + kv + lw = 0$.
- Two planes ($h_1k_1l_1$) and ($h_2k_2l_2$) are normal if $h_1h_2 + k_1k_2 + l_1l_2=0$.
- Two directions ($u_1v_1w_1$) and ($u_2v_2w_2$) are normal if $u_1u_2 + v_1v_2 + w_1w_2=0$
- Angle between two planes is given by:

$$\cos\theta = \frac{h_1h_2 + k_1k_2 + l_1l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2}\sqrt{h_2^2 + k_2^2 + l_2^2}}$$

The same equation applies for two directions.

Why Miller indices are calculated in that way?

- Using reciprocals spares us the complication of infinite intercepts.
- Formulas involving Miller indices are very similar to related formulas from analytical geometry.
- Specifying dimensions in unit cell terms means that the same label can be applied to any plane with a similar stacking pattern, regardless of the crystal class of the crystal. Plane (111) always steps the same way regardless of crystal system.

2.3.2 Miller-Bravis indices

Though Miller indices can describe all possible planes through any crystal, Miller-Bravis indices are used in hexagonal crystal systems. This is because they reveal hexagonal symmetry more clearly. Although partially redundant, they are used exclusively for hexagonal systems.

Direction indices are obtained as above where first three indices are representative of projections of the direction over three co-planar axes in the plane called basal plane while the last index denotes the projection over the axis perpendicular to the basal plane. Miller-Bravis indices for a plane are denoted as *[uvtw]*, where $t = -(u+v)$

In the same procedure, planes in a hexagonal crystal are denoted by *(hkil)*, where $i = -(h+k)$.

2.3.3 Anisotropy

It's been agreed that many of the materials properties depend on the crystal structure. However, crystals are not symmetric in all directions, or not the crystal planes same with respect to atomic density/packing. Different directions in the crystal have different packing. For instance, atoms along the edge of FCC crystals are more separated than along its face diagonal. This causes properties to be different in different directions. This directionality of properties is termed as *Anisotropy*.

Substances in which measured properties are independent of direction in which they are measured are called *isotropic*. Though, in polycrystalline materials, the crystallographic orientations of individual grains are random, specimen may behave isotropically.

2.3.4 Elastic Behavior of Composites

The idea is that by combining two or more distinct materials one can engineer a new material with the desired combination of properties (e.g., light, strong, corrosion resistant). The idea that a better combination of properties can be achieved is called the *principle of combined action*.

For example, Pearlitic steel that combines hard and brittle Cementite with soft and ductile ferrite to get a superior material.

A composite is defined as a artificially made muti-phase material, where constituent phases are chemically dissimilar and separated by a distinct interface. Thus, composites shall have at least two or more constituent phases. Many of the composites are made of two phases – one is termed as *matrix*, which is continuous and surrounds the other phase, called *dispersed phase*. Dispersed phase can exist in many forms like particulates, short-/long- fibers.

Properties of composites depend on

- Properties of the constituent phases.
- Geometry of dispersed phase (particle size, size distribution, orientation).
- Amount of each constituent phase.

Classification of composites based on geometry of dispersed phase:

- Particle-reinforced (large-particle and dispersion-strengthened)
- Fiber-reinforced (continuous (aligned) and short fibers (aligned or random)
- Structural (laminates and sandwich panels)

Classification of composites based on matrix phase:

- Metal Matrix Composites.
- Polymer Matrix Composites.
- Ceramic Matrix Composites.

Composite properties can be calculated using *rule of mixtures*. For example, elastic modulus of a particle reinforced composite bound by the limits given by

$$E_c(u) = E_m V_m + E_p V_p$$

$$E_c(l) = \frac{E_m E_p}{E_m V_p + E_p V_m}$$

where E and V denote the elastic modulus and volume fraction respectively; c, m, and p represent composite, matrix, and particulate phases.

However, in case of continuous fiber reinforced composites upper bound is applicable to condition of longitudinal loading, while the lower bound is applicable to transverse loading condition. The above equations can be simplified using the following relations:

$$V_m + V_p = 1$$

When fiber orientation is random, or short and discontinuous fibers are used, rule of mixtures will be modified as follows to take care of the randomness of fibers:

$$E_{cd} = E_m V_m + K E_f V_f$$

where K – is fiber efficiency parameter that depends on V_f and E_f/E_m ratio. K's values are less than unity, usually attains a value in the range of 0.1 to 0.6.

Many applications, like in aircraft parts, there is a need for high strength per unit weight (specific strength). This can be achieved by composites consisting of a low-density (and soft) matrix reinforced with stiff fibers.

2.4 Structure and properties of polymers

Polymers are common in nature, in the form of wood, rubber, cotton, leather, wood, silk, proteins, enzymes, starches, cellulose. Artificial polymers are made mostly from oil. Their use has grown exponentially, especially after WW2 (World War-2). The key factor is the very low production cost and useful properties (e.g., combination of transparency and flexibility, long elongation, etc.).

Most polymers are organic, and formed from hydrocarbon molecules. These molecules can have single, double, or triple carbon bonds. A *saturated hydrocarbon* is one where all bonds are single, i.e. the number of atoms is maximum (or saturated). Among this type are the paraffin compounds, C_nH_{2n+2}. In contrast, non-saturated hydrocarbons contain some double and triple bonds.

Isomers are molecules that contain the same molecules but in a different arrangement. An example is butane and iso-butane. Some physical properties of hydrocarbons depend on the isomeric state.

2.4.1 Polymer molecules

Polymer molecules are huge, macromolecules that have internal covalent bonds. For most polymers, these molecules form very long chains. The backbone is a string of carbon atoms, often single bonded. Polymers are composed of basic structures called ***mer*** units. A molecule with just one mer is a monomer. Within each molecule / mer atoms are bonding together by strong covalent bonds. When many mers are together, they form polymer. Bi-functional monomers may bond with two other units in forming 2-D chain-like structures; while Tri-functional monomers can form three active bonds, and thus 3-D molecular network. Examples of polymers are polyvinyl chloride (PVC), poly-tetra-fluoro-ethylene (PTFE or Teflon), polypropylene, nylon and polystyrene. When all the mers are the same, the molecule is called a *homopolymer*. When there is more than one type of mer present, the molecule is a *copolymer*.

The mass of a polymer is not fixed, but is distributed around a mean value, since not all polymer chains will grow same extent. The average molecular weight can be obtained by averaging the masses with the fraction of times they appear (*number-average*) or with the weight fraction of the molecules (*weight-average*). Another representation of average chain size is *degree of polymerization* (n) – average number of mer units in a chain. It is obtained by dividing the average mass of the polymer by the mass of a mer unit. Numbers of polymer characteristics are affected by the magnitude of the molecular weight. Short chain polymers usually exist in form of gases or liquids at room temperature; where as medium range polymers are waxy solids and soft resins. Solid polymers are commonly having weights ranging between 10K and several million g/mol.

2.4.2 Polymer structures

Polymers consist of large number of molecular chains which are usually not linear; bending and rotations can occur around single C-C bonds (double and triple bonds are very rigid). Random kinks and coils in chains along with bending of chains lead to intertwining and entanglement of neighboring chains, situation like in the spaghetti structure. These characteristic entanglements are responsible for a number of properties specific to polymers, e.g.: large elastic extension. However, physical properties of polymers depend not only on molecular weight and shape, but also on differences in structure of the chains. It should be remembered that polymers are not usually of only one distinctive structural type, though they are classified into different groups. Typical polymer chain structures are: (a) *linear*, where mer units are joined together end to end in single chains. E.g.: PVC, nylon. (b) *branched*, where side-branch chains are connected to main ones. Branching of polymers lowers polymer density because of lower packing efficiency. (c) *cross-linked*, where chains are joined one to another at various positions by covalent bonds. This cross-linking is usually achieved at elevated temperatures by additive atoms. E.g.: vulcanization of rubber. (d) *network*, trifunctional mer units with 3-D networks comes under this category. E.g.: epoxies, phenol-formaldehyde.

2.4.3 Polymer crystallinity

Crystallinity in polymers is more complex than in metals. Polymer molecules are often partially crystalline (*semicrystalline*), with crystalline regions dispersed within amorphous material. The degree of crystallinity may range from completely amorphous to almost entirely crystalline; on the other hand metals are almost always crystalline whereas ceramics are either completely crystalline or noncrystalline. The degree of crystallinity of a polymer depends on cooling path, and also on chain configuration. For copolymers, the more irregular and random the mer arrangement, the greater is probability for noncrystalline nature. Crystalline polymers are denser than amorphous polymers, so the degree of crystallinity can be obtained from the measurement of density.

Different models have been proposed to describe the arrangement of molecules in semicrytalline polymers. In the fringed-micelle model, the crystallites (micelles) are embedded in an amorphous matrix. Polymer single crystals grown are shaped in regular platelets (lamellae). Spherulites are chain-folded crystallites in an amorphous matrix that grow radially in spherical shape “grains”. These are considered to be the polymer analogue of grains in polycrystalline metals and ceramics. Many semicrystalline polymers form spherulites; each spherulite consists of a collection of ribbonlike chain-folded lamellar crystallites that radiate outward from its center. E.g.: polyethelene, PVC.

A polymer’s response to mechanical forces under elevated temperatures is related to its molecular structure. Based on this response, polymers are classified as: *thermoplasts* (soften when heated and harden when cooled), and *thermosets* (become permanently hard when heat is applied and do not soften upon subsequent heating). Thermosets are generally harder and stronger than thermoplasts, and have better dimensional stability. Most of the cross-linked and network polymers are thermosets; whereas linear and some branched polymers are thermoplasts.

2.4.4 Properties of polymers

Fluids and amorphous solids undergo viscous flow when external forces are applied. It is well known that polymers exhibit very high viscosity in order of 10^{12} Pa.s at room temperature. Polymers are non-Newtonian in nature, and formed into plastic products at a temperature above their glass-transition temperature. It is evident that temperature has very strong influence on mechanical behavior of polymers. Elastic strain occurs simultaneously with viscous flow, resulting in visco-elastic deformation of polymers under externally applied loads. Below the glass transition temperature elastic deformation dominates and the material behaves rigid. In the range of glass temperature, the materials is leathery; in the rubber plateau, polymers deform readily but quickly regain their previous shape if the stress in removed. At still higher temperatures, under sustained loads, the polymer deforms extensively by viscous flow. Figure below depicts temperature effect on deformation behavior of polymers.

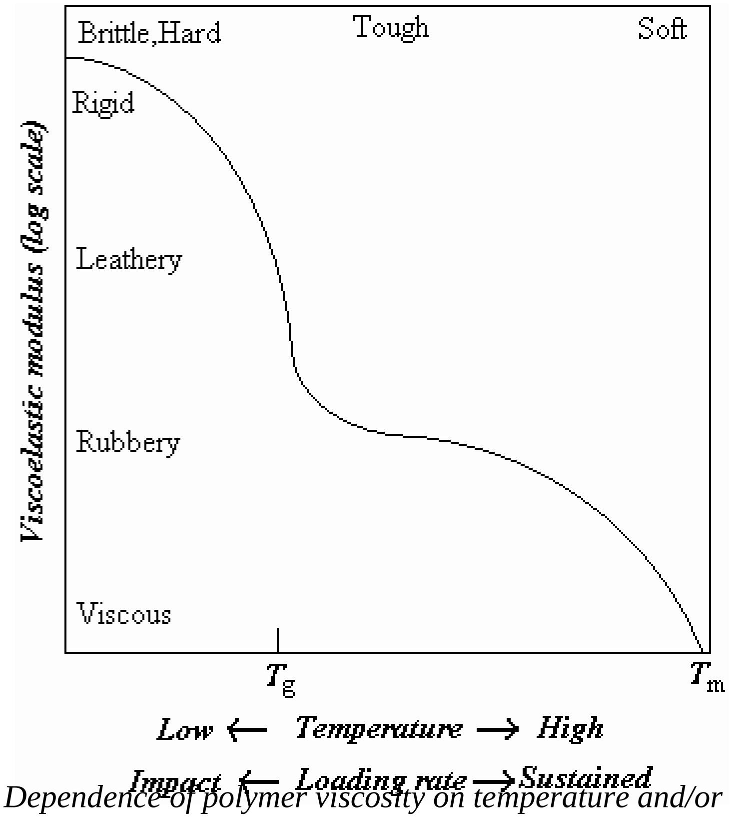


Figure 2.3: *Dependence of polymer viscosity on temperature and/or loading rate.*

2.5 Structure and properties of ceramics

Ceramics are inorganic and non-metallic materials that are commonly electrical and thermal insulators, brittle and composed of more than one element (e.g., two in Al_2O_3). As ceramics are composed of two or more elements, their crystal structures are generally more complex than those of metals. Ceramic bonds are mixed, ionic and covalent, with a proportion that depends on the particular ceramics. The ionic character is given by the difference of electronegativity between the cations (+) and anions (-). Covalent bonds involve sharing of valence electrons. Very ionic crystals usually involve cations which are alkalis or alkaline-earths (first two columns of the periodic table) and oxygen or halogens as anions.

The building criteria for the ceramic crystal structure are as follows:

- maintain neutrality (charge balance dictates chemical formula)
- achieve closest packing

The crystal stability condition i.e. condition of minimum energy implies maximum attraction and minimum repulsion. This leads to contact and configurations such that anions have the highest number of cation neighbors (coordination number) and vice versa. The coordination number is dependent on cation-anion radius ratio, which can be determined from geometric relations. Table 2.2 presents relevant coordination numbers and radius ratios.

Table 2.2. Co-ordination number dependency on cation-anion radius ratio.

Cation-anion radius ratio (r_c/r_a)	< 0.155	0.155 – 0.225	0.225 – 0.414	0.414 – 0.732	0.732 – 1.000	> 1.000
Coordination number	2	3	4	6	8	12

Figure-2.4 presents schematic arrangement of cations and anions with respective coordinate positions for different radius ratios.

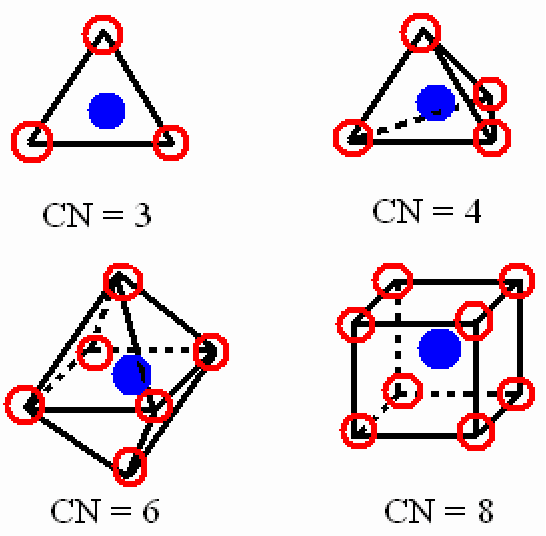


Figure 2.4: *Ion arrangements for different coordination numbers.*

2.5.1 Ceramic crystal structures

<u>AX-type ceramic crystal structures</u>: Most common ceramics are made of equal number of cations and anions, and are referred to as AX compounds (A-cation, and X-anion). These ceramics assume many different structures, named after a common material that possesses the particular structure.

Rock salt structure: here the coordination number is 6, i.e. r_c/r_a = 0.414-0.732. This structure can be viewed as an FCC of anions with cations occupying center of each edge and the center of the cell. Thus it can be said that lattice is made of two interpenetrating FCC lattices, one composed of cations, and the other of anions. E.g.: NaCl, MgO, FeO.

Cesium Chloride structure: here the coordination number is 8. Crystal structure consists of anions at corners of a cube while a cation occupies the center, and vice versa. E.g.: CsCl.

Zinc Blende structure: here the coordination number is 4. Unit cell is composed of one kind of ions occupying corners and face centers of a cube, while the other kind of ions occupies the interior tetrahedral positions. E.g.: ZnS, SiC.

A_mX_p- type structures: when the charges of cation and anions are not the same, to maintain the neutrality, ceramic structures with chemical formula would exist. For example – CaF_2 with r_c/r_a = 0.8, and thus coordination number of 8. It can be expected that crystal structure could be the same as that of CsCl. However cations are half many as anions, thus only half the center positions are occupied. One unit cell shall be made of eight cubes. E.g.: UO_2, ThO_2, PuO_2.

$A_mB_nX_p$-type structures: it is possible that ceramics do have more than one kind of cations. E.g.: $BaTiO_3$. unit cell is made of cube where Ba^{2+} ions occupies all eight corners, T^{4+} occupies cube center, while O^{2-} are at center of each face. This structure is called *perovskite crystal structure*.

It is worth to understand and know more about some common most common ceramic in nature. For example: silicates and carbon.

2.5.2 Silicates

Oxygen and Silicon are the most abundant elements in Earth's crust. Their combination (silicates) occurs in rocks, soils, clays and sand. The bond is weekly ionic, with Si^{4+} as the cation and O^{2-} as the anion. However the bonds in silicates are strongly of covalent character with strong directional Si-O bonds. Basic unit of silicates structures are thus is SiO_4^{4-} tetrahedron that consists of four oxygen atoms at corners of tetrahedron, and silicon atom at the center of it. Various silicate structures consists SiO_4^{4-} unit bonded in 1-, 2-, and 3- dimensions.

In silica (SiO_2) every oxygen atom is shared by adjacent tetrahedra. Silica can either be crystalline (e.g., quartz) or amorphous, as in glass. Crystalline forms of silica are known to be complicated and comparatively open, thus of low densities compared with amorphous glasses. Soda glasses melt at lower temperature than amorphous SiO_2 because the addition of Na_2O (soda) that act as *network modifier* breaks the tetrahedral network. Addition of *intermediates* such as Al_2O_3, TiO_2 substitute of silicon atoms and become part of stabilized network. Addition of network modifiers and intermediates lowers melting point, and thus it is easy to form glass, for instance, bottles.

In complicated silicate structure, corner oxygen atom of basic unit is shared by other tetrahedra, resulting in formulas such as SiO_4^{4-}, $Si_2O_7^{6-}$, $Si_3O_9^{6-}$, etc. The repeating unit of 2-D sheet or layered structure is represented as $Si_2O_5^{2-}$. Such layered structures are characteristics of clays and other minerals.

2.5.3 Carbon

Carbon is not really a ceramic, but one of its allotropic form, diamond may be considered as a ceramic. *Diamond* has very interesting and even unusual properties such as:

- possesses diamond-cubic structure (like Si, Ge)
- consists covalent C-C bonds
- having highest hardness of any material known
- very high thermal conductivity (unlike ceramics)
- transparent in the visible and infrared, with high index of refraction
- semiconductor (can be doped to make electronic devices)
- meta-stable (transforms to carbon when heated)

Synthetic diamonds are made by application of high temperatures and pressures or by chemical vapor deposition. Future applications of this latter, cheaper production method include hard coatings for metal tools, ultra-low friction coatings for space applications, and microelectronics.

Graphite, another allotropic form of carbon, has a layered structure with very strong hexagonal bonding within the planar layers (using 3 of the 3 bonding electrons) and weak, van der Waals bonding between layers using the fourth electron. This leads to easy inter-planar cleavage and applications as a lubricant and for writing (pencils). Graphite is a good electrical conductor and chemically stable even at high temperatures. Applications include furnaces, rocket nozzles, electrodes in batteries, etc.

Recently (1985) discovered allotropic form of carbon is the C_{60} molecule, also known as *fullerene* or *bucky-ball* (after the architect Buckminster Fuller who designed the geodesic structure that C_{60} resembles.). Structure of this form resembles a hallow spherical cluster of 60 atoms, and is found to consist of 20 hexagons and 12 pentagons where no two pentagons share a common edge. Fullerenes and related structures like nanotubes are exceptionally stiff, strong, and ductile. Future applications of fullerenes are as a structural material and possibly in microelectronics, due to the unusual properties that result when fullerenes are doped with other atoms.

2.5.4 Imperfections in ceramics

Imperfections in ceramics include point defects and impurities. Their formation is strongly affected by the condition of charge neutrality (creation of unbalanced charges requires the expenditure of a large amount of energy). Both vacancies and interstitials are possible in ceramics as in metals; however as ceramics have more then one element these defects can be associated with each of these elements. Neutral charge defects include the Frenkel and Schottky defects. A *Frenkel-defect* is a vacancy- interstitial pair of cations (placing large anions in an interstitial position requires a lot of energy in lattice distortion). A *Schottky-defect* is a pair of nearby cation and anion vacancies.

Non-stoichiometry refers to a change in composition so that the elements in the ceramic are not in the proportion appropriate for the compound (condition known as stoichiometry). To minimize energy, the effect of non-stoichiometry is a redistribution of the atomic charges. Introduction of impurity atoms in the lattice is likely in conditions where the charge is maintained. This is the case of electronegative impurities that substitute lattice anions or electropositive substitutional impurities. This is more likely for similar ionic radii since this minimizes the energy required for lattice distortion. Defects will appear if the charge of the impurities is not balanced.

2.5.5 Mechanical response of ceramics

Brittle Fracture of Ceramics: The brittle fracture of ceramics limits their engineering applications. It occurs due to the unavoidable presence of microscopic flaws (micro-cracks, internal pores, and atmospheric contaminants) that result during cooling from the melt. The flaws need to crack initiation, and crack propagation (perpendicular to the applied stress) is usually trans-granular, along cleavage planes. The flaws cannot be closely controlled in manufacturing; this leads to a large scatter in the fracture strength of ceramic materials.

Under compressive stresses, however, flaws do not associate with amplification of stress. Hence, the compressive strength of ceramics is typically ten times to their tensile strength. This makes ceramics good structural materials under compression (e.g., bricks in houses, stone blocks in the pyramids), but not in conditions of tensile stress, such as under flexure.

Plastic deformation in crystalline ceramics is by slip, which is difficult due to the structure and the strong local (electrostatic) potentials. There is very little plastic deformation before fracture. Non-crystalline ceramics, like common glass, deform by viscous flow (like very high-density liquids) as deformation can not occur by slip because of absence of regular atomic structure, which occurs in metals. Characteristic property of viscous flow, viscosity, is a measure of non-crystalline material’s resistance to deformation. It was found to decrease with increasing temperature. At room temperature, the viscosity of many non-crystalline ceramics is extremely high.

References

1. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
2. B. D. Cullity and S. R. Stock, Elements of X-Ray Diffraction, Third Edition, Prentice Hall, Upper Saddle River, NJ, 2001
3. Lawrence H. Van Vlack, Elements of Materials Science and Engineering, sixth edition, Addison Wesley Longman, Inc. New York, 1998

Chapter 3. Imperfections in Solids

3.1 Theoretical yield strength, Point defects, Line defects and Dislocations

3.1.1 Theoretical strength

Atoms are orderly arranged without any disruptions in ideal solids. Under mechanical loads, all solids tend to reshape themselves. This occurs usually by process called slip, the translation of one plane of atoms over another under implied shear stresses.

Consider two planes of atoms subjected a shear stress under applied load. Let’s assume, as shown in *figure-3.1*, distance between atoms in slip direction is *b*, and spacing between the two planes is *a*.

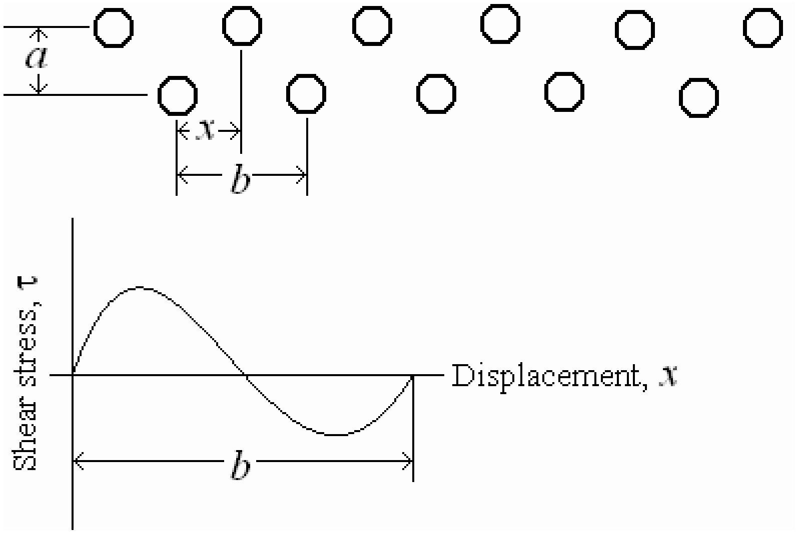


Figure-3.1: *Shear displacement of atom planes and variation of shear stress.*

The shear stress is zero initially and also once planes moved one identity distance, *b*. In between, it can be said that shear stress is a periodic function of displacement as each atom is attracted toward the nearest atom of the other row. Thus the relation of shear stress with displacement can be expressed as follows:

$$\tau = \tau_m \sin \frac{2\Pi x}{b}$$

where is the amplitude of sine wave, and b is the period. For small values of x/b, after simplification,

$$\tau \approx \tau_m \frac{2\Pi x}{b}$$

It is safe to approximate at small values of displacement Hooke's law is applicable. Thus,

$$\tau = G\gamma = \frac{Gx}{a}$$

Combing the above two equations, maximum shear stress at which slip should occur is

$$\tau_m = \frac{G}{2\Pi}\frac{b}{a}$$

If it is assumed as $b \approx a$, approximate value of shear stress needed for initiation of slip is

$$\tau_m = \frac{G}{2\Pi}$$

Shear modulus, G, of solids is in the range of 20-150 GPa, and thus theoretical shear strength could be in range 3-30 GPa. However, actual values of the shear stress required for plastic deformation is only about 0.5-10 MPa i.e. about 100 times less than the theoretical value. Because of this discrepancy, it must be said that a mechanism other than shearing of planes must be operative under application of loads thus causing plastic deformation.

3.1.2 Point defects

Defects exists any all solid materials. For ease of their characterization, defects are classified on the basis of their geometry, which is realistic as defects are disrupted region in a volume of a solid. Defects are:

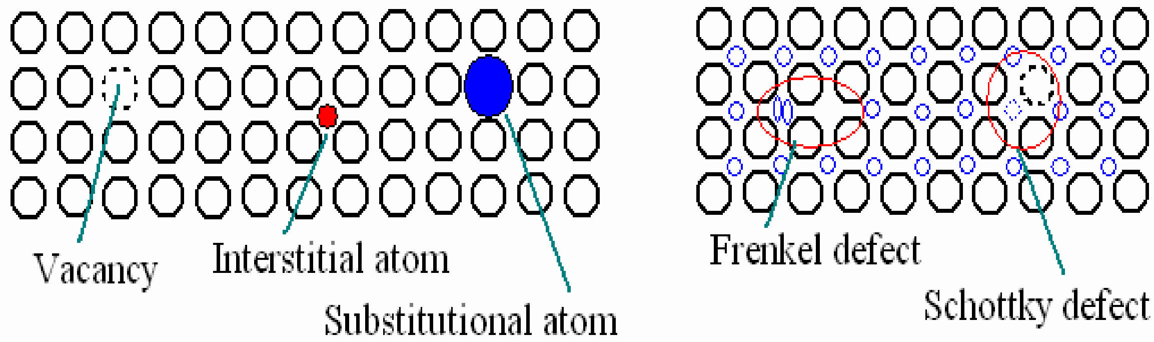
a. Point defects (zero-dimensional)
b. Line defects (single dimensional)
c. Surface defects (two dimensional)
d. Volume defects (three dimensional)

Point defects, as the name implies, are imperfect point-like regions in the crystal. Typical size of a point defect is about 1-2 atomic diameters. Different point defects are explained in the following paragraphs. *Figure-3.2* depicts various point defects.

A *vacancy* is a vacant lattice position from where the atom is missing. It is usually created when the solid is formed by cooling the liquid. There are other ways of making a vacancy, but they also occur naturally as a result of thermal excitation, and these are thermodynamically stable at temperatures greater than zero. At equilibrium, the fraction of lattice sites that are vacant at a given temperature (*T*) are:

$$\frac{n}{N} = e^{-Q/kT}$$

where *n* is the number of vacant sites in *N* lattice positions, *k* is gas or Boltzmann's constant, *T* is absolute temperature in kelvins, and *Q* is the energy required to move an atom from the interior of a crystal to its surface. It is clear from the equation that there is an exponential increase in number of vacancies with temperature. When the density of vacancies becomes relatively large, there is a possibility for them to cluster together and form voids.



Figur-3.2: Schematic depiction of various point defects

An *interstitial atom* or *interstitialcy* is an atom that occupies a place outside the normal lattice position. It may be the same type of atom as the rest surrounding it (self interstitial) *or* a foreign impurity atom. Interstitialcy is most probable if the atomic packing factor is low.

Another way an impurity atom can be fitted into a crystal lattice is by substitution. A *substitutional atom* is a foreign atom occupying original lattice position by displacing the parent atom.

In the case of vacancies and foreign atoms (both interstitial and substitutional), there is a change in the coordination of atoms around the defect. This means that the forces are not balanced in the same way as for other atoms in the solid, which results in lattice distortion around the defect.

In ionic crystals, existence of point defects is subjected to the condition of charge neutrality. There are two possibilities for point defects in ionic solids.

> - when an ion displaced from a regular position to an interstitial position creating a vacancy, the pair of vacancy-interstitial is called *Frenkel defect*. Cations are usually smaller and thus displaced easily than anions. Closed packed structures have fewer interstitials and displaced ions than vacancies because additional energy is required to force the atoms into the interstitial positions.
>
> - a pair of one cation and one anion can be missing from an ionic crystal, without violating the condition of charge neutrality when the valency of ions is equal. The pair of vacant sites, thus formed, is called *Schottky defect*. This type of point defect is dominant in alkali halides. These *ion-pair vacancies*, like single vacancies, facilitate atomic diffusion.

3.1.3 Line defects or Dislocations:

Line imperfections (one-dimensional defects) are also called Dislocations. They are abrupt changes in the regular ordering of atoms along a line (dislocation line) in the solid. They occur in high densities and strongly influence the mechanical properties of material. They are characterized by the Burgers vector (***b***), whose direction and magnitude can be determined by constructing a loop around the disrupted region and noticing the extra inter-atomic spacing needed to close the loop. The Burgers vector in metals points in a close packed lattice direction. It is unique to a dislocation.

Dislocations occur when an extra incomplete plane is inserted. The dislocation line is at the end of the plane. Dislocations can be best understood by referring to two limiting cases - Edge dislocation *and* Screw dislocation.

Edge dislocation or *Taylor-Orowan dislocation* is characterized by a Burger’s vector that is perpendicular to the dislocation line. It may be described as an edge of an extra plane of atoms within a crystal structure. Thus regions of compression and tension are associated with an edge dislocation. Because of extra incomplete plane of atoms, the atoms above the dislocation line are squeezed together and are in state of compression whereas atoms below are pulled apart and experience tensile stresses. Edge dislocation is considered *positive* when compressive stresses present above the dislocation line, and is represented by $\perp$. If the stress state is opposite i.e. compressive stresses exist below the dislocation line, it is considered as negative edge dislocation, and represented by $\top$. A schematic view of edge dislocations are shown in *figure 3.3.*

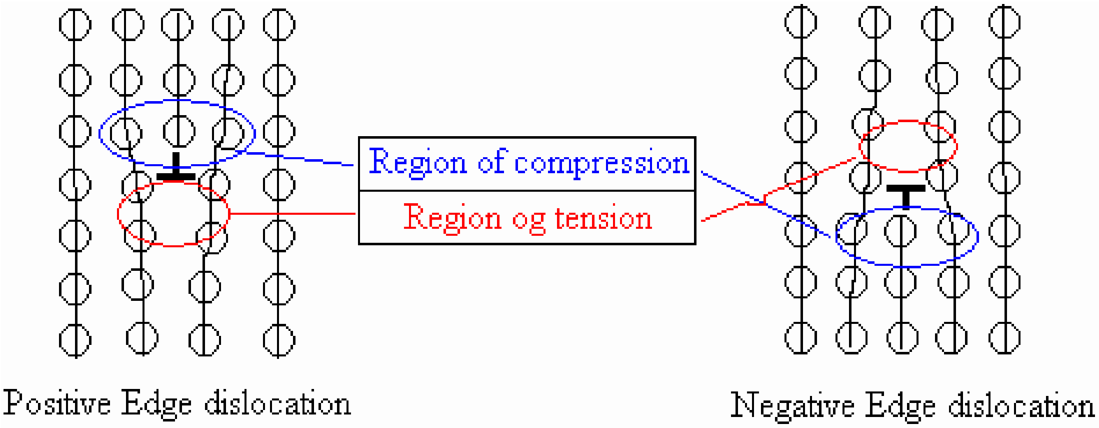
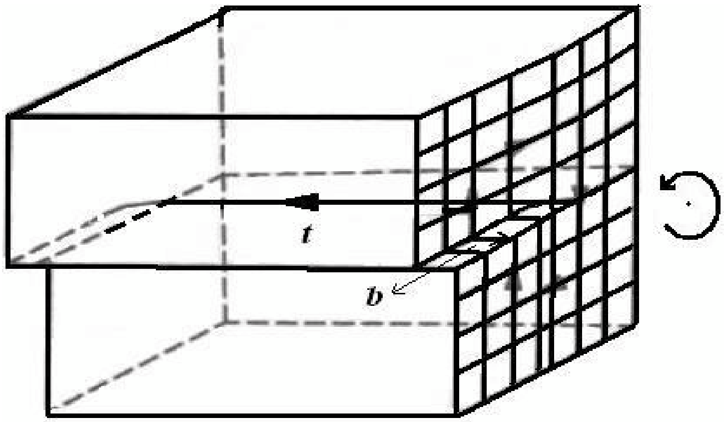


Figure-3.3: *Edge dislocations.*

A pure edge dislocation can glide or slip in a direction perpendicular to its length i.e. along its Burger’s vector in the slip plane (made of ***b*** and ***t*** vectors), on which dislocation moves by slip while conserving number of atoms in the incomplete plane. It may move vertically by a process known as climb, if diffusion of atoms or vacancies can take place at appropriate rate. Atoms are added to the incomplete plane for negative climb i.e. the incomplete plane increases in extent downwards, and vice versa. Thus climb motion is considered as non-conservative, the movement by climb is controlled by diffusion process.

Screw dislocation or *Burgers dislocation* has its dislocation line parallel to the Burger’s vector. A screw dislocation is like a spiral ramp with an imperfection line down its axis. Screw dislocations result when displacing planes relative to each other through shear. Shear stresses are associated with the atoms adjacent to the screw dislocation; therefore extra energy is involved as it is in the case of edge dislocations. Screw dislocation is considered positive if Burger’s vector and *t*-vector or parallel, and vice versa. (***t***-vector – an unit vector representing the direction of the dislocation line). A positive screw dislocation is represented by “‘↻’ a dot surrounded by circular direction in clock-wise direction”, whereas the negative screw dislocation is represented by “‘↺’ a dot surrounded by a circular direction in anti-clock-wise direction”. A schematic view of a negative screw dislocation is shown in *figure 3.4.*



Fifure-3.4: *Negative* s*crew dislocation*.

A screw dislocation does not have a preferred slip plane, as an edge dislocation has, and thus the motion of a screw dislocation is less restricted than the motion of an Edge dislocation. As there is no preferred slip plane, screw dislocation can cross-slip on to another plane, and can continue its glide under favorable stress conditions. However, screw dislocation can not move by climb process, whereas edge dislocations can not cross-slip.

Dislocations more commonly originate during plastic deformation, during solidification, and as a consequence of thermal stresses that result from rapid cooling. Edge dislocation arises when there is a slight mismatch in the orientation of adjacent parts of the growing crystal. A screw dislocation allows easy crystal growth because additional atoms can be added to the 'step' of the screw. Thus the term screw is apt, because the step swings around the axis as growth proceeds. Unlike point defects, these are not thermodynamically stable. They can be removed by heating to high temperatures where they cancel each other or move out through the crystal to its surface. Virtually all crystalline materials contain some dislocations. The density of dislocations in a crystal is measures by counting the number of points at which they intersect a random cross-section of the crystal. These points, called etch-pits, can be seen under microscope. In an annealed crystal, the dislocation density is the range of 10^{8}-10^{10} m^{-2}.

Any dislocation in a crystal is a combination of edge and screw types, having varying degrees of edge and screw character. *Figure 3.5* depicts a schematic picture of a common dislocation.

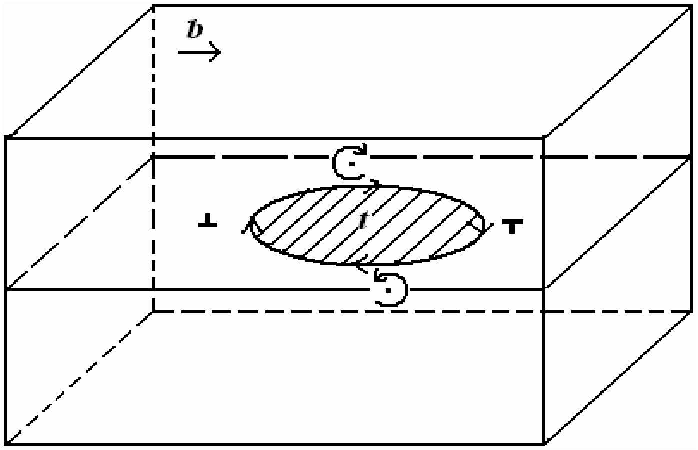


Figure-3.5: *Typical dislocation in a crystal.*

As shown in *figure 3.5*, the Burger's vector will have the same magnitude and direction over the whole length of dislocation irrespective of the character of the dislocation (edge, screw, or mixed) i.e. Burger's vector is invariant. Other geometrical characteristics of dislocations are:

- vectorial sum of Burger's vectors of dislocations meeting at a point, called node, must be zero.

- t-vectors of all the dislocations meeting at a node must either point towards it or away from it.

- a dislocation line can not end abruptly within the crystal. It can close on itself as a loop, or ends either at a node or at the surface.

Dislocations have distortional energy associated with them as is evident from the presence of tensile/compressive/shear stresses around a dislocation line. Strains can be expected to be in the elastic range, and thus stored elastic energy per unit length of the dislocation can be obtained from the following equation:

$$E \cong {Gb^2}/{2}$$

where G – shear modulus and $\boldsymbol{b}$ – Burger's vector.

Dislocations in the real crystal can be classified into two groups based on their geometry – full dislocations and partial dislocations. Partial dislocation's Burger's vector will be a fraction of a lattice translation, whereas Burger's vector is an integral multiple of a lattice translation for full dislocation. As mentioned above, elastic energy associated with a dislocation is proportional to square of its Burger's vector; dislocation will tend to have as small a Burger's vector as possible. This explains the reason for separation of dislocations that tend to stay away from each other!

3.2 Interfacial defects, Bulk or Volume defects and Atomic vibrations

3.2.1 Interfacial defects

Interfacial defects can be defined as boundaries that have two dimensional imperfections in crystalline solids, and have different crystal structures and/or crystallographic orientations on either side of them. They refer to the regions of distortions that lie about a surface having thickness of a few atomic diameters. For example: external surfaces, grain boundaries, twin boundaries, stacking faults, and phase boundaries. These imperfections are not thermodynamically stable, rather they are meta-stable imperfections. They arise from the clustering of line defects into a plane.

External surface: The environment of an atom at a surface differs from that of an atom in the bulk; especially the number of neighbors (coordination) at surface is less. Thus the unsaturated bonds of surface atoms give rise to a surface energy. This result in *relaxation* (the lattice spacing is decreased) or *reconstruction* (the crystal structure changes). To reduce the energy, materials tend to minimize, if possible, the total surface area.

Grain boundaries: Crystalline solids are, usually, made of number of grains separated by grain boundaries. Grain boundaries are several atoms distances wide, and there is mismatch of orientation of grains on either side of the boundary as shown in *figure-3.6*. When this misalignment is slight, on the order of few degrees (< 10˚), it is called *low angle grain boundary*. These boundaries can be described in terms of aligned dislocation arrays. If the low grain boundary is formed by edge dislocations, it is called *tilt boundary*, and *twist boundary* if formed of screw dislocations. Both tilt and twist boundaries are planar surface imperfections in contrast to high angle grain boundaries. For *high angle grain boundaries*, degree of disorientation is of large range (> 15˚). Grain boundaries are chemically more reactive because of grain boundary energy. In spite of disordered orientation of atoms at grain boundaries, polycrystalline solids are still very strong as cohesive forces present within and across the boundary.

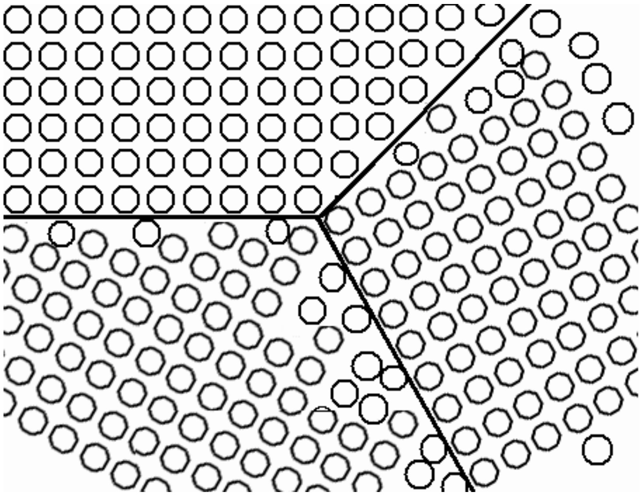


Figure-3.6: *Schematic presentation of grain boundaries*.

Twin boundaries: It is a special type of grain boundary across which there is specific mirror lattice symmetry. Twin boundaries occur in pairs such that the orientation change introduced by one boundary is restored by the other (*figure-3.7*). The region between the pair of boundaries is called the twinned region. Twins which forms during the process of recrystallization are called *annealing twins*, whereas *deformation twins* form during plastic deformation. Twinning occurs on a definite crystallographic plane and in a specific direction, both of which depend on the crystal structure. Annealing twins are typically found in metals that have FCC crystal structure (and low stacking fault energy), while mechanical/deformation twins are observed in BCC and HCP metals. Annealing twins are usually broader and with straighter sides than mechanical twins. Twins do not extend beyond a grain boundary.

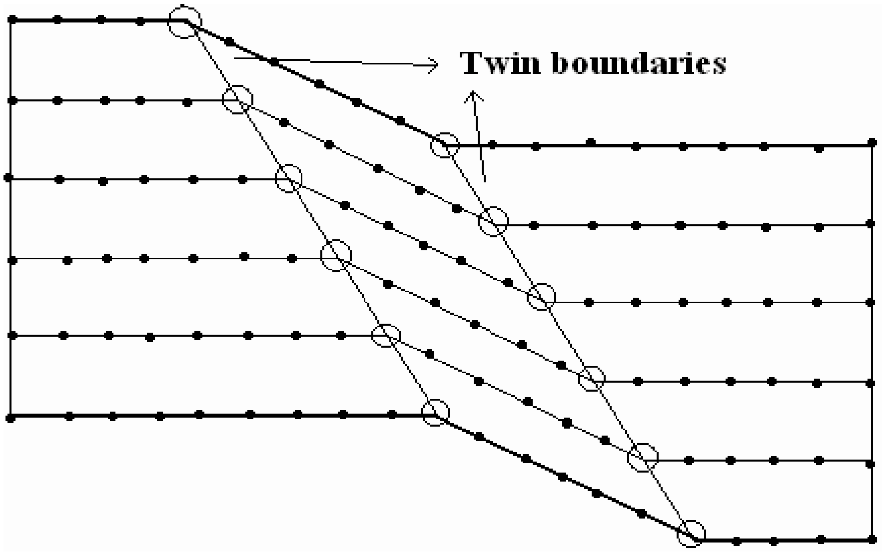


Figure-3.7: *A pair of twin boundaries.*

Stacking faults: They are faults in stacking sequence of atom planes. Stacking sequence in an FCC crystal is ABC ABC ABC …, and the sequence for HCP crystals is AB AB AB…. When there is disturbance in the stacking sequence, formation of stacking faults takes place. Two kinds of stacking faults in FCC crystals are: (a) ABC AC ABC…where CA CA represent thin HCP region which is nothing but stacking fault in FCC, (b) ABC ACB CABC is called *extrinsic* or *twin stacking fault*. Three layers ACB constitute the twin. Thus stacking faults in FCC crystal can also be considered as submicroscopic twins. This is why no microscopic twins appear in FCC crystals as formation of stacking faults is energetically favorable. Stacking fault energy varies in range 0.01-0.1 J/m^2. Lower the stacking fault energy, wider the stacking fault, metal strain hardens rapidly and twin easily. Otherwise, metals of high stacking fault energy i.e. narrower stacking faults show a deformation structure of banded, linear arrays of dislocations.

Phase boundaries exist in multiphase materials across which there is sudden change in physical/chemical characteristics.

3.2.2 Bulk or Volume defects

Volume defects as name suggests are defects in 3-dimensions. These include pores, cracks, foreign inclusions and other phases. These defects are normally introduced during processing and fabrication steps. All these defects are capable of acting as stress raisers, and thus deleterious to parent metal's mechanical behavior. However, in some cases foreign particles are added purposefully to strengthen the parent material. The procedure is called dispersion hardening where foreign particles act as obstacles to movement of dislocations, which facilitates plastic deformation. The second-phase particles act in two distinct ways – particles are either may be cut by the dislocations or the particles resist cutting and dislocations are forced to bypass them. Strengthening due to ordered particles is responsible for the good high-temperature strength on many super-alloys. However, pores are detrimental because they reduce effective load bearing area and act as stress concentration sites.

3.2.3 Atomic vibrations

Atomic vibrations occur, even at zero temperature (a quantum mechanical effect) and increase in amplitude with temperature. In fact, the temperature of a solid is really just a measure of average vibrational activity of atoms and molecules. Vibrations displace *transiently* atoms from their regular lattice site, which destroys the perfect periodicity. In a sense, these atomic vibrations may be thought of as imperfections or defects. At room temperature, a typical vibrational frequency of atoms is of the order of 10^{13} vibrations per second, whereas the amplitude is a few thousandths of a nanometer. Many properties and processes in solids are manifestations of this vibrational atomic motion. For example: melting occurs once the atomic bonds are overcome by vigorous vibrations.

References

1. ASM handbook, Metallography and Microstructures, Vol. 9, ASM International, Materials Park, OH, 1985
2. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
3. V. Raghavan, Materials Science and Engineering, third edition, Prentice Hall of India Private Limited, New Delhi, 1990.

Chapter 4. Mechanical Properties of Metals

Most of the materials used in engineering are metallic in nature. The prime reason simply is the versatile nature of their properties those spread over a very broad range compared with other kinds of materials. Many engineering materials are subjected to forces both during processing/fabrication and in service. When a force is applied on a solid material, it may result in translation, rotation, or deformation of that material. Aspects of material translation and rotation are dealt by *engineering dynamics*. We restrict ourselves here to the subject of material deformation under forces. Deformation constitutes both change in shape, *distortion*, and change in size/volume, *dilatation*. Solid material are defined such that change in their volume under applied forces in very small, thus deformation is used as synonymous to distortion. The ability of material to with stand the applied force without any deformation is expressed in two ways, i.e. strength and hardness. Strength is defined in many ways as per the design requirements, while the hardness may be defined as resistance to indentation of scratch.

Material deformation can be permanent *or* temporary. Permanent deformation is irreversible i.e. stays even after removal of the applied forces, while the temporary deformation disappears after removal of the applied forces i.e. the deformation is recoverable. Both kinds of deformation can be function of time, or independent of time. Temporary deformation is called *elastic deformation*, while the permanent deformation is called *plastic deformation*. Time dependent recoverable deformation under load is called *anelastic deformation*, while the characteristic recovery of temporary deformation after removal of load as a function of time is called *elastic aftereffect*. Time dependent i.e. progressive permanent deformation under constant load/stress is called *creep*. For *visco-elastic* materials, both recoverable and permanent deformations occur together which are time dependent. When a material is subjected to applied forces, first the material experiences elastic deformation followed by plastic deformation. Extent of elastic- and plastic- deformations will primarily depend on the kind of material, rate of load application, ambient temperature, among other factors. Change over from elastic state to plastic state is characterized by the yield strength (σ_0) of the material.

Forces applied act on a surface of the material, and thus the force intensity, force per unit area, is used in analysis. Analogous to this, deformation is characterized by percentage change in length per unit length in three distinct directions. Force intensity is also called *engineering stress* (or simply *stress*, *s*), is given by force divided by area on which the force is acting. *Engineering strain* (or simply *strain*, *e*) is given by change in length divided by original length. Engineering strain actually indicates an average change in length in a particular direction. According to definition, *s* and *e* are given as

$$s = \frac{P}{A_0}, e = \frac{L - L_0}{L_0}$$

where *P* is the load applied over area *A*, and as a consequence of it material attains the final length *L* from its original length of L_0.

Because material dimensions changes under application of the load continuously, engineering stress and strain values are not the true indication of material deformation characteristics. Thus the need for measures of stress and strain based on instantaneous dimensions arises. Ludwik first proposed the concept of, and defined the *true strain* or *natural strain* (ε) as follows:

$$\varepsilon = \sum \frac{L_1 - L_0}{L_0} + \frac{L_2 - L_1}{L_1} + \frac{L_3 - L_2}{L_2} + \ldots$$

$$\varepsilon = \int_{L_0}^{L} \frac{dL}{L} = \ln \frac{L}{L_0}$$

As material volume is expected to be constant i.e. $A_0L_0=AL$, and thus

$$\varepsilon = \ln \frac{L}{L_0} = \ln \frac{A_0}{A} = \ln(e+1)$$

There are certain advantages of using true strain over conventional strain or engineering strain. These include (i) equivalent absolute numerical value for true strains in cases of tensile and compressive for same intuitive deformation and (ii) total true strain is equal to the sum of the incremental strains. As shown in *figure-4.1*, if L_1=2 L_0 and L_2=1/2 L_1=L_0, absolute numerical value of engineering strain during tensile deformation (1.0) is different from that during compressive deformation (0.5). However, in both cases true strain values are equal (*ln [2]*).

True stress (σ) is given as load divided by cross-sectional area over which it acts at an instant.

$$\sigma = \frac{P}{A} = \frac{P}{A_0}\frac{A_0}{A} = s(e+1)$$

It is to be noted that engineering stress is equal to true stress up to the elastic limit of the material. The same applies to the strains. After the elastic limit i.e. once material starts deforming plastically, engineering values and true values of stresses and strains differ. The above equation relating engineering and true stress-strains are valid only up to the limit of uniform deformation i.e. up to the onset of necking in tension test. This is because the relations are developed by assuming both constancy of volume and homogeneous distribution of strain along the length of the tension specimen. Basics of both elastic and plastic deformations along with their characterization will be detailed in this chapter.

4.1 Elastic deformation and Plastic deformation

4.1.1 Elastic deformation

Elastic deformation is reversible i.e. recoverable. Up to a certain limit of the applied stress, strain experienced by the material will be the kind of recoverable i.e. elastic in nature. This elastic strain is proportional to the stress applied. The proportional relation between the stress and the elastic strain is given by *Hooke's law*, which can be written as follows:

$$\sigma \propto \varepsilon$$

$$\sigma = E\varepsilon$$

where the constant E is the modulus of elasticity or Young's modulus,

Though Hooke's law is applicable to most of the engineering materials up to their elastic limit, defined by the critical value of stress beyond which plastic deformation occurs, some materials won't obey the law. E.g.: Rubber, it has nonlinear stress-strain relationship and still satisfies the definition of an elastic material. For materials without linear elastic portion, either *tangent modulus* or *secant modulus* is used in design calculations. The tangent modulus is taken as the slope of stress-strain curve at some specified level, while secant module represents the slope of secant drawn from the origin to some given point of the σ-ε curve, as shown in *figure-4.1*.

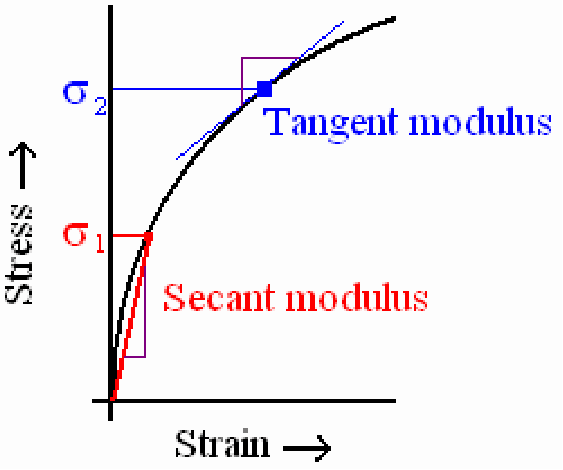


Figure-4.1: *Tangent and Secant moduli for non-linear stress-strain relation.*

If one dimension of the material changed, other dimensions of the material need to be changed to keep the volume constant. This lateral/transverse strain is related to the applied longitudinal strain by empirical means, and the ratio of transverse strain to longitudinal strain is known as *Poisson's ratio* (*ν*). Transverse strain can be expected to be opposite in nature to longitudinal strain, and both longitudinal and transverse strains are linear strains. For most metals the values of *ν* are close to 0.33, for polymers it is between 0.4 – 0.5, and for ionic solids it is around 0.2.

Stresses applied on a material can be of two kinds – normal stresses, and shear stresses. Normal stresses cause linear strains, while the shear stresses cause shear strains. If the material is subjected to torsion, it results in torsional strain. Different stresses and corresponding strains are shown in *figure-4.2*.

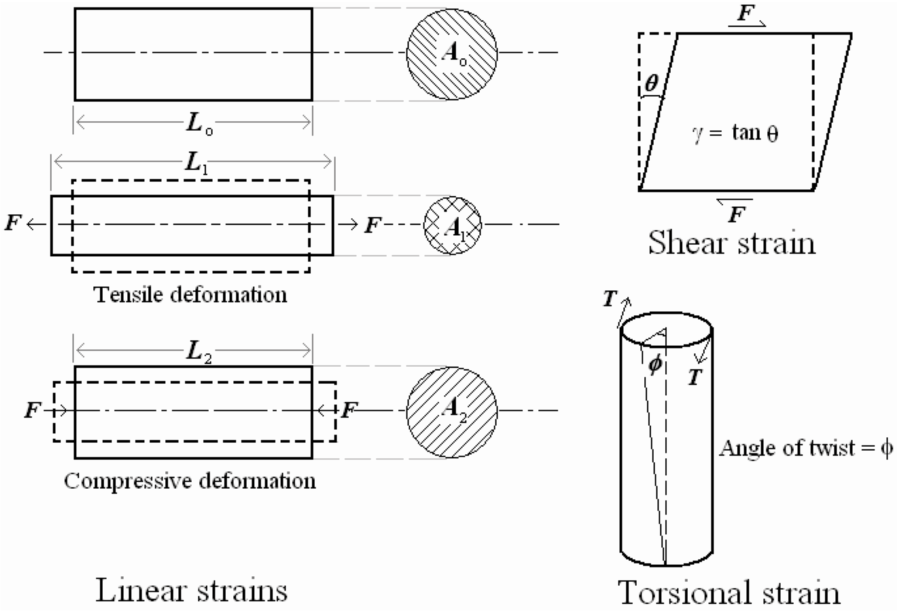


Figure 4.2: *Schematic description of different kinds of deformations/strains.*

Analogous to the relation between normal stress and linear strain defined earlier, shear stress (τ) and shear strain (γ) in elastic range are related as follows:

$$\tau = G\gamma$$

where G is known as *Shear modulus* of the material. It is also known as *modulus of elasticity in shear*. It is related with Young's modulus, E, through Poisson's ratio, ν, as

$$G = \frac{E}{2(1+\nu)}$$

Similarly, the *Bulk modulus* or *volumetric modulus of elasticity* K, of a material is defined as the ratio of hydrostatic or *mean stress* (σ_m) to the volumetric strain (Δ). The relation between E and K is given by

$$K = \frac{\sigma_m}{\Delta} = \frac{E}{3(1-2\nu)}$$

Let σ_x, σ_y *and* σ_z are linear stresses and ε_x, ε_y *and* ε_z are corresponding strains in X-, Y- and Z- directions, then

$$\sigma_m = \frac{\sigma_x + \sigma_y + \sigma_z}{3}$$

Volumetric strain or *cubical dilatation* is defined as the change in volume per unit volume.

$$\Delta = (1+\varepsilon_x)(1+\varepsilon_y)(1+\varepsilon_z) - 1 \approx \varepsilon_x + \varepsilon_y + \varepsilon_z, \ \Delta = 3\varepsilon_m$$

where ε_m is *mean strain* or *hydrostatic (spherical) strain* defined as

$$\varepsilon_m = \frac{\varepsilon_x + \varepsilon_y + \varepsilon_z}{3}$$

An engineering material is usually subjected to stresses in multiple directions than in just one direction. If a cubic element of a material is subjected to normal stresses σ_x, σ_y, and σ_z, strains in corresponding directions are given by

$$\varepsilon_x = \frac{1}{E}\left[\sigma_x - \nu(\sigma_y + \sigma_z)\right], \varepsilon_y = \frac{1}{E}\left[\sigma_y - \nu(\sigma_x + \sigma_x)\right], \text{and}\ \varepsilon_z = \frac{1}{E}\left[\sigma_z - \nu(\sigma_x + \sigma_y)\right]$$

$$\text{and } \sigma_x + \sigma_y + \sigma_z = \frac{E}{1-2\nu}(\varepsilon_x + \varepsilon_y + \varepsilon_z)$$

The strain equation can be modified as

$$\varepsilon_x = \frac{1}{E}\left[\sigma_x - \nu(\sigma_y + \sigma_z)\right] = \frac{1+\nu}{E}\sigma_x - \frac{\nu}{E}(\sigma_x + \sigma_y + \sigma_z)$$

After substituting the sum-of-stress into the above equation, stress can be related to strains as follows:

$$\sigma_x = \frac{E}{1+\nu}\varepsilon_x + \frac{\nu E}{(1+\nu)(1-2\nu)}(\varepsilon_x + \varepsilon_y + \varepsilon_z) = \frac{E}{1+\nu}\varepsilon_x + \lambda(\varepsilon_x + \varepsilon_y + \varepsilon_z) = 2G\varepsilon_x + \lambda\Delta$$

where λ is called Lame's constant.

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}$$

Using the above equations, it is possible to find strains from stresses and vice versa in elastic range.

The basis for elastic deformation is formed by reversible displacements of atoms from their equilibrium positions. On an atomic scale, elastic deformation can be viewed as small changes in the inter-atomic distances by stretching of inter-atomic bonds i.e. it involve small changes in inter-atomic distances. Elastic moduli measure the *stiffness* of the material. They are related to the *second* derivative of the inter-atomic potential, or the first derivative of the inter-atomic force vs. inter-atomic distance (*dF/dr*) (*figure-4.3*). By examining these curves we can tell which material has a higher modulus.

Elastic modulus can also be said as a measure of the resistance to separation of adjacent atoms, and is proportional to the slope of the inter-atomic fore Vs inter-atomic distance curve (*figure-4.3*). Hence values of modulus of elasticity are higher for ceramic materials which consist of strong covalent and ionic bonds. Elastic modulus values are lower for metal when compared with ceramics, and are even lower for polymers where only weak covalent bonds present.

Moreover, since the inter-atomic forces will strongly depend on the inter-atomic distance as shown in *figure-4.3*, the elastic constants will vary with direction in the crystal lattice i.e. they are anisotropic in nature for a single crystal. However, as a material consists of number of randomly oriented crystals, elastic constants of a material can be considered as isotropic.

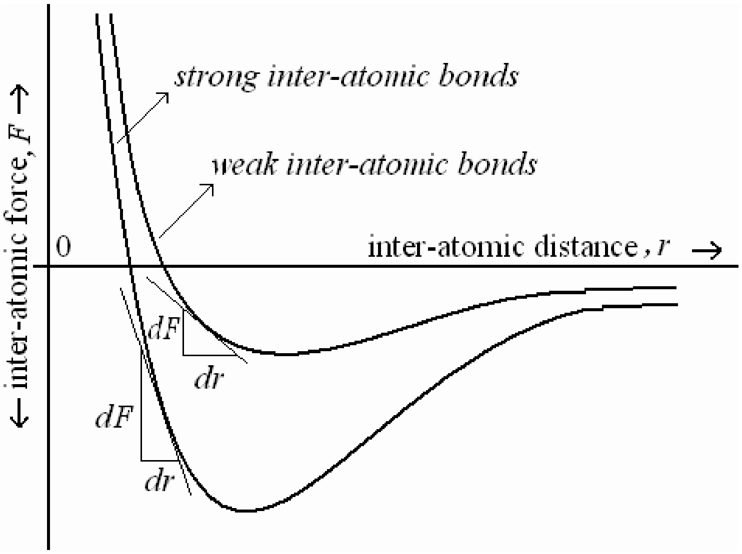


Figure-4.3: *Graph showing variation of inter-atomic forces against inter-atomic distance for both weak and strong inter-atomic bonds.*

The elastic moduli are usually measured by direct static measurements in tension or torsion tests. For more precise measurements, dynamic techniques are employed. These tests involve measurement of frequency of vibration or elapsed time for an ultrasonic pulse to travel down and back in a specimen. Because strain cycles occur at very high rates, no time for heat transfer and thus elastic constants are obtained under adiabatic conditions. Elastic modulus obtained under isothermal and adiabatic conditions are related as follows:

$$E_{adiabatic} = \frac{E_{iso}}{1 - \frac{E_{iso} T \alpha^2}{9c}}$$

where α is volume coefficient of thermal expansion, and c is the specific heat. It can be observed that with increasing temperature, the modulus of elasticity diminishes. This is because, intensity of thermal vibrations of atoms increases with temperature which weakens the inter-atomic bonds.

4.1.2 Plastic deformation

When the stress applied on a material exceeds its elastic limit, it imparts permanent non-recoverable deformation called plastic deformation in the material. Microscopically it can be said of plastic deformation involves breaking of original atomic bonds, movement of atoms and the restoration of bonds i.e. plastic deformation is based on irreversible displacements of atoms through substantial distances from their equilibrium positions. The mechanism of this deformation is different for crystalline and amorphous materials. For crystalline materials, deformation is accomplished by means of a process called slip that involves motion of dislocations. In amorphous materials, plastic deformation takes place by viscous flow mechanism in which atoms/ions slide past one another under applied stress without any directionality.

Plastic deformation is, as elastic deformation, also characterized by defining the relation between stresses and the corresponding strains. However, the relation isn't simpler as in case of elastic deformation, and in fact it is much more complex. It is because plastic deformation is accomplished by substantial movement of atomic planes, dislocations which may encounter various obstacles. This movement becomes more complex as number slip systems may get activated during the deformation.

The analysis of plastic deformation, and the large plastic strains involved is important in many manufacturing processes, especially forming processes. It is very difficult to describe the behavior of metals under complex conditions. Therefore, certain simplifying assumptions are usually necessary to obtain an amenable mathematical solution. Important assumptions, thus, involved in theory of plasticity are neglecting the following aspects:

(i) Anelastic strain, which is time dependent recoverable strain.

(ii) Hysteresis behavior resulting from loading and un-loading of material.

(iii) Bauschinger effect – dependence of yield stress on loading path and direction.

The relations describing the state of stress and strain are called *constitutive equations* because they depend on the material behavior. These relations are applicable to any material whether it is elastic, plastic or elastic-plastic. Hooke's law which states that strain is proportional to applied stress is applicable in elastic range where deformation is considered to be uniform. However, plastic deformation is indeed uniform *but* only up to some extent of strain value, where after plastic deformation is concentrated the phenomenon called *necking*. The change over from uniform plastic deformation to non-uniform plastic deformation is characterized by ultimate tensile strength (σ_u). As a result of complex mechanism involved in plastic deformation and its non-uniform distribution before material fractures, many functional relations have been proposed to quantify the stress-strain relations in plastic range.

A true stress-strain relation plotted as a curve is known as *flow curve* because it gives the stress (σ) required to cause the material to flow plastically to any given extent of strain (ε) under a set of conditions. Other important parameters affecting the stress-strain curve are: rate at which the load is applied / strain rate ($\dot{\varepsilon}$), and the temperature of the material (T, in K) i.e.

$$\sigma = fn(\varepsilon, \dot{\varepsilon}, T, microstructure)$$

Following are the most common equations that describe the material flow behavior:

$$\sigma = K\varepsilon^n$$

where *K* – is strength coefficient, and *n* – is strain hardening exponent. The strain-hardening exponent may have values from *n=0* (perfectly plastic solid) to *n=1* (elastic solid). For most metals n has values between 0.10 and 0.50. The *Power law equation* described above is also known as *Holloman-Ludwig equation*. Another kind of power equation which describes the material behavior when strain rate effect is prominent:

$$\sigma = K\dot{\varepsilon}^m$$

where *m* – is the index of strain-rate sensitivity. If *m*=0, the stress is independent of strain rate. *m*=0.2 for common metals. If *m*=0.4-0.9, the material may exhibit super-plastic behavior – ability to deform by several hundred percent of strain without necking. If *m*=1, the material behaves like a viscous liquid and exhibits Newtonian flow.

Deviations from the above power equations are frequently observed, especially at low strains ($\varepsilon<10^{-3}$) *or* high strains ($\varepsilon>>1,0$). One common type of deviation is for a log-log plot of Power equation to result in two straight lines with different slopes. For data which do not follow the Power equation, following equation could be used

$$\sigma = K(\varepsilon_0 + \varepsilon)^n$$

where ε_0 - is the strain material had under gone before the present characterization.

Another common variation from Power law equation, also known as *Ludwig equation*, is:

$$\sigma = \sigma_o + K\varepsilon^n$$

where σ_o - is the yield strength of the material.

Many other expressions for flow curve are available in literature. The true strain used in the above equations should actually be the plastic strain value, given by the following equation, where elastic strains can be safely neglected because plastic strains are very much when compared with elastic strain values.

$$\varepsilon_p = \varepsilon - \varepsilon_e = \varepsilon - \frac{\sigma_0}{E} \approx \varepsilon$$

Both temporary elastic deformation and permanent plastic deformation are compared in *tabel-4.1.*

Table-4.1: *Elastic deformation Vs Plastic deformation.*

Elastic deformation	Plastic deformation
Reversible	Not reversible
Depends on initial and final states of	Depends on loading path

stress and strain	
Stress is proportional to strain	No simple relation between stress and strain
No strain hardening effects	Strain hardening effects

4.2 Interpretation of tensile stress-strain curves

It is well known that material deforms under applied loads, and this deformation can be characterized by stress-strain relations. The stress-strain relation for a material is usually obtained experimentally. Many kinds of experiments those differ in way of loading the material are standardized. These include tension test, compression test (upsetting), plane strain compression test, torsion test, etc.

The engineering tension test is commonly used to provide basic design information on the strength characteristics of a material. Standardized test procedure is explained by ASTM standard E0008-04. In this test a specimen is subjected to a continually increasing uni-axial tensile force while measuring elongation simultaneously. A typical plot of load-elongation is given in the *figure-4.4*. The curve also assumes the shape of an engineering stress – engineering strain curve after dividing the load with initial area and the elongation with initial length. The two curves are frequently used interchangeably as they differ only by constant factors. The shape and relative size of the engineering stress-strain curve depends on material composition, heat treatment, prior history of plastic deformation, strain rate, temperature and state of stress imposed on specimen during the test.

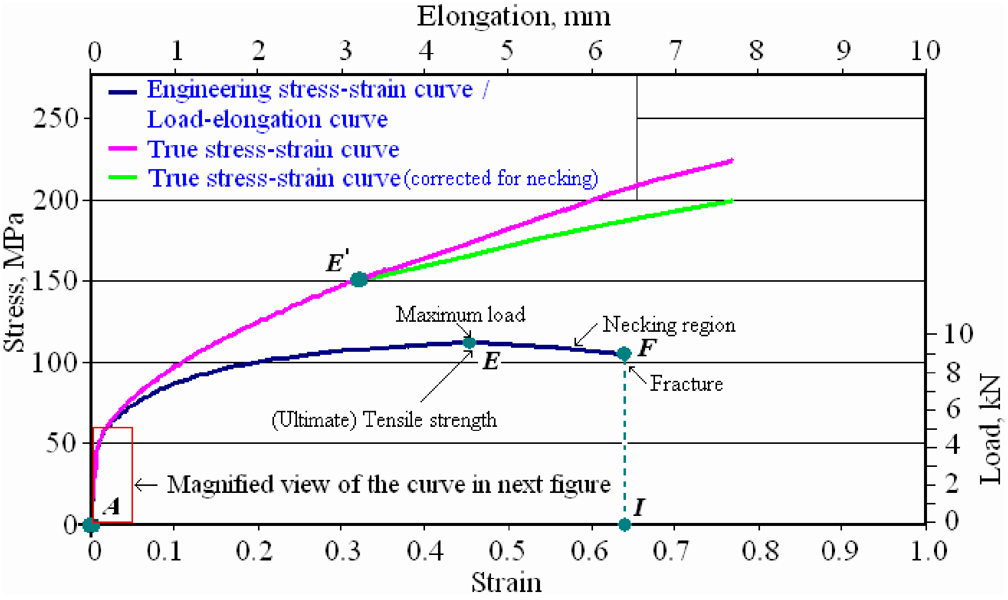


Figure-4.4: *Typical load – elongation / engineering stress – engineering strain / true stress – true strain curve.*

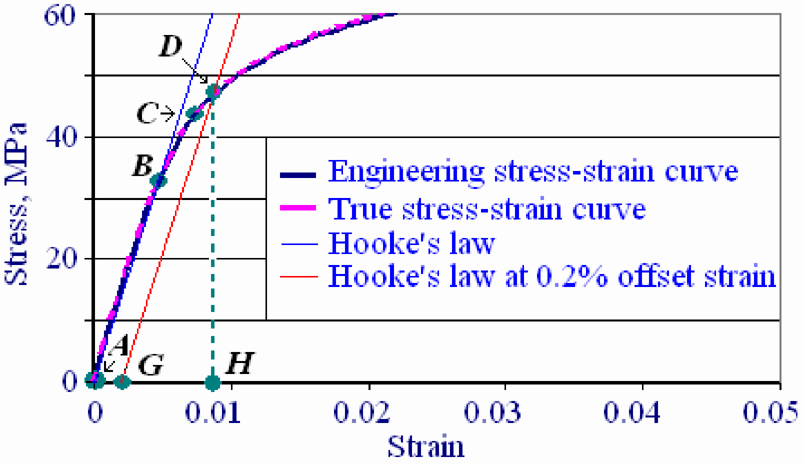


Figure-4.5: *Magnified view of initial part of stress-strain curve.*

As shown in *figure-4.4*, at initial stages load is proportional to elongation to a certain level (elastic limit), and then increases with elongation to a maximum (uniform plastic deformation), followed by decrease in load due to necking (non-uniform plastic deformation) before fracture of the specimen occurs. Magnified view of the initial stage of the curve is shown in *figure-4.5*. Along the segment *AB* of the curve, engineering stress is proportional to engineering strain as defined by Hooke's law, thus *point-B* is known as *proportional limit*. Slope of the line AB gives the elastic modulus of the material. With further increase in stress up to *point-C*, material can still be elastic in nature. Hence *point-C* is known as *elastic limit*. It is to be noted that there is no point where exactly material starts deform plastically i.e. there is no sharp point to indicate start of the yield. Otherwise *point-C* can also be called yield point. Thus it is common to assume that stress value at 0.2% offset strain as *yield strength* (σ_0), denoted by *point-D*. This offset yield strength is also called *proof stress*. Proof stress is used in design as it avoids the difficulties in measuring proportional or elastic limit. For some materials where there is essentially no initial linear portion, offset strain of 0.5% is frequently used. In *figure-4.5* distance between *points-B, C and D* is exaggerated for clarity. For many materials it is difficult to make any difference between *points-B and C*, and *point-D* also coincides with *point-B/C*. After the yield point, *point-D*, stress reaches a maximum at *point-E* (*tensile strength* (σ_f) as *figure-4.4*) till where plastic deformation is uniform along the length of the specimen. Stress decreases hereafter because of onset of necking that result in non-uniform plastic deformation before specimen fractures at *point-F*, *fracture limit*. Yield strength and tensile strength are the parameters that describe the material's strength, while percent elongation and reduction in cross-sectional area are used to indicate the material's ductility – extent of material deformation under applied load before fracture.

Percent elongation (*e*) and reduction is area (*r*) are related as follows:

Volume constancy $A_0 L_0 = AL$

Percent elongation $e = \frac{L - L_0}{L_0}$

Reduction in area $r = \frac{A_0 - A}{A_0}, \ \frac{A_0}{A} = \frac{1}{1 - r}$

Relation $e = \frac{L - L_0}{L_0} = \frac{L}{L_0} - 1 = \frac{A_0}{A} - 1 = \frac{1}{1 - r} - 1 = \frac{r}{1 - r}$

Other important parameters from the engineering stress-strain curve are – *resilience* and *toughness*. Resilience is defined as ability of a material to absorb energy when deformed elastically and to return it when unloaded. Approximately this is equal to the area under elastic part of the stress-strain curve, and equal to *area-ADH* in *figure-4.5*. This is measured in terms of modulus of resilience (U_r)– strain energy per unit volume required to stress the material from zero stress to yield stress ($\sigma_0 = s_0$).

$$U_r = \frac{1}{2} s_0 e_0 = \frac{1}{2} s_0 \frac{s_0}{E} = \frac{s_0^2}{2E}$$

where e_0 – it elastic strain limit. Toughness (U_t) of the materials is defined as its ability to absorb energy in the plastic range. In other terms, it can be said to equal to work per unit volume which can be done on the material without causing it to rupture. It is considered that toughness of a material gives an idea about both strength and ductility of that material. Experimentally toughness is measured by either *Charpy* or *Izod impact tests*. Numerically the toughness value is equal to the area under the stress-strain curve, *area-AEFI*, approximately of rectangular shape.

$$U_t \approx s_u e_f \approx \frac{s_0 + s_u}{2} e_f$$

For brittle materials (those have high elastic modulus and low ductility), stress-strain curve is considered to assume the shape of parabola, thus

$$U_t \approx \frac{2}{3} s_u e_f$$

where s_u – ultimate tensile strength and e_f – strain at fracture.

As explained in earlier section, engineering stress-strain curve is not true representative of the material behavior. But, the flow curve (true stress-true strain curve) represents the basic plastic-flow characteristics of the material. More upon, special feature of the flow curve is that any point on the curve can be considered as yield point i.e. if load is removed and then reapplied, material will behave elastically throughout the entire range of reloading. It can be said from the relations between engineering stress-strain and true stress-strains the true stress-true strain curve is always to the left of the engineering curve until the maximum load is reached. *Point-E'* on represents the corresponding location on true stress-strain curve to ultimate tensile stress *point-E* on engineering stress-strain curve. After the *point-E'*, flow curve is usually linear up to fracture, and in some cases its slope decreases continually up to fracture.

In correspondence to different variables defined based on engineering stress-strain curve, following parameters are defined based on flow curve:

True stress at maximum load $\sigma_u = \frac{P_{max}}{A_u}, s_u = \frac{P_{max}}{A_o}, \varepsilon_u = \ln\frac{A_0}{A_u} \Rightarrow \sigma_u = s_u\frac{A_0}{A_u} = s_u e^{\varepsilon_u}$

True fracture strain $\varepsilon_f = \ln\frac{A_0}{A_f} = \ln\frac{1}{1-r}$

It is not possible to calculate ε_f from estimated values of e_f because the relation between them is not valid beyond the onset of necking.

True uniform strain $\varepsilon_u = \ln\frac{A_0}{A_u}$

The uniform strain is useful in estimating the formability of metals from the results of a tension test. The condition '$\varepsilon_u = n$' represents the onset of necking.

True local necking strain $\varepsilon_n = \ln\frac{A_u}{A_f}$

It is important to note that the rate of strain hardening ($d\sigma/d\varepsilon$) is not identical to the strain-hardening exponent (n). They differ as follows:

$$n = \frac{d(\log\sigma)}{d(\log\varepsilon)} = \frac{d(\ln\sigma)}{d(\ln\varepsilon)} = \frac{\varepsilon}{\sigma}\frac{d\sigma}{d\varepsilon} \qquad or \qquad \frac{d\sigma}{d\varepsilon} = n\frac{\sigma}{\varepsilon}$$

4.3 Yielding under multi-axial stress, Yield criteria, Macroscopic aspects of plastic deformation and Property variability & Design considerations

4.3.1 Yielding under multi-axial stress

Once the necking starts to form i.e. material starts to deform plastically but in non-uniform mode, uni-axial state of stress turns into multi-axial (tri-axial) stress state. The necked region is in effect a mild notch. The chief effect of the notch is not in introducing a stress concentration but in producing a tri-axial state of stress i.e. introduction of transverse stresses. As a result of tri-axial state of stress, yield stress becomes greater than the uni-axial yield stress, σ_0, because it is more difficult to spread the yielded zone in the presence of tri-axial stresses. Thus, the average true stress at the neck is higher than the stress which would be required to cause flow in simple tension prevailed.

Bridgman put forwarded mathematical analysis to calculate the true stress from measured stress in axial direction under tri-axial stress condition. His analysis is based on the following assumptions: counter of the neck is approximated by the arc of a circle; cross-section of the necked region remains circular; von Mises yield criterion is applicable; strains are constant over the cross-section of the neck. Bridgman's correction is applicable from the onset of necking for flow curve i.e. from *point-E'* shown in *figure-4.4*. Corrected yield stress under tri-axial state of stress is given as follows:

$$\sigma = \frac{(\sigma_x)_{avg}}{(1+2R/a)\left[\ln(1+a/2R)\right]}$$

where $(\sigma_x)_{avg}$ measured stress in the axial direction, a – smallest radius in the neck region, R – radius of the curvature of neck (*figure-4.6*).

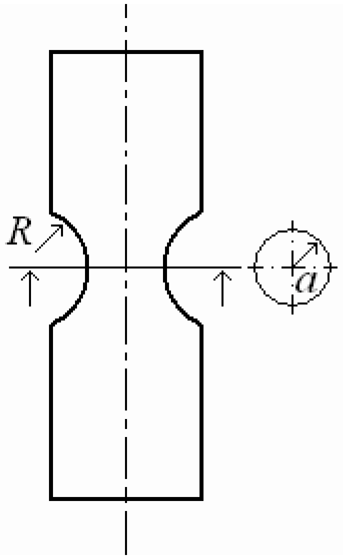


Figure-4.6: *Geometry of necked region in cylindrical specimen under tensile load.*

4.3.2 Yield criteria

It is known that material yields under condition of applied stress(es). In uni-axial loading, like in a tension test, yield occurs i.e. macroscopic plastic flow starts at yield stress, σ_0. However the situation is much more complicated in presence of multi-axial stresses. It could be expected that yielding condition under multi-axial stresses to be a function of particular combination of principal stresses.

Presently available yield criteria are essentially empirical relations. Thus, it needs to satisfy some experimental observations. These include: hydrostatic component of stress state should not influence the stress where the yield occurs; it must be an invariant function i.e. independent of the choice of axes. These lead to the statement that yield criterion must be some function of invariants of stress deviator.

There are two generally accepted criteria are in use for predicting the onset of yielding in ductile materials: *von Mises* or *Distortion-energy criterion* and *Maximum-shear-stress* or *Tresca criterion*.

von Mises *or* **Distortion-energy criterion**: it states that yielding occur when the second invariant of the stress deviator J_2 exceeded some critical value.

$$J_2 = k^2$$

where $J_2 = \frac{1}{6}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right]$; σ_1, σ_2 *and* σ_3 are principal stresses.

It implies that yield condition in not dependent on any particular stress, but instead it depends on all three principal stresses. And as the criterion is based on differences of normal stresses, it is independent of hydrostatic stress component. In energy terms, it can be said that yielding occurs when the *distortion energy* reaches a critical value. Distortion energy is part of total strain energy per unit volume that is involved in change of shape as opposed to a change in volume.

To evaluate the constant *k*, let's consider the yield in uni-axial tension test i.e. $\sigma_1 = \sigma_0$, $\sigma_2 = \sigma_3 = 0$. Thus,

$$\frac{1}{6}(\sigma_0^2 + \sigma_0^2) = k^2 \Rightarrow \sigma_0 = \sqrt{3}k$$

$$\Rightarrow \sigma_0 = \frac{1}{\sqrt{2}}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right]^{1/2}$$

To identify the constant *k*, by considering the state of stress in pure shear (torsion test): σ_1 =- σ_3 = τ, σ_2 = 0.

$$\frac{1}{6}(\sigma_1^2 + \sigma_1^2 + 4\sigma_1^2) = k^2 \Rightarrow \sigma_1 = k$$

i.e. *k* represents the yield stress under pure shear, whereas σ_0 represents the yield stress under uni-axial tension. These two yield stresses can be related as follows:

$$k = \frac{1}{\sqrt{3}}\sigma_0 = 0.577\sigma_0$$

von Mises yield criterion can also be interpreted as yielding occurs if octahedral shear stress reaches a critical value. This is shear stress on octahedral plane which makes equal angles with all three principal axes. It also represents the mean square of the shear stress averaged over all orientations in the solid.

Cosine of angle between normal to a face of octahedron and a nearest principal axis is 1/√3, i.e. the angle is 54 44’. Normal octahedral stress (σ_{oct}) is equivalent to hydrostatic component of the stress system. Thus it can result in yielding, but shear octahedral stress (τ_{oct}) do.

$$\sigma_{oct} = \sigma_m = \frac{\sigma_1 + \sigma_2 + \sigma_3}{3}$$

$$\tau_{oct} = \frac{1}{3}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right]^{1/2}$$

$$\Rightarrow \tau_{oct} = \frac{\sqrt{2}}{3}\sigma_0 = 0.471\sigma_0$$

Corresponding octahedral strains are given as follows:

$$\varepsilon_{oct} = \frac{\varepsilon_1 + \varepsilon_2 + \varepsilon_3}{3}, \gamma = \frac{2}{3}\left[(\varepsilon_1 - \varepsilon_2)^2 + (\varepsilon_2 - \varepsilon_3)^2 + (\varepsilon_3 - \varepsilon_1)^2\right]^{1/2}$$

Maximum-shear-stress *or* **Tresca criterion**: this criterion states that yielding occurs once the maximum shear stress of the stress system reaches the value of shear stress in uni-axial tension test.

If σ_1, σ_2 *and* σ_3 are principal stresses arranged in descending order, maximum shear stress is given as

$$\tau_{\max} = \frac{\sigma_1 - \sigma_3}{2}$$

Though this criterion is simpler than the von Mises yield criterion, here it is necessary to know before-hand the maximum and minimum principal stresses from the stress system.

As with the earlier criterion, under uni-axial tension test conditions ($\sigma_1 = \sigma_0, \sigma_2 = \sigma_3 = 0$),

$$\tau_{\max} = \frac{\sigma_1 - \sigma_3}{2} = \tau_0 = \frac{\sigma_0}{2} \Rightarrow \sigma_1 - \sigma_3 = \sigma_0$$

Under pure shear stress conditions ($\sigma_1 = -\sigma_3 = k, \sigma_2 = 0$),

$$k = \frac{\sigma_1 - \sigma_3}{2} = \frac{1}{2}\sigma_0$$

Thus the other mathematical form for the criterion can be

$$\sigma_1 - \sigma_3 = \sigma_1^{'} - \sigma_3^{1} = k$$

Note that the yield locus for the Tresca criterion falls inside of the von Mises yield ellipse. The two yielding criteria predict the same yield stress for conditions of uni-axial stress and balanced bi-axial stress ($\sigma_1 = \sigma_3$). The greatest divergence between the two criteria occurs for pure shear ($\sigma_1 = - \sigma_3$) where yield stress from von Mises criterion ($\frac{1}{\sqrt{3}}\sigma_0$) is 15.5% greater than the yield stress from Tresca criterion ($\frac{1}{2}\sigma_0$).

4.3.3 Macroscopic aspects of plastic deformation

In single crystal, plastic deformation is accomplished by the process called slip, and sometimes by twinning. The extent of slip depends on many factors including external load and the corresponding value of shear stress produced by it, the geometry of crystal structure, and the orientation of active slip planes with the direction of shearing stresses generated. Schmid first recognized that single crystals at different orientations but of same material require different stresses to produce slip. The dependence of various factors has been summarized using a parameter – critical resolved shear stress, τ_R, given as

$$\tau = \frac{P\cos\lambda}{A/\cos\varphi} = \frac{P}{A}\cos\varphi\cos\lambda = \sigma\cos\varphi\cos\lambda$$

$$\Rightarrow m = \cos\varphi\cos\lambda$$

where P – external load applied, A – cross-sectional area over which the load applied, λ – angle between slip direction and tensile axis, ϕ – angle between normal to the slip plane and the tensile axis and m – Schmid factor.

Shear stress is maximum for the condition where $\lambda = \phi = 45°$. If either of the angles are equal to 90°, resolved shear stress will be zero, and thus no slip occurs. If the conditions are such that either of the angles is close to 90°, crystal will tend to fracture rather than slip. Single crystal metals and alloys are used mainly for research purpose and only in a few cases of engineering applications.

Almost all engineering alloys are polycrystalline. Gross plastic deformation of a polycrystalline specimen corresponds to the comparable distortion of the individual grains by means of slip. This involves generation, movement and (re-)arrangement of dislocations. During deformation, mechanical integrity and coherency are maintained along the grain boundaries; that is, the grain boundaries are constrained, to some degree, in the shape it may assume by its neighboring grains. Let's assume that the grains are equi-axed before deformation, or have approximately the same dimension in all directions. If the material is imparted with deformation in a particular direction, the grains become elongated along the direction in which the specimen was extended, as shown in *figure-4.7*.

Another noticeable aspect of plastic deformation is arrangement of dislocations forming cell like configurations with clear areas in the centers of the cells. With increasing deformation, the cell structure becomes denser. In early stages of deformation, slip is essentially confined to primary glide planes and thus dislocations form coplanar arrays. As deformation proceeds, cross slip comes into picture, leading to tangled network of dislocations forming cell walls, *figure-4.8*. Cell size decreases with increase in deformation but reaches a fixed size depending on the material, the strain, the strain rate and the temperature of deformation. In materials with low stacking fault energy where cross-slip id difficult, development of cell structure is less pronounced.

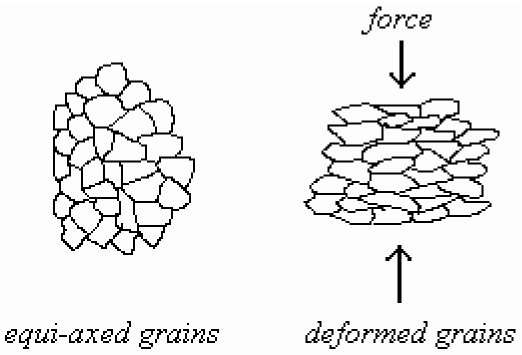


Figure-4.7: *Grain elongation under applied external forces.*

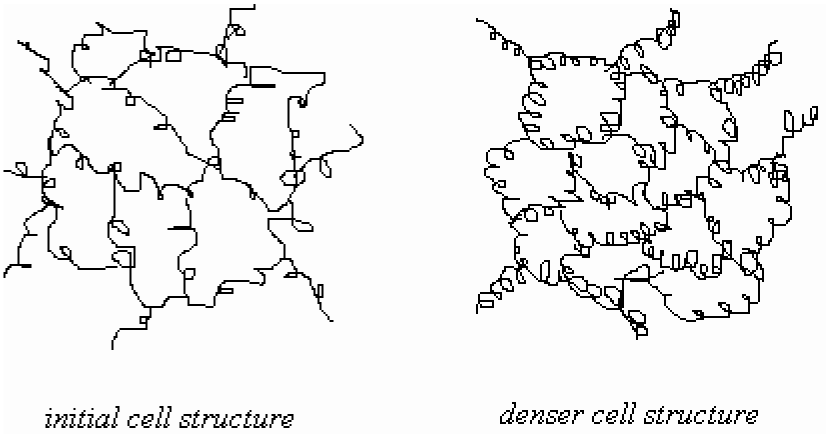


Figure-4.8: *Schematic view of cell structure.*

Schematic presentation of macroscopic observation of a tensile specimen is shown in *figure-4.9*. As explained earlier, initially specimen undergoes elastic deformation, followed by uniform plastic deformation. After attaining maximum load capacity, non-uniform plastic deformation sets-in. this leads to formation of macroscopic defects, like cracks, which causes eventual fracture failure of the sample.

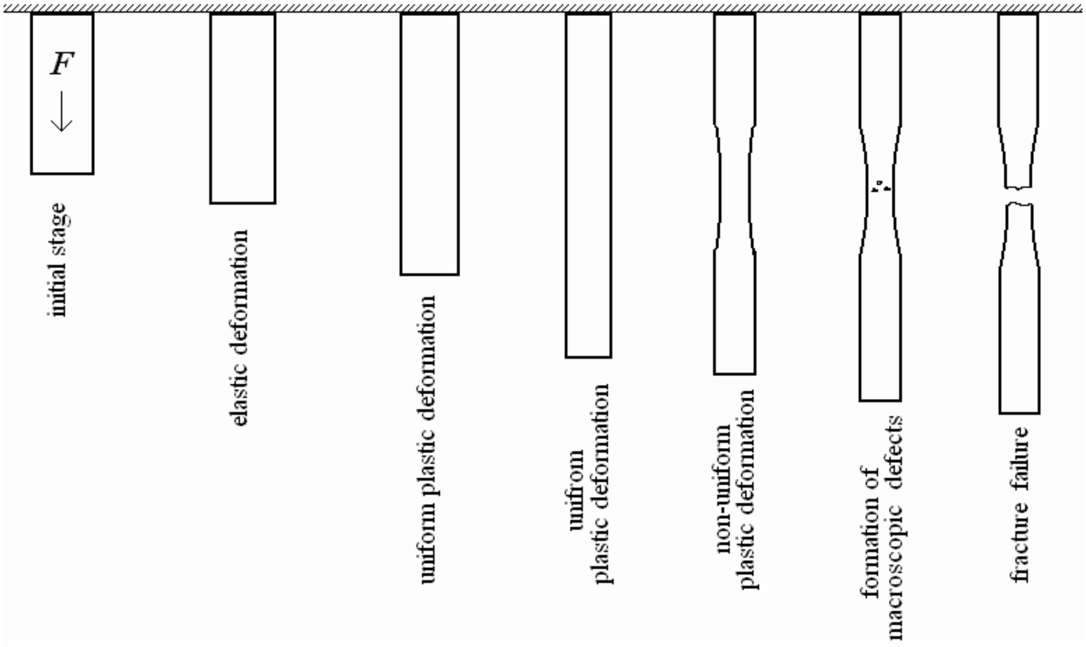


Figure-4.9: *Schematic macroscopic presentation of plastic deformation in tensile specimen.*

4.3.4 Property variability & Design consideration

Scatter in measured properties of engineering materials is inevitable because of number of factors such as test method, specimen fabrication procedure, operator bias, apparatus calibration etc. In spite of property variation, some typical value is desirable. Most commonly used typical value is by taking an average of the data. The average ($\bar{x}$) of the a parameter, x_i, is given as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where n is the number of samples, and x_i is the discrete measurement.

In some instances, it is desirable to have an idea about the degree of variability, scatter, of the measured data. Most common measure of this degree of variability is the standard deviation, s, which is given by

$$s = \left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \right]^{1/2}$$

A large value for s means a high degree of scatter. Scatter is usually represented in graphical form using error bars. If a parameter is averaged to $\bar{x}$, and the corresponding standard deviation is *s*, the upper error bar limit is given by ($\bar{x}$ +*s*), while the lower error bar is equal to ($\bar{x}$ -*s*).

To account for variability of properties, designers use, instead of an average value of, say, the tensile strength, the probability that the yield strength is above the minimum value tolerable. This leads to the use of a *safety factor* $N > 1$. Thus, a working stress or safe stress would be $\sigma_w = \sigma_y / N$.

Alternatively, a *design factor* (*N'*) is defined to account for allowances thus protect against unanticipated failure. If calculated yield stress is σ_c, then the design stress σ_d can be given as $\sigma_d = N'\sigma_c$, where *N'* is greater than unity. Thus the material to be used for a particular application is chosen so as to have strength at least as high as the value of σ_d.

Utilization of design stress is usually preferred since it is based on the anticipated maximum applied stress instead of the yield strength of the material. In design it is important to use an appropriate value for *N*. In general, values for factor of safety range from 1.2 to 4.0. Higher the value of *N*, lesser will be the efficiency of design or material usage i.e. either too much material or a material having a higher than necessary strength will be used.

Selection of *N* will depend on a number of factors, including economics, previous experience, the accuracy with which mechanical forces and material properties may be determined and most importantly on the consequences of failure in terms of loss of life or property damage. If failure would result in loss of life, the factor of safety should be increased. In military equipment where light weight may be a prime consideration, the factor of safety may be lower than in commercial equipment. For static loading, as in a building, the factor of safety would be lower than in a machine, which is subjected to vibration and fluctuating stresses.

References

1. G. E. Dieter, Mechanical Metallurgy, Third Edition, McGraw-Hill, New York, 1986.
2. M. F. Ashby and D. R. H. Jones, Engineering Materials 1, An introduction to Their Properties and Applications, second edition, Butterworth-Heinemann, Woburn, UK, 1996
3. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.

Chapter 5. Diffusion

Diffusion is the process by which atoms move in a material. Many reactions in solids and liquids are diffusion dependent. Structural control in a solid to achieve the optimum properties is also dependent on the rate of diffusion.

Atoms are able to move throughout solids because they are not stationary but execute rapid, small-amplitude vibrations about their equilibrium positions. Such vibrations increase with temperature and at any temperature a very small fraction of atoms has sufficient amplitude to move from one atomic position to an adjacent one. The fraction of atoms possessing this amplitude increases markedly with rising temperature. In jumping from one equilibrium position to another, an atom passes through a higher energy state since atomic bonds are distorted and broken, and the increase in energy is supplied by thermal vibrations. As might be expected defects, especially vacancies, are quite instrumental in affecting the diffusion process on the type and number of defects that are present, as well as the thermal vibrations of atoms.

Diffusion can be *defined* as the mass flow process in which atoms change their positions relative to neighbors in a given phase under the influence of thermal and a gradient. The gradient can be a compositional gradient, an electric or magnetic gradient, or stress gradient. In this chapter we discuss diffusion because of concentration gradient only.

5.1 Diffusion mechanisms, Steady-state and Non-steady-state diffusion

5.1.1 Diffusion mechanisms

In pure metals self-diffusion occurs where there is no net mass transport, but atoms migrate in a random manner throughout the crystal. In alloys inter-diffusion takes place where the mass transport almost always occurs so as to minimize compositional differences. Various atomic mechanisms for self-diffusion and inter-diffusion have been proposed. *Figure-5.1* presents schematic view of different atomic diffusion mechanisms.

The most energetically favorable process involves an interchange of places by an atom and a neighboring vacancy – *vacancy diffusion*. This process demands not only the motion of vacancies, but also the presence of vacancies. The unit step in vacancy diffusion is an atom breaks its bonds and jumps into neighboring vacant site. In *interstitial diffusion*, solute atoms which are small enough to occupy interstitial sites diffuse by jumping from one interstitial site to another. The unit step here involves jump of the diffusing atom from one interstitial site to a neighboring site. Hydrogen, Carbon, Nitrogen and Oxygen diffuse interstitially in most metals, and the activation energy for diffusion is only that associated with motion since the number of occupied, adjacent interstitial sites usually is large. *Substitutional diffusion* generally proceeds by the vacancy mechanism. Thus interstitial diffusion is faster than substitutional diffusion by the vacancy mechanism. During *self-diffusion* or *ring mechanism* or *direct-exchange mechanism*, three or four atoms in the form of a ring move simultaneously round the ring, thereby interchanging their positions. This mechanism is untenable because exceptionally high activation energy would be required. A *self-interstitial* is more mobile than a vacancy as only small activation energy is required for self-interstitial atom to move to an equilibrium atomic position and simultaneously displace the neighboring atom into an interstitial site. However, the equilibrium number of self-interstitial atoms present at any temperature is negligible in comparison to the number of vacancies. This is because the energy to form a self-interstitial is extremely large.

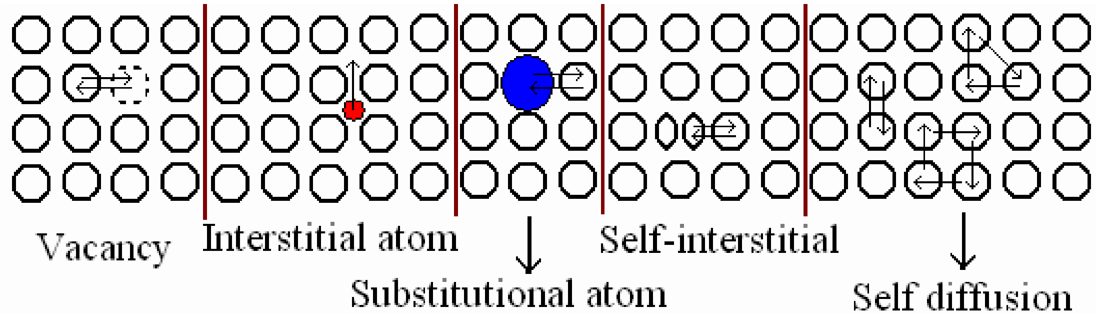


Figure-5.1: *Diffusion mechanisms.*

Diffusion in most ionic solids occurs by a vacancy mechanism. In ionic crystals, Schottky and Frankel defects assist the diffusion process. When Frenkel defects (pair of vacancy-interstial) dominate in an ionic crystal, the cation interstitial of the Frenkel defect carries the diffusion flux. If Schottky defects (pair of vacant sites) dominate, the cation vacancy carries the diffusion flux. In thermal equilibrium, in addition to above defects, ionic crystal may have defects generated by impurities and by deviation from stochiometry. Thus imperfections in ionic materials that influence diffusion arise in two ways: (1) intrinsic point defects such as Frenkel and schottky defects whose number depends on temperature, and (2) extrinsic point defects whose presence is due to impurity ions of different valance than the host ions. The former is responsible for temperature dependence of diffusion similar to that for self-diffusion in metals, while the latter result in a temperature dependence of diffusion which is similar to that for interstitial solute diffusion in metals.

For example: Cd^{+2} cation in NaCl crystal will results in a cation vacancy. As Schottky defects form easily in NaCl crystal and thus cation vacancies carry the diffusion flux, even small fraction of Cd^{+2} increases the diffusivity of NaCl by several orders. Excess Zn^{+2} interstitials present in a non-stoichiometric ZnO compound increase the diffusivity of Zn^{+2} ions significantly. It is same with non-stoichiometric FeO.

In addition to diffusion through the bulk of a solid (volume diffusion), atoms may migrate along external or internal paths that afford lower energy barriers to motion. Thus diffusion can occur along dislocations, grain boundaries or external surfaces. The rates of diffusion along such *short-circuit paths* are significantly higher than for volume diffusion. However, most cases of mass transport are due to volume diffusion because the effective cross-sectional areas available for short-circuit processes are much smaller than those for volume diffusion.

There is a difference between diffusion and net diffusion. In a homogeneous material, atoms also diffuse but this motion is hard to detect. This is because atoms move randomly and there will be an equal number of atoms moving in one direction than in another. In inhomogeneous materials, the effect of diffusion is readily seen by a change in concentration with time. In this case there is a *net* diffusion. Net diffusion occurs because, although all atoms are moving randomly, there are more atoms moving away from regions where their concentration is higher.

5.1.2 Steady-state diffusion

Diffusional processes can be either steady-state or non-steady-state. These two types of diffusion processes are distinguished by use of a parameter called flux. It is defined as net number of atoms crossing a unit area perpendicular to a given direction per unit time. For steady-state diffusion, flux is constant with time, whereas for non-steady-state diffusion, flux varies with time. A schematic view of concentration gradient with distance for both steady-state and non-steady-state diffusion processes are shown in *figure-5.2*.

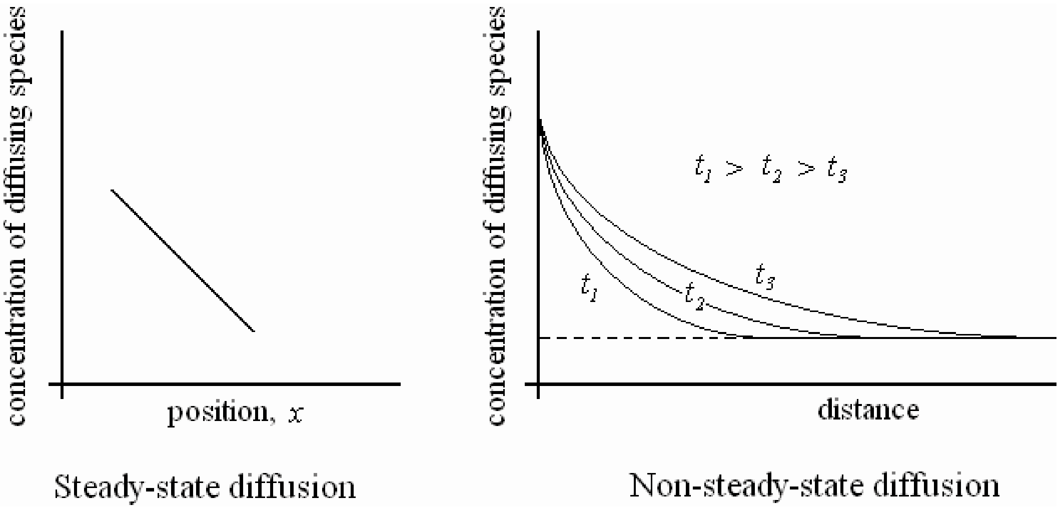


Figure-5.2: *Steady-state and Non-steady-state diffusion processes.*

Steady-state diffusion is described by Fick's first law which states that flux, *J*, is proportional to the concentration gradient. The constant of proportionality is called diffusion coefficient (diffusivity), *D* (cm^2/sec). diffusivity is characteristic of the system and depends on the nature of the diffusing species, the matrix in which it is diffusing, and the temperature at which diffusion occurs. Thus under steady-state flow, the flux is independent of time and remains the same at any cross-sectional plane along the diffusion direction. for the one-dimensional case, *Fick's first law* is given by

$$J_x = -D\frac{dc}{dx} = \frac{1}{A}\frac{dn}{dt}$$

and

$$J_x \neq f(x,t)$$

where *D* is the diffusion constant, *dc/dx* is the gradient of the concentration *c*, *dn/dt* is the number atoms crossing per unit time a cross-sectional plane of area *A*. The minus sign in the equation means that diffusion occurs down the concentration gradient. Although, the concentration gradient is often called the *driving force* for diffusion (but it is not a force in the mechanistic sense), it is more correct to consider the reduction in total free energy as the driving force.

An example of steady-state diffusion is provided by the permeation of hydrogen atoms through a sheet of palladium with different imposed hydrogen gas pressures on either side the slab. This process has been used to purify the hydrogen gas as other gases like nitrogen, oxygen and water vapor can not diffuse through palladium.

5.1.3 Non-steady-state diffusion

Most interesting cases of diffusion are non-steady-state processes since the concentration at a given position changes with time, and thus the flux changes with time. This is the case when the diffusion flux depends on time, which means that a type of atoms accumulates in a region or depleted from a region (which may cause them to accumulate in another region). *Fick's second law* characterizes these processes, which is expressed as:

$$\frac{dc}{dt} = -\frac{dJ}{dx} = \frac{d}{dx}\left(D\frac{dc}{dx}\right)$$

where *dc/dt* is the time rate of change of concentration at a particular position, *x*. If *D* is assumed to be a constant, then

$$\frac{dc}{dt} = D\frac{d^2c}{dx^2}$$

Solution to the above expression is possible when meaningful boundary conditions are specified. One common set of boundary conditions can be written as

For $t = 0$, $C = C_0$ at $0 \le x \le \infty$

For $t > 0$, $C = C_s$ at $x=0$
$C = C_0$ at $x = \infty$

And the solution is

$$\frac{C_x - C_0}{C_s - C_0} = 1 - erf\left(\frac{x}{2\sqrt{Dt}}\right)$$

where C_x represents the concentration at depth x after time t. The term *erf* stands for Gaussian error function. Corresponding error function values for a variable are usually found from standard mathematical tables. The above equation demonstrates the relationship between concentration, position, and time. Thus the equation can be used to explain many practical industrial problems like corrosion resistance of duralumin, carburization and de-carburization of steel, doping of semi-conductors, etc.

5.2 Factors that influence diffusion and Non-equilibrium transformation & microstructure

5.2.1 Factors that influence diffusion

Ease of a diffusion process is characterized by the parameter *D*, diffusivity. The value of diffusivity for a particular system depends on many factors as many mechanisms could be operative.

Diffusing species: If the diffusing species is able to occupy interstitial sites, then it can easily diffuse through the parent matrix. On the other hand if the size of substitutional species is almost equal to that of parent atomic size, substitutional diffusion would be easier. Thus size of diffusing species will have great influence on diffusivity of the system.

Temperature: Temperature has a most profound influence on the diffusivity and diffusion rates. It is known that there is a barrier to diffusion created by neighboring atoms those need to move to let the diffusing atom pass. Thus, atomic vibrations created by temperature assist diffusion. Empirical analysis of the system resulted in an Arrhenius type of relationship between diffusivity and temperature.

$$D = D_0 \exp\left(-\frac{Q}{RT}\right)$$

where D_0 is a pre-exponential constant, Q is the activation energy for diffusion, R is gas constant (Boltzmann's constant) and T is absolute temperature. From the above equation it can be inferred that large activation energy means relatively small diffusion coefficient. It can also be observed that there exists a linear proportional relation between (lnD) and $(1/T)$. Thus by plotting and considering the intercepts, values of Q and D_0 can be found experimentally.

Lattice structure: Diffusion is faster in open lattices or in open directions than in closed directions.

Presence of defects: As mentioned in earlier section, defects like dislocations, grain boundaries act as short-circuit paths for diffusing species, where the activation energy is diffusion is less. Thus the presence of defects enhances the diffusivity of diffusing species.

5.2.2 Non-equilibrium transformation & microstructure

During the processing of metallic materials, they are subjected to different conditions and thus transformation of its structure. During casting process, liquid metal is allowed to cool to become solid component. However, during cooling conditions can be such that they are in equilibrium or non-equilibrium state. Phases and corresponding microstructures, usually shown in a phase diagram, are generated during equilibrium solidification under the conditions that are realized only for extremely slow cooling rates. This is because with change in temperature, there must be readjustments in the compositions of liquids and solid phases in accordance with the phase diagram. These readjustments are accomplished by diffusional processes in both solid and liquid phases and also across the solid-liquid interface. But, it is well understood that diffusion is time-dependent phenomenon, and moreover diffusion in solid phases are much lower than in liquid phases, equilibrium solidification requires extremely longer times those are impractical. Thus virtually all practical solidification takes place under non-equilibrium conditions, leading to compositional gradients and formation of meta-stable phases.

As a consequence of compositional gradients during non-equilibrium cooling, segregation (concentration of particular, usually impurity elements, along places like grain boundaries) and coring (gradual compositional changes across individual grains) may occur. Coring is predominantly observed in alloys having a marked difference between liquidus and solidus temperatures. It is often being removed by subsequent annealing (incubation at relatively high temperatures that are close to lower solidus temperature, enhances diffusion in solids) and/or hot-working, and is exploited in zone-refining technique to produce high-purity metals. Segregation is also put to good use in zone refining, and also in the production of rimming steel. Micro-segregation is used to describe the differences in composition across a crystal or between neighboring crystals. On the other hand, macro-segregation is used to describe more massive heterogeneities which may result from entrapment of liquid pockets between growing solidifying zones. Micro-segregation can often be removed by prolonged annealing or by hot-working; but macro-segregation persists through normal heating and working operations.

In most situations cooling rates for equilibrium solidification are impractically slow and unnecessary; in fact, on many occasions non-equilibrium conditions are desirable. Two non equilibrium effects of practical importance are (1) the occurrence of phase changes or transformations at temperatures other than those predicted by phase boundary lines on the phase diagram, and (2) the existence of non-equilibrium phases at room temperature that do not appear on the phase diagram. These will be detailed in next chapters about phase transformation.

References

1. G. E. Dieter, Mechanical Metallurgy, Third Edition, McGraw-Hill, New York, 1986.
2. E. A. Brandes and G. B. Brook, Smithells Metals Reference Book, Seventh Edition, Butterworth-Heinemann, Oxford, 1992.
3. Shewmon, P. G., Diffusion in Solids, Second Edition, The Minerals, Metals and Materials Society, Warrendale, PA, 1989

Chapter 7. Dislocations and Strengthening Mechanisms

6.1 Dislocations & Plastic deformation and Mechanisms of plastic deformation in metals

6.1.1 Dislocations & Plastic deformation

While some materials are elastic in nature up point of fracture, many engineering materials like metals and thermo-plastic polymers can undergo substantial permanent deformation. This characteristic property of materials makes it feasible to shape them. However, it imposes some limitations on the engineering usefulness of such materials. Permanent deformation is due to process of shear where particles change their neighbors. During this process inter-atomic or inter-molecular forces and structure plays important roles, although the former are much less significant than they are in elastic behavior. Permanent deformation is broadly two types – plastic deformation and viscous flow. Plastic deformation involves the relative sliding of atomic planes in organized manner in crystalline solids, while the viscous flow involves the switching of neighbors with much more freedom that does not exist in crystalline solids.

It is well known that dislocations can move under applied external stresses. Cumulative movement of dislocations leads to the gross plastic deformation. At microscopic level, dislocation motion involves rupture and reformation of inter-atomic bonds. The necessity of dislocation motion for ease of plastic deformation is well explained by the discrepancy between theoretical strength and real strength of solids, as explained in chapter-3. It has been concluded that one-dimensional crystal defects – dislocations – plays an important role in plastic deformation of crystalline solids. Their importance in plastic deformation is relevant to their characteristic nature of motion in specific directions (slip-directions) on specific planes (slip-planes), where edge dislocation move by slip and climb while screw dislocation can be moved by slip and cross-slip.

The onset of plastic deformation involves start of motion of existing dislocations in real crystal, while in perfect crystal it can be attributed to generation of dislocations and subsequently their motion. During the motion, dislocations will tend to interact among themselves. Dislocation interaction is very complex as number of dislocations moving on number of slip planes in various directions. When they are in the same plane, they repel each other if they have the same sign, and annihilate if they have opposite signs (leaving behind a perfect crystal). In general, when dislocations are close and their strain fields add to a larger value, they repel, because being close increases the potential energy (it takes energy to strain a region of the material). When unlike dislocations are on closely spaced neighboring slip planes, complete annihilation cannot occur. In this situation, they combine to form a row of vacancies *or* an interstitial atom.

An important consequence interaction of dislocations that are not on parallel planes is that they intersect each other or inhibit each others motion. Intersection of two dislocations results in a sharp break in the dislocation line. These breaks can be of two kinds:

(a) A jog is break in dislocation line moving it out of slip plane.
(b) A kink is break in dislocation line that remains in slip plane.

Other hindrances to dislocation motion include interstitial and substitutional atoms, foreign particles, grain boundaries, external grain surface, and change in structure due to phase change. Important practical consequences of hindrance of dislocation motion are that dislocations are still movable but at higher stresses (or forces), and in most instances that leads to generation of more dislocations. Dislocations can spawn from existing dislocations, and from defects, grain boundaries and surface irregularities. Thus, the number of dislocations increases dramatically during plastic deformation. As further motion of dislocations requires increase of stress, material can be said to be strengthened i.e. materials can be strengthened by controlling the motion of dislocation.

6.1.2 Mechanisms of plastic deformation in metals

Plastic deformation, as explained in earlier section, involves motion of dislocations. There are two prominent mechanisms of plastic deformation, namely ***slip*** and ***twinning***.

Slip is the prominent mechanism of plastic deformation in metals. It involves sliding of blocks of crystal over one other along definite crystallographic planes, called slip planes. In physical words it is analogous to a deck of cards when it is pushed from one end. Slip occurs when shear stress applied exceeds a critical value. During slip each atom usually moves same integral number of atomic distances along the slip plane producing a step, but the orientation of the crystal remains the same. Steps observable under microscope as straight lines are called slip lines.

Slip occurs most readily in specific directions (slip directions) on certain crystallographic planes. This is due to limitations imposed by the fact that single crystal remains homogeneous after deformation. Generally slip plane is the plane of greatest atomic density, and the slip direction is the close packed direction within the slip plane. It turns out that the planes of the highest atomic density are the most widely spaced planes, while the close packed directions have the smallest translation distance. Feasible combination of a slip plane together with a slip direction is considered as a slip system. The common slip systems are given in *table-6.1*.

Table-6.1: *Slip systems for different crystal structures.*

Crystal	**Occurrence**	**Slip planes**	**Slip directions**
FCC		{111}	<110>
BCC	More common Less common	{110} {112},{123}	<111>
HCP	More common Less common	Basal plane Prismatic & Pyramidal planes	Close packed directions
NaCl		{110}	<110>

In a single crystal, plastic deformation is accomplished by the process called slip, and sometimes by twinning. The extent of slip depends on many factors including external load and the corresponding value of shear stress produced by it, the geometry of crystal structure, and the orientation of active slip planes with the direction of shearing stresses generated. Schmid first recognized that single crystals at different orientations but of same material require different stresses to produce slip. The dependence of various factors has been summarized using a parameter – critical resolved shear stress, τ_R, given as

$$\tau_R = \frac{P\cos\lambda}{A/\cos\varphi} = \frac{P}{A}\cos\varphi\cos\lambda = \sigma\cos\varphi\cos\lambda$$

$$\Rightarrow m = \cos\varphi\cos\lambda$$

where *P* – external load applied, *A* – cross-sectional area over which the load applied, *λ* – angle between slip direction and tensile axis, *ø* – angle between normal to the slip plane and the tensile axis and *m* – Schmid factor.

Shear stress is maximum for the condition where *λ* = ø = 45°. If either of the angles are equal to 90°, resolved shear stress will be zero, and thus no slip occurs. If the conditions are such that either of the angles is close to 90°, crystal will tend to fracture rather than slip. Single crystal metals and alloys are used mainly for research purpose and only in a few cases of engineering applications.

Almost all engineering alloys are polycrystalline. Gross plastic deformation of a polycrystalline specimen corresponds to the comparable distortion of the individual grains by means of slip. Although some grains may be oriented favorably for slip, yielding cannot occur unless the unfavorably oriented neighboring grains can also slip. Thus in a polycrystalline aggregate, individual grains provide a mutual geometrical constraint on one other, and this precludes plastic deformation at low applied stresses. That is to initiate plastic deformation, polycrystalline metals require higher stresses than for equivalent single crystals, where stress depends on orientation of the crystal. Much of this increase is attributed to geometrical reasons.

Slip in polycrystalline material involves generation, movement and (re-)arrangement of dislocations. Because of dislocation motion on different planes in various directions, they may interact as well. This interaction can cause dislocation immobile or mobile at higher stresses. During deformation, mechanical integrity and coherency are maintained along the grain boundaries; that is, the grain boundaries are constrained, to some degree, in the shape it may assume by its neighboring grains. Once the yielding has occurred, continued plastic deformation is possible only if enough slip systems are simultaneously operative so as to accommodate grain shape changes while maintaining grain boundary integrity. According to von Mises criterion, a minimum of five independent slip systems must be operative for a polycrystalline solid to exhibit ductility and maintain grain boundary integrity. This arises from the fact that an arbitrary deformation is specified by the six components of strain tensor, but because of requirement of constant volume, there are only independent strain components. Crystals which do not possess five independent slip systems are never ductile in polycrystalline form, although small plastic elongation may be noticeable because of twinning or a favorable preferred orientation.

The second important mechanism of plastic deformation is twinning. It results when a portion of crystal takes up an orientation that is related to the orientation of the rest of the untwined lattice in a definite, symmetrical way. The twinned portion of the crystal is a mirror image of the parent crystal. The plane of symmetry is called twinning plane. Each atom in the twinned region moves by a homogeneous shear a distance proportional to its distance from the twin plane. The lattice strains involved in twinning are small, usually in order of fraction of inter-atomic distance, thus resulting in very small gross plastic deformation. The important role of twinning in plastic deformation is that it causes changes in plane orientation so that further slip can occur. If the surface is polished, the twin would be still visible after etching because it possesses a different orientation from the untwined region. This is in contrast with slip, where slip lines can be removed by polishing the specimen.

Twinning also occurs in a definite direction on a specific plane for each crystal structure. However, it is not known if there exists resolved shear stress for twinning. Twinning generally occurs when slip is restricted, because the stress necessary for twinning is usually higher than that for slip. Thus, some HCP metals with limited number of slip systems may preferably twin. Also, BCC metals twin at low temperatures because slip is difficult. Of course, twinning and slip may occur sequentially or even concurrently in some cases. Twinning systems for some metals are given in table-6.2.

Table-6.2: *Twin systems for different crystal structures*.

Crystal	**Example**	**Twin plane**	**Twin direction**
FCC	Ag, Au, Cu	(111)	[112]
BCC	α-Fe, Ta	(112)	[111]
HCP	Zn, Cd, Mg, Ti	$(10\bar{1}2)$	$[\bar{1}011]$

Figure-6.1 presents schematic movement of atoms during plastic deformation in slip and during twinning.

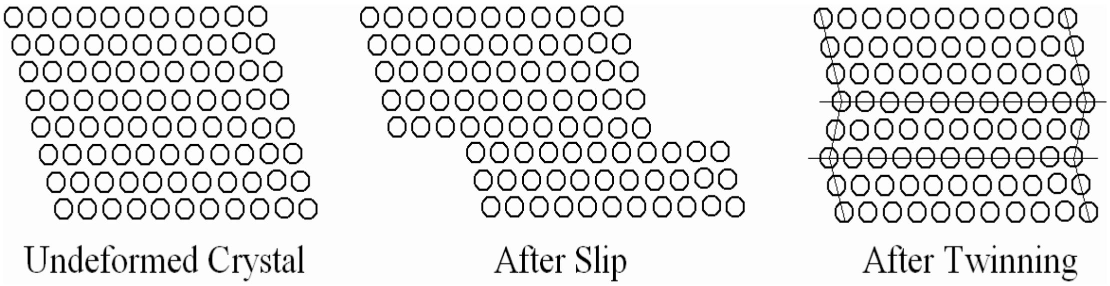


Figure-6.1: *Schematic presentation of different plastic deformation mechanism.*

In *table-6.3*, both the mechanisms of plastic deformations are compared with respect to their characteristics.

Table-6.3: *Comparison of mechanism of plastic deformation.*

	during/in slip	**during/in twinning**
Crystal orientation	Same above and below the slip plane	Differ across the twin plane
Size (in terms of inter-atomic distance)	Multiples	Fractions
Occurs on	Widely spread planes	Every plane of region involved
Time required	Milli seconds	Micro seconds
Occurrence	On many slip systems simultaneously	On a particular plane for each crystal

6.2 Strengthening mechanisms in Metals

Ability of a metal to deform plastically depends on ease of dislocation motion under applied external stresses. As mentioned in earlier section, strengthening of a metal consist hindering dislocation motion. Dislocation motion can be hindered in many ways, thus are strengthening mechanisms in metals. Strengthening by methods of grain-size reduction, solid-solution alloying and strain hardening applies for single-phase metals. Precipitation hardening, dispersion hardening, fiber strengthening and Martensite strengthening are applicable to multi-phase metallic materials.

6.2.1 Strengthening by Grain Size Reduction

This strengthening mechanism is based on the fact that crystallographic orientation changes abruptly in passing from one grain to the next across the grain boundary. Thus it is difficult for a dislocation moving on a common slip plane in one crystal to pass over to a similar slip plane in another grain, especially if the orientation is very misaligned. In addition, the crystals are separated by a thin non-crystalline region, which is the characteristic structure of a large angle grain boundary. Atomic disorder at the boundary causes discontinuity in slip planes. Hence dislocations are stopped by a grain boundary and pile up against it. The smaller the grain size, the more frequent is the pile up of dislocations. A twin boundary can also act as an obstacle to dislocation motion.

A grain boundary can hinder the dislocation motion in two ways: (1) by forcing the dislocation to change its direction of motion and (2) discontinuity of slip plane because of disorder. Effectiveness of grain boundary depends on its characteristic misalignment, represented by an angle. The ordinary high-angle grain boundary (misalignment > 5) represents a region of random misfit between the grains on each side of the boundary. This structure contains grain-boundary dislocations which are immobile. However they group together within the boundary to form a step or grain boundary ledge. These ledges can act as effective sources of dislocations as the stress at end of slip plane may trigger new dislocations in adjacent grains. Small angle grain boundaries (misalignment < 1) are considered to be composed of a regular array of dislocations, and are not effective in blocking dislocations.

With decrease in grain size, the mean distance of a dislocation can travel decreases, and soon starts pile up of dislocations at grain boundaries. This leads to increase in yield strength of the material. E.O.Hall and N.J.Petch have derived the following relation, famously known as *Hall-Petch relation*, between yield strength (σ_y) and grain size (d):

$$\sigma_y = \sigma_i + kd^{-1/2}$$

where σ_i is the 'friction stress', representing the overall resistance of the crystal lattice to dislocation movement, k is the 'locking parameter' that measures the relative hardening contribution of the grain boundaries and d is the average grain diameter. Friction stress is interpreted as the stress needed to move unlocked dislocations along the slip plane. It depends strongly on temperature, strain, alloy and impurity content. Locking parameter is known to be independent of temperature. Thus friction stress and locking parameters are constants for particular material.

It is important to note that the above relation is not valid for both very large grain and extremely fine grain sizes. Grain size reduction improves not only strength, but also the toughness of many alloys. Grain size can be controlled by rate of cooling, and also by plastic deformation followed by appropriate heat treatment.

Grain size is usually measured using a light microscope to observe a polished specimen by counting the number of grains within a given area, by determining the number of grains that intersect a given length of random line, or by comparing with standard-grain-size charts. If d is average grain diameter, S_v is grain boundary area per unit volume, N_L is mean number of intercepts of grain boundaries per unit length of test line, N_A is number of grains per unit area on a polished surface; the all these are related as follows:

$$S_v = 2N_L, \quad d = \frac{3}{S_v} = \frac{3}{2N_L} \quad \text{and} \quad d = \sqrt{\frac{6}{\pi N_A}}$$

Another common method of measuring the grain size is by comparing the grains at a fixed magnification with standard grain size charts. Charts are coded with ASTM grain size number, G, and is related with n_a – number of grains per mm^2 at 1X magnification as

$$G = -2.9542 + 1.4427 \ln n_a$$

G represents number of grains per square inch (645 mm^2) at a magnification of 100X is equal to 2^{G-1}. Higher the ASTM grain number, smaller is the grain diameter. Grain diameter, D (in mm), and ASTM number, G, can be related as follows:

$$D = \frac{1}{100}\sqrt{\frac{645}{2^{G-1}}}$$

6.2.2 Solid Solution Strengthening

Adding atoms of another element that those occupy interstitial or substitutional positions in parent lattice increases the strength of parent material. This is because stress fields generated around the solute atoms interact with the stress fields of a moving dislocation, thereby increasing the stress required for plastic deformation i.e. the impurity atoms cause lattice strain which can "anchor" dislocations. This occurs when the strain caused by the alloying element compensates that of the dislocation, thus achieving a state of low potential energy. Since solid-solution alloy additions affect the entire stress-strain curve, it can be said that solute atoms have more influence on the frictional resistance to dislocation motion than on the static locking of dislocations. Pure metals are almost always softer than their alloys. Solute strengthening effectiveness depends on two factors – size difference between solute and parent atoms, and concentration of solute atoms.

Solute atoms are two categories with respect to their relative strengthening effect – (1) those produce non-spherical distortions, such as most interstitial atoms, have a relative strengthening effect per unit concentration of about three times their shear modulus, (2) those produce spherical distortion, such as substitutional atoms, have a relative strengthening of about G/10.

Solute atoms interact with dislocations in many ways, namely: elastic interaction; modulus interaction; stacking-fault interaction; electrical interaction; short-range order interaction; and long-range order interaction. Elastic, modulus, and long-range order interactions are of long-range i.e. they are relatively insensitive to temperature and continue to act about 0.6 T_m where T_m is the melting temperature in absolute Kelvin degrees.

Elastic interaction results from mutual interaction of elastic stress fields, while *modulus interaction* occurs if the presence of a solute atom locally alters the modulus of the crystal. *Stacking-fault interactions* arise because solute atoms may segregate to the stacking-faults, thus lowering stacking-fault energy and widening partial dislocations. This interaction is also called as Suzuki or Chemical interaction. *Electrical interaction* arises if solute atoms of dissimilar valence interact with dislocations which have electrical dipoles. *Short-range order interaction* arises from the tendency for solute atoms to arrange themselves so that they have more then the equilibrium number of dissimilar neighbors. The opposite of short-range order is clustering. *Long-range order interaction* arises in alloys which form super-lattices, in which long-range periodic arrangement of dissimilar atoms gets disturbed, to form anti-phase boundaries, because of dislocation motion which leads to dissociation of dislocation into pairs of ordinary dislocations.

Some polycrystalline metals, such as mild steel, display a discrete yield point type of behavior where a higher stress is necessary to initiate plastic flow than to continue it. Thus, there exists a localized, heterogeneous type of transition from elastic to plastic deformation which produces a yield point in the stress-strain curve i.e. elastic-plastic transition is very well demarked and occurs abruptly in what is called *yield-point phenomenon*. During loading the load increases steadily with elastic strain, drops suddenly at the upper yield point where plastic deformation gets initiated. Continued deformation fluctuates slightly about some constant stress value – lower yield point. Subsequently stress increase with increasing strain. Elongation that occurs at constant load is called the yield-point elongation. Yield strength for metals with this phenomenon is taken as average of lower yield point, thus it is not necessary to employ strain offset method. A schematic picture in *figure-6.2* presents the exaggerated version of yield-point phenomenon.

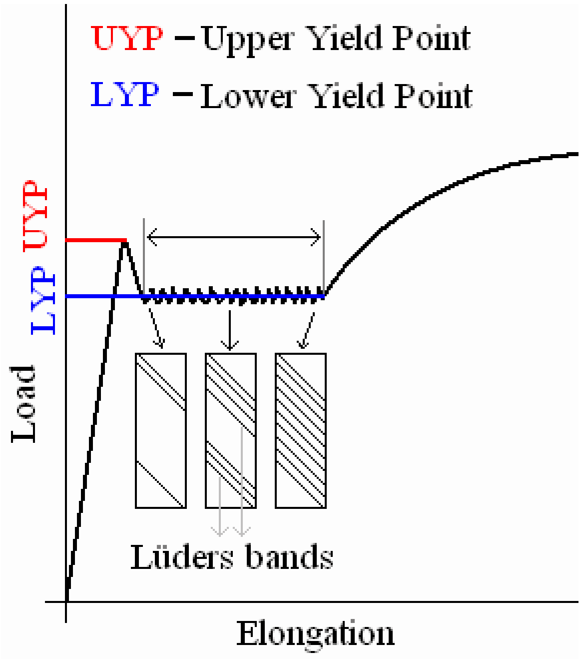


Figure-6.2: *Yield point phenomenon.*

At the upper yield point, deformed part of metal forms a discrete band usually visible to eye at a stress concentration. Many bands may form at the same time at different stress concentration sites. Each band then propagates along the length of the specimen, causing yield-point elongation. These bands are called Lüders bands / Hartmann lines / stretcher stains, and generally are approximately 45 to the tensile axis. Occurrence of yield point is associated with presence of small amounts of interstitial or substitutional impurities. It's been found that either unlocking of dislocations by a high stress for the case of strong pinning or generation of new dislocations are the reasons for yield-point phenomenon. Thus when the dislocation line is pulled free from the influence of solute atoms, slip can occur at a lower stress. On the other hand, when dislocations are strongly pinned, new dislocations will generate followed by drop in flow stress. This is the origin of upper yield stress. Released dislocations tend to pile-up at grain boundaries, producing stress concentration which in addition to applied stress unlocks sources in next grain. This is the way band propagates. Magnitude of yield-point effect will depend on energy of interaction between solute atoms and dislocations and on the concentration of solute atoms at the dislocations.

6.2.3 Strain Hardening

Two most important industrial processes used to harden metals or alloys are: strain hardening and heat treatment. Strain hardening is used for hardening/strengthening materials that are not responsive to heat treatment. The phenomenon where ductile metals become stronger and harder when they are deformed plastically is called strain hardening *or* work hardening.

Intensity of strain hardening can be gaged from the slope of the flow curve, defined by the parameter strain hardening exponent, *n*. It is measure of the ability of a metal to strain harden. For a given amount of plastic strain, higher the value of n, greater is the strain hardening. Increasing temperature lowers the rate of strain hardening, and thus the treatment is given, usually, at temperatures well below the melting point of the material. Thus the treatment is also known as *cold working*. Most metals strain hardens at room temperature. The consequence of strain hardening a material is improved strength and hardness but material's ductility will be reduced.

As mentioned in earlier chapters, with plastic deformation dislocation density increases because under applied stress dislocation sources, like Frank-Reed source, becomes active. Thus lightly cold worked material may have a dislocation density in order of 10^{12} m^{-2}, while a very heavily cold worked material may have 10^{16} m^{-2} against 10^{10} m^{-2} for annealed material. High density of dislocations and thus increased interaction under applied loads during cold working is the cause for increase in strength of a material. In general rate of strain hardening is lower for HCP metals than for cubic metals where higher number of slip systems is active at an instant of time. Empirically shear stress (τ) to move a dislocation that increases with increase in dislocation density (ρ) is given as:

$$\tau = \tau_0 + A\sqrt{\rho}$$

It is convenient to express the degree of plastic deformation as percent cold work, defined as:

$$\%CW = \left[\frac{A_0 - A_d}{A_0}\right] x100$$

where A_0 is the original cross sectional area that experiences deformation, and A_d is the area after deformation.

Strain hardening is used commercially to enhance the mechanical properties of metals during fabrication procedures. In addition to mechanical properties, physical properties of a material also changes during cold working. There is usually a small decrease in density, an appreciable decrease in electrical conductivity, small increase in thermal coefficient of expansion and increased chemical reactivity (decrease in corrosion resistance).

The cold worked state is a condition of higher internal energy than the un-deformed metal. Although cold worked dislocation cell structure is mechanically stable, it is not thermodynamically stable. With increase in temperature state becomes more unstable, eventually reverts to strain-free condition. This process of heating to attain strain-free condition is called annealing heat treatment where effects of strain hardening may be removed. Annealing process can be divided into three distinct processes: recovery, recrystallization and grain growth. All these steps of the heat treatment process are explained in next section. It is usual industrial practice to use alternate cycles of strain hardening and annealing to deform most metals to a very great extent.

6.2.4-5 Precipitation Hardening and *Dispersion strengthening*

Small second-phase particles distributed in a ductile matrix can hinder the dislocation motion and thus increase the strength of a material. Second-phase particles either can be introduced by mixing and consolidation (dispersion strengthening) or precipitated in solid state (precipitation hardening).

Precipitation hardening or age hardening is produced by solution treating and quenching an alloy. Term ‘Age hardening’ is used to describe the process because strength develops with time. Requisite for precipitation hardening to take place is that second phase must be soluble at an elevated temperature but precipitates upon quenching and aging at a lower temperature. This limits the alloy systems which can be strengthened by precipitation hardening. For example: Al-alloys, Cu-Be alloys, Mg-Al alloys, Cu-Sn alloys. If the precipitation occurs at normal ambient temperatures, it is called natural aging. Some alloy systems needed to be aged at higher temperatures and the process is known as artificial aging. Most precipitation hardened alloys are limited in their maximum service temperatures, which may lose their strength at elevated temperatures due to over-aging.

In dispersion strengthening, hard particles are mixed with matrix powder and consolidated and processed by powder metallurgy techniques. Here second phase shall have very little solubility in the matrix, even at elevated temperatures. Because there is very little solubility, the particles resist growth or over-aging to a much greater extent than the second phase particles in a precipitation hardening system. Theoretically, at least, it is possible to produce infinite number of dispersion-hardened systems by mixing finely divided metallic powders and second phase particles (oxides, carbides, nitrides, borides, etc).

Dislocations moving through the matrix, which is either precipitation hardened or dispersion strengthened, have two alternatives. They can either cut through the precipitate particles *or* bend around and bypass them. The first alternative is possible only when the slip plane is continuous from the matrix through the precipitate particle and when the stress to move a dislocation in precipitate is comparable to that in matrix. Cutting of particles is easier for small particles which can be considered as segregated solute atoms. There are many particle properties that can dictate the ease of shearing, for example: coherency strains, stacking-fault energy, ordered structure, modulus effect, interfacial energy, morphology and lattice friction stress.

Cutting of particles is not possible when there is an interface or an abrupt change in orientation. Under such instances, dislocations have to bend around them, and bypass. The mechanism involved in this instance is similar to the operation of a Frank-Reed source. Stress required to bend a dislocation is inversely proportional to the average interspacing (λ) of particles. Effective strengthening is achieved in the bending process, when the particles are submicroscopic in size.

$$\tau = {}^{Gb}\!/_{\lambda}$$

The degree of strengthening from second phase particles depends on particle distribution in the matrix. Particle dispersion along with the shape can be expressed by specifying the volume fraction, average particle diameter, and mean interspacing of particles which are interrelated. A simple expression for interspacing of particles is:

$$\lambda = \frac{4(1-f)r}{3f}$$

where f is the volume fraction of spherical particles of radius r.

The spacing between second phase particles should be typically a few hundred angstroms. Optimum strengthening occurs during aging once the right interspacing of particles is achieved. Smaller the particles, dislocations can cut through them at lower stresses. On the other hand, larger the particles they will be distributed at wider distances. Thus it is necessary to arrest the growth of particles in precipitation hardened alloys at right interspacing. Over-aging allows the fine particles to coalesce into larger and fewer particles. The interspacing is increased thereby and the yield stress decreases.

6.2.6 Fiber strengthening

Second phase material can also be introduced into matrix in form of fibers to strengthen it. However, mechanism of strengthening is different from either precipitation hardening or dispersion strengthening where second phase is introduced as fine particles.

Prerequisites are materials to be used as fibers include high strength and/or high strength-to-weight ratio. Fibers usually, thus, have high strength and high modulus while the matrix must be ductile and non-reactive with the fibers. Fibers may be long and continuous or they may be discontinuous. Examples for fiber material: Al_2O_3, boron, graphite, metal, glass, etc. Examples for matrix material: metals, polymers. Fiber reinforced materials are an important group of materials known as composite materials.

In fiber-reinforced materials, high modulus fibers carry essentially the entire load while the matrix serves to transmit the load to the fibers. Matrix also protects fibers from surface damage, serves to blunt cracks which arise from fiber breakage while it also serves to separate the fibers. It is understood that analysis of precipitation/dispersion strengthening is based on dislocation theory. On the other hand, the analysis of strength of fiber-reinforced material involves the direct application of continuum principles at the microscopic level as the material behavior is essentially elastic.

To achieve any benefit from presence of fibers, empirically it can be shown that there is a need for critical fiber volume which must be exceeded for fiber strengthening to occur.

$$f_{critical} = \frac{\sigma_{mu} - \sigma'_m}{\sigma_{fu} - \sigma'_m}$$

where σ_{mu} – strength of strain hardened matrix, σ'_m – flow stress of matrix at a strain equal to fiber breaking stress, σ_{fu} – ultimate tensile strength of the fiber.

However for small values of critical fiber volume, number of fibers present may not be sufficient to effectively restrain the elongation of the matrix. Thus a minimum volume fraction of fiber which must be exceeded to have real reinforcement is defined as follows:

$$f_{\min} = \frac{\sigma_{mu} - \sigma'_m}{\sigma_{fu} + \sigma_{mu} - \sigma'_m}$$

When fiber-reinforced material is loaded, load transmits from matrix to fibers via interface. Thus interface between matrix and fiber play an important role in fiber strengthening.

Usually fibers are arranged in uni-directional manner, thus it is highly anisotropic in nature. One of the consequences of the anisotropy of fiber composites is that they exhibit shear coupling. This means that an axial load produces shear stresses, and shear load produces axial strains. In an isotropic material a uni-axially applied load produces only axial and transverse normal strains. Shear strains in fiber reinforced composites are compensated practically by using a cross-ply laminate.

6.2.7 Martensite Strengthening

This strengthening can be achieved in systems where a diffusion-controlled invariant transformation can be suppressed by rapid cooling. The Martensite strengthening process, thus, basically is a diffusion-less and displacive reaction. The martensitic phase is formed from the retained high temperature phase at temperatures lower then the equilibrium invariant transformation temperature. It occurs by a process of lattices shearing. Martensite under microscope appears as lenticular plates which divide and subdivide the grains of the parent phase. Always touching but never crossing one another. The characteristic lenticular shape minimizes the elastic distortion in the matrix. These platelets grow at about one-third the velocity of sound. This high speed means activation energy for growth is very low, and thus activation energy for nucleation determines the amount of Martensite that can form under given conditions.

Martensite platelets attain their shape by two successive shear displacements contained in boundaries coherent with the parent phase. The first displacement is a homogeneous shear throughout the plate which occurs parallel to a specific plane in the parent phase known as the habit plane. The second displacement, the lesser of the two, can take place by one of two mechanisms: slip as in Fe-C Martensite or twinning as in Fe-Ni Martensite. Martensite formation occurs in many systems like Fe-C, Fe-Ni, Fe-Ni-C, Cu-Zn, Au-Cd, and even in pure metals like Li, Zr and Co. However, only the alloys based on Fe and C show a pronounced strengthening effect.

High strength of Martensite is attributed to its typical crystal structure i.e. effective barriers to slip are provided by the fine twin structure or the high dislocation density. In Fe-C system, carbon atoms are also involved in strengthening. Super saturated carbon atoms strain the ferrite lattice, and this strain can be relieved by redistribution of carbon atoms by diffusion at room temperature. One result is that a strong binding is set up between dislocations and the carbon atoms. This hinders the motion of dislocations, thus increasing the strength.

This strengthening mechanism is one of the most common processes used in engineering materials. The name basically arises from the phase *Martensite* that forms in steels when they are quenched from the solutionizing temperature.

6.3 Recovery, Recrystallization and Grain Growth

As mentioned in earlier sections, annealing is an important industrial process to relieve the stresses from cold working. During cold working grain shape changes, while material strain hardens because of increase in dislocation density. Between 1-10% of the energy of plastic deformation is stored in material in the form of strain energy associated with point defects and dislocations. On annealing i.e. on heating the deformed material to higher temperatures and holding, material tends to lose the extra strain energy and revert to the original condition before deformation by the processes of recovery and recrystallization. Grain growth may follow these in some instances.

6.3.1 Recovery

This is the first stage of restoration after cold working where physical properties of the cold-worked material are restored without any observable change in microstructure. The properties that are mostly affected by recovery ate those sensitive to point defects, for example – thermal and electrical conductivities. During recovery, which takes place at low temperatures of annealing, some of the stored internal energy is relieved by virtue of dislocation motion as a result of enhanced atomic diffusion. There is some reduction, though not substantial, in dislocation density as well apart from formation of dislocation configurations with low strain energies. Excess point defects that are created during deformation are annihilated either by absorption at grain boundaries or dislocation climbing process. Stored energy of cold work is the driving force for recovery.

6.3.2 Recrystallization

This stage of annealing follows after recovery stage. Here also driving force is stored energy of cold work. Even after complete recovery, the grains are still in relatively high strain energy state. This stage, thus, involves replacement of cold-worked structure by a new set of strain-free, approximately equi-axed grains i.e. it is the process of nucleation and growth of new, strain-free crystals to replace all the deformed crystals. It starts on heating to temperatures in the range of 0.3-0.5 T_m, which is above the recovery stage. There is no crystal structure change during recrystallization. This process is characterized by recrystallization temperature which is defined as the temperature at which 50% of material recrystallizes in one hour time. The recrystallization temperature is strongly dependent on the purity of a material. Pure materials may recrystallizes around 0.3 T_m, while impure materials may recrystallizes around 0.5-0.7 T_m.

There are many variables that influence recrystallization behavior, namely amount of prior deformation, temperature, time, initial grain size, composition and amount of recovery prior to the start of the recrystallization. This dependence leads to following empirical laws:

- A minimum amount of deformation is needed to cause recrystallization.
- Smaller the degree of deformation, higher will be the recrystallization temperature.
- The finer is the initial grain size; lower will be the recrystallization temperature.
- The larger the initial grain size, the greater degree of deformation is required to produce an equivalent recrystallization temperature.
- Greater the degree of deformation and lower the annealing temperature, the smaller will be the recrystallized grain size.
- The higher is the temperature of cold working, the less is the strain energy stored and thus recrystallization temperature is correspondingly higher.
- The recrystallization rate increases exponentially with temperature.

During recrystallization, the mechanical properties that were changes during deformation are restored to their pre-cold-work values. Thus material becomes softer, weaker and ductile. During this stage of annealing impurity atoms tend to segregate at grain boundaries, and retard their motion and obstruct the processes of nucleation and growth. This *solute drag effect* can be used to retain cold worked strength at higher service temperatures. Presence of second phase particles causes slowing down of recrystallization – *pinning action of the particles*.

6.3.3 Grain growth

This stage follows complete crystallization if the material is left at elevated temperatures. However, grain growth does not need to be preceded by recovery and recrystallization; it may occur in all polycrystalline materials. During this stage newly formed strain-free grains tend to grow in size. This grain growth occurs by the migration of grain boundaries. Driving force for this process is reduction in grain boundary energy i.e. decreasing in free energy of the material. As the grains grow larger, the curvature of the boundaries becomes less. This results in a tendency for larger grains to grow at the expense of smaller grains. In practical applications, grain growth is not desirable. Incorporation of impurity atoms and insoluble second phase particles are effective in retarding grain growth.

Because the driving force for grain growth is lower than the driving force for recrystallization, grain growth occurs slowly at a temperature where recrystallization occurs at substantially high speeds. However, grain growth is strongly temperature dependent.

References

1. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
2. ASM Handbook, Alloy Phase Diagrams, ASM International, Materials Park, OH, 1992.
3. ASM Handbook, Metallography and Microstructures, ASM International, Materials Park, OH, 1985.

Chapter 6. Phase Diagrams

Many of the engineering materials possess mixtures of phases, e.g. steel, paints, and composites. The mixture of two or more phases may permit interaction between different phases, and results in properties usually are different from the properties of individual phases. Different components can be combined into a single material by means of solutions or mixtures. *A solution (liquid or solid) is phase with more than one component; a mixture is a material with more than one phase*. Solute does not change the structural pattern of the solvent, and the composition of any solution can be varied. In mixtures, there are different phases, each with its own atomic arrangement. It is possible to have a mixture of two different solutions!

A pure substance, under equilibrium conditions, may exist as either of a phase namely vapor, liquid or solid, depending upon the conditions of temperature and pressure. A *phase* can be defined as a homogeneous portion of a system that has uniform physical and chemical characteristics i.e. it is a physically distinct from other phases, chemically homogeneous and mechanically separable portion of a system. In other words, a phase is a structurally homogeneous portion of matter. When two phases are present in a system, it is not necessary that there be a difference in both physical and chemical properties; a disparity in one or the other set of properties is sufficient.

There is only one vapor phase no matter how many constituents make it up. For pure substance there is only one liquid phase, however there may be more than one solid phase because of differences in crystal structure. A liquid solution is also a single phase, even as a liquid mixture (e.g. oil and water) forms two phases as there is no mixing at the molecular level. In the solid state, different chemical compositions and/or crystal structures are possible so a solid may consist of several phases. For the same composition, different crystal structures represent different phases. A solid solution has atoms mixed at atomic level thus it represents a single phase. A single-phase system is termed as homogeneous, and systems composed of two or more phases are termed as mixtures or heterogeneous. Most of the alloy systems and composites are heterogeneous.

It is important to understand the existence of phases under various practical conditions which may dictate the microstructure of an alloy, thus the mechanical properties and usefulness of it. Phase diagrams provide a convenient way of representing which state of aggregation (phase or phases) is stable for a particular set of conditions. In addition, phase diagrams provide valuable information about melting, casting, crystallization, and other phenomena.

***Useful terminology*:-**

Component – is either pure metal and/or compounds of which an alloy is composed. The components of a system may be elements, ions or compounds. They refer to the independent chemical species that comprise the system.

System – it can either refer to a specific body of material under consideration or it may relate to the series of possible alloys consisting of the same components but without regard to alloy composition.

Solid solution – it consists of atoms of at least two different types where solute atoms occupy either substitutional or interstitial positions in the solvent lattice and the crystal structure of the solvent is maintained.

Solubility limit – for almost all alloy systems, at a specific temperature, a maximum of solute atoms can dissolve in solvent phase to form a solid solution. The limit is known as solubility limit. In general, solubility limit changes with temperature. If solute available is more than the solubility limit that may lead to formation of different phase, either a solid solution or compound.

Phase equilibrium – it refers to the set of conditions where more than one phase may exist. It can be reflected by constancy with time in the phase characteristics of a system. In most metallurgical and materials systems, phase equilibrium involves just solid phases. However the state of equilibrium is never completely achieved because of very slow rate of approach of equilibrium in solid systems. This leads to non-equilibrium or meta-stable state, which may persist indefinitely and of course, has more practical significance than equilibrium phases. An equilibrium state of solid system can be reflected in terms of characteristics of the microstructure, phases present and their compositions, relative phase amounts and their spatial arrangement or distribution.

Variables of a system – these include two external variables namely temperature and pressure along with internal variable such as composition (C) and number of phases (P). Number of independent variables among these gives the degrees of freedom (F) or variance. All these are related for a chosen system as follows:

$$P + F = C + 2$$

which is known as *Gibbs Phase rule*. The degrees of freedom cannot be less than zero so that we have an upper limit to the number of phases that can exist in equilibrium for a given system. For practical purpose, in metallurgical and materials field, pressure can be considered as a constant, and thus the *condensed phase rule* is given as follows:

$$P + F = C + 1$$

7.1 Equilibrium Phase Diagrams, Particle strengthening by precipitation and precipitation reactions

7.1.1 Equilibrium Phase Diagrams

A diagram that depicts existence of different phases of a system under equilibrium is termed as *phase diagram*. It is also known as *equilibrium* or *constitutional diagram*. Equilibrium phase diagrams represent the relationships between temperature and the compositions and the quantities of phases at equilibrium. In general practice it is sufficient to consider only solid and liquid phases, thus pressure is assumed to be constant (1 *atm*.) in most applications. These diagrams do not indicate the dynamics when one phase transforms into another. However, it depicts information related to microstructure and phase structure of a particular system in a convenient and concise manner. Important information, useful for the scientists and engineers who are involved with materials development, selection, and application in product design, obtainable from a phase diagram can be summarized as follows:

- To show phases are present at different compositions and temperatures under slow cooling (equilibrium) conditions.
- To indicate equilibrium solid solubility of one element/compound in another.
- To indicate temperature at which an alloy starts to solidify and the range of solidification.
- To indicate the temperature at which different phases start to melt.
- Amount of each phase in a two-phase mixture can be obtained.

A phase diagram is actually a collection of solubility limit curves. The phase fields in equilibrium diagrams depend on the particular systems being depicted. Set of solubility curves that represents locus of temperatures above which all compositions are liquid are called *liquidus*, while *solidus* represents set of solubility curves that denotes the locus of temperatures below which all compositions are solid. Every phase diagram for two or more components must show a liquidus and a solidus, and an intervening freezing range, except for pure system, as melting of a phase occurs over a range of temperature. Whether the components are metals or nonmetals, there are certain locations on the phase diagram where the liquidus and solidus meet. For a pure component, a contact point lies at the edge of the diagram. The liquidus and solidus also meet at the other invariant positions on the diagram. Each invariant point represents an invariant reaction that can occur only under a particular set of conditions between particular phases, so is the name for it!

Phase diagrams are classified based on the number of components in the system. Single component systems have unary diagrams, two-component systems have binary diagrams, three-component systems are represented by ternary diagrams, and so on. When more than two components are present, phase diagrams become extremely complicated and difficult to represent. This chapter deals mostly with binary phase diagrams.

Unary diagrams: In these systems there is no composition change (*C*=1), thus only variables are temperature and pressure. Thus in region of single phase two variables (temperature and pressure) can be varied independently. If two phases coexist then, according to Phase rule, either temperature or pressure can be varied independently, but not both. At triple points, three phases can coexist at a particular set of temperature and pressure. At these points, neither temperature nor the pressure can be changed without disrupting the equilibrium i.e. one of the phases may disappear. *Figure-1* depicts phase diagram for water.

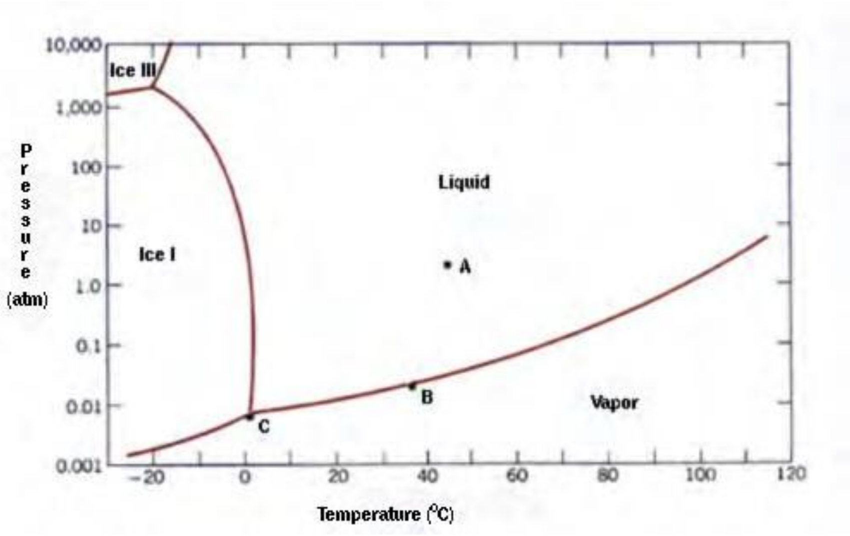


Figure-1: Unary phase diagram for water.

Binary diagrams: These diagrams constitutes two components, e.g.: two metals (Cu and Ni), or a metal and a compound (Fe and Fe_3C), or two compounds (Al_2O_3 and Si_2O_3), etc. In most engineering applications, as mentioned before, condensed phase rule is applicable. It is assumed that the same is applicable for all binary diagrams, thus the presentation of binary diagrams becomes less complicated. Thus binary diagrams are usually drawn showing variations in temperature and composition only. It is also to be noted that all binary systems consist only one liquid phase i.e. a component is completely soluble in the other component when both are in liquid state.

Hence, binary systems are classified according to their solid solubility. If both the components are completely soluble in each other, the system is called *isomorphous* system. E.g.: Cu-Ni, Ag-Au, Ge-Si, Al_2O_3-Cr_2O_3. Extent solid solubility for a system of two metallic components can be predicted based on Hume-Ruthery conditions, summarized in the following:

- Crystal structure of each element of solid solution must be the same.
- Size of atoms of each two elements must not differ by more than 15%.

- Elements should not form compounds with each other i.e. there should be no appreciable difference in the electro-negativities of the two elements.
- Elements should have the same valence.

All the Hume-Rothery rules are not always applicable for all pairs of elements which show complete solid solubility.

In systems other than isomorphous systems i.e. in case of limited solid solubility, there exist solid state miscibility gaps; number of invariant reactions can take place; intermediate phases may exist over a range of composition (*intermediate solid solutions*) or only at relatively fixed composition (*compound*). These intermediate phases may undergo polymorphic transformations, and some may melt at a fixed temperature (*congruent transformations*, in which one phase changes to another of the same composition at definite temperature). A solid solution based on a pure component and extending to certain finite compositions into a binary phase diagram is called a *terminal solid solution*, and the line representing the solubility limit of a terminal solid solution *w.r.t* a two-phase solid region is called a *solvus* line (*figure-4*).

<u>*Isomorphous system*</u>: *Figure-2* depicts a typical phase diagram for an isomorphous system made of two metallic elements ***A*** and ***B***. As cited earlier, any phase diagram can be considered as a map. A set of coordinates – a temperature and a composition – is associated with each point in the diagram. If the alloy composition and temperature specified, then the phase diagram allows determination of the phase or phases that will present under equilibrium conditions. There are only two phases in the phase diagram, the liquid and the solid phases. These single-phases regions are separated by a two-phase region where both liquid and solid co-exist. The area in the *figure-2* above the line marked liquidus (***A'bB'***) corresponds to the region of stability of the liquid phase, and the area below the solidus line (***A'dB'***) represents the stable region for the solid phase.

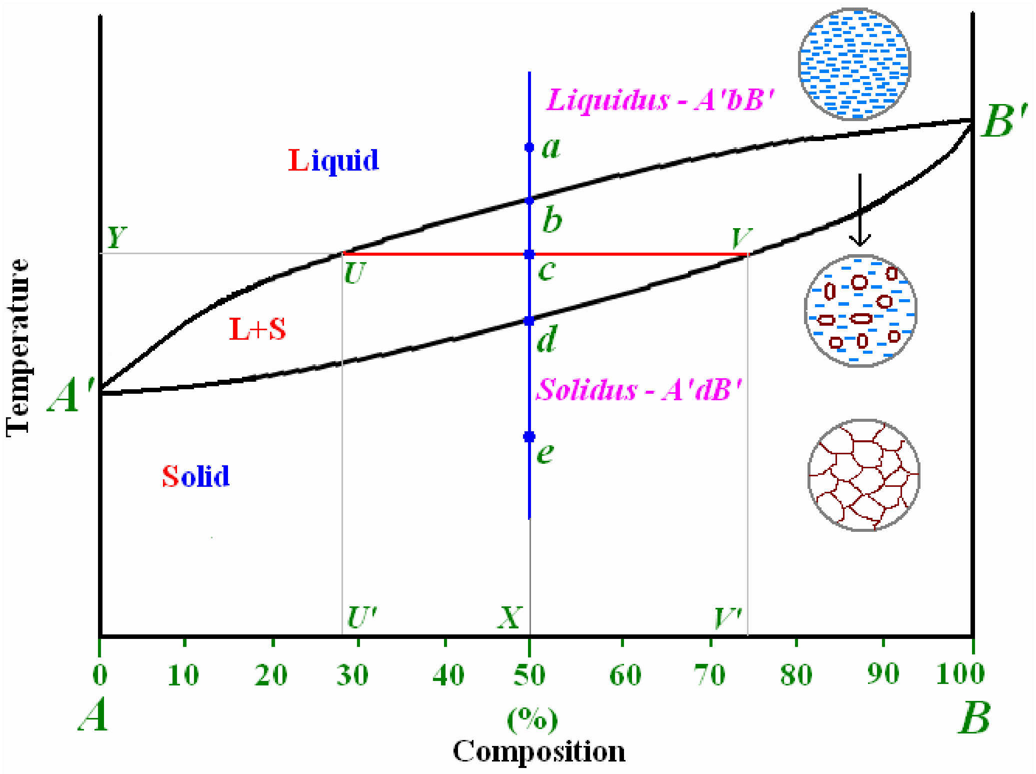


Figure-2: *Phase diagram for typical isomorphous binary system.*

For the interpretation of the phase diagram, let's consider the vertical line ***ae*** drawn corresponding to composition of 50%***A*** +50%***B*** and assume that the system is undergoing equilibrium cooling. The point ***a*** on the line ***ae*** signifies that for that particular temperature and composition, only liquid phase is stable. This is true up to the point ***b*** which lies on the liquidus line, representing the starting of solidification. Completion of solidification of the alloy is represented by the point, ***d***. Point ***e*** corresponds to single-phase solid region up to the room temperature. Point ***c*** lies in the two-phase region made of both liquid and solid phases. Corresponding micro-structural changes are also shown in *figure-2*. As shown in figure-2, above liquidus only a liquid phase exists, and below the solidus single solid phase exists as completely solidified grains. Between these two lines, system consist both solid crystals spread in liquid phase. It is customary to use ***L*** to represent liquid phase(s) and Greek alphabets (***α***, ***β***, ***γ***) for representing solid phases.

Between two extremes of the horizontal axis of the diagram, cooling curves for different alloys are shown in *figure-3* as a function of time and temperature. Cooling curves shown in *figure-3* represent ***A***, ***U'***, ***X***, ***V'*** and ***B*** correspondingly in *figure-2*. Change in slope of the cooling curve is caused by heat of fusion. In fact these changes in slope are nothing but points on either solidus or liquidus of a phase diagram. An experimental procedure where repeated cooling/heating of an alloy at different compositions, and corresponding changes in slope of cooling curves will be used to construct the phase diagram.

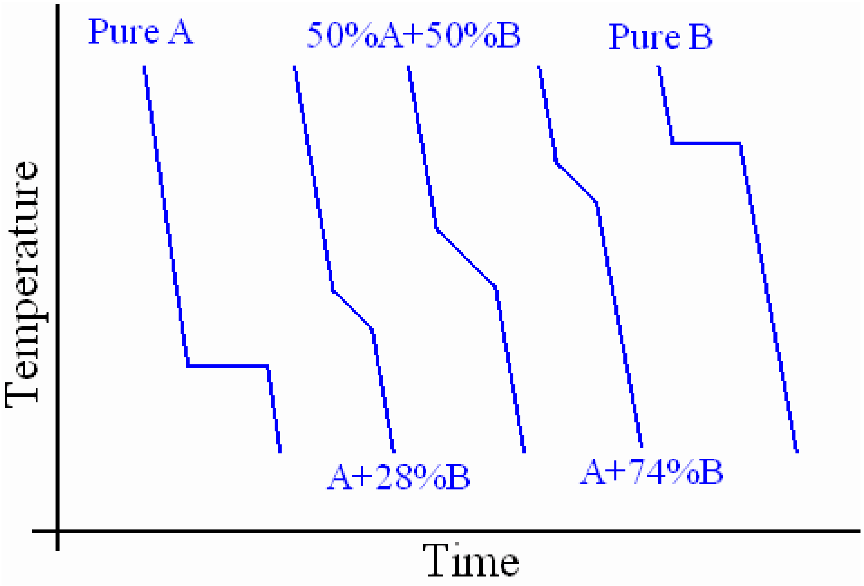


Figure-3: *Cooling curves for isomorphous binary system.*

Another important aspect of interpreting phase diagrams along with phases present is finding the relative amount of phases present and their individual composition.

Procedure to find equilibrium concentrations of phases:

- A *tie-line* or *isotherm* (***UV***) is drawn across two-phase region to intersect the boundaries of the region.
- Perpendiculars are dropped from these intersections to the composition axis, represented by ***U'*** and ***V'*** in *figure-2*, from which each of each phase is read. ***U'*** represents composition of liquid phase and ***V'*** represents composition of solid phase as intersection ***U*** meets liquidus line and ***V*** meets solidus line.

Procedure to find equilibrium relative amounts of phases (*lever rule*):

- A tie-line is constructed across the two phase region at the temperature of the alloy to intersect the region boundaries.
- The relative amount of a phase is computed by taking the length of tie line from overall composition to the phase boundary for the other phase, and dividing by the total tie-line length. From *figure-2*, relative amounts of liquid and solid phases is given respectively by

  $C_L = \frac{cV}{UV}$, $C_S = \frac{Uc}{UV}$, and it is to be noted that $C_L + C_S = 1$.

<u>*Eutectic system*</u>: Many binary systems have components which have limited solid solubility, e.g.: Cu-Ag, Pb-Sn. The regions of limited solid solubility at each end of a phase diagram are called *terminal solid solutions* as they appear at ends of the diagram.

Many of the binary systems with limited solubility are of eutectic type, which consists of specific alloy composition known as *eutectic composition* that solidifies at a lower temperature than all other compositions. This low temperature which corresponds to the lowest temperature at which the liquid can exist when cooled under equilibrium conditions is known as *eutectic temperature*. The corresponding point on the phase diagram is called *eutectic point*. When the liquid of eutectic composition is cooled, at or below eutectic temperature this liquid transforms simultaneously into two solid phases (two terminal solid solutions, represented by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$). This transformation is known as *eutectic reaction* and is written symbolically as:

Liquid ($\boldsymbol{L}$) ↔ solid solution-1 ($\boldsymbol{\alpha}$) + solid solution-2 ($\boldsymbol{\beta}$)

This eutectic reaction is called invariant reaction as it occurs under equilibrium conditions at a specific temperature and specific composition which can not be varied. Thus, this reaction is represented by a thermal horizontal arrest in the cooling curve of an alloy of eutectic composition. A typical eutectic type phase diagram is shown in *figure-4* along with a cooling curve.

As shown in *figure-4*, there exist three single phase regions, namely liquid ($\boldsymbol{L}$), $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ phases. There also exist three two phase regions: $\boldsymbol{L+\alpha}$, $\boldsymbol{L+\beta}$ and $\boldsymbol{\alpha+\beta}$. These three two phase regions are separated by horizontal line corresponding to the eutectic temperature. Below the eutectic temperature, the material is fully solid for all compositions. Compositions and relative amount of the phases can be determined using tie-lines and lever rule. Compositions that are on left-hand-side of the eutectic composition are known as *hypo-eutectic compositions* while compositions on right-hand-side of the eutectic composition are called *hyper-eutectic compositions*. Development of micro-structure and respective cooling curves for eutectic alloys are shown in *figure-5, 6, 7 and 8* for different compositions. The phase that forms during cooling but before reaching eutectic temperature is called *pro-eutectic phase*.

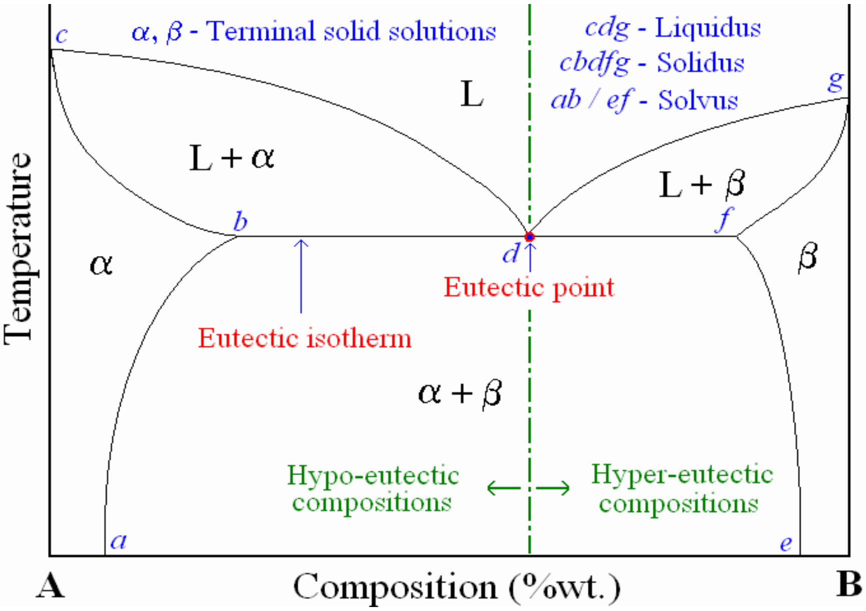


Figure-4: *Typical phase diagram for a binary eutectic system.*

In many systems, solidification in the solid + liquid region may lead to formation of layered (cored) grains, even at very slow cooling rates. This is as a result of very slow or no-diffusion in solid state compared with very high diffusion rates in liquids. The composition of the liquid phase evolves by diffusion, following the equilibrium values that can be derived from the tie-line method. However, new layers that solidify on top of the grains have the equilibrium composition at that temperature but once they are solid their composition does not change.

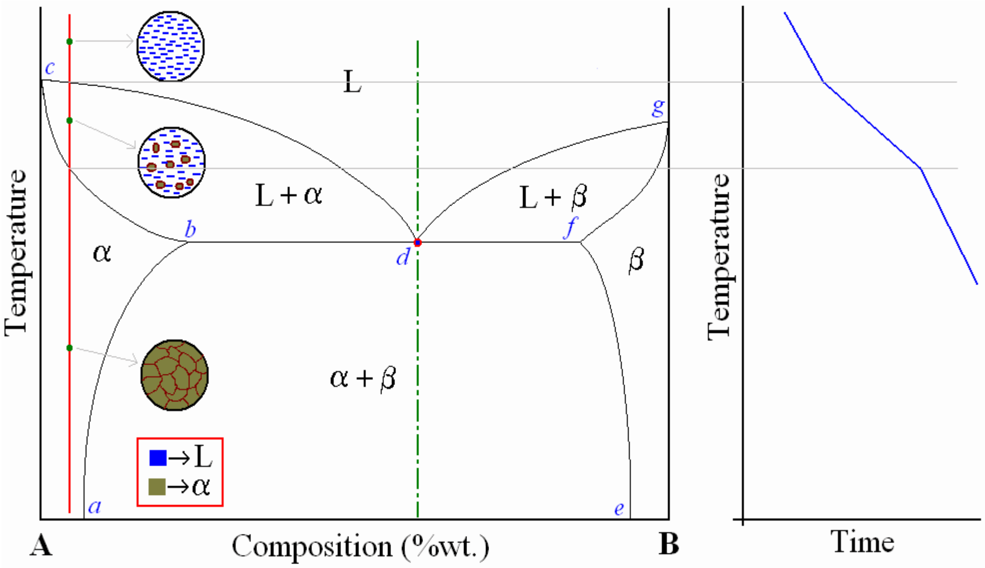


Figure-5: *Cooling curve and micro-structure development for eutectic alloy that passes mainly through terminal solid solution.*

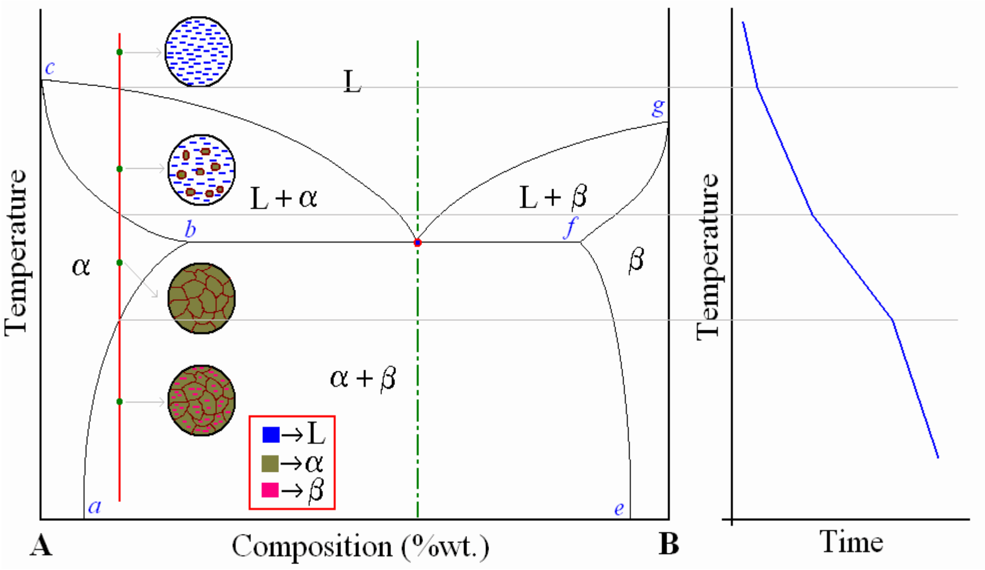


Figure-6: *Cooling curve and micro-structure development for eutectic alloy that passes through terminal solid solution without formation of eutectic solid.*

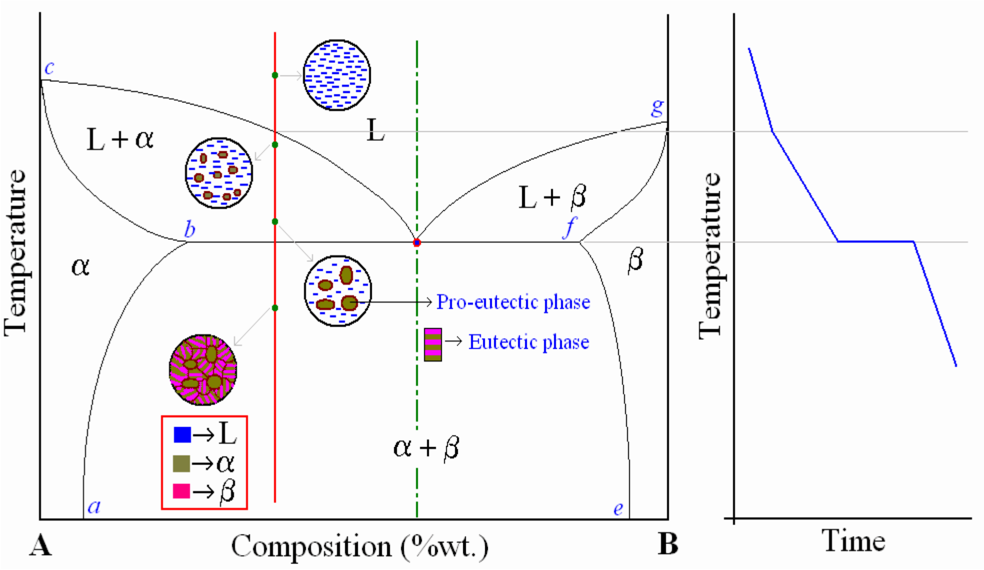


Figure-7: *Cooling curve and micro-structure development for eutectic alloy that passes through hypo-eutectic region.*

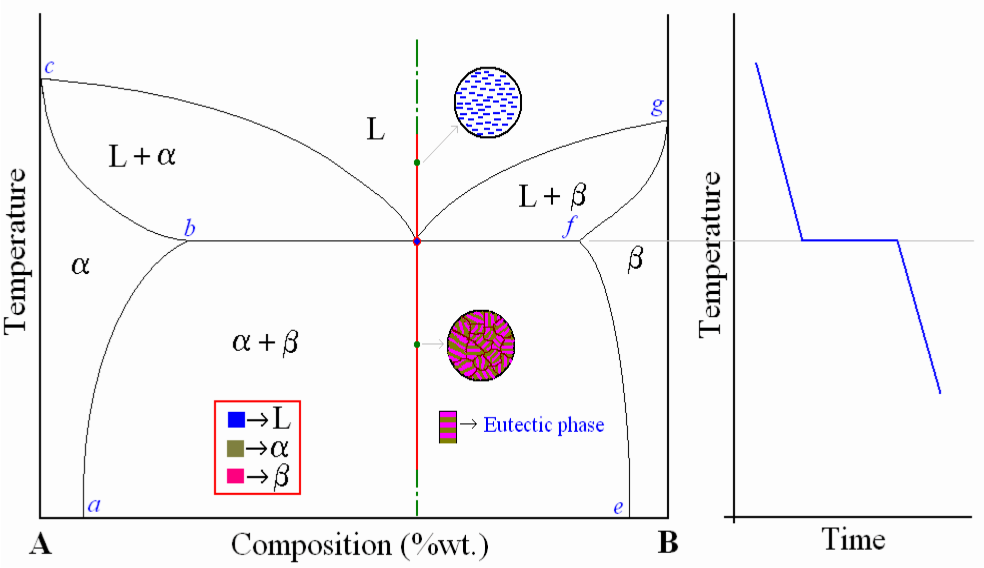


Figure-8: *Cooling curve and micro-structure development for eutectic alloy that passes through eutectic-point.*

Invariant reactions: The *eutectic reaction*, in which a liquid transforms into two solid phases, is just one of the possible three-phase invariant reactions that can occur in binary systems those are not isomorphous. Schematically it can be shown as in *figure-9*. It represents that a liquid phase, ***L***, transforms into two different solids phases (***α*** and ***β***) upon cooling during the eutectic reaction.

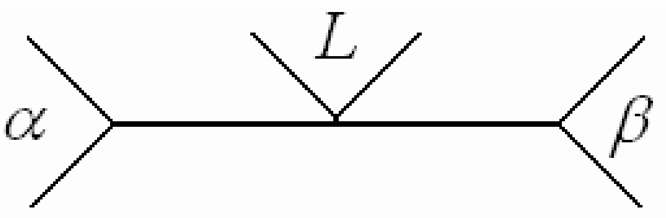
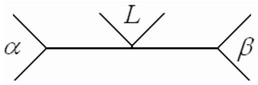
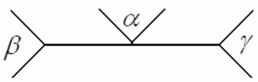
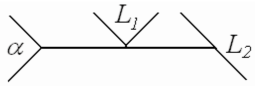
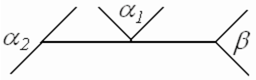
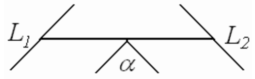


Figure-9: *Schematic of eutectic invariant reaction.*

In the solid state analog of a eutectic reaction, called a *eutectoid reaction*, one solid phase having eutectoid composition transforms into two different solid phases. Another set of invariant reactions that occur often in binary systems are - *peritectic reaction* where a solid phase reacts with a liquid phase to produce a new solid phase, and in *peritectoid reaction*, two solid phases react to form a new solid phase. Peritectic reaction is commonly present as part of more-complicated binary diagrams, particularly if the melting points of the two components are quite different. Peritectic and peritectoid reactions do not give rise to micro-constituents as the eutectic and eutectoid reactions do. Another invariant reaction that involves liquid phase is *monotectic reaction* in which a liquid phase transforms into a solid phase and a liquid phase of different composition. Over a certain range of compositions the two liquids are immiscible like oil and water and so constitute individual phases, thus monotectic reaction can said to be associated with miscibility gaps in the liquid state. Example system for monotectic reaction: Cu-Pb at 954 C and 36%Pb. Analog to monotectic reaction in solid state is *monotectoid reaction* in which a solid phase transforms to produce two solid phases of different compositions. Another notable invariant reaction that is associated with liquid immiscibility is *syntectic reaction* in which two liquid phases react to form a solid phase. All the invariant reactions are summarized in the *table-1* showing both symbolic reaction and schematic part of phase diagram.

Table-1: *Summary of invariant reactions in binary systems.*

Reaction	Symbolic equation	Schematic presentation	Example
Eutectic	$L \leftrightarrow \alpha + \beta$		Fe-C, 4.27% C, 1147 C
Eutectoid	$\alpha \leftrightarrow \beta + \gamma$		Fe-C, 0.80% C, 723 C
Peritectic	$L + \alpha \leftrightarrow \beta$		Fe-C, 0.16%C, 1495 C
Peritectoid	$\alpha + \beta \leftrightarrow \gamma$		
Monotectic	$L_1 \leftrightarrow L_2 + \alpha$		Fe-C, 0.51%C, 1495 C

Monotectoid	$\boldsymbol{\alpha_1 \leftrightarrow \alpha_2 + \beta}$		
Syntectic	$\boldsymbol{L_1 + L_2 \leftrightarrow \alpha}$		

<u>*Intermediate phases*</u>: An intermediate phase may occur over a composition range (intermediate solid solution) or at a relatively fixed composition (compound) inside the phase diagram and are separated from other two phases in a binary diagram by two phase regions. Many phase diagrams contain intermediate phases whose occurrence cannot be readily predicted from the nature of the pure components. Intermediate solid solutions often have higher electrical resistivities and hardnesses than either of the two components. Intermediate compounds form relatively at a fixed composition when there exists a stoichiometric relationship between the components, for example: Mg_2Ni and $MgNi_2$ in Mg-Ni system. These are called *inter-metallic compounds*, and differ from other chemical compounds in that the bonding is primarily metallic rather than ionic or covalent, as would be found with compounds in certain metal-nonmetal or ceramic systems. Some metal-nonmetal compounds, Fe_3C, are metallic in nature, whereas in others, MgO and Mg_2Si, bonding is mainly covalent. When using the lever rules, inter-metallic compounds are treated like any other phase, except they appear not as a wide region but as a vertical line.

Number of phase transformations may takes place for each system. Phase transformations in which there are no compositional alternations are said to be *congruent transformations*, and during *incongruent transformations* at least one of the phases will experience a change in composition. Examples for (1) congruent transformations: allotropic transformations, and melting of pure materials (2) incongruent transformations: all invariant reactions, and also melting of alloy that belongs to an isomorphous system. Intermediate phases are sometimes classified on the basis of whether they melt congruently or incongruently. $MgNi_2$, for example, melts congruently whereas Mg_2Ni melts incongruently since it undergoes peritectic decomposition.

7.1.2 Particle strengthening by precipitation and precipitation reactions

As explained in an earlier chapter (*chapter-6: Dislocations and Strengthening Mechanisms*), by obstructing dislocation motion in different means, material's strength can be increased. One of the methods that are applicable to multi-phase material is particle strengthening in which second phase particles are introduced into the matrix by either mixing-and-consolidation (dispersion strengthening) or precipitated in solid state (precipitation hardening).

The object of the precipitation strengthening is to create in a heat-treated alloy a dense and fine dispersion of precipitated particles in a matrix of deformable metal. The particles act as obstacles to dislocation motion. In order for an alloy system to be able to precipitation-strengthened for certain alloy compositions; there must be a terminal solid solution which has a decreasing solid solubility as the temperature decreases. For example: Au-Cu in which maximum solid solubility of Cu in Al is 5.65% at 548 C that decreases with decreasing temperature.

The precipitation strengthening process involves the following three basic steps:

- Solutionizing (solution heat treatment), where the alloy is heated to a temperature between solvus and solidus temperatures and kept there till a uniform solid-solution structure is produced.
- Quenching, where the sample is rapidly cooled to a lower temperature (room temperature) and the cooling medium is usually water. Alloy structure in this stage consists of supersaturated solid solution.
- Aging is the last but critical step. During this heat treatment step finely dispersed precipitate particle will form. Aging the alloy at room temperature is called natural aging, whereas at elevated temperatures is called artificial aging. Most alloys require artificial aging, and aging temperature is usually between 15-25% of temperature difference between room temperature and solution heat treatment temperature.

Precipitation strengthening and reactions that occur during precipitation can be best illustrated using the Al-4%Cu (duralumin) system. *Figure-10* depicts the Al-rich end of the Al-Cu phase diagram. It can be observed that the alloy with 4%Cu exists as a single phase ***α***-solid solution at around 550 C, and at room temperature as a mixture of ***α*** (with less than 0.5%Cu) and an inter-metallic compound, $CuAl_2$ (***θ***) with 52%Cu. On slow cooling ***α*** rejects excess Cu as precipitate particles of θ. These particles relatively coarse in size and can cause only moderate strengthening effect.

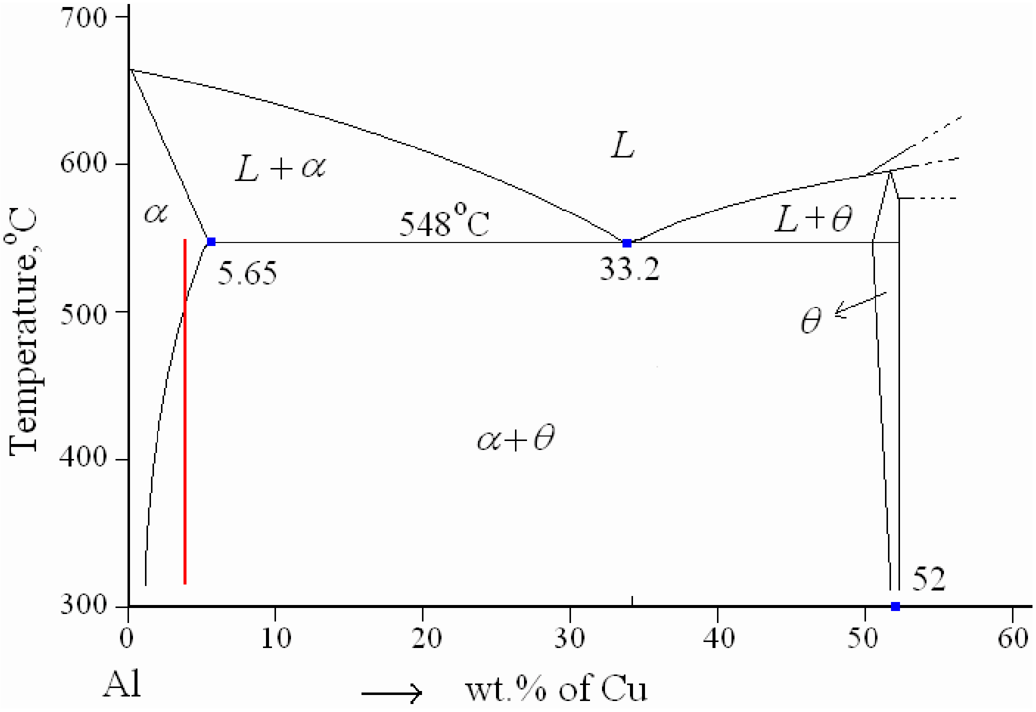


Figure-10: *Aluminium rich end of Al-Cu phase diagram.*

By rapidly cooling the alloy, a supersaturated solution can be obtained at room temperature. As a function of time at room temperature, and at higher temperatures up to 200 C, the diffusion of Cu atoms may take place and the precipitate particles can form. For this particular alloy, Al-4%Cu, five sequential structures can be identified: (a) supersaturated solid solution $\boldsymbol{\alpha}$, (b) GP1 zones, (c) GP2 zones ($\boldsymbol{\theta''}$ phase), (d) $\boldsymbol{\theta'}$ phase and (e) $\boldsymbol{\theta}$ phase, $CuAl_2$. Not all these phases can be produced at all aging temperatures. GP1 and GP2 zones are produced at lower temperatures, and $\boldsymbol{\theta'}$ and $\boldsymbol{\theta}$ phases occur at higher temperatures. The initial stages of precipitation are the most difficult to analyze because of the extremely small size of the particles and their relatively uniform distribution. GP zones meant for *Guinier-Preston* zones which have a definite composition and structure that is not the same as that of the final stable precipitate. Evidently these particles are easier to nucleate than the final precipitate, as a result, form first. Eventually they disappear as later more stable phases appear. $\boldsymbol{\theta''}$ and $\boldsymbol{\theta'}$ are meta-stable transition precipitates with distinct crystal structure of their own, while $\boldsymbol{\theta}$ is the equilibrium stable precipitate of $CuAl_2$.

GP1 zones:- These zones are created by Cu atoms segregating in $\boldsymbol{\alpha}$, and the segregated regions are of disk shape with thickness of 0.4-0.6 nm, and 8-10 nm in diameter and form on the {100} cubic planes of the matrix. As Cu atoms which replace Al atoms are smaller in diameter, matrix lattice strains tetragonally. These zones are said to be coherent with the matrix lattice.

GP2 zones / $\boldsymbol{\theta''}$ phase:- With additional aging, ordering of larger clumps of Cu atoms on {100} occurs. These zones have tetragonal structure which therefore introduces coherency in the lattice with {100} planes of the matrix, accompanied by further hardening. However, their size ranges from 1-4 nm thick and 10-100 nm in diameter as aging proceeds.

$\boldsymbol{\theta'}$ phase:- This phase nucleates heterogeneously especially on dislocations. It has tetragonal structure but is partially coherent with the matrix. This phase forms platelets with thickness 10-150 nm.

$\boldsymbol{\theta}$ phase:- With still further aging the equilibrium phase $CuAl_2$ or $\boldsymbol{\theta}$ is formed from the transition lattice $\boldsymbol{\theta'}$ or directly from the matrix accompanied by a reduction in hardness. It has a BCT (body-centered-tetragonal) structure, and is incoherent with the matrix. As these particles are no longer coherent with the matrix, hardness is lower than at the stage when coherent was present. Over-aging continues with the growth of these particles controlled by diffusion. Variation of hardness with aging time is shown in *figure-11*.

The general sequence of precipitation in binary Al-Cu alloys can represented as:

Supersaturated $\boldsymbol{\alpha}$ → GP1 zones → GP2 zones ($\boldsymbol{\theta''}$ phase) → $\boldsymbol{\theta'}$ phase → $\boldsymbol{\theta}$ phase ($CuAl_2$)

Most precipitation-hardening systems operate in a similar way, peak hardness usually being attained in the later stages of coherency or at the onset of incoherency. It is quite common for a coherent precipitate to form and then lose coherency when the particle grows to a critical size. However, in some systems there is no evidence of coherency strains, and the fine particles appear to act alone as impediments to dislocation movements, for example – systems with dispersion strengthening.

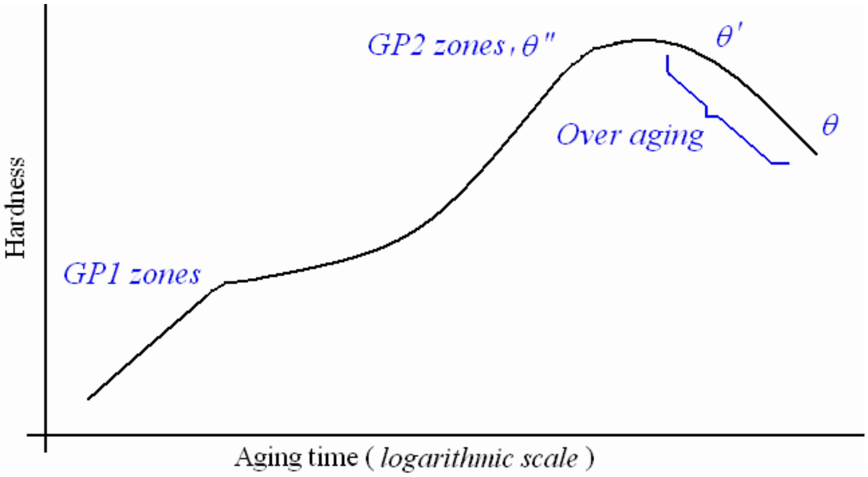


Figure-11: *Correlation of structures and hardness for Al-4%Cu alloy*.

7.2 Kinetics of nucleation and growth

Structural changes in metallic systems usually take place by nucleation and growth whether it is just a phase change within one of the three states, or a simple structural rearrangement within a single phase, or a phase transformation. An equilibrium phase diagram presents the phases and phase changes expected under equilibrium conditions, but it provides no information about the rates of transformation. Although changes in pressure, composition, or temperature can cause phase transformations, it is temperature changes that are more important. From a micro structural standpoint, the first process to accompany a phase transformation is nucleation (i.e. the formation of very small particles or nuclei of the product phase from the parent phase) of the new phase particles which are capable of growing. The second stage is growth, in which the nucleated particles increase their size. The transformation reaches completion if growth of these new phase particles is allowed to proceed until the equilibrium fraction is attained.

Both nucleation and growth require that the accompanying free energy change be negative. Consequently, the super-heating or super-cooling that is necessary for a phase change is to be expected. That is a transformation cannot tale place precisely at the equilibrium transformation temperature because at that temperature free energies of phases are equal. In addition to temperature, two other factors that affect transformation rate – first, diffusion controlled rearrangement of atoms because of compositional and/or crystal structural differences; second, difficulty encountered in nucleating small particles via change in surface energy associated with the interface. Diffusion limits both the nucleation and growth rates in many cases.

With the nucleation of new particle, new interface is created between the particle and liquid. This interface will have positive energy that must be supplied during the transformation process. A tiny particle has a large surface area to volume ratio and therefore be unstable. Thus energy of the surface can effectively prevent the initial formation of a tiny particle. A particle said to have nucleated when it becomes stable and will not disappear due to thermal fluctuations. After a particle attained a critical size, it can grow further with a continuous decrease in energy. The surface energy is no longer a dominant factor in the growth process.

7.2.1 Nucleation kinetics

In *homogeneous nucleation*, the probability of nucleation occurring at any given site is identical to that at any other site within the volume of the parent phase. When a pure liquid metal is cooled below its equilibrium freezing temperature to a sufficient degree, numerous homogeneous nuclei are created by slow-moving atoms bonding together. Homogeneous nucleation usually requires a considerable amount of *undercooling* (cooling a material below the equilibrium temperature for a given transformation without the transformation occurring). Undercooling enhances the formation of nuclei that eventually grow. If Δf is the free energy change accompanying the formation of a spherical new phase particle,

$$\Delta f = \frac{4}{3}\pi r^3 \Delta g + 4\pi r^2 \gamma$$

where r is the radius of the particle, Δg is the Gibbs free energy change per unit volume and γ is the surface energy of the interface. As surface energy, γ, is always positive, and Δg is negative, passes through a maximum. From calculus, critical values can be found from the following:

$$r^* = -\frac{2\gamma}{\Delta g} \; ; \; \Delta f^*_{\text{hom}} = \frac{16}{3}\pi\gamma^3/(\Delta g)^2$$

Particles which are smaller than the critical size are called embryos; those larger than the critical size are called nuclei. As Δg becomes more negative with a lowering of the temperature, the critical values of Δf and r becomes smaller as shown in *figure-12*. At sufficiently low temperatures, nucleation can be triggered by a few atoms statistically clustering as a nucleus, so a small critical radius is exceeded. With added growth, the new phase attains stability. Of course atom movements are sluggish at low temperatures, so growth is generally slow.

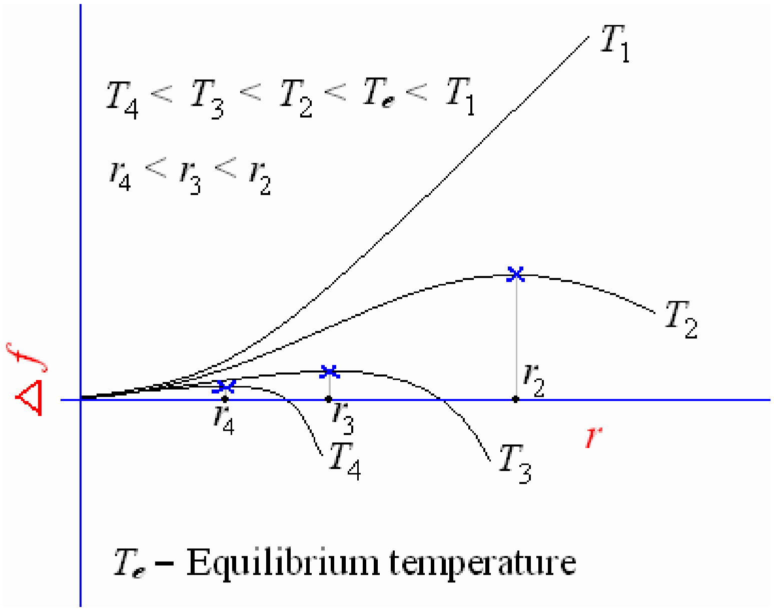


Figure-12: *Effect of temperature on free energy change and particle radius.*

The greater the degree of undercooling below the equilibrium melting temperature of the metal, the greater the change in volume free energy, however, the change in free energy due to surface energy does not change much with temperature. Thus, the critical size of nuclei is mainly determined by volume free energy. Near the freezing temperature, critical nucleus size must be infinite since ΔT approaches zero. As the amount of undercooling increases, critical size decreases, and are related as follows:

$$r^* = \frac{2\gamma T_m}{\Delta H_f \Delta T}$$

where T_m – freezing temperature (in K), ΔH_f – latent heat of fusion, ΔT – amount of undercooling at which nucleus is formed.

In *heterogeneous nucleation*, the probability of nucleation occurring at certain preferred sites is much greater than that at other sites. During solidification, inclusions of foreign particles (*inoculants*), walls of container holding the liquid provide preferred sites. Irregularities in crystal structure such as point defects and dislocations possess strain energy. In solid-solid transformation, foreign inclusions, grain boundaries, interfaces, stacking faults and dislocations can act as preferred sites for nucleation as the strain energy associated with them will be reduced. The released strain energy can reduce the energy requirements for free energy change, Δf. Therefore, nucleation proceeds with a smaller critical radius. A majority of reactions are initiated by some type of heterogeneous nucleation which is common among the two types.

For example, consider the nucleation of β from α occurring on a foreign inclusion, δ, as shown in *figure-13*. Considering the force equilibrium in surface tension terms,

$$\gamma_{\alpha\delta} = \gamma_{\alpha\beta}\cos\theta + \gamma_{\beta\delta}$$

where θ is the contact angle. An expression for Δf can be written in terms of volume energy and surface energies as follows:

$$\Delta f^*_{het} = \frac{4\pi\gamma^3_{\alpha\beta}}{3(\Delta g)^2}(2 - 3\cos\theta + \cos^3\theta) = \Delta f^*_{\mathrm{hom}}\ \frac{2 - 3\cos\theta + \cos^3\theta}{4}$$

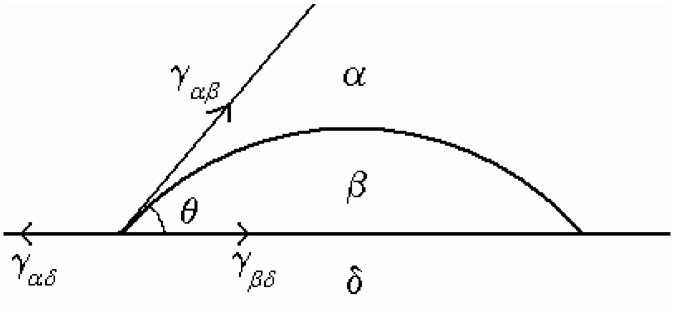


Figure-13: *Schematic of heterogeneous nucleation.*

By comparing the free energy terms for homogeneous and heterogeneous nucleation processes for various contact conditions:

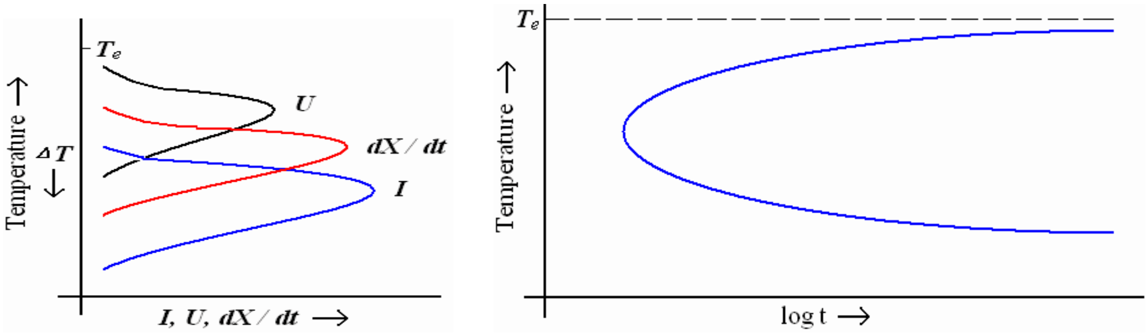
- When product particle makes only a point contact with the foreign surface, i.e. θ = 180°, the foreign particle does not play any role in the nucleation process → $\Delta f^*_{het} = \Delta f^*_{\mathrm{hom}}$
- If the product particle completely wets the foreign surface, i.e. θ = 0°, there is no barrier for heterogeneous nucleation → $\Delta f^*_{het} = 0$
- In intermediate conditions such as where the product particle attains hemispherical shape, θ = 0 → $\Delta f^*_{het} = \frac{1}{2}\Delta f^*_{\mathrm{hom}}$

The above derivations are helpful in selecting a heterogeneous nucleation agent. It shows that a small contact angle is very helpful in heterogeneous nucleation. For a system of ***α* – *β*** interface, θ can be minimized by choosing ***δ*** such that energy of ***β*- *δ*** interface is kept to minimum. If the crystal structure two phases are similar and their lattice parameters are nearly equal, energy of the interface between those two phases will be minimum. This criterion is useful in selecting a agent for heterogeneous nucleation.

7.2.2 Growth kinetics

Many transformations occur as a result of continuous formation of critical nuclei in the parent phase and the subsequent growth of the particles. Growth is the increase in size of the particle after it has nucleated i.e. growth kinetics become important once an embryo has exceeded the critical size and become a stable nucleus. Growth may proceed in two radically different manners. In one type of growth, individual atoms move independently from the parent to the product phase, thus it is diffusion controlled and is thermally activated. In the other type of growth that occurs in solid-solid transformations many atoms move cooperatively without thermal assistance. Growth that is diffusion controlled is more common the other.

Growth usually occurs by the thermally activated jump of atoms from the parent phase to the product phase. The unit step in the growth process thus consists of an atom leaving the parent phase and jumping across the interface to join the product phase. At the equilibrium temperature, both phases have the same free energy, hence the frequency of jumps from parent phase to product phase will be equal to that from product phase to parent phase i.e. the net growth rate is zero. At lower temperatures, product phase is expected to have lower free energy, and thus a net flow of atoms from parent phase to product phase. This net flux of atoms results in interface motion i.e. growth rate is taken as the rate of increase of a linear dimension of a growing particle. As a function of temperature, the growth rate first increases with increasing degree of supercooling, but eventually slows-down as thermal energy decreases. This is same as for nucleation; however the maximum in the growth rate usually occurs at a higher temperature than the maximum in the nucleation rate. *Figure-14(a)* depicts the temperature dependence of nucleation rate (*U*), growth rate (*I*) and overall transformation rate (*dX/dt*) that is a function of both nucleation rate and growth rate i.e. *dX/dt= fn (U, I)*. On the other-hand, the time required for a transformation to completion has a reciprocal relationship to the overall transformation rate. Temperature dependence of this time is shown in *figure-14 (b)*. The *C-curve* shown in *figure-14(b)* is characteristic of all thermally activated nucleation and growth transformations involving the transformation of a high-temperature phase to low-temperature phase. This curve is also known as *time-temperature-transformation (TTT) curve*. The nose of the C-curve corresponds to the minimum time for a specified fraction of transformation. It is also the place where overall transformation rate is a maximum.



(a) **(b)**

Figure-14: *(a) Temperature dependence of rates, (b) Time dependence of transformation as a function of temperature.*

In many investigations, the fraction of transformation that has occurred is measured as a function of time, while the temperature is maintained constant. Transformation progress is usually ascertained by either microscopic examination *or* measurement of some physical property. Transformation data are plotted as the fraction of transformed material versus the logarithm of time, which results in characteristic *S-curve* (*figure-15*), largely because diffusion plays such an important role in both nucleation and growth. Conditions under which transformation may take place greatly affect the nature of the resulting microstructure. At small degrees of supercooling, where slow nucleation and rapid growth prevail, relatively coarse particles appear; at larger degrees of supercooling, relatively fine particles result. The time dependence of the transformation rate is an important consideration in the heat treatment of materials.

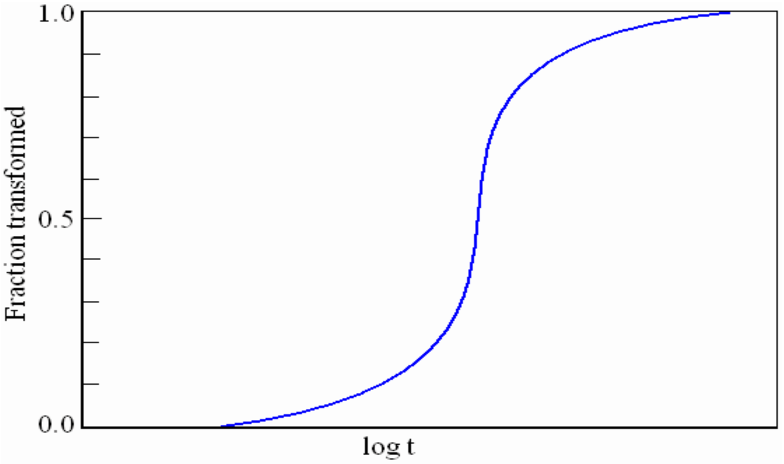


Figure-15: *Fraction of transformation Vs the logarithm of time at constant temperature.*

The other kind of growth involves congruent transformation which is considered as diffusion-less because it takes place at a rate approaching the speed of sound. It can be visualized as a cooperative type of process in which, without aid of thermal activation, atoms move into new locations because of the strain energy resulting from like movements of adjacent atoms. However the strains set-up in the parent phase may impede the further transformation, thus a lower temperature or mechanical deformation may be required to complete this *martensitic* transformation. The cooperative displacement of atoms here resembles a shear process during which, for example, FCC structure of Co transforms into HCP-Co *or* FCC-austenite into BCT-martensite. This merely requires that atoms in the FCC-phase move a fraction of an inter-atomic distance.

Because of its crystallographic nature, a martensitic transformation only occurs in the solid state. In addition, the crystal structure of the product phase must be easily generated from that of the parent phase without diffusive motion of atoms. This is true for most allotropic transformations in metals that occur at low temperatures or for high-temperature transformations of metals brought about by a quench. Reasons for martensitic transformation: (a) the free energy difference between the high-temperature phase and low-temperature phases becomes increasingly negative with decreasing temperature (b) the crystal structures of allotropes of a metal are relatively simple and share similar features with each other. It is also said that diffusion-controlled nucleation and growth and a martensitic change are competitive processes in many cases.

The martensitic transformation starts at a temperature designated M_s, which is generally below the equilibrium temperature, T_e. The transformation is completed at a lower temperature, M_f. The amount of parent phase transformed into product phase depends on temperature only, and is independent of time. Furthermore, in most cases, M_s temperature and the fractional amount of product phase as a function of temperature are independent of quenching rate. Consequently, M_s and M_f are presented as horizontal lines on a TTT diagram. Catalytic effect of cold working can be used to make M_s approach T_e. Martensitic transformations in Fe-C alloys and Ti are of great technological importance. Fe-C alloy transformations are dealt in detail in the following sections.

7.3 The iron – carbon system, phase transformations

A study of iron-carbon system is useful and important in many respects. This is because (1) steels constitute greatest amount of metallic materials used by man (2) solid state transformations that occur in steels are varied and interesting. These are similar to those occur in many other systems and helps explain the properties.

Iron-carbon phase diagram shown in *figure-16* is not a complete diagram. Part of the diagram after 6.67 wt% C is ignored as it has little commercial significance. The 6.67%C represents the composition where an inter-metallic compound, cementite (Fe_3C), with solubility limits forms. In addition, phase diagram is not true equilibrium diagram because cementite is not an equilibrium phase. However, in ordinary steels decomposition of cementite into graphite never observed because nucleation of cementite is much easier than that of graphite. Thus cementite can be treated as an equilibrium phase for practical purposes.

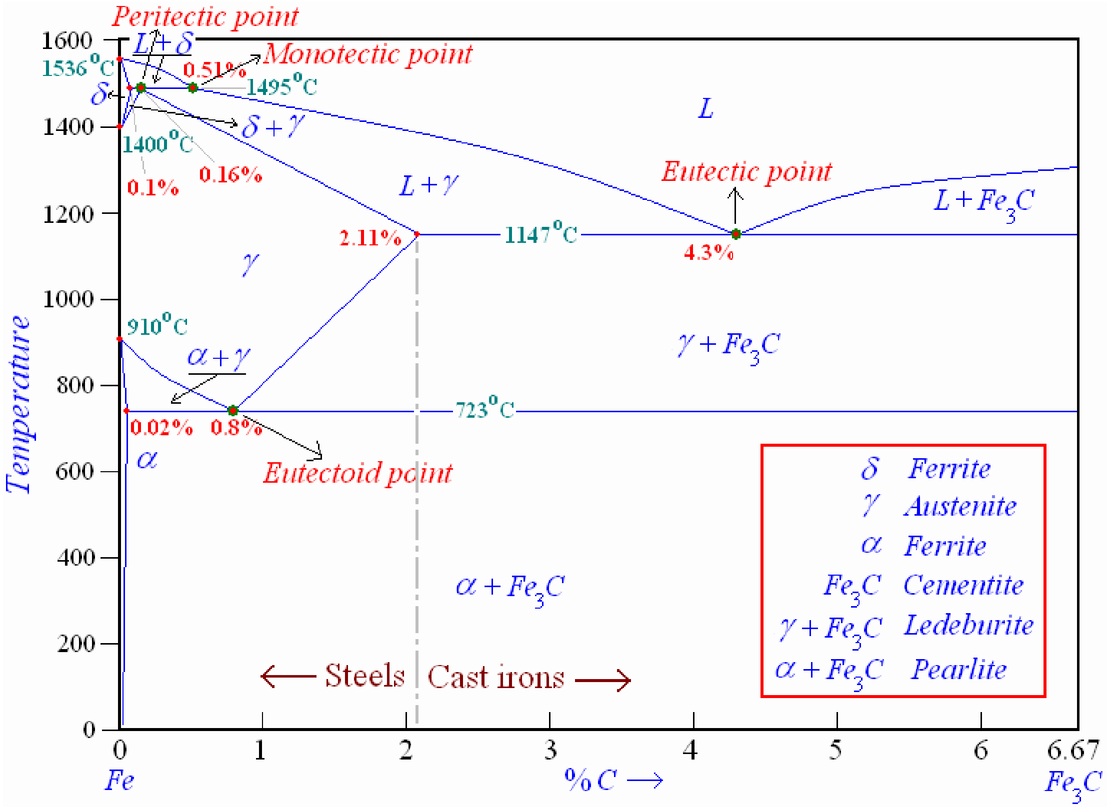


Figure-16: *Iron – Iron carbide phase diagram.*

The Fe-Fe_3C is characterized by five individual phases and four invariant reactions. Five phases that exist in the diagram are: ***α***–ferrite (BCC) Fe-C solid solution, ***γ***-austenite (FCC) Fe-C solid solution, ***δ***-ferrite (BCC) Fe-C solid solution, Fe_3C (iron carbide) *or* cementite - an inter-metallic compound and liquid Fe-C solution. Four invariant reactions that cause transformations in the system are namely eutectoid, eutectic, monotectic and peritectic.

As depicted by left axes, pure iron upon heating exhibits two allotropic changes. One involves ***α***–ferrite of BCC crystal structure transforming to FCC austenite, ***γ***-iron, at 910 C. At 1400 C, austenite changes to BCC phase known as ***δ***-ferrite, which finally melts at 1536 C.

Carbon present in solid iron as interstitial impurity, and forms solid solution with ferrites / austenite as depicted by three single fields represented by ***α***, ***γ*** and ***δ***. Carbon dissolves least in ***α***–ferrite in which maximum amount of carbon soluble is 0.02% at 723 C. This limited solubility is attributed to shape and size of interstitial position in BCC ***α***–ferrite. However, carbon present greatly influences the mechanical properties of ***α***–ferrite. ***α***–ferrite can be used as magnetic material below 768 C. Solubility of carbon in ***γ***-iron reaches its maximum, 2.11%, at a temperature of 1147 C. Higher solubility of carbon in austenite is attributed to FCC structure and corresponding interstitial sites. Phase transformations involving austenite plays very significant role in heat treatment of different steels. Austenite itself is non-magnetic. Carbon solubility in ***δ***-ferrite is maximum (0.1%) at 1495 C. As this ferrite exists only at elevated temperatures, it is of no commercial importance. Cementite, Fe_3C an inter-metallic compound forms when amount of carbon present exceeds its solubility limit at respective temperatures. Out of these four solid phases, cementite is hardest and brittle that is used in different forms to increase the strength of steels. $\boldsymbol{\alpha}$–ferrite, on the other hand, is softest and act as matrix of a composite material. By combining these two phases in a solution, a material's properties can be varied over a large range.

For technological convenience, based on %C dissolved in it, a Fe-C solution is classified as: *commercial pure irons* with less than 0.008%C; *steels* having %C between 0.008-2.11; while *cast irons* have carbon in the range of 2.11%-6.67%. Thus commercial pure iron is composed of exclusively $\boldsymbol{\alpha}$–ferrite at room temperature. Most of the steels and cast irons contain both $\boldsymbol{\alpha}$–ferrite and cementite. However, commercial cast irons are not simple alloys of iron and carbon as they contain large quantities of other elements such as silicon, thus better consider them as ternary alloys. The presence of Si promotes the formation of graphite instead of cementite. Thus cast irons may contain carbon in form of both graphite and cementite, while steels will have carbon only in combined from as cementite.

As shown in *figure-16*, and mentioned earlier, Fe-C system constitutes four invariant reactions:

- peritectic reaction at 1495 °C and 0.16%C, $\boldsymbol{\delta}$-ferrite + $\boldsymbol{L} \leftrightarrow \boldsymbol{\gamma}$-iron (austenite)
- monotectic reaction 1495 °C and 0.51%C, $\boldsymbol{L} \leftrightarrow \boldsymbol{L} + \boldsymbol{\gamma}$-iron (austenite)
- eutectic reaction at 1147 °C and 4.3 %C, $\boldsymbol{L} \leftrightarrow \boldsymbol{\gamma}$-iron + Fe_3C (cementite) [ledeburite]
- eutectoid reaction at 723 °C and 0.8%C, $\boldsymbol{\gamma}$-iron $\leftrightarrow$ $\boldsymbol{\alpha}$–ferrite + Fe_3C (cementite) [pearlite]

Product phase of eutectic reaction is called ledeburite, while product from eutectoid reaction is called pearlite. During cooling to room temperature, ledeburite transforms into pearlite and cementite. At room temperature, thus after equilibrium cooling, Fe-C diagram consists of either $\boldsymbol{\alpha}$–ferrite, pearlite and/or cementite. Pearlite is actually not a single phase, but a micro-constituent having alternate thin layers of $\boldsymbol{\alpha}$–ferrite (~88%) and Fe_3C, cementite (~12%). Steels with less than 0.8%C (mild steels up to 0.3%C, medium carbon steels with C between 0.3%-0.8% i.e. hypo-eutectoid Fe-C alloys) i.e. consists pro-eutectoid $\boldsymbol{\alpha}$–ferrite in addition to pearlite, while steels with carbon higher than 0.8% (high-carbon steels i.e. hyper-eutectoid Fe-C alloys) consists of pearlite and pro-eutectoid cementite. Phase transformations involving austenite i.e. processes those involve eutectoid reaction are of great importance in heat treatment of steels.

In practice, steels are almost always cooled from the austenitic region to room temperature. During the cooling upon crossing the boundary of the single phase $\boldsymbol{\gamma}$-iron, first pro-eutectoid phase (either $\boldsymbol{\alpha}$–ferrite or cementite) forms up to eutectoid temperature. With further cooling below the eutectoid temperature, remaining austenite decomposes to eutectoid product called pearlite, mixture of thin layers of $\boldsymbol{\alpha}$–ferrite and cementite. Though pearlite is not a phase, nevertheless, a constituent because it has a definite appearance under the microscope and can be clearly identified in a structure composed of several constituents. The decomposition of austenite to form pearlite occurs by nucleation and growth. Nucleation, usually, occurs heterogeneously and rarely homogeneously at grain boundaries. When it is not homogeneous, nucleation of pearlite occurs both at grain boundaries and in the grains of austenite. When austenite forms pearlite at a constant temperature, the spacing between adjacent lamellae of cementite is very nearly constant. For a given colony of pearlite, all cementite plates have a common orientation in space, and it is also true for the ferrite plates. Growth of pearlite colonies occurs not only by the nucleation of additional lamellae but also through an advance at the ends of the lamellae. Pearlite growth also involves the nucleation of new colonies at the interfaces between established colonies and the parent austenite. The thickness ratio of the ferrite and cementite layers in pearlite is approximately 8 to 1. However, the absolute layer thickness depends on the temperature at which the isothermal transformation is allowed to occur.

The temperature at which austenite is transformed has a strong effect on the inter-lamellar spacing of pearlite. The lower the reaction temperature, the smaller will be inter-lamellar spacing. For example, pearlite spacing is in order of 10^{-3} mm when it formed at 700 C, while spacing is in order of 10^{-4} mm when formed at 600 C. The spacing of the pearlite lamellae has a practical significance because the hardness of the resulting structure depends upon it; the smaller the spacing, the harder the metal. The growth rate of pearlite is also a strong function of temperature. At temperatures just below the eutectoid, the growth rate increases rapidly with decreasing temperature, reaching a maximum at 600 C, and then decreases again at lower temperatures.

Additions of alloying elements to Fe-C system bring changes (alternations to positions of phase boundaries and shapes of fields) depends on that particular element and its concentration. Almost all alloying elements causes the eutectoid concentration to decrease, and most of the alloying elements (e.g.: Ti, Mo, Si, W, Cr) causes the eutectoid temperature to increase while some other (e.g.: Ni, Mn) reduces the eutectoid temperature. Thus alloying additions alters the relative amount of pearlite and pro-eutectoid phase that form.

Fe-C alloys with more than 2.11% C are called cast irons. Phase transformations in cast irons involve formation of pro-eutectic phase on crossing the liquidus. During the further cooling, liquid of eutectic composition decomposes in to mixture of austenite and cementite, known as ledeburite. On further cooling through eutectoid temperature, austenite decomposes to pearlite. The room temperature microstructure of cast irons thus consists of pearlite and cementite. Because of presence of cementite, which is hard, brittle and white in color, product is called *white cast iron*. However, depending on cooling rate and other alloying elements, carbon in cast iron may be present as graphite or cementite. *Gray cast iron* contains graphite in form of flakes. These flakes are sharp and act as stress risers. Brittleness arising because of flake shape can be avoided by producing graphite in spherical nodules, as in *malleable cast iron* and *SG (spheroidal graphite) cast iron*. Malleable cast iron is produced by heat treating white cast iron (Si < 1%) for prolonged periods at about 900 ̊C and then cooling it very slowly. The cementite decomposes and temper carbon appears approximately as spherical particles. SG iron is produced by adding inoculants to molten iron. In these Si content must be about 2.5%, and no subsequent heat treatment is required.

7.4 Transformation rate effects and TTT diagrams, Microstructure and Property Changes in Fe-C Alloys

Solid state transformations, which are very important in steels, are known to be dependent on time at a particular temperature, as shown in *figure-14(b).* Isothermal transformation diagram, also known as TTT diagram, measures the rate of transformation at a constant temperature i.e. it shows time relationships for the phases during isothermal transformation. Information regarding the time to start the transformation and the time required to complete the transformation can be obtained from set of TTT diagrams. One such set of diagram for reaction of austenite to pearlite in steel is shown in *figure-17*. The diagram is not complete in the sense that the transformations of austenite that occur at temperatures below about 550 °C are not shown.

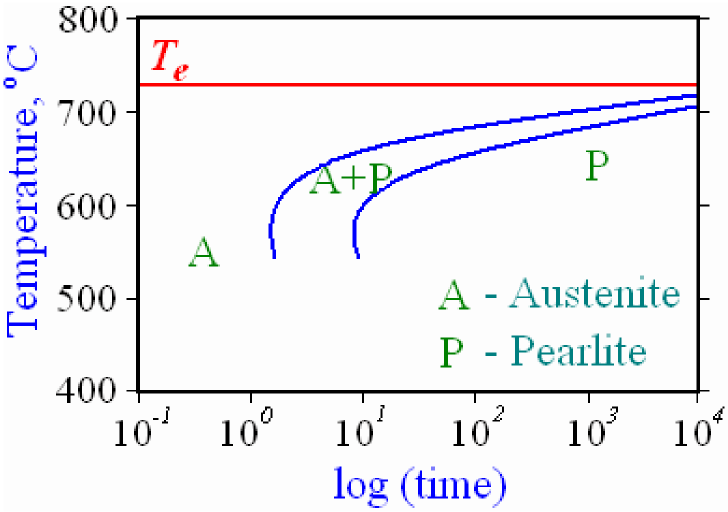


Figure-17: *Partial TTT diagram for a eutectoid Fe-C alloy.*

As mentioned in previous section, thickness of layers in pearlite depends on the temperature at which the transformation occurred. If the transformation took place at a temperature that is just below the eutectoid temperature, relatively thick layers of ***α***–ferrite and cementite are produced in what is called *coarse pearlite*. This is because of high diffusion rates of carbon atoms. Thus with decreasing transformation temperature, sluggish movement of carbon results in thinner layers ***α***–ferrite and cementite i.e. *fine pearlite* is produced.

At transformation temperatures below 550 C, austenite results in different product known as bainite. Bainite also consists of ***α***–ferrite and cementite phases i.e. transformation is again diffusion controlled but morphologically it consists of very small particles of cementite within or between fine ferrite plates. Bainite forms needles or plates, depending on the temperature of the transformation; the microstructural details of bainite are so fine that their resolution is only possible using electron microscope. It differs from pearlite in the sense that different mechanism is involved in formation ob bainite which does not have alternating layers of ***α***–ferrite and cementite. In addition, because of equal growth rates in all directions pearlite tends to form spherical colonies, whereas bainite grows as plates and has a characteristic acicular (needlelike) appearance. *Upper bainite,* formed at the upper end of the temperature range (550 C-350 C), is characterized by relatively coarse, irregular shaped cementite particles in ***α***–ferrite plates. If the transformation is taking place at lower temperatures (350 C-250 C), the ***α***–ferrite plates assume a more regular needlelike shape, and the transformation product is called *lower bainite*. At the same time carbide particles become smaller in size and appear as cross-striations making an angle of about 55 to the axis of the ***α***–ferrite plate. Upper bainite has large rod-like cementite regions, whereas lower bainite has much finer cementite particles as a result of sluggish diffusion of carbon atoms at lower temperatures. Lower bainite is considerably harder than upper bainite. Another characteristic of bainite is that as it has crystallographic orientation that is similar to that found in simple ferrite nucleating from austenite, it is believed that bainite is nucleated by the formation of ferrite. This is in contrast to pearlite which is believed to be nucleated by formation of cementite.

Basically, bainite is a transformation product that is not as close to equilibrium as pearlite. The most puzzling feature of the bainite reaction is its dual nature. In a number of respects, it reveals properties that are typical of a nucleation and growth type of transformation such as occurs in the formation pearlite and also a mixture of ***α***–ferrite and cementite though of quite different morphology (no alternate layers), but at the same time it differs from the Martensite as bainite formation is athermal and diffusion controlled though its microstructure is characterized by acicular (needlelike) appearance.

The time-temperature dependence of the bainite transformation can also be presented using TTT diagram. It occurs at temperatures below those at which pearlite forms i.e. it does not form until the transformation temperature falls below a definite temperature, designated as B_s. Above this temperature austenite does not form bainite except under external stresses. Below B_s, austenite does not transform completely to bainite. The amount of bainite formed increases as the isothermal reaction temperature is lowered. By reaching a lower limiting temperature, B_f, it is possible to transform austenite completely to bainite. The B_s and B_f temperatures are equivalent to the M_s and M_f temperatures for Martensite.

In simple eutectoid steels, pearlite and bainite transformations overlap, thus transition from the pearlite to bainite is smooth and continuous i.e. knees of individual pearlite and bainite curves are merged together. However each of the transformations has a characteristic C-curve, which can be distinguishable in presence of alloying elements. As shown in complete TTT diagram for eutectoid steel in *figure-18*, above approximately 550 C-600 C, austenite transforms completely to pearlite. Below this range up to 450 C, both pearlite and bainite are formed. Finally, between 450 °C and 210 C, the reaction product is bainite only. Thus bainite transformation is favored at a high degree of supercooling, and the pearlite transformation at a low degree of supercooling. In middle region, pearlitic and bainitic transformations are competitive with each other.

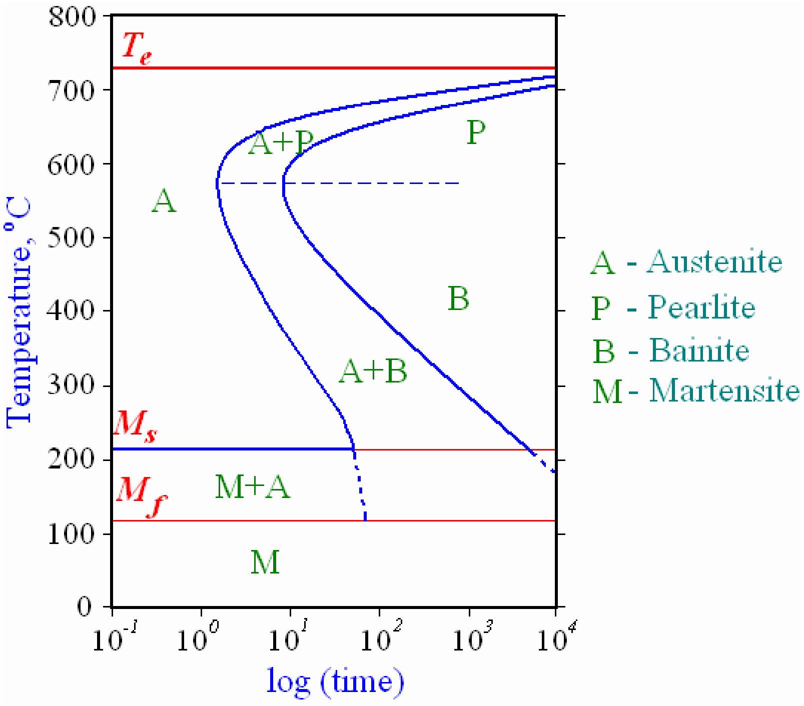


Figure-18: *Complete TTT (isothermal transformation) diagram for eutectoid steel.*

As explained in earlier section, martensitic transformation can dominate the proceedings if steel is cooled rapid enough so that diffusion of carbon can be arrested. Transformation of austenite to Martensite is diffusion-less, time independent and the extent of transformation depends on the transformation temperature. Martensite is a *meta-stable* phase and decomposes into ferrite and pearlite but this is extremely slow (and not noticeable) at room temperature. Alloying additions retard the formation rate of pearlite and bainite, thus rendering the martensitic transformation more competitive. Start of the transformation is designated by M_s, while the completion is designated by M_f in a transformation diagram. Martensite forms in steels possesses a body centered tetragonal crystal structure with carbon atoms occupying one of the three interstitial sites available. This is the reason for characteristic structure of steel Martensite instead of general BCC. Tetragonal distortion caused by carbon atoms increases with increasing carbon content and so is the hardness of Martensite. Austenite is slightly denser than Martensite, and therefore, during the phase transformation upon quenching, there is a net volume increase. If relatively large pieces are rapidly quenched, they may crack as a result of internal stresses, especially when carbon content is more than about 0.5%.

Mechanically, Martensite is extremely hard, thus its applicability is limited by brittleness associated with it. Characteristics of steel Martensite render it unusable for structural applications in the as-quenched form. However, structure and thus the properties can be altered by *tempering*, heat treatment observed below eutectoid temperature to permit diffusion of carbon atoms for a reasonable period of time. During tempering, carbide particles attain spherical shape and are distributed in ferrite phase – structure called s*pheroidite*. Spheroidite is the softest yet toughest structure that steel may have. At lower tempering temperature, a structure called *tempered Martensite* forms with similar microstructure as that of spheroidite except that cementite particles are much, much smaller. The tempering heat treatment is also applicable to pearlitic and bainitic structures. This mainly results in improved machinability. The mechanism of tempering appears to be first the precipitation of fine particles of hexagonal ε-carbide of composition about $Fe_{2.4}C$ from Martensite, decreasing its tetragonality. At higher temperatures or with increasing tempering times, precipitation of cementite begins and is accompanied by dissolution of the unstable ε-carbide. Eventually the Martensite loses its tetragonality and becomes BCC ferrite, the cementite coalesces into spheres. A schematic of possible transformations involving austenite decomposition are shown in *figure-19*.

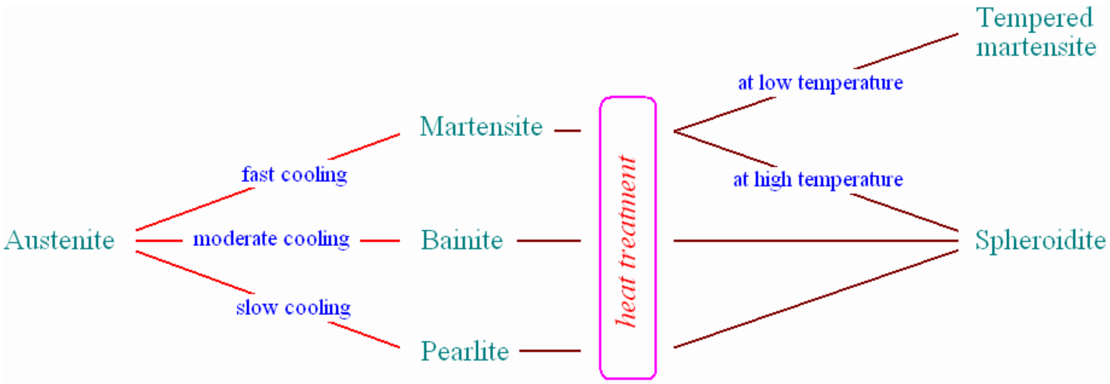


Figure-19: *Possible transformation involving austenite decomposition*.

Tempering of some steels may result in a reduction of toughness what is known as *temper embrittlement*. This may be avoided by (1) compositional control, and/or (2) tempering above 575 *or* below 375, followed by quenching to room temperature. The effect is greatest in Martensite structures, less severe in bainitic structures and least severe in pearlite structures. It appears to be associated with the segregation of solute atoms to the grain boundaries lowering the boundary strength. Impurities responsible for temper brittleness are: P, Sn, Sb and As. Si reduces the risk of embrittlement by carbide formation. Mo has a stabilizing effect on carbides and is also used to minimize the risk of temper brittleness in low alloy steels.

TTT diagrams are less of practical importance since an alloy has to be cooled rapidly and then kept at a temperature to allow for respective transformation to take place. However, most industrial heat treatments involve continuous cooling of a specimen to room temperature. Hence, Continuous Cooling Transformation (CCT) diagrams are generally more appropriate for engineering applications as components are cooled (air cooled, furnace cooled, quenched etc.) from a processing temperature as this is more economic than transferring to a separate furnace for an isothermal treatment. CCT diagrams measure the extent of transformation as a function of time for a continuously decreasing temperature. For continuous cooling, the time required for a reaction to begin and end is delayed, thus the isothermal curves are shifted to longer times and lower temperatures.

Both TTT and CCT diagrams are, in a sense, phase diagrams with added parameter in form of time. Each is experimentally determined for an alloy of specified composition. These diagrams allow prediction of the microstructure after some time period for constant temperature and continuous cooling heat treatments, respectively. Normally, bainite will not form during continuous cooling because all the austenite will have transformed to pearlite by the time the bainite transformation has become possible. Thus, as shown in *figure-20*, region representing austenite-pearlite transformation terminates just below the nose.

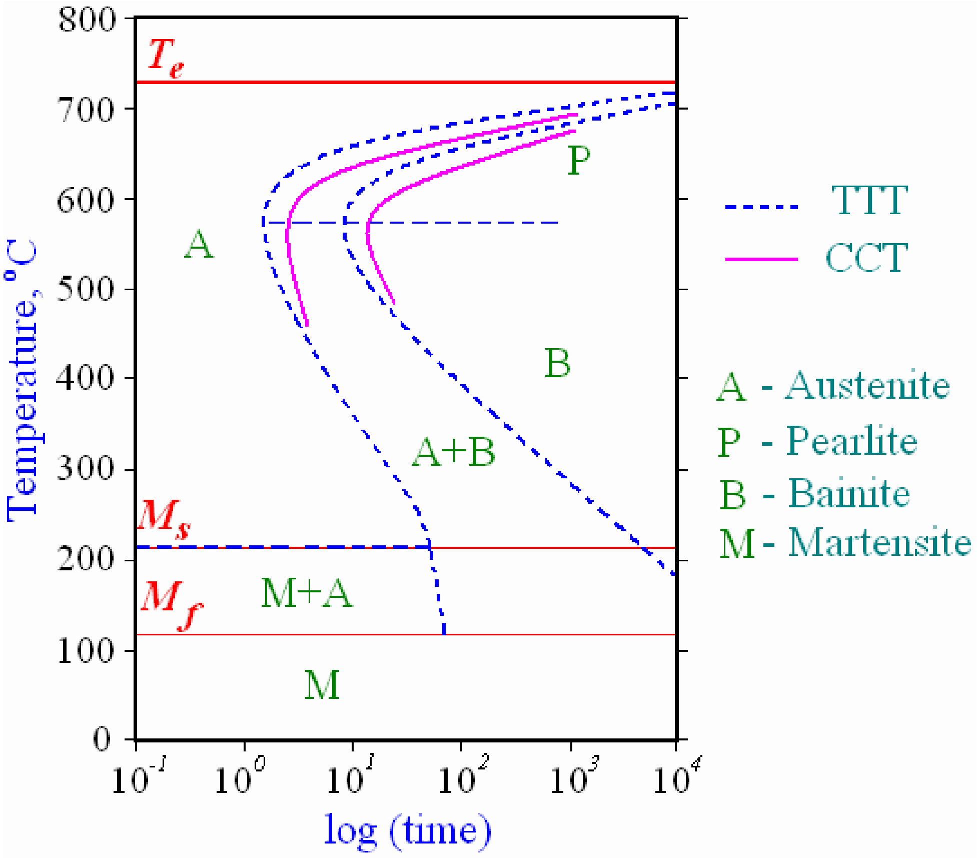


Figure-20: *Superimposition of TTT and CCT diagrams for a eutectoid steel*.

References

1. G. E. Dieter, Mechanical Metallurgy, Third Edition, McGraw-Hill, New York, 1986.
2. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
3. D. Hull, Introduction of Dislocations, Third Edition, Butterworth-Heinemann, Woburn, UK, 1984

Chapter 8. Failure

Failure can be defined, in general, as an event that does not accomplish its intended purpose. Failure of a material component is the loss of ability to function normally. Components of a system can fail one of many ways, for example excessive deformation, fracture, corrosion, burning-out, degradation of specific properties (thermal, electrical, or magnetic), etc. Failure of components, especially, structural members and machine elements can lead to heavy loss of lives, wealth and even may jeopardize the society! This chapter deals with the study of failures by mechanical means i.e. application stresses.

Even though the causes of failure are known, prevention of failure is difficult to guarantee. Causes for failure include: improper materials selection, improper processing, inadequate design, misuse of a component, and improper maintenance. It's the engineer's responsibility to anticipate and prepare for possible failure; and in the event of failure, to assess its cause and then take preventive measures.

Structural elements and machine elements can fail to perform their intended functions in three general ways: excessive elastic deformation, excessive plastic deformation or yielding, and fracture. Under the category of failure due to *excessive elastic deformation*, for example: too flexible machine shaft can cause rapid wear of bearing. On the other hand sudden *buckling* type of failure may occur. Failures due to excessive elastic deformation are controlled by the modulus of elasticity, not by the strength of the material. The most effective way to increase stiffness of a component is by tailoring the shape *or* dimensions. *Yielding* or *plastic deformation* may render a component useless after a certain limit. This failure is controlled by the yield strength of the material. At room temperature, continued loading over the yielding point may lead to strain hardening followed by fracture. However at elevated temperatures, failure occurs in form of time-dependent yielding known as *creep*. *Fracture* involves complete disruption of continuity of a component. It starts with initiation of a crack, followed by crack propagation. Fracture of materials may occur in three ways – brittle/ductile fracture, fatigue or progressive fracture, delayed fracture. *Ductile/brittle fracture* occurs over short period of time, and distinguishable. *Fatigue* failure is the mode in which most machine parts fail. Fatigue, which is caused by a critical localized tensile stress, occurs in parts which are subjected to alternating or fluctuating stress. *Stress-rupture* occurs when a metal has been statically loaded at an elevated temperature for a long time, and is best example for delayed fracture.

8.1 Fracture, Ductile and Brittle fracture

8.1.1 Fracture

Fracture is a form of failure, and is defined as the separation or fragmentation of a solid body into two or more parts under the action of stress. Fracture that occurs over a very short time period and under simple loading conditions (static i.e. constant or slowly changing) is considered here. Fracture under complex condition, for example alternating stress, is considered in later sections.

The process of fracture can be considered to be made up of two components, crack initiation followed by crack propagation. Fractures are classified *w.r.t.* several characteristics, for example, strain to fracture, crystallographic mode of fracture, appearance of fracture, etc. *Table-8.1* gives a brief summary of different fracture modes.

Table-8.1: *Different fracture modes.*

characteristic	***terms used***	
Strain to fracture	Ductile	Brittle
Crystallographic mode	Shear	Cleavage
Appearance	Fibrous and gray	Granular and bright
Crack propagation	Along grain boundaries	Through grains

Shear fracture, promoted by shear stresses, occurs as result of extensive slip on active slip plane. On the other hand, *cleavage fracture* is controlled by tensile stresses acting normal to cleavage plane. A shear fracture surface appears gray and fibrous, while a cleavage fracture surface appears bright or granular. Actual fracture surfaces often appear as mixture of fibrous and granular mode. Based on metallographic examination of fracture surfaces of polycrystalline materials, they are classified as either transgranular or intergranular. *Transgranular fracture*, as the name go by, represents crack propagation through the grains, whereas *intergranular fracture* represents the crack that propagated along the grain boundaries.

The fracture is termed ductile or brittle depending on the ability of a material to undergo plastic deformation during the fracture. A *ductile fracture* is characterized by considerable amount of plastic deformation prior to and during the crack propagation. On the other hand, *brittle fracture* is characterized by micro-deformation or no gross deformation during the crack propagation. Plastic deformation that occurs during ductile fracture, if monitored, can be useful as warning sign to the fracture that may occur in later stages. Thus brittle fracture shall be avoided as it may occur without warning!

Since deformation of a material depends on many conditions such as stress state, rate of loading, ambient temperature, crystal structure; *ductile and brittle are relative terms*. Thus the boundary between a ductile and brittle fracture is arbitrary and depends on the situation being considered. A change from the ductile to brittle type of fracture is promoted by a decrease in temperature, an increase in the rate of loading, and the presence of complex state of stress (for example, due to a notch). Under the action of tensile stresses, most metallic materials are ductile, whereas ceramics are mostly brittle, while polymers may exhibit both types of fracture. Materials with BCC *or* HCP crystal structure can be expected to experience brittle fracture under normal conditions, whereas materials with FCC crystal structure are expected to experience ductile fracture.

Figure-8.1 depicts characteristic macroscopic fracture profiles. The profile shown in *figure-8.1(a)* is representative of very high ductility represented by close to 100% reduction in cross-sectional area. This kind of failure is usually called *rupture*. It is observed in very soft metals such as pure gold and lead at room temperature and other metals, polymers, glasses at elevated temperatures. Most ductile metals fracture preceded by a moderate amount of necking, followed by formation of voids, cracks and finally shear. This gives characteristic *cup-and-cone fracture* as shown by *figure-8.1(b)*. In this central interior region has an irregular and fibrous appearance. *Figure-8.1(c)* presents the typical profile of brittle fracture which is usually transgranular. It occurs in most ceramics and glasses at room temperature, long-chain polymers below their glass transition temperatures, certain metals and alloys below their ductile-to-brittle transition temperatures.

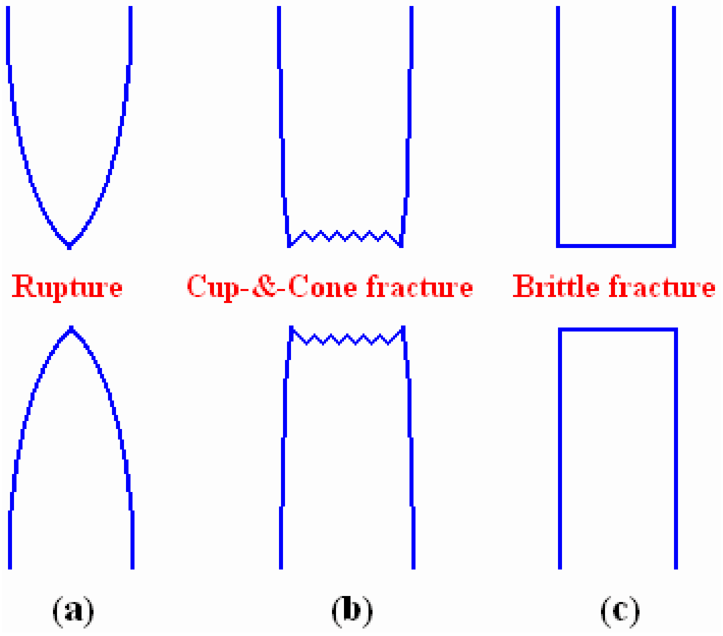


Figure-8.1: *Fracture profiles.*

Detailed and important information on the mechanism of fracture can be obtained from microscopic examination of fracture surfaces. This study is known as *fractography*. This study is most commonly done using SEM (scanning electron microscope). Common microscopic modes of fracture observed include cleavage, quasi-cleavage, and dimpled rupture. Characteristic feature of cleavage fracture is flat facets, and these exhibit *river marking* caused by crack moving through the crystal along number of parallel planes which form a series of plateaus and connecting ledges. Quasi-cleavage fracture is related but distinct from cleavage in the sense that fracture surfaces are not true cleavage planes. This often exhibit dimples and tear ridges around the periphery of the facets. Dimpled rupture is characterized by cup-like depressions whose shape is dependent on stress state. The depressions may be equi-axial, parabolic, or elliptical. This dimpled rupture represents a ductile fracture. *Table-8.2* distinguishes two common modes of fracture.

Table-8.2: *Ductile Vs Brittle fracture*.

Parameter	***Ductile fracture***	***Brittle fracture***
Strain energy required	Higher	Lower
Stress, during cracking	Increasing	Constant
Crack propagation	Slow	Fast
Warning sign	Plastic deformation	None
Deformation	Extensive	Little
Necking	Yes	No
Fractured surface	Rough and dull	Smooth and bright
Type of materials	Most metals (not too cold)	Ceramics, Glasses, Ice

8.1.2 Ductile fracture

Most often ductile fracture in tension occurs after appreciable plastic deformation. It occurs by a slow tearing of the metal with the expenditure of considerable energy. It can be said that ductile fracture in tension is usually preceded by a localized reduction in cross-sectional area, called necking. Further it exhibits three stages - (1) after on set of necking, cavities form, usually at inclusions at second-phase particles, in the necked region because the geometrical changes induces hydrostatic tensile stresses, (2) the cavities grow, and further growth leads to their coalesce resulting in formation of crack that grows outward in direction perpendicular to the application of stress, (3) final failure involves rapid crack propagation at about 45 to the tensile axis. This angle represents the direction of maximum shear stress that causes shear slip in the final stage. During the shear slip, crack propagates at a rapid speed around the outer perimeter of neck leaving one surface in form of cup, and the other in form of cone. Thus it is known as *cup-and-cone fracture*. In this central interior region has an irregular and fibrous appearance, which signifies plastic deformation. Different progressive stages of ductile fracture are shown in *figure-8.2*.

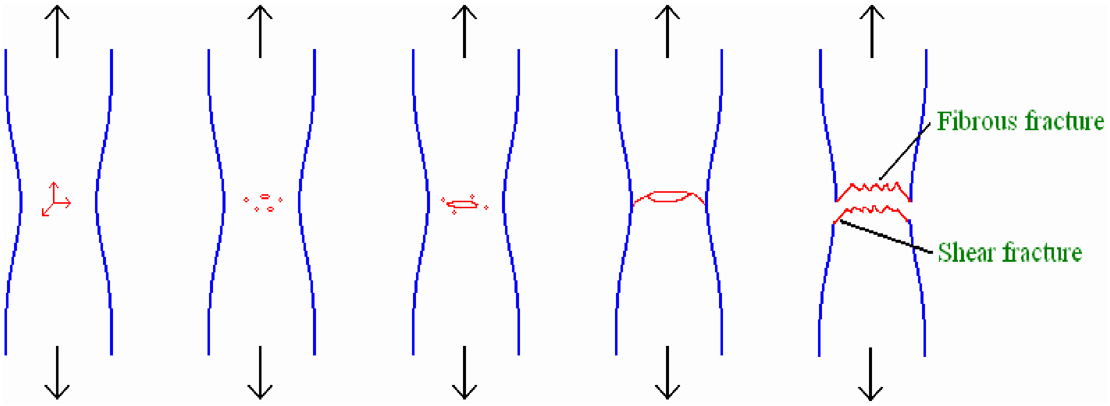


Figure-8.2: *Stages of ductile tensile fracture.*

The voids are thought to be nucleated heterogeneously at sites where further deformation is difficult. These preferred sites mainly consists of foreign inclusions, second-phase particles like oxide particles, or even voids those can form at grain boundary triple points in high-purity metals. It has been observed that concentration of nucleating sites had a strong influence on ductile fracture as true strain to fracture decreases rapidly with increasing volume fraction of second phase particles. In addition, particle shape also has an important influence. When the particles are more spherical than plate-like, cracking is more difficult and the ductility is increased. This is because dislocations can cross slip around spherical particles with ease than around plate-like particles thus avoids buildup of high stresses.

More details of fracture mechanism can be obtained from fractographic study of the fracture surface. At high magnification under microscope, numerous spherical dimples separated by thin walls are found on the ductile fractured surface. This is an indication that surface had formed from numerous holes which were separated by thin walls until it fractures. Dimples formed on shear lip of cup-and-cone fracture will be elongated attaining parabolic shape which is indication that shear failure took place.

Ductile fracture is not particularly important in terms of mechanical behavior because it usually is associated with good toughness. However, McClintock's analytical treatment of ductile fracture resulted in the following equation, which gives strain to fracture, ε_f,

$$\varepsilon_f = \frac{(1-n)\ln(l_0/2b_0)}{\sinh\left[(1-n)(\sigma_a+\sigma_b)\big/(2\bar{\sigma}/\sqrt{3})\right]}$$

for a material with a stress-strain curve given by $\sigma = K\varepsilon^n$, and σ_a, σ_b are stresses parallel and perpendicular to the axis of the cylindrical holes respectively, $\bar{\sigma}$ is the true flow stress, b_0 is initial radius of cylindrical holes and l_0 is the average spacing of holes. The equation indicates that ductility decreases as void fraction increases, as the strain-hardening exponent, n, decreases. This is the consequence of change of stress state from uni-axial to tri-axial tension.

8.1.2 Brittle fracture

The other common mode of fracture is known as brittle fracture that takes place with little *or* no preceding plastic deformation. It occurs, often at unpredictable levels of stress, by rapid crack propagation. The direction of crack propagation is very nearly perpendicular to the direction of applied tensile stress. This crack propagation corresponds to successive and repeated breaking to atomic bonds along specific crystallographic planes, and hence called cleavage fracture. This fracture is also said to be transgranular because crack propagates through grains. Thus it has a grainy or faceted texture. Most brittle fractures occur in a transgranular manner. However, brittle fracture can occur in intergranular manner i.e. crack propagates along grain boundaries. This happens only if grain boundaries contain a brittle film or if the grain-boundary region has been embrittled by the segregation of detrimental elements.

In analogy to ductile fracture, as supported by number of detailed experiments, the brittle fracture in metals is believed to take place in three stages - (1) plastic deformation that causes dislocation pile-ups at obstacles, (2) micro-crack nucleation as a result of build-up of shear stresses, (3) eventual crack propagation under applied stress aided by stored elastic energy.

As mentioned earlier, brittle fracture occurs without any warning sign, thus it needs to be avoided. Hence brittle fracture and its mechanism have been analyzed to a great extent compared to ductile fracture. Brittle fracture usually occurs at stress levels well below those predicted theoretically from the inherent strength due to atomic or molecular bonds. This situation in some respects is analogous to the discrepancy between the theoretical strength shear strength of perfect crystals and their observed lower yield strength values.

When a crystal is subjected to tensile force, separation between atoms will be increased. The repulsive force decreases more rapidly with this than the attractive force, so that a net force between atoms balances the tensile force. With further increase in tensile force, repulsive force continues to decrease. At an instant, repulsive force becomes negligible because of increased separation, which corresponds to theoretical cohesive strength of the material. Assume that inter-atomic spacing of atoms in unstrained condition is ***a***, and ***x*** is change in mean inter-atomic distance. Strain, $\boldsymbol{\varepsilon}$, is given by

$$\varepsilon = \frac{x}{a}$$

And according to Hooke’s law, if ***E*** – Young’s modulus

$$\sigma = E\varepsilon = \frac{Ex}{a}$$

Cohesive force can be approximated with little or no error using a sine curve as follows (refer to chapter-3: theoretical strength):

$$\sigma = \sigma_{th} \sin \frac{2\pi x}{\lambda}$$

where ***λ*** is wave length of the curve. For small values of ***x***, *sin* ***x*** ≈ ***x***, and thus

$$\sigma = \sigma_{th} \frac{2\pi x}{\lambda}$$

If Hooke’s law equation is substituted in the above equation,

$$\sigma_{th} = \frac{\lambda}{2\pi} \frac{E}{a} \approx \frac{E}{\pi}$$

In a brittle material, when fracture occurs, work expended goes into creation of two new surfaces, each with a surface energy, ***γ***. Work done per unit area of fracture surface, ***W***, is the area under the stress-strain curve.

$$W = \int_0^{\lambda/2} \sigma_{th} \sin \frac{2\pi x}{\lambda} dx = \frac{\sigma_{th}\lambda}{\pi} = 2\gamma$$

$$\Rightarrow \quad \lambda = \frac{2\pi\gamma}{\sigma_{th}} \qquad \Rightarrow \quad \sigma_{th} = \left(\frac{E\gamma}{a} \right)^{1/2}$$

As brittle fracture involves, crack propagation, let’s assume that a material has an interior crack of length 2***c*** (*or* a surface crack of length ***c***) with radius of curvature as ***ρ*** as shown in *figure-8.3*.

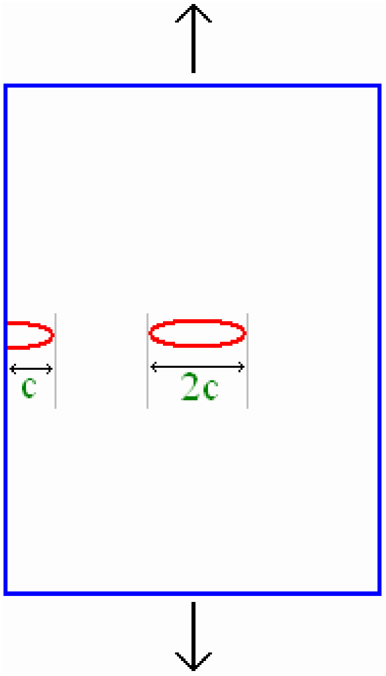


Figure-8.3: *Schematic presentation of interior and surface cracks*.

The maximum stress at the tip of crack is

$$\sigma_{max} = \sigma\left[1 + 2\left(\frac{c}{\rho}\right)^{1/2}\right] \approx 2\sigma\left(\frac{c}{\rho}\right)^{1/2}$$

Before fracture starts, maximum stress at the crack tip shall rise to theoretical value of cohesive strength. Once both are equal, crack propagates. The stress is then can be called *nominal fracture stress*, σ_f, is given by

$$\sigma_f \approx \left(\frac{E\gamma\rho}{4ac}\right)^{1/2}$$

The sharpest crack (i.e. maximum stress concentration) can be represented by $\boldsymbol{\rho} = \boldsymbol{a}$. Thus,

$$\sigma_f \approx \left(\frac{E\gamma}{4c}\right)^{1/2}$$

Griffith theory: Griffith proposed that a brittle material contains number of micro-cracks which causes stress rise in localized regions at a nominal stress which is well below the theoretical value. When one of the cracks spreads into a brittle fracture, it produces an increase in the surface energy of the sides of the crack. Source of the increased surface energy is the elastic strain energy, released as crack spreads. *Griffith's criteria* for propagation of crack include these terms as: a crack will propagate when the decrease in elastic energy is at least equal to the energy required to create the new crack surface. This criterion is useful in determining the tensile stress which will just cause a critical sized crack to propagate as a brittle fracture.

Elastic energy stored under tensile stress will be released as crack propagates. Part of this energy is expended in forming the surface of the crack, while the remaining is transformed into kinetic energy. According to Griffith, such as crack will propagate and produce brittle fracture when an incremental increase in its length doe not change the net energy of the system. Strain energy released in a thin plate of unit thickness is given by Inglis as follows:

$$U_e = \frac{\pi\sigma^2 c^2}{E}$$

where ***E*** is Young's modulus, and is the applied stress. At the same time, surface energy gained by the system due to new surfaces formed as a crack, of length 2***c***, propagates can be given as

$$U_s = 4\gamma c$$

Griffith's criterion can be expressed as follows for incremental change in systems energy with crack length:

$$\frac{\partial U_e}{\partial c} = \frac{\partial U_s}{\partial c} \quad \Rightarrow \quad \frac{2\pi\sigma^2 c}{E} = 4\gamma \quad \Rightarrow \quad \sigma = \left(\frac{2E\gamma}{c\pi}\right)^{1/2}$$

The equation gives the stress required to propagate a crack in a thin plate under plane stress. The stress necessary to cause fracture varies inversely with length of existing cracks, thus largest crack determines the strength of material. The sensitivity of the fracture of solids to surface conditions has been termed *Joffe effect*. For a plate which is thick compared with crack size (i.e. plane strain condition), the stress is given as

$$\sigma = \left(\frac{2E\gamma}{(1-\nu^2)c\pi}\right)^{1/2}$$

where ***ν*** is Poisson's ratio. The Griffith theory applies only to completely brittle materials. In crystalline materials which fracture in an apparently brittle mode, plastic deformation usually occurs next to fracture surface. Orowan, thus, modified the Griffith equation to make it more compatible by including plastic energy term as follows:

$$\sigma = \left(\frac{2E(\gamma + p)}{c\pi}\right)^{1/2} \approx \left(\frac{Ep}{c}\right)^{1/2}$$

where p is the work of plastic deformation at the tip of the growing crack. The surface energy term is neglected in the above equation since plastic work values are in order of 10^2-10^3 J m^{-2} as compared to surface energy values whose range is 1-2 J m^{-2}.

8.2 Fracture mechanics

It is a relatively new section of materials study under mechanical loading conditions. Using fracture mechanics concept it possible to determine whether a crack of given length in a material with known toughness is dangerous at a given stress level. This mechanics section can also provides guide lines for selection of materials and design against fracture failures.

Orowan modified equation is further modified by Irwin to replace the hard to measure plastic work term with other term that was directly measurable. New form of the equation

$$\sigma = \left(\frac{EG_c}{c\pi} \right)^{1/2}$$

where G_c is the critical value of *crack-extension force*.

$$\Rightarrow \qquad G = \frac{\pi\sigma^2 c}{E}$$

G has units of J m^{-2}, and is actually *strain-energy release rate* that is experimentally measurable. The critical value of crack extension force, G_c, makes the crack propagate to fracture, and is considered as one form of *fracture toughness* of the material. Fracture toughness is defined as fracture resistance of a material in the presence of cracks.

Fracture occurs due to *stress concentration* at flaws, like surface scratches, voids, etc. If $\boldsymbol{c}$ is the length of the crack and ρ the radius of curvature at crack tip, the enhanced stress (σ_m) near the crack tip is given by:

$$\sigma_m = 2\sigma \left(\frac{c}{\rho} \right)^{1/2}$$

The above equation states that smaller the radius, higher is the stress enhancement. Another parameter, often used to describe fracture toughness is known as critical *stress concentration factor*, K, and is defined as follows for an infinitely wide plate subjected to tensile stress perpendicular to crack faces:

$$K = \sigma \sqrt{c\pi}$$

This relation holds for specific conditions, and here it is assumed that the plate is of infinite width having a through-thickness crack. It is worth noting that K has the unusual units of MPa√m. It is a material property in the same sense that yield strength is a material property. The stress intensity factor K is a convenient way of describing the stress distribution around a flaw. For the general case the stress intensity factor is given by

$$K = \alpha\sigma\sqrt{c\pi}$$

where $\boldsymbol{\alpha}$ is a parameter that depends on the specimen and crack sizes and geometries, as well as the manner of load application. For example, for a plate of width $\boldsymbol{w}$ loaded in tension with a centrally located crack of length $2\boldsymbol{c}$,

$$K = \alpha\sigma\sqrt{c\pi} = \sigma\sqrt{c\pi}\left(\frac{w}{c\pi}\tan\frac{c\pi}{w}\right)^{1/2}$$

In this regard, it is important to realize that there are three basic modes of fracture, as shown in *figure-8.4*. *Mode-I* corresponds to fracture where the crack surfaces are displaced normal to themselves. This is a typical tensile type of fracture. In *mode-II*, crack surfaces are sheared relative to each other in a direction normal to the edge of the crack. In *mode-III*, shearing action is parallel to the edge of the crack. To indicate different modes, it is normal practice to add the corresponding subscript. The above example described is obviously of the mode-I. There are two extreme cases for mode-I loading. With thin plate-type specimens the stress state is plane stress, while with thick specimens there is a plane-strain condition. The plane-strain condition represents the more severe stress state and the values of K are lower than for plane-stress specimens. The fracture toughness measured under plane-strain conditions is obtained under maximum constraint or material brittleness. The plane strain fracture toughness is designated as K_{Ic}, and is a true material property.

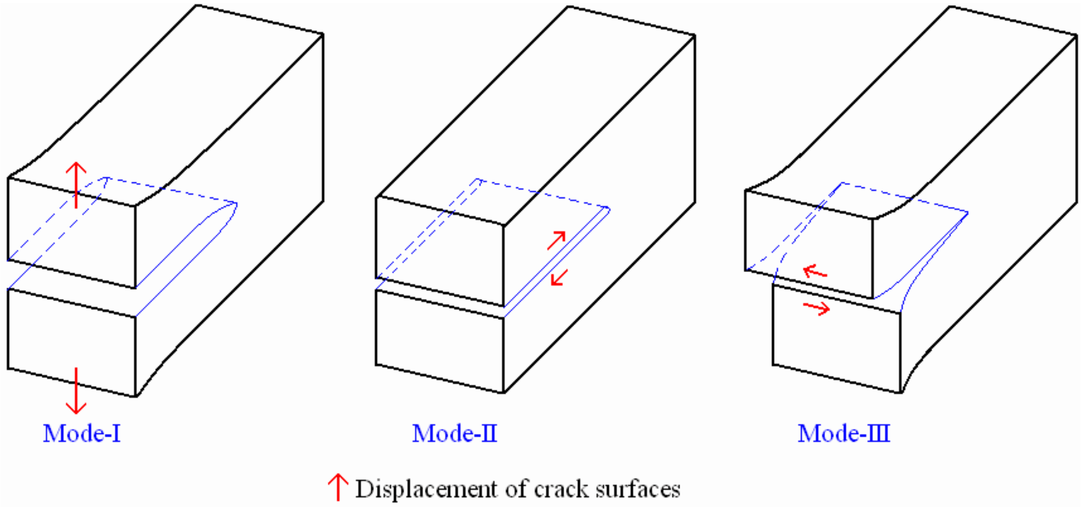


Figure-8.4: *Crack deformation modes.*

While crack extension force, *G*, has more direct physical significance to the fracture process, the stress intensity factor *K* is preferred in working with fracture mechanics because it is more amenable to analytical determination. Both these parameters are related as:

$K^2 = GE$ for plane stress (e.g. thin plate)

$K^2 = GE/(1-\nu^2)$ for plane strain (e.g. thick plate)

Fracture toughness for mode-I, plane strain fracture toughness K_{Ic}, depends on many factors, the most influential of which are temperature, strain rate, and microstructure. The value of K_{Ic} decreases with increasing strain rate and decreasing temperature. Normally values of K_{Ic} increases with reduction in grain size. In addition, an enhancement in yield strength generally produces a corresponding decrease in K_{Ic}. The minimum thickness to achieve plane-strain conditions and valid K_{Ic} measurements is

$$B = 2.5\left(\frac{K_{Ic}}{\sigma_0}\right)^2$$

where σ_0 is the 0.2% offset yield strength.

8.3 Impact fracture testing, Ductile-to-Brittle transition

8.3.1 Impact fracture testing

As mentioned in earlier section, three basic factors contribute to a brittle-cleavage type of fracture. They are (1) tri-axial state of stress (2) a low temperature and (3) a high strain rate or rapid rate of loading. A tri-axial state of stress, such as exists at a notch, and low temperature are responsible for most service failures of brittle type. Since, these effects are accented at a high rate of loading, various types of notched-bar impact tests have been used to determine the susceptibility of materials to brittle fracture.

The changes produced by the introduction of a notch have important consequences in the fracture process. The chief effect of the notch is not in introducing a stress concentration but in producing a tri-axial state of stress at the notch. As a result of this tri-axial state of stress, yield strength of a notched specimen is greater than the uni-axial yield strength because it is more difficult to spread the yielded zone in the presence of tri-axial stresses. The ratio of notched-to-unnotched strength is referred as the *plastic-constraint factor*, ***q***. Unlike stress concentration factor, which can reach values in excess of 10 as the notch is made sharper and deeper, Orowan has shown that the plastic constraint factor cannot exceed a value of 2.57. It results in notch-strengthening of a ductile material. But in a material prone to brittle fracture the increased tensile stresses from the plastic constraint can exceed the critical value for fracture before the material undergoes general plastic yielding. A notch, thus, increases the tendency for brittle fracture in four important ways: (a) By producing high local stresses, (b) by introducing a tri-axial state of stress, (c) by producing high local strain hardening and cracking, and (d) by producing a local magnification to the strain rate.

Two classes of specimens have been standardized for notch-impact testing: (i) Charpy specimen. It has a square cross section and contains 45 V notch, 2 mm deep with a 0.25 mm root radius. The specimen is supported as a horizontal beam, and loaded behind the notch by the impact of heavy swinging pendulum with velocity about 5 m/sec. The energy expended is measured in terms of change in potential energy (height) of the pendulum. The specimen is forced to bend and fracture at a high strain rate in order of 10^3 sec^{-1}. The V-notch in the specimen provides triaxiality of stress, and the high hammer velocity insures a high strain rate. (ii) Izod specimen. It has either circular *or* square cross section and contains V notch near the clamped end. It is supported as over hanging vertical beam. Load is again applied using swinging pendulum but in front of the notch. *Figure-8.5* depicts loading of Charpy and Izod specimens for notch-impact tests.

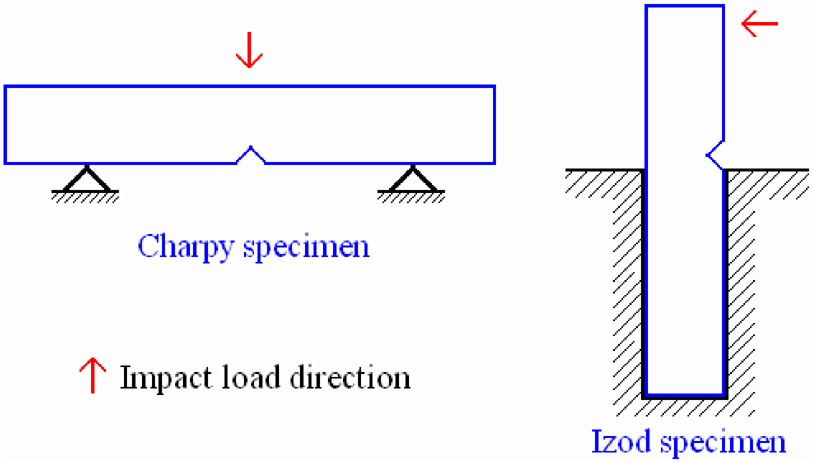


Figure-8.5: *Loading of Charpy and Izod notched-bar impact testing specimens*.

The principal measurement from the impact test is the energy absorbed in fracturing the specimen. Energy expended during fracture is some times known as *notch toughness*. The energy expended will be high for complete ductile fracture, while it is less for brittle fracture. However, it is important to note that measurement of energy expended is only a relative energy, and can not be used directly as design consideration. Another common result from the Charpy test is by examining the fracture surface. It is useful in determining whether the fracture is fibrous (shear fracture), granular (cleavage fracture), or a mixture of both. A third measurement that can be made is the ductility, indicated by the percent contraction of the specimen at the notch. The notched-bar impact test is most useful when conducted over a range of temperatures. Thus it is possible to know at which temperature the ductile-to-brittle transition is taking place. Thus the chief engineering use of Charpy test is in selecting materials which are resistant to brittle fracture by means of transition temperature curves. When selecting a material for design purposes which might be susceptible to brittle fracture, the material with the lowest transition temperature is to be preferred.

Both plane strain fracture toughness and notched-bar impact tests determine the fracture properties of materials. The former are quantitative in nature, in that a specific property is determined. On the other hand, the results of the impact tests are qualitative in nature, thus are of little use fro design purposes. Plane strain fracture toughness tests are not as simple to perform as impact tests, and equipment and specimens are more expensive.

8.3.2 Ductile-to-Brittle transition

It is well understood that ductile and brittle are relative, and thus interchange between these two modes of fracture is achievable with ease. The term *Ductile-to-Brittle transition* (DBT) is used in relation to the temperature dependence of the measured impact energy absorption. For a material, as the temperature is lowered, the impact energy drops suddenly over a relatively narrow temperature range, below which the energy has a considerably lower value as a representative of brittle fracture.

The temperature where DBT occurs is termed as *Ductile-to-Brittle Transition Temperature* (DBTT). A typical variation of energy expended as a function of temperature is shown in *figure-8.6*. From the figure it can be concluded that there is no single criterion that defines the transition temperature. Above the DBTT, the yield strength (σ_y) is lower than the tensile stress necessary to cause brittle failure (σ_f) i.e. $\sigma_y < \sigma_f$. With decreasing temperature, the yield strength increases rapidly to the point where it equals the tensile stress for brittle failure, and below this temperature, fracture usually occurs in brittle/cleavage mode. So, at and below the DBTT, $\sigma_y = \sigma_f$. At the transition temperature, the micro-cracks that form are of critical size for crack propagation, and at lower temperatures these cracks exceed the critical size. Over a temperature range just above DBTT, micro-cracks formed are initially sub-critical so that further plastic deformation and strain hardening must proceed before the tensile stress level becomes sufficient to cause crack propagation. With further increase in temperature, micro-cracks no longer form and fracture mode changes from cleavage to ductile.

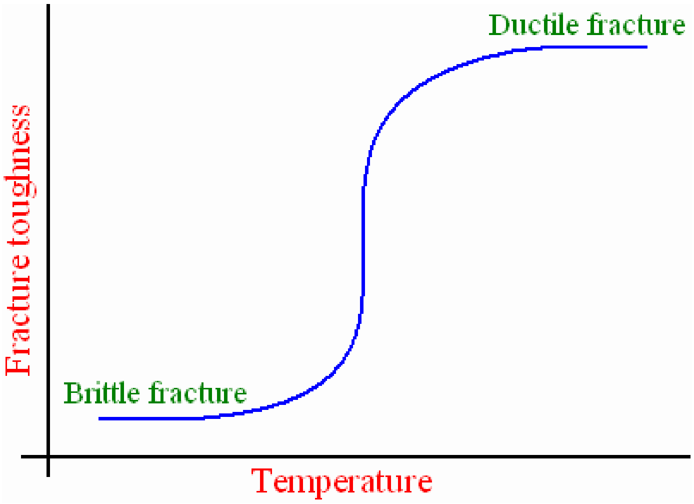


Figure-8.6: *Typical variation of fracture toughness as a function of temperature.*

As shown in the above figure, BCC metals possess a distinct DBTT compared with other metals. Common BCC metals are to become brittle at low temperatures or at extremely high rates of strain. Many FCC metals, on the other hand, remain ductile even at very low temperatures. In metals DBTT is around 0.1-0.2 T_m while in ceramics it is about 0.5-0.7 T_m, where T_m represents absolute melting temperature. The crack propagation stress usually is relatively insensitive to temperature. Strain rate which increases the yield strength but not the crack propagation stress increases the DBTT. As mentioned earlier, DBTT does not have a unique value for a given material, and is a function of several other variables such as flaw size, strain rate, triaxiality of stress, etc. Sharp notches in the specimen provide stress concentration centers and thus increase the DBTT. Fine grained materials have a lower transition temperature as compared to coarse grained materials.

There are other factors which affect the DBTT for a material, for example metallurgical factors. One of the most important is microstructure, which in turn can depend on heat treatment practice and temperature, for example, for steels. The composition also has a very pronounced effect on DBTT. In steels, it is found that impurities like P, Si, Mo, Cr along with C increase the DBTT, while Mn and Ni have reverse effect.

8.4 Fatigue, Crack initiation and propagation, Crack propagation rate

8.4.1 Fatigue

Failures occurring under conditions of dynamic or alternating loading are called *fatigue failures*, presumably because it is generally observed that these failures occur only after a considerable period of service. Fatigue failure usually occurs at stresses well below those required for yielding, or in some cases above the yield strength but below the tensile strength of the material. These failures are dangerous because they occur without any warning. Typical machine components subjected to fatigue are automobile crank-shaft, bridges, aircraft landing gear, etc. Fatigue failures occur in both metallic and non-metallic materials, and are responsible for a large number fraction of identifiable service failures of metals.

A typical fatigue-fracture surface looks like the one shown in *figure-8.7.* The fatigue crack nucleates at the stress concentration. Generally, the fatigue fracture surface is perpendicular to the direction of an applied stress. A fatigue failure can be recognized from the appearance of the fracture surface, which shows a smooth and polished surface that corresponds to the slow growth of crack, when the crack faces smoothen out by constant rubbing against each other and a rough/granular region corresponds to the stage of fast growth, after critical conditions is attained where member has failed in a ductile manner when cross section was no longer able to carry the applied load. The region of a fracture surface that formed during the crack propagation step may be results in characteristic pattern of concentric rings spread over the smooth region of the fracture surface, known as *beach marks or* striations, radiating outward from the point of initiation of the failure, as shown in *figure-8.7*. Beach marks (also known as *clamshell pattern*) are macroscopic dimensions and may be observed with the unaided eye. These markings are found for components that experienced interruptions during the crack propagation stage. Each beach mark band represents a period of time over which crack growth occurred. On the other hand fatigue striations are microscopic in size and subject to observation with the electron microscope (either TEM or SEM). The relatively widely spaced striations are caused by variations in the stress amplitude during the life of the component. On a much finer level, a large number of striations may be sometimes being seen. The width of each striation here is equal to the distance by which the crack grows during one cycle. Any point with stress concentration such as sharp corner *or* notch *or* metallurgical inclusion can act as point of initiation of fatigue crack.

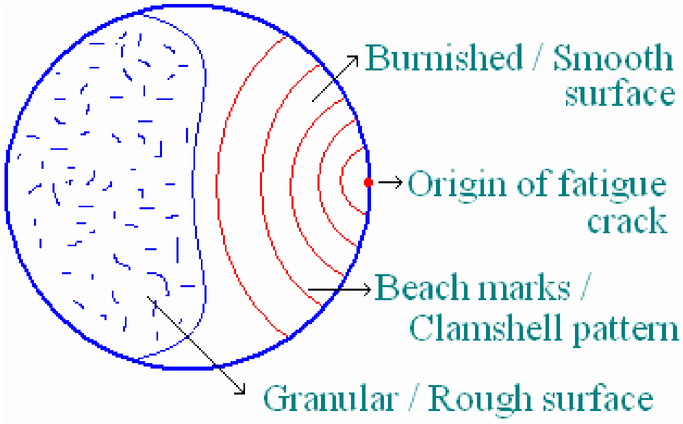


Figure-8.7: *Schematic of fatigue fracture surface.*

Three basic requisites for occurrence of fatigue fracture are: (a) a maximum tensile stress of sufficiently high value (b) a large enough variation or fluctuation in the applied stress and (c) a sufficiently large number of cycles of applied stress. The stress cycles that are evident in fatigue studies are characterized using many parameters, such as mean stress, alternating stress, stress ratio and amplitude ratio. If the applied stress varies between σ_{max} and σ_{min},

Range of stress, $\sigma_r = \sigma_{max} - \sigma_{min}$

Alternating stress, $\sigma_a = \sigma_r/2 = (\sigma_{max} - \sigma_{min})/2$

Mean stress, $\sigma_m = (\sigma_{max} + \sigma_{min})/2$

Stress ratio, $R = \sigma_{min} / \sigma_{max}$

Amplitude ratio, $A = \sigma_a / \sigma_m = (1-R) / (1+R)$

There are as many ways of conducting a test to measure fatigue as there are ways of applying repeated stresses to a metal. A specimen of rotating beam type is often used because of its simplicity. In a fatigue test, the premium output is number of cycles required to fracture the specimen at a given stress. If the maximum tensile stress applied is only slightly less than the yield strength (*or* tensile strength), test will run only a few cycles. Continues reduction of the stress greatly increases the life of the specimen, hence fatigue data is usually presented by plotting maximum stress (*S*) against number of cycles to fracture (*N*), using a logarithmic scale for the latter variable. This form of curve, *S-N* curve, shown in *figurer-8.8*, is significant because there is a stress below which the specimens do not fracture. This limiting stress is called *fatigue limit* or *endurance limit* (σ_e), below which fatigue will never occur. Fatigue limit can be related to tensile strength of the material, and the ratio of fatigue limit to tensile strength is known as *endurance ratio*. Most steels have distinct fatigue limit, and is usually about 0.4-0.5 of tensile strength i.e. endurance ratio=0.4-0.5. Unlike steels, most nonferrous metals do not have a fatigue limit i.e. *S-N* curve continues to fall steadily with decreasing stress, though at a decreasing rate. Thus, fatigue will ultimately occur regardless of the magnitude of the applied stress. Fatigue response of these materials is specified for a number of stress cycles, normally 10^7, and is known as *fatigue strength*. Another important parameter that characterizes a material's fatigue behavior is *fatigue life*, N_f, number of cycles to cause fatigue failure at a specified stress level.

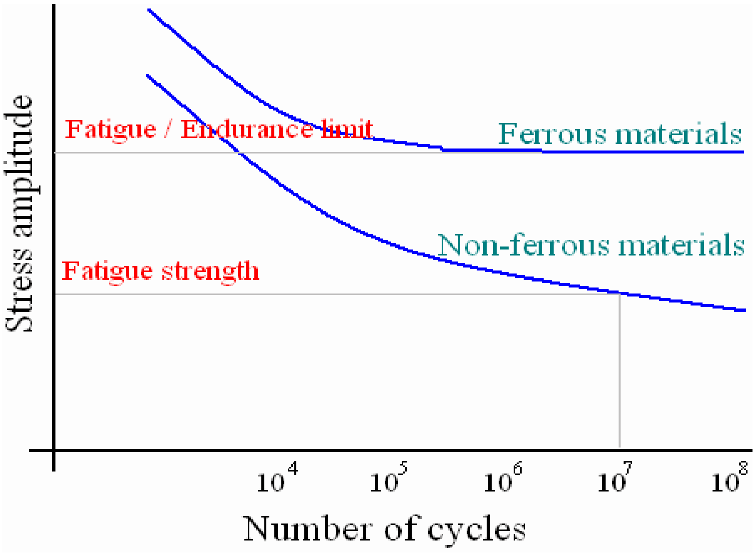


Figure-8.8: *Typical S-N curves for ferrous and non-ferrous metals.*

The S-N curve in the high-cycle region is some times described by the Basquin equation:

$$N\sigma_a^p = C$$

where σ_a is the alternating stress or stress amplitude and p and C are empirical constants. It will be generally found that there is a considerable amount of scatter in the results. Thus fatigue life and fatigue strength are considered as statistical quantities. It has been observed that scatter in fatigue life decreases with increase in stress. The statistical problem of accurately determining the fatigue limit is complicated by the fact that complete S-N curve cannot obtainable using a single specimen as specimen cannot be rested during the test.

With the exception of rotating shafts, most structural members are not subjected to symmetrical stress cycles having a mean stress of zero. When (R = -1 for the case of completely reversed stress) R becomes more positive i.e. with increasing mean stress, the measured fatigue limit becomes greater (*figure-8.9*). For a given stress amplitude, a mean tensile stress reduces the fatigue life of a material. It has been observed that as the mean stress becomes more positive the allowable alternating stress decreases (*figure-8.10*).

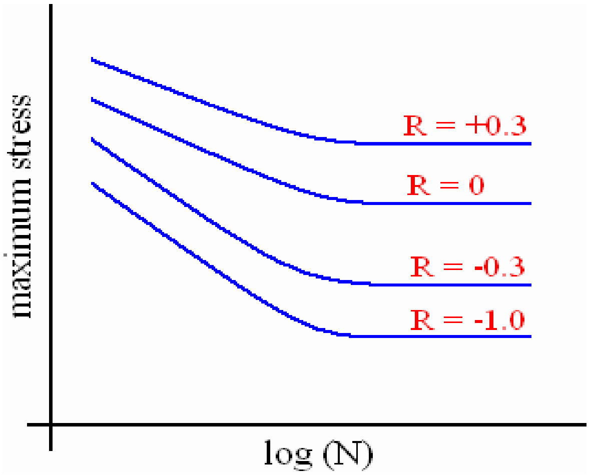


Figure-8.9: *Dependence of fatigue limit on stress ratio.*

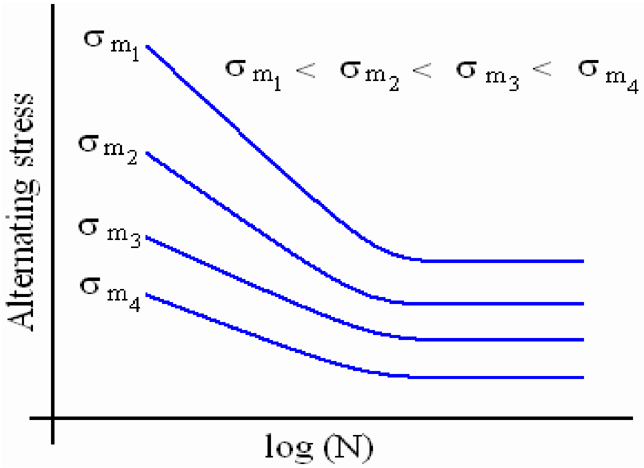


Figure-8.10: *Dependence of alternating stress on mean stress.*

The Goodman diagram presents the dependence of allowable stress ranges on mean stress for a material. As shown in *figure-8.11*, allowable stress range increases with increasing compressive mean stress i.e. compressive stress increases the fatigue limit.

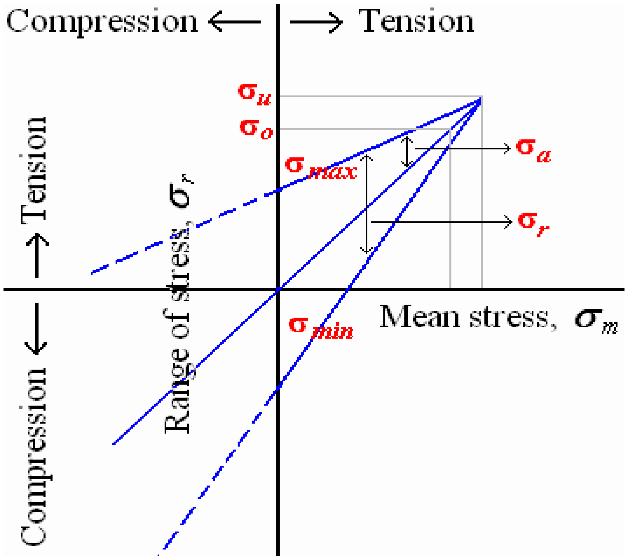


Figure-8.11: *Goodman diagram.*

An alternative method of presenting mean stress data is by using Heig-Soderberg diagram in which alternating stress is plotted against the mean stress. As *figure-8.12* depicts, Goodman's criterion appears as a straight line. Test data for ductile metals, however, follows parabolic curve proposed by Gerber. Both these criteria can be expressed as:

$$\sigma_a = \sigma_e \left[1 - \left(\frac{\sigma_m}{\sigma_u}\right)^x\right]$$

where x=1 for Goodman line, x=2 for the Gerber parabola, and σ_e is the fatigue limit for completely reversed loading. Soderberg line presents the data when the design is based on yield strength (σ_0).

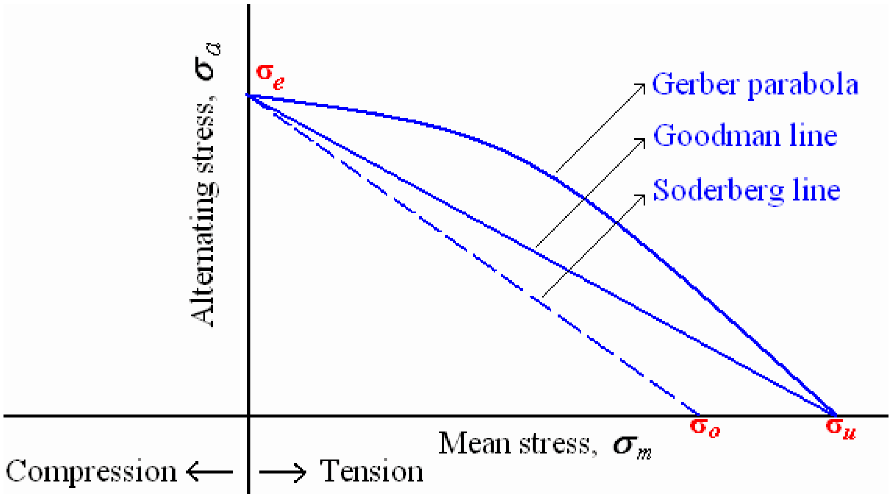


Figure-8.12: *Heig-Soderberg diagram.*

8.4.2 Crack initiation and propagation

Based on structural changes that occur during fatigue, fatigue failure process can be considered as made of following stages: (a) crack initiation – includes the early development of fatigue damage that can be removed by suitable thermal anneal (b) slip-band crack growth – involves the deepening of initial crack on planes of high shear stress. This is also known as stage-I crack growth. (c) crack growth on planes of high tensile stress – involves growth of crack in direction normal to maximum tensile stress, called stage-II crack growth (d) final ductile failure – occurs when the crack reaches a size so that the remaining cross-section cannot support the applied load.

Fatigue failures usually are found to initiate at a free surface or at internal flaws such as inclusions where the local stress causes some heterogeneous permanent flow leading to formation of a small crack. Fatigue failures start as small microscopic cracks and, accordingly, are very sensitive to even minute stress raisers. It has been observed that diffusion processes are not necessary to the formation of fatigue cracks. The initiation of a fatigue crack does not lead to immediate failure, rather, the crack propagates slowly and discontinuously across the specimen under the action of cyclic stress. The amount of crack motion per cycle depends on the material and the stress level; high stresses favor larger crack growth increments per cycle. Eventually, the crack propagates to the point where the remaining intact cross section of material no longer can support the applied load, and further crack propagation is rapid, leading to catastrophic failure. The final fracture surface is composed of an area over which there was slow crack propagation and an area where the crack moved rapidly. Final fracture can be either ductile or brittle type.

In polycrystalline metals, during a fatigue test slip lines appear first on crystal whose slip planes have the highest resolved shear stress. As time goes on and the number of stress cycles increases, the size and number of slip bands (clusters of slip lines) increase. The extent and number of slip bands are also a function of the amplitude of the applied stress; higher stresses give larger values. In fatigue, under cyclic loading, the slip bands tend to group into packets or striations in a slip band. Each striation represents the successive position of an advancing crack front that is normal to the greatest tensile stress. Ridge kinds of striations are called *extrusions* while, while crevice striations are known as *intrusions*, and both tend to be formed depending on the crystal orientation. It has been shown that fatigue cracks initiate at intrusions and extrusions. These surface disturbances are approximately 10^{-3} to 10^{-4} mm in height and appear as early as 1/10 of the total life of a specimen. *Figure-8.13* depicts the slip bands forming under static load and cyclic load. *Table-8.3* summarizes deformation features under static and cyclic loading.

With increasing numbers of cycles, the surface grooves deepen and the crevices *or* intrusions take on the nature of a crack. When this happens, stage-I of the crack-growth process has begun i.e. stage-I crack propagates along persistent slip bands, and can continue for a large fraction of the fatigue life. Low applied stresses and deformation by slip on a single slip plane favor stage-I growth. On the other hand multiple-slip conditions favor stage-II growth. The transition from stage-I to stage-II is often induced when a slip-plane crack meets an obstacle such as a grain boundary. The rate of crack propagation in stage-I is generally very low on the order of nm per cycle compared with that in stage-II where it is in order of μm per cycle. Thus, from a practical viewpoint, stage-II is of importance than stage-I. Stage-I growth follows a slip plane, whereas stage-II growth does not have this crystallographic character. The fracture surface of stage-I is practically featureless. On the other hand, stage-II shows a characteristic pattern of striations, which occurs by repetitive plastic blunting and sharpening of the crack tip. *Table-8.4* distinguishes stage-I from stage-II crack growth.

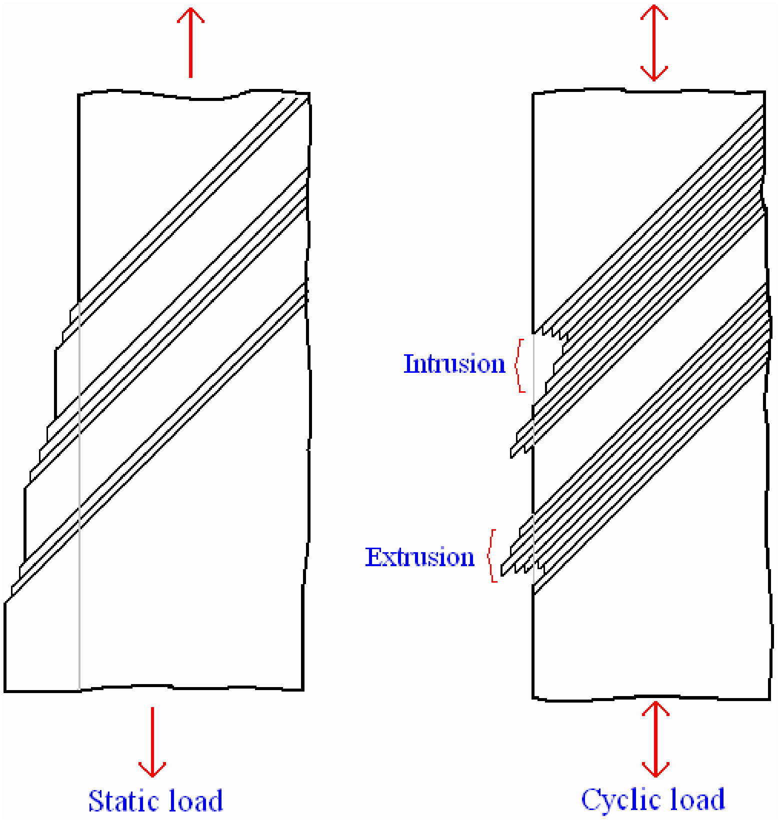


Figure-8.13: *Comparison of slip bands formed under (a) static loading and (b) cyclic loading*

Table-8.3: *Deformation under static and cyclic loads*.

Feature	***Static load***	***Cyclic load***
Slip (*nm*)	1000	1-10
Deformation feature	Contour	Extrusions & Intrusions
Grains involved	All grains	Some grains
Vacancy concentration	Less	Very high
Necessity of diffusion	Required	Not necessary

Table-8.4: *Fatigue crack growth: stage-I Vs stage-II.*

	Stage-I	***Stage-II***
Stresses involved	Shear	Tensile
Crystallographic orientation	Yes	No
Crack propagation rate	Low (nm/cycle)	High (μm/cycle)
Slip on	Single slip plane	Multiple slip planes
Feature	Feature less	Striations

The region of fatigue fracture surface that formed during the crack propagation step can be characterized by two types of markings termed *beach marks* and *striations*. Both of these features indicate the position of the crack tip at some point in time and appear as concentric rings that expand away from the crack initiation site(s), frequently in a circular or semicircular pattern. Beach marks (some times also called *clamshell marks*) are of macroscopic dimensions, found for components that experienced interruptions during the crack propagation stage. Each beach mark band represents a period of time over which crack growth occurred. Striations are microscopic in size, and each of it is thought to represent the advance distance of a crack front during a single load cycle. Striation width depends on, and increases with, increasing stress range. There may be literally thousands of striations within a single beach mark. The presence of beach marks/striations on a fracture surface confirms that the cause of failure in fatigue. Nevertheless, the absence of either or both does not exclude fatigue as the case of failure.

8.4.3 Crack propagation rate

As pointed out in earlier section, crack propagation in stage-II, where it is faster, has greater practical importance. Considerable research has gone into studying the crack propagation in this stage as the results from this can be used as fail-safe design considerations.

To distinguish the crack propagation time from crack initiation time, pre-existing flaws or cracks are introduced into a fatigue specimen. Crack propagation rate (*da/dN*) is studied in terms of increase in crack growth which is a proportional to number of applied stress cycles and length of the crack i.e.

$$\frac{da}{dN} = fn(\sigma, a) = C\sigma_a^m a^n$$

where C – a constant, σ_a – the alternating stress, a – the crack length, m=2-4, and n=1-2.

It has been found that crack growth rate can be related to stress-intensity factor, K (mode-I) of fracture mechanics which itself is a combination of stress and crack length. Thus,

$$\frac{da}{dN} = A(\Delta K)^{p}$$

where ΔK – stress-intensity factor range., A and p – constants that are functions of material, environment, frequency, temperature and stress ratio.

$$\Delta K = K_{\max} - K_{\min} = \sigma_{\max}\sqrt{a\pi} - \sigma_{\min}\sqrt{a\pi} = \sigma_{r}\sqrt{a\pi}$$

Since the stress intensity factor is not defined for compressive stresses, if σ_{min} is in compression, zero value is assigned for K_{min}.

Fatigue data presented in terms of log (crack growth rate) against log (ΔK) is shown in *figure-8.14*. The curve can be divided into three regions. In region-I, which is bound by a threshold value ΔK_{th}, no observable crack growth occurs i.e. crack growth rate is very slow in order of 0.25 nm per cycle or less. The straight line part of curve in region-II can be represented by the power law (also known as *Paris law*)

$$\frac{da}{dN} = A(\Delta K)^{p}$$

and slope of the straight line will give value for p. The values of p are the range of 3 for steels, and in the range of 3-4 for aluminium alloys. Region-III of the curve represents the rapid unstable crack growth to ultimate fracture. Here K_{max} approaches K_c, the fracture toughness of the material. When K is known under relevant loading conditions, the fatigue life can be estimated by integrating the above equation between limits of initial crack size and final crack size. One such equation is given below when $p \neq 2$:

$$N_{f} = \frac{a_{f}^{-(m/2)+1} - a_{0}^{-(m/2)+1}}{A\sigma^{p}\pi^{p/2}\alpha^{p}(-(p/2)+1)}$$

N_f – number of cycles required for a crack to grow from a_0 to a_f. α - is a parameter that depends on the specimen and crack sizes and geometries, as well as the manner of load application.

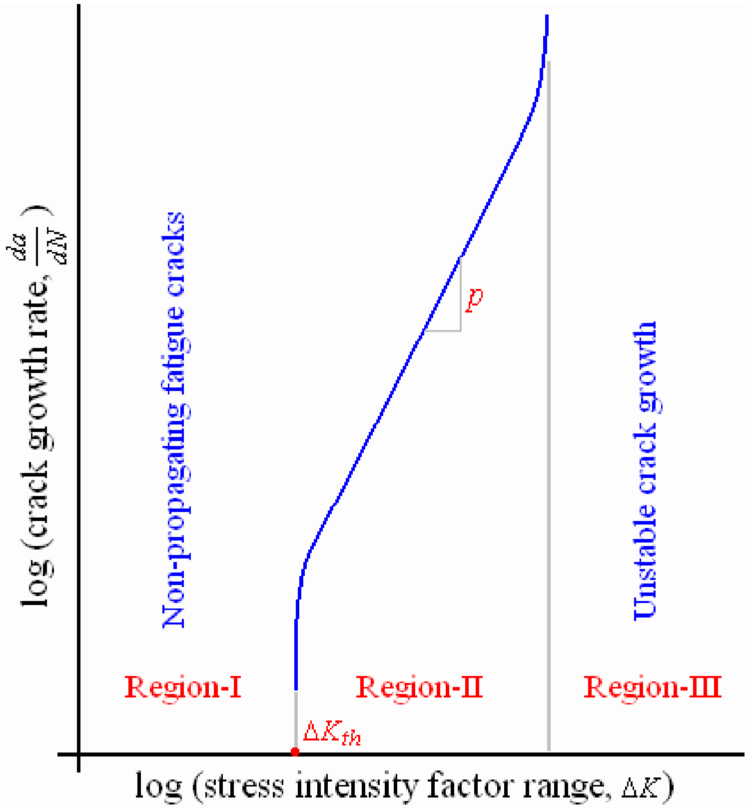


Figure-8.14: *Schematic presentation of fatigue crack growth.*

Increasing mean stress, *R*, in the fatigue cycle has a tendency to increase the crack growth rates in all portions of the curve. The effect of increasing *R* is less in region-II than in regions I and III. The influence of R on the region-II is given by:

$$\frac{da}{dN} = \frac{A(\Delta K)^p}{(1-R)K_c - \Delta K}$$

where K_c – fracture toughness, *R* – stress ratio. Fatigue life estimation is usually carried out when R=-1 i.e. fully reversed stress conditions, whereas fatigue crack growth data is determined for R=0 i.e. pulsating tension without compression. Compression loads are avoided because during compression loading the crack is closed and the stress intensity factor is zero.

8.5 Creep, Generalized creep behavior, Stress and Temperature effects

8.5.1 Creep

It is well known and observed number of times that shear modulus of a material increases with decreasing temperature, and this has a corresponding effect on the flow-stress, since the stress associated with a dislocation is always proportional to the shear modulus. Thus σ/E can be expected to be constant with varying temperature. However, this ratio decreases with increase in temperature. Since, the dislocation-density contribution of the flow stress is effectively constant, it was concluded that there must be a second basic component of flow stress that is temperature-dependent i.e.

$$\sigma = \sigma^{*} + \sigma_{E}$$

where σ^* is thermal component and σ_E is athermal component of flow stress. The fact that the flow stress contains a component that responds to thermal activation implies that plastic deformation can occur while both the temperature and the stress are maintained constant. Deformation that occurs under these conditions which is time-dependent is known as *creep*. Creep deformation (constant stress) is possible at all temperatures above absolute zero. However, it is extremely sensitive to temperature. Hence, creep in usually considered important at elevated temperatures (temperatures greater than 0.4 T_m, T_m is absolute melting temperature).

8.5.2 Generalized creep behavior

Creep behavior of a material is studied using creep test. In creep test, the tensile specimen is subjected to a constant load *or* stress at a constant temperature. Most creep tests are conducted at constant load in analogous to engineering application, whereas creep tests at constant stress are necessary for understanding of mechanism of creep. During the creep test, strain (change in length) is measured as a function of elapsed time. Creep test data is presented as a plot between time and strain known as creep curve. *Figure-8.15* depicts a typical creep curve. The slope of the creep curve is designated as creep rate.

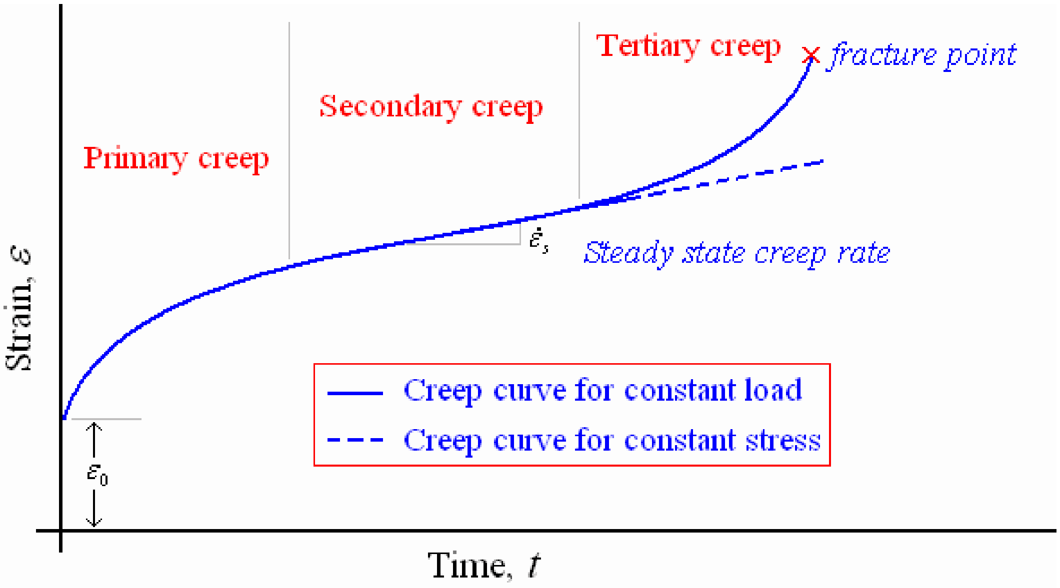


Figure-8.15: *Typical creep curve.*

As shown in the above figure, upon loading the specimen, there is an instantaneous deformation (ε_0) that is mostly elastic. Actual creep curve follows this elastic deformation. Based on the variation of creep rate with time, creep curve is considered to be consists of three portions, each of which has its own distinctive strain-time feature. After initial rapid elongation, ε_0, the creep rate decreases continuously with time, and is known as *primary* or t*ransient creep*. Primary creep is followed by *secondary* or *steady-state* or *viscous creep,* which is characterized by constant creep rate. This stage of creep is often the longest duration of the three modes. Finally, a third stage of creep known as, *tertiary creep* occurs that is characterized by increase in creep rate. Creep curve could be represented by the following equation, according to Andrade:

$$\varepsilon = \varepsilon_0 (1 + \beta t^{1/3}) e^{kt}$$

where ε is strain in time t, and β and k are constants. A better equation is proposed by Garofalo:

$$\varepsilon = \varepsilon_0 + \varepsilon_t (1 - e^{-rt}) + \dot{\varepsilon}_s t$$

where ε_0 – instantaneous strain on loading, ε_t – limit for transient creep, r – ratio of transient creep rate to the transient creep strain, $\cdot\varepsilon_s$ – steady-state creep rate.

It is suggested that during primary creep, material strain hardens thus increases its creep resistance. Constant creep rate during secondary creep is believed to be due to balance between the competing processes of strain hardening and recovery. The average value of creep rate during the secondary creep is called the minimum creep rate. Third stage creep occurs in constant load tests at high stresses at high temperatures. This stage is greatly delayed in constant stress tests. Tertiary creep is believed to occur because of either reduction in cross-sectional area due to necking or internal void formation. Third stage is often associated with metallurgical changes such as coarsening of precipitate particles, recrystallization, or diffusional changes in the phases that are present.

For metallic materials most creep tests are conducted in uni-axial tensile mode. However, uni-axial compression tests are used for brittle materials to avoid stress amplification and corresponding crack propagation. For most materials creep properties are independent of loading direction.

The minimum creep rate is the most important design parameter derived from the creep curve. It is the engineering design parameter that is considered for long-life applications, e.g. nuclear power plant components. On the other hand, for short-life components (e.g. turbine blades, rocket motor nozzles), time to rupture or rupture lifetime is the dominant factor in design. It is found from test conducted to the point of failure, and the test is known as *stress-rupture test*. This is basically similar to a creep test except that the test is always carried out at higher loads. In a creep test total strain is often less than 0.5%, while in stress-rupture test it is around 50%.

8.5.2 Stress and temperature effects

Creep properties of a material are usually characterized at a constant stress and a constant temperature. A change in either of these parameters causes change in creep behavior of the material. With increase in either stress or temperature (a) instantaneous elastic strain increases (b) steady state creep rate increases and (c) rupture lifetime decreases.

Influence of stress on steady state creep rate can be written as

$$\dot{\varepsilon}_s = K_1 \sigma^n$$

where K_1 and n are material constants. Value of these parameters can found from logarithmic plot between creep rate and stress after conducting tests at a constant temperature. Similarly temperature influence on steady state creep rate can be characterized. Influence of combined action of temperature and stress on steady state creep rate can be presented as follows:

$$\dot{\varepsilon}_s = K_2 \sigma^n e^{-\frac{Q_c}{RT}}$$

where K_2 and Q_c (activation energy for creep) are constants. Values of n for a specific material are dependent mechanism of creep operational in those particular conditions of stress and temperature. *Figure-8.16* presents the influence of stress/temperature on creep behavior.

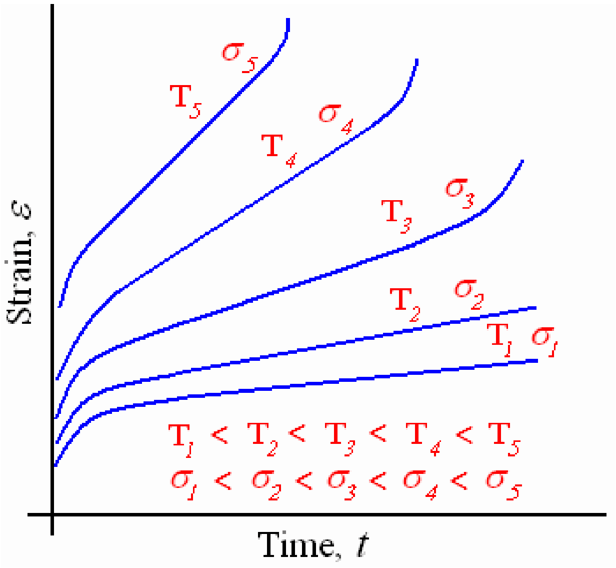


Figure-8.16: *Influence of stress/temperature on creep behavior.*

Creep data presented in form of stress-temperature diagrams are termed *deformation mechanism maps* which indicates stress-temperature regimes over which various mechanisms operate. The chief creep deformation mechanisms are:

Dislocation glide – involves dislocation moving along slip planes by overcoming barrier with help of thermal activation. This occurs at high stresses, $\sigma/G > 10^{-2}$.

Dislocation creep – involves movement of dislocations which overcome barriers by thermally assisted mechanisms like diffusion of vacancies or interstitials. Occurs at moderate stresses, $10^{-2} > \sigma/G > 10^{-4}$.

Diffusion creep – involves flow of vacancies and interstitials under the influence of applied stress. Occurs for $\sigma/G < 10^{-4}$.

Grain boundary sliding – involves sliding of grains against each other.

It is to be noted that, quite often, more than one creep mechanism will operate at a time. When several mechanisms are in operation at a time simultaneously, i.e. independent of each other, fastest mechanism will dominate the proceedings. On the other hand, when they are operating in series, slowest mechanism will control the creep deformation.

References

1. G. E. Dieter, Mechanical Metallurgy, Third Edition, McGraw-Hill, New York, 1986.
2. R. W. Hertzberg, Deformation and Fracture Mechanics of Engineering Materials, Fourth Edition, Wiley, New York, 1996.
3. T. H. Courtney, Mechanical Behavior of Materials, Second Edition, McGraw-Hill, New York, 2000.

Chapter 9. Applications and Processing of Metals and Alloys

In the materials world we are living in, when making a new device/component, most often we come across a very familiar problem. This is nothing but select the right material. As learnt in earlier chapter, selection of material can play very important role preventing failures. Selection of material for a specific purpose depends on many factors. Some of the important ones are: strength, ease of forming, resistance to environmental degradation, etc. Another dimension an engineer should be aware of it is how to tailor the required properties of materials.

As introduced in one of the earlier chapters, materials can be are broadly classified as metals, ceramics and plastics. This chapter introduces different classes of metallic materials, common fabrication methods, and means to alter their properties on purpose. Following chapters deal with ceramic materials and plastic materials.

9.1 Types of metals and alloys

Metallic materials are broadly of two kinds – *ferrous* and *non-ferrous* materials. This classification is primarily based on tonnage of materials used all around the world. Ferrous materials are those in which iron (***Fe***) is the principle constituent. All other materials are categorized as non-ferrous materials. Another classification is made based on their formability. If materials are hard to form, components with these materials are fabricated by casting, thus they are called cast alloys. If material can be deformed, they are known as wrought alloys. Materials are usually strengthened by two methods – cold work and heat treatment. Strengthening by heat treatment involves either precipitation hardening or martensitic transformation, both of which constitute specific heat treating procedure. When a material can not be strengthened by heat treatment, it is referred as non-heat-treatable alloys.

9.1.1 Ferrous materials

Ferrous materials are produced in larger quantities than any other metallic material. Three factors account for it: (a) availability of abundant raw materials combined with economical extraction, (b) ease of forming and (c) their versatile mechanical and physical properties. One main drawback of ferrous alloys is their environmental degradation i.e. poor corrosion resistance. Other disadvantages include: relatively high density and comparatively low electrical and thermal conductivities. In ferrous materials the main alloying element is carbon (***C***). Depending on the amount of carbon present, these alloys will have different properties, especially when the carbon content is either less/higher than 2.14%. This amount of carbon is specific as below this amount of carbon, material undergoes eutectoid transformation, while above that limit ferrous materials undergo eutectic transformation. Thus the ferrous alloys with less than 2.14% C are termed as *steels*, and the ferrous alloys with higher than 2.14% C are termed as *cast irons*.

9.1.1.a Steels

Steels are alloys of iron and carbon plus other alloying elements. In steels, carbon present in atomic form, and occupies interstitial sites of Fe microstructure. Alloying additions are necessary for many reasons including: improving properties, improving corrosion resistance, etc. Arguably steels are well known and most used materials than any other materials.

Mechanical properties of steels are very sensitive to carbon content. Hence, it is practical to classify steels based on their carbon content. Thus steels are basically three kinds: low-carbon steels (% wt of C < 0.3), medium carbon steels (0.3 <% wt of C < 0.6) and high-carbon steels (% wt of C > 0.6). The other parameter available for classification of steels is amount of alloying additions, and based on this steels are two kinds: (plain) carbon steels and alloy-steels.

<u>Low carbon steels</u>: These are arguably produced in the greatest quantities than other alloys. Carbon present in these alloys is limited, and is not enough to strengthen these materials by heat treatment; hence these alloys are strengthened by cold work. Their microstructure consists of ferrite and pearlite, and these alloys are thus relatively soft, ductile combined with high toughness. Hence these materials are easily machinable and weldable. Typical applications of these alloys include: structural shapes, tin cans, automobile body components, buildings, etc.

A special group of ferrous alloys with noticeable amount of alloying additions are known as HSLA (high-strength low-alloy) steels. Common alloying elements are: Cu, V, Ni, W, Cr, Mo, etc. These alloys can be strengthened by heat treatment, and yet the same time they are ductile, formable. Typical applications of these HSLA steels include: support columns, bridges, pressure vessels.

<u>Medium carbon steels</u>: These are stronger than low carbon steels. However these are of less ductile than low carbon steels. These alloys can be heat treated to improve their strength. Usual heat treatment cycle consists of austenitizing, quenching, and tempering at suitable conditions to acquire required hardness. They are often used in tempered condition. As hardenability of these alloys is low, only thin sections can be heat treated using very high quench rates. Ni, Cr and Mo alloying additions improve their

hardenability. Typical applications include: railway tracks & wheels, gears, other machine parts which may require good combination of strength and toughness.

High carbon steels: These are strongest and hardest of carbon steels, and of course their ductility is very limited. These are heat treatable, and mostly used in hardened and tempered conditions. They possess very high wear resistance, and capable of holding sharp edges. Thus these are used for tool application such as knives, razors, hacksaw blades, etc. With addition of alloying element like Cr, V, Mo, W which forms hard carbides by reacting with carbon present, wear resistance of high carbon steels can be improved considerably.

Stainless steels: The name comes from their high resistance to corrosion i.e. they are rust-less (stain-less). Steels are made highly corrosion resistant by addition of special alloying elements, especially a minimum of 12% Cr along with Ni and Mo. Stainless steels are mainly three kinds: ferritic & hardenable Cr steels, austenitic and precipitation hardenable (martensitic, semi-austenitic) steels. This classification is based on prominent constituent of the microstructure. Typical applications include cutlery, razor blades, surgical knives, etc.

Ferritic stainless steels are principally Fe-Cr-C alloys with 12-14% Cr. They also contain small additions of Mo, V, Nb, and Ni.

Austenitic stainless steels usually contain 18% Cr and 8% Ni in addition to other minor alloying elements. Ni stabilizes the austenitic phase assisted by C and N. Other alloying additions include Ti, Nb, Mo (prevent weld decay), Mn and Cu (helps in stabilizing austenite).

By alloying additions, for martensitic steels ***Ms*** is made to be above the room temperature. These alloys are heat treatable. Major alloying elements are: Cr, Mn and Mo.

Ferritic and austenitic steels are hardened and strengthened by cold work because they are not heat treatable. On the other hand martensitic steels are heat treatable. Austenitic steels are most corrosion resistant, and they are produced in large quantities. Austenitic steels are non-magnetic as against ferritic and martensitic steels, which are magnetic.

9.1.1.b Cast irons

Though ferrous alloys with more than 2.14 wt.% C are designated as cast irons, commercially cast irons contain about 3.0-4.5% C along with some alloying additions. Alloys with this carbon content melt at lower temperatures than steels i.e. they are responsive to casting. Hence casting is the most used fabrication technique for these alloys.

Hard and brittle constituent presented in these alloys, cementite is a meta-stable phase, and can readily decompose to form α-ferrite and graphite. In this way disadvantages of brittle phase can easily be overcome. Tendency of cast irons to form graphite is usually controlled by their composition and cooling rate. Based on the form of carbon present, cast irons are categorized as gray, white, nodular and malleable cast irons.

Gray cast iron: These alloys consists carbon in form graphite flakes, which are surrounded by either ferrite or pearlite. Because of presence of graphite, fractured surface of these alloys look grayish, and so is the name for them. Alloying addition of Si (1-3wt.%) is responsible for decomposition of cementite, and also high fluidity. Thus castings of intricate shapes can be easily made. Due to graphite flakes, gray cast irons are weak and brittle. However they possess good damping properties, and thus typical applications include: base structures, bed for heavy machines, etc. they also show high resistance to wear.

White cast iron: When Si content is low (< 1%) in combination with faster cooling rates, there is no time left for cementite to get decomposed, thus most of the brittle cementite retains. Because of presence of cementite, fractured surface appear white, hence the name. They are very brittle and extremely difficult to machine. Hence their use is limited to wear resistant applications such as rollers in rolling mills. Usually white cast iron is heat treated to produce malleable iron.

Nodular (*or* ductile) cast iron: Alloying additions are of prime importance in producing these materials. Small additions of Mg / Ce to the gray cast iron melt before casting can result in graphite to form nodules or sphere-like particles. Matrix surrounding these particles can be either ferrite or pearlite depending on the heat treatment. These are stronger and ductile than gray cast irons. Typical applications include: pump bodies, crank shafts, automotive components, etc.

Malleable cast iron: These formed after heat treating white cast iron. Heat treatments involve heating the material up to 800-900 C, and keep it for long hours, before cooling it to room temperature. High temperature incubation causes cementite to decompose and form ferrite and graphite. Thus these materials are stronger with appreciable amount of ductility. Typical applications include: railroad, connecting rods, marine and other heavy-duty services.

9.1.2 Non-ferrous materials

Non-ferrous materials have specific advantages over ferrous materials. They can be fabricated with ease, high relatively low density, and high electrical and thermal conductivities. However different materials have distinct characteristics, and are used for specific purposes. This section introduces some typical non-ferrous metals and their alloys of commercial importance.

Aluminium alloys: These are characterized by low density, high thermal & electrical conductivities, and good corrosion resistant characteristics. As Al has FCC crystal structure, these alloys are ductile even at low temperatures and can be formed easily. However, the great limitation of these alloys is their low melting point (660 C), which restricts their use at elevated temperatures. Strength of these alloys can be increased by both cold and heat treatment – based on these alloys are designated in to two groups, cast and wrought. Chief alloying elements include: Cu, Si, Mn, Mg, Zn. Recently, alloys of Al and other low-density metals like Li, Mg, Ti gained much attention as there is much concern about vehicle weight reduction. Al-Li alloys enjoy much more attention especially as they are very useful in aircraft and aerospace industries. Common applications of Al alloys include: beverage cans, automotive parts, bus bodies, aircraft structures, etc. Some of the Al alloys are capable of strengthening by precipitation, while others have to be strengthened by cold work or solid solution methods.

<u>Copper alloys</u>: As history goes by, bronze has been used for thousands of years. It is actually an alloy of Cu and Sn. Unalloyed Cu is soft, ductile thus hard to machine, and has virtually unlimited capacity for cold work. One special feature of most of these alloys is their corrosion resistant in diverse atmospheres. Most of these alloys are strengthened by either cold work or solid solution method. Common most Cu alloys: Brass, alloys of Cu and Zn where Zn is substitutional addition (e.g.: yellow brass, catridge brass, muntz metal, gilding metal); Bronze, alloys of Cu and other alloying additions like Sn, Al, Si and Ni. Bronzes are stronger and more corrosion resistant than brasses. Mention has to be made about Beryllium coppers who possess combination of relatively high strength, excellent electrical and corrosion properties, wear resistance, can be cast, hot worked and cold worked. Applications of Cu alloys include: costume jewelry, coins, musical instruments, electronics, springs, bushes, surgical and dental instruments, radiators, etc.

<u>Magnesium alloys</u>: The most sticking property of Mg is its low density among all structural metals. Mg has HCP structure, thus Mg alloys are difficult to form at room temperatures. Hence Mg alloys are usually fabricated by casting or hot working. As in case of Al, alloys are cast or wrought type, and some of them are heat treatable. Major alloying additions are: Al, Zn, Mn and rare earths. Common applications of Mg alloys include: hand-held devices like saws, tools, automotive parts like steering wheels, seat frames, electronics like casing for laptops, camcoders, cell phones etc.

<u>Titanium alloys</u>: Ti and its alloys are of relatively low density, high strength and have very high melting point. At the same time they are easy to machine and forge. However the major limitation is Ti’s chemical reactivity at high temperatures, which necessitated special techniques to extract. Thus these alloys are expensive. They also possess excellent corrosion resistance in diverse atmospheres, and wear properties. Common applications include: space vehicles, airplane structures, surgical implants, and petroleum & chemical industries.

<u>Refractory metals</u>: These are metals of very high melting points. For example: Nb, Mo, W and Ta. They also possess high strength and high elastic modulus. Common applications include: space vehicles, x-ray tubes, welding electrodes, and where there is a need for corrosion resistance.

<u>Noble metals</u>: These are eight all together: Ag, Au, Pt, Pa, Rh, Ru, Ir and Os. All these possess some common properties such as: expensive, soft and ductile, oxidation resistant.

Ag, Au and Pt are used extensively in jewelry, alloys are Ag and Au are employed as dental restoration materials; Pt is used in chemical reactions as a catalyst and in thermo couples.

9.2 Fabrication of metals

Metals are fabricated by different means to achieve metals and alloys of desired characteristics. There been many kinds of fabrication techniques, and for a particular metal use of these depend on properties of metal, product shape-size-properties, cost, etc. Metal fabrication techniques are mainly four kinds: Casting - to give a shape by pouring in liquid metal into a mold that holds the required shape, and letting harden the metal without external pressure; Forming – to give shape in solid state by applying pressure; Machining – in which material is removed in order to give it the required shape; and Joining – where different parts are joined by various means. One of the most important miscellaneous techniques is powder metallurgy.

9.2.1 Metal casting

This technique is employed when (a) product is large and/or complicated shape (b) particular material is low in ductility. This is also employed as it is usually economical compared with other techniques. Different casting techniques include: sand, die, investment, continuous casting

<u>Sand casting</u>: The common casting method where sand is used as casting material. A two piece mold (cope and drag) is formed by compact packing of sand around a pattern of required shape. An addition gating is provided for proper distribution of liquid metal.

<u>Die casting</u>: Here metal is forced into mold by external pressure at high velocities. Usually a permanent two-piece mold made of steel is used. In this technique rapid cooling rates are achieved, thus inexpensive.

<u>Investment casting</u>: In this pattern is made of wax. Then fluid slurry of casting material is poured over which eventually hardens and holds the required shape. Subsequently, pattern material is heated to leave behind the cavity. This technique is employed when high dimensional accuracy, reproduction of fine details, and an excellent finish are required. For example: jewelry, dental crowns, and gas turbine blades jet engine impellers.

<u>Continuous casting</u>: After refining metals are usually in molten state, which are later solidified into ingots for further processing like forming. In continuous casting, solidification and primary forming process are combined, where refined metal is cast directly into a continuous strand which is cooled by water jets. This technique is highly automated and more efficient. Uniform composition through-out the casting is achievable when compared with ingot-cast products.

9.2.2 Metal forming

In these techniques, a metallic piece is subjected to external pressures (in excess of yield strength of the material) to induce deformation, thus material acquires a desired shape. These are basically two types – one that performed at relatively low temperatures, cold working; ad the other performed at high temperatures, hot working. Hot working is responsible mainly for substantial change in cross section without material getting strengthened, while during cold working, fine details are achieved along with material getting strengthened. Most common forming techniques are: forging, rolling, extrusion, and drawing. *Figure-9.1* illustrates different forming processes.

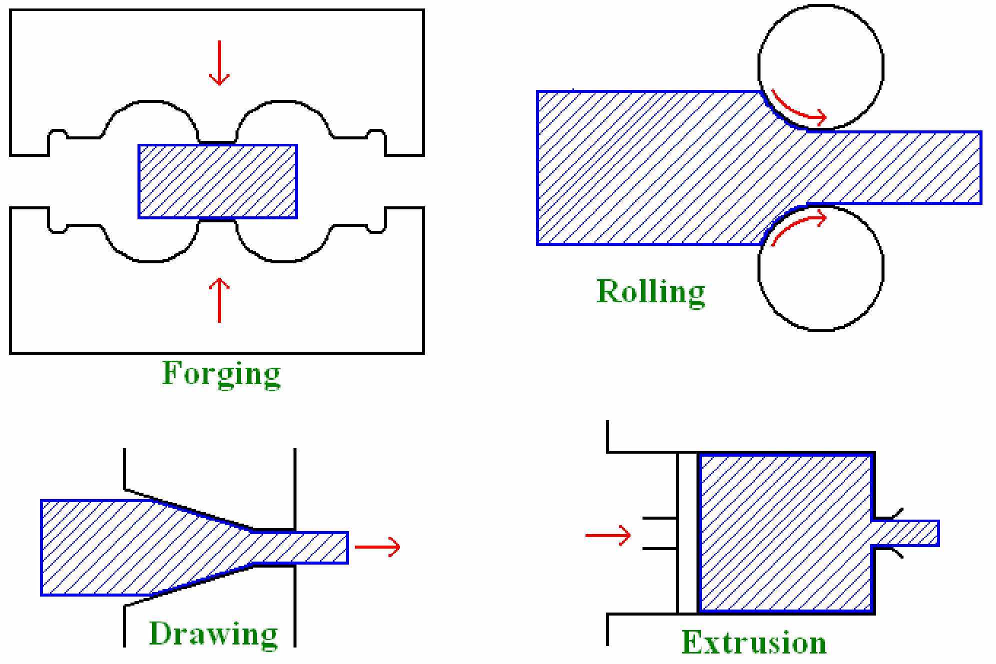


Figure-9.1: *Different forming processes.*

Forging: This involves deforming a single piece of metal, usually, by successive blows or continuous squeezing. In open die forging, two dies having same shape is employed, usually, over large work-pieces; while in closed die forging, there may be more than two pieces of die put together having finished shape. Forged products have outstanding grain structures and very good mechanical properties. Typical products include: crane hooks, wrenches, crank shafts, connecting rods.

Rolling: Most widely used forming technique because of high production rate and close dimensional control of final product. It involves passing a piece of metal between two rotating rolls. Deformation is terms of reduction in thickness resulting from applied compressive forces. This technique is typically employed to produce sheets, strips, foil, I-beams, rails, etc.

Extrusion: In this technique a piece of material is forced though a die orifice by a compressive force. Final product emerging from die will have the desired shape and reduced cross sectional area, and will constant cross-section over very long lengths. Two varieties of extrusion are direct extrusion and indirect extrusion, where distinction limits to movement of tool and final product and consequent changes in required force. Typical extrusion products are: rods, (seamless) tubes, complicated shapes for domestic purpose.

Drawing: It is pulling of material though die orifice using tensile forces. Again a reduction in cross-section results with corresponding change in length. Drawing die entrance is at angle against to extrusion die which is usually rectangular. Typical drawing strand includes number of dies in a series sequence. Rods, wire, and tubes are commonly produced using drawing technique.

9.2.3 Machining

This technique employs removable of metal from selected areas of the workpiece to give final shape to the product. This is in direct contrast with metal forming where metal is moved and volume is conserved. Machining usually is employed to produce shapes with high dimensional tolerance, good surface finish, and often with complex geometry. And another important note is that when number of product pieces required is small, machining is preferred over forming as special tool cost will be less.

9.2.4 Joining

There been many joining techniques, especially for metallic materials. These include: welding, brazing, soldering, and riveting. In these techniques, two pieces are joined together either by adhesive/cohesive bonding and/or mechanical locking. Welding, brazing, and soldering involve melting of either parent metal or external metallic liquid (filler material) which upon cooling provides cohesive bonds. In riveting, pieces are put together by mechanical locking. These techniques are employed to join two pieces of same metal with complicated shapes, or of different metals because of difficulty in fabricating them using one of the previous methods. This may be employed when on-part fabrication is expensive or inconvenient.

9.2.5 Powder metallurgy

In this technique, metal powders or mixture of metal powers at desired relative amounts are compacted into the desired shape, followed by sintering in controlled atmosphere to produce a denser product. It is makes it possible to produce a virtually non-porous product where diffusional processes controls the efficiency of the process. It is suitable, especially, for metals with low ductility/high melting points. Other advantage include: close dimensional tolerance of complicated shapes. Usually products are less dense than wrought products because of porosity. However, it is advantageous as pores can retain oil for self-lubrication of bushes, and high damping capacity.

9.3 Thermal processing of metals and alloys

Apart from mechanical processing, metals are very often subjected to thermal processing for various reasons, like: to refine grain structure/size, to minimize residual stresses, to impart phase changes, to develop special phases over external surfaces, etc. Metals and alloys develop requisite properties by thermal processing either through grain refinement of phase changes. Thermal processing is also known as heat treatment. Heat treatment originated as an ancient art in man’s attempts to improve the performance of materials in their practical applications. In present day metallurgical practice, heat treatment has become very important for obvious reasons. There has been tremendous progress over centuries in the systematic understanding of materials structure and structure-property relationships that eliminated the empiricism in thermal processing. Properly designed and implemented thermal processing can result in optimum modifications in the composition and distribution of phases, corresponding changes in physical, chemical and mechanical properties at substantial levels. However, most of the thermal processes are aimed to improving mechanical characteristics of materials. Thus it is possible to extend the service performance of materials considerably within constraints of available resources.

All metals can be subjected to thermal processing. But the effect of it may differ from one metal to another. Metals are subjected to heat treatment for one or more of the following purposes: improvement in ductility; relieving internal stresses; grain size refinement; increase of strength; improvement in machinability, toughness; etc.

Heat treatment of materials involves number of factors – temperature up to which material is heated, length of time that the material is held at the elevated temperature, rate of cooling, and the surrounding atmosphere under the thermal treatment. All these factors depend on material, pre-processing of the material’s chemical composition, size and shape of the object, final properties desired, material’s melting point/liquidus, etc.

Thermal processes may be broadly classified into two categories based on cooling rates from elevated temperatures – annealing and quenching & tempering. Annealing involved cooling the material from elevated temperatures slowly, while quenching means very fast cooling of the material using cooling medium like water/oil bath. Quenching is done to retain the phases of elevated temperatures at room temperature.

9.3.1 Annealing processes

The term annealing was used by craftsmen who discovered the beneficial effects of heating the material at elevated temperatures followed by slow cooling of it to room temperature. *Annealing* can be defined as a heat treatment process in which the material is taken to a high temperature, kept there for some time and then cooled. High temperatures allow diffusion processes to occur fast. The time at the high temperature (soaking time) must be long enough to allow the desired transformation to occur. Cooling is done slowly to avoid the distortion (warping) of the metal piece, or even cracking, caused by stresses induced by differential contraction due to thermal inhomogeneities. Benefits of annealing are:

- relieve stresses
- increase softness, ductility and toughness
- produce a specific microstructure

Depending on the specific purpose, annealing is classified into various types: process annealing, stress relief, full annealing and normalizing.

Process annealing is primarily applied to cold worked metals to negate the effects of cold work. During this heat treatment, material becomes soft and thus its ductility will be increased considerably. It is commonly sandwiched between two cold work operations. During this, recovery and recrystallization are allowed whereas grain growth was restricted.

Stress relief operation removes the stresses that might have been generated during plastic deformation, non-uniform cooling, or phase transformation. Unless removed, these stresses may cause distortion of components. Temperature used is normally low such that effects resulting from cold working are not affected.

Full annealing is normally used for products that are to be machined subsequently, such as transmission gear blanks. After heating and keeping at an elevated temperature, components are cooled in furnace to effect very slow cooling rates. Typically, the product receives additional heat treatments after machining to restore hardness and strength.

Normalizing is used to refine the grains and produce a more uniform and desirable size distribution. It involves heating the component to attain single phase (e.g.: austenite in steels), then cooling in open air atmosphere.

9.3.2 Quenching and Tempering processes

Quenching is heat treatment process where material is cooled at a rapid rate from elevated temperature to produce Martensite phase. This process is also known as *hardening*. Rapid cooling rates are accomplished by immersing the components in a quench bath that usually contains quench media in form of either water or oil, accompanied by stirring mechanism.

Quenching process is almost always followed by tempering heat treatment. *Tempering* is the process of heating martensitic steel at a temperature below the eutectoid transformation temperature to make it softer and more ductile. During the tempering process, Martensite transforms to a structure containing iron carbide particles in a matrix of ferrite.

Martempering is a modified quenching procedure used to minimize distortion and cracking that may develop during uneven cooling of the heat-treated material. It involves cooling the austenized steel to temperature just above M_s temperature, holding it there until temperature is uniform, followed by cooling at a moderate rate to room temperature before austenite-to-bainite transformation begins. The final structure of martempered steel is tempered Martensite.

Austempering is different from martempering in the sense that it involves austenite-to-bainite transformation. Thus, the structure of austempered steel is bainite. Advantages of austempering are – improved ductility; decreased distortion and disadvantages are – need for special molten bath; process can be applied to limited number of steels.

9.4 Case Hardening

In case hardening, the surface of the steel is made hard and wear resistant, but the core remains soft and tough. Such a combination of properties is desired in applications such as gears.

9.4.1. Induction hardening

Here, an alternating current of high frequency passes through an induction coil enclosing the steel part to be heat treated. The induced *emf* heats the steel. The depth up to which the heat penetrates and raises the temperature above the elevated temperature is inversely proportional to the square root of the *ac* frequency. In induction hardening, the heating time is usually a few seconds. Immediately after heating, water jets are activated to quench the surface. Martensite is produced at the surface, making it hard and wear resistant. The microstructure of the core remains unaltered. Induction hardening is suitable for mass production of articles of uniform cross-section.

9.4.2. Flame hardening

For large work pieces and complicated cross-sections induction heating is not easy to apply. In such cases, flame hardening is done by means of an oxyacetylene torch. Heating should be done rapidly by the torch and the surface quenched, before appreciable heat transfer to the core occurs

9.4.3. Laser hardening

In this case, a laser beam can be used for surface hardening. As laser beams are of high intensity, a lens is used to reduce the intensity by producing a defocused spot of size ranging from 0.5 to 25 mm. Proper control of energy input is necessary to avoid melting. Laser hardening has the advantage of precise control over the area to be hardened, an ability to harden reentrant surfaces, very high speed of hardening and no separate quenching step. The disadvantage is that the hardening is shallower than in induction and flame hardening

9.4.4. Carburizing

Carburizing is the most widely used method of surface hardening. Here, the surface layers of low carbon steel are enriched with carbon up to 0.8-1.0%. The source of carbon may be a solid medium, a liquid or a gas. In all cases, the carbon enters the steel at the surface and diffuses into the steel as a function of time at an elevated temperature. Carburizing is done at 920-950^o C. This fully austenitic state is essential. If carburizing is done in the ferritic region, the carbon, with very limited solubility in ferrite, tends to form massive cementite particles near the surface, making the subsequent heat treatment difficult. For this reason, carburizing is always done in the austenitic state, even though longer times are required due to the diffusion rate of carbon in austenite being less that in ferrite at such temperatures.

9.4.5. Cyaniding

Cyaniding is done in a liquid bath of NaCN, with the concentration varying between 30 and 97%. The temperature used for cyaniding is lower than that for carburizing and is in the range of 800-870^o C. The time of cyaniding is 1-3 hr to produce a case depth of 0.25 mm or less

9.4.5. Nitriding

Nitriding is carried out in the ferritic region. No phase change occurs after nitriding. The part to be nitrided should posses the required core properties prior to nitriding. During nitriding, pure ammonia decomposes to yield nitrogen which enters the steel. The solubility of nitrogen in ferrite is small. Most of the nitrogen, that enters the steel, forms hard nitrides (e.g., Fe_3N). The temperature of nitriding is 500-590^o C. The time for a case depth of 0.02 mm is about 2 hr. In addition to providing outstanding wear resistance, the nitride layer increases the resistance of carbon steel to corrosion in moist atmospheres.

References

1. G. E. Dieter, Mechanical Metallurgy, Third Edition, McGraw-Hill, New York, 1986.
2. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
3. ASM Handbook, Heat treating, Vol. 4, ASM International, Materials Park, OH, 1991.

Chapter 10. Applications and Processing of Ceramics

Ceramics form an important part of materials group. Ceramics are compounds between metallic and nonmetallic elements for which the inter-atomic bonds are either ionic or predominantly ionic. The term ceramics comes from the Greek word *keramikos* which means ‘burnt stuff’. Characteristic properties of ceramics are, in fact, optimized through thermal treatments. They exhibit physical properties those are different from that of metallic materials. Thus metallic materials, ceramics, and even polymers tend to complement each other in service.

10.1 Types and applications of ceramics

Ceramics greatly differ in their basic composition. The properties of ceramic materials also vary greatly due to differences in bonding, and thus found a wide range of engineering applications. Classification of ceramics based on their specific applications and composition are two most important ways among many.

Based on their composition, ceramics are classified as:

Oxides,
Carbides,
Nitrides,
Sulfides,
Fluorides, etc.

The other important classification of ceramics is based on their application, such as:

Glasses,
Clay products,
Refractories,
Abrasives,
Cements,

Advanced ceramics.

In general, ceramic materials used for engineering applications can be divided into two groups: traditional ceramics, and the engineering ceramics. Typically, traditional ceramics are made from three basic components: clay, silica (flint) and feldspar. For example bricks, tiles and porcelain articles. However, engineering ceramics consist of highly pure compounds of aluminium oxide (Al_2O_3), silicon carbide (SiC) and silicon nitride (Si_3N_4).

Glasses: glasses are a familiar group of ceramics – containers, windows, mirrors, lenses, etc. They are non-crystalline silicates containing other oxides, usually CaO, Na_2O, K_2O and Al_2O_3 which influence the glass properties and its color. Typical property of glasses that is important in engineering applications is its response to heating. There is no definite temperature at which the liquid transforms to a solid as with crystalline materials. A specific temperature, known as glass transition temperature or fictive temperature is defined based on viscosity above which material is named as super cooled liquid or liquid, and below it is termed as glass.

Clay products: clay is the one of most widely used ceramic raw material. It is found in great abundance and popular because of ease with which products are made. Clay products are mainly two kinds – structural products (bricks, tiles, sewer pipes) and white-wares (porcelain, chinaware, pottery, etc.).

Refractories: these are described by their capacity to withstand high temperatures without melting or decomposing; and their inertness in severe environments. Thermal insulation is also an important functionality of refractories.

Abrasive ceramics: these are used to grind, wear, or cut away other material. Thus the prime requisite for this group of materials is hardness or wear resistance in addition to high toughness. As they may also exposed to high temperatures, they need to exhibit some refractoriness. Diamond, silicon carbide, tungsten carbide, silica sand, aluminium oxide / corundum are some typical examples of abrasive ceramic materials.

Cements: cement, plaster of paris and lime come under this group of ceramics. The characteristic property of these materials is that when they are mixed with water, they form slurry which sets subsequently and hardens finally. Thus it is possible to form virtually any shape. They are also used as bonding phase, for example between construction bricks.

Advanced ceramics: these are newly developed and manufactured in limited range for specific applications. Usually their electrical, magnetic and optical properties and combination of properties are exploited. Typical applications: heat engines, ceramic armors, electronic packaging, etc.

Some typical ceramics and respective applications are as follows:

Aluminium oxide / Alumina (Al_2O_3): it is one of most commonly used ceramic material. It is used in many applications such as to contain molten metal, where material is operated at very high temperatures under heavy loads, as insulators in spark plugs, and in some unique applications such as dental and medical use. Chromium doped alumina is used for making lasers.

Aluminium nitride (AlN): because of its typical properties such as good electrical insulation but high thermal conductivity, it is used in many electronic applications such as in electrical circuits operating at a high frequency. It is also suitable for integrated circuits. Other electronic ceramics include – barium titanate ($BaTiO_3$) and Cordierite ($2MgO\text{-}2Al_2O_3\text{-}5SiO_2$).

Diamond (C): it is the hardest material known to available in nature. It has many applications such as industrial abrasives, cutting tools, abrasion resistant coatings, etc. it is, of course, also used in jewelry.

Lead zirconium titanate (PZT): it is the most widely used piezoelectric material, and is used as gas igniters, ultrasound imaging, in underwater detectors.

Silica (SiO_2): is an essential ingredient in many engineering ceramics, thus is the most widely used ceramic material. Silica-based materials are used in thermal insulation, abrasives, laboratory glassware, etc. it also found application in communications media as integral part of optical fibers. Fine particles of silica are used in tires, paints, etc.

Silicon carbide (SiC): it is known as one of best ceramic material for very high temperature applications. It is used as coatings on other material for protection from extreme temperatures. It is also used as abrasive material. It is used as reinforcement in many metallic and ceramic based composites. It is a semiconductor and often used in high temperature electronics. Silicon nitride (Si_3N_4) has properties similar to those of SiC but is somewhat lower, and found applications in such as automotive and gas turbine engines.

Titanium oxide (TiO_2): it is mostly found as pigment in paints. It also forms part of certain glass ceramics. It is used to making other ceramics like $BaTiO_3$.

Titanium boride (TiB_2): it exhibits great toughness properties and hence found applications in armor production. It is also a good conductor of both electricity and heat.

Uranium oxide (UO_2): it is mainly used as nuclear reactor fuel. It has exceptional dimensional stability because its crystal structure can accommodate the products of fission process.

Yttrium aluminium garnet (YAG, $Y_3Al_5O_{12}$): it has main application in lasers (Nd-YAG lasers).

Zirconia (ZrO_2): it is also used in producing many other ceramic materials. It is also used in making oxygen gas sensors, as additive in many electronic ceramics. Its single crystals are part of jewelry.

10.2 Fabrication and processing of ceramics

Ceramics melt at high temperatures and they exhibit a brittle behavior under tension. As a result, the conventional melting, casting and thermo-mechanical processing routes are not suitable to process the polycrystalline ceramics. Inorganic glasses, though, make use of lower melting temperatures due to formation of eutectics. Hence, most ceramic products are made from ceramic powders through powder processing starting with ceramic powders. The powder processing of ceramics is very close to that of metals, powder metallurgy. However there is an important consideration in ceramic-forming that is more prominent than in metal forming: it is dimensional tolerance. Post forming shrinkage is much higher in ceramics processing because of the large differential between the final density and the as-formed density.

Glasses, however, are produced by heating the raw materials to an elevated temperature above which melting occurs. Most commercial glasses are of the silica-soda-lime variety, where silica is supplied in form of common quartz sand, soda (Na_2O) in form of soda ash (Na_2CO_3) while the lime (CaO) is supplied in form of limestone ($CaCO_3$). Different forming methods- pressing, blowing, drawing and fiber forming- are widely in practice to fabricate glass products. Thick glass objects such as plates and dishes are produced by pressing, while the blowing is used to produce objects like jars, bottles and light bulbs. Drawing is used to form long objects like tubes, rods, fibers, whiskers etc. The pressing and blowing process is shown in figure 10.1

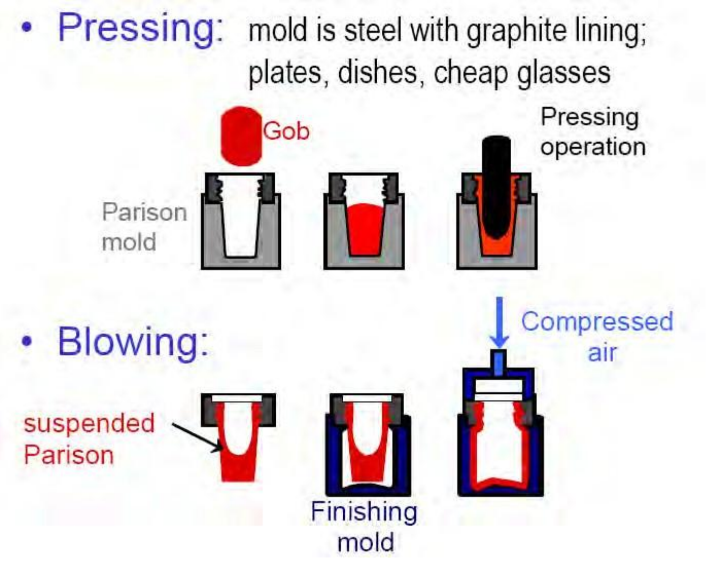


Figure 10.1: *Schematic diagram of pressing and blowing processes*

Ceramic powder processing consists of powder production by milling/grinding, followed by fabrication of green product, which is then consolidated to obtain the final product. A powder is a collection of fine particles. Synthesis of powder involves getting it ready for shaping by crushing, grinding, separating impurities, blending different powders, drying to form soft agglomerates. Different techniques such as compaction, tape casting, slip casting, injection molding and extrusion are then used to convert processed powders into a desired shape to form what is known as green ceramic. The green ceramic is then consolidated further using a high-temperature treatment known as sintering or firing.

As-mined raw materials are put through a milling or grinding operation in which particle size is reduced to and physically ‘liberate’ the minerals of interest from the rest of the ‘gangue’ material. Wet milling is much more common with ceramic materials than with metals. The combination of dry powders with a dispersant such as water is called slurry. Ball- and vibratory- milling is employed to further reduce the size of minerals and to blend different powders.

Ceramic powders prepared are shaped using number of techniques, such as casting, compaction, extrusion/hydro-plastic forming, injection molding. **Tape casting**, also known as *doctor blade process*, is used for the production of thin ceramic tapes. The schematic diagram of tape casting process is shown in figure 10.2. In this technique slurry containing ceramic particles, solvent, plasticizers, and binders is then made to flow under a blade and onto a plastic substrate. The shear thinning slurry spreads under the blade. The tape is then dried using clean hot air. Later-on the tape is subjected to binder burnout and sintering operations. Tape thickness normally range between 0.1 and 2 mm. Commercially important electronic packages based on alumina substrates and barium titanate capacitors are made using this technique. A schematic diagram of doctor blade process is shown in the figure.

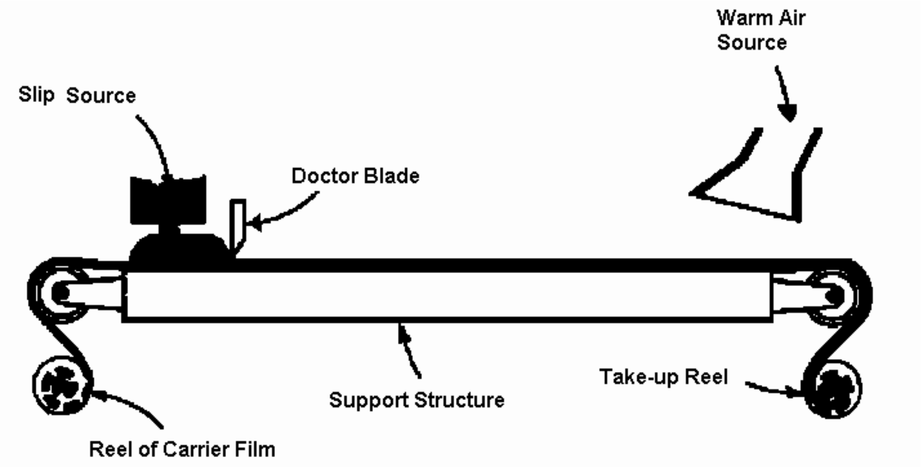


Figure 10.2: *Schematic diagram of tape casting process*

Slip casting is another casting technique widely used. This technique uses aqueous slurry, also known as slip, of ceramic powder. The slip is poured into a plaster of Paris ($CaSO_4$:$2H_2O$) mold. As the water from slurry begins to move out by capillary action, a thick mass builds along the mold wall. When sufficient product thickness is built, the rest of the slurry is poured out (drain casting). It is also possible to continue to pour more slurry in to form a solid piece (solid casting). The schematic diagram of slip casting process is shown in figure 10.3

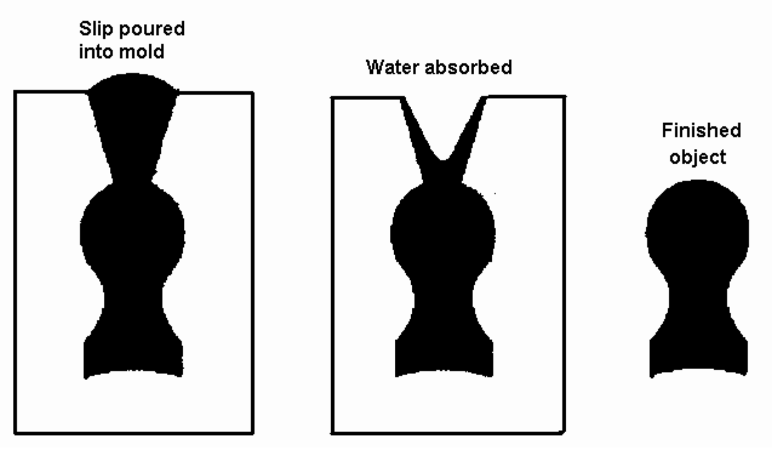


Figure 10.3: *Schematic diagram of slip casting process*

Extrusion and injection molding techniques are used to make products like tubes, bricks, tiles etc. The basis for **extrusion** process is a viscous mixture of ceramic particles, binder and other additives, which is fed through an extruder where a continuous shape of green ceramic is produced. The product is cut to required lengths and then dried and sintered. **Injection molding** of ceramics is similar to that of polymers. Ceramic powder is mixed with a plasticizer, a thermoplastic polymer, and additives. Then the mixture is injected into a die with use of an extruder. The polymer is then burnt off and the rest of the ceramic shape is sintered at suitable high temperatures. Ceramic injection molding is suitable for producing complex shapes. Figure 10.4 shows schematically the injection molding process

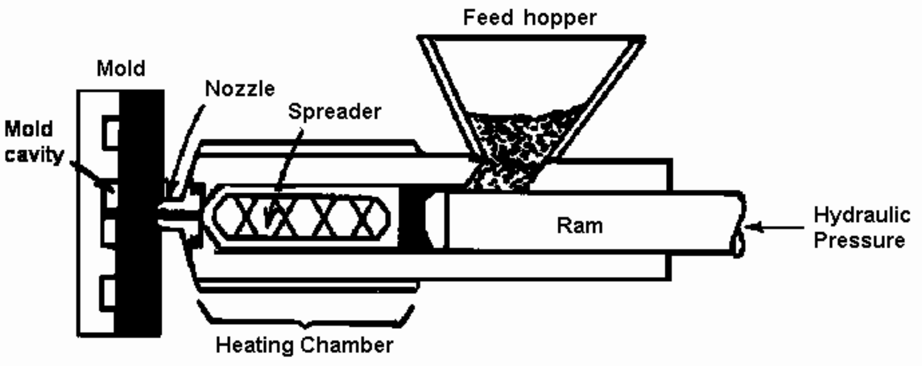


Figure 10.4: Schematic diagram of Injection molding

Most popular technique to produce relatively simple shapes of ceramic products in large numbers is combination of compaction and sintering. For example: electronic ceramics, magnetic ceramics, cutting tools, etc. **Compaction** process is used to make green ceramics that have respectable strength and can be handled and machined. Time for compaction process varies from within a minute to hours depending on the complexity and size of the product. Basically compaction process involves applying equal pressure in all directions to a mixture ceramic powder to increase its density. In some cases, compaction involves application of pressure using oil/fluid at room temperatures, called cold iso-static pressing (**CIP**). Then the green ceramic is sintered with or without pressure. CIP is used to achieve higher ceramic density or where the compaction of more complex shapes is required. In some instances, parts may be produced under conditions in which compaction and sintering are conducted under pressure at elevated temperatures. This technique is known as hot iso-static pressing (**HIP**), and is used for refractory and covalently bonded ceramics that do not show good bonding characteristics under CIP. HIP is also used when close to none porosity is the requirement. Another characteristic feature of HIP is high densities can be achieved without appreciable grain growth.

Sintering is the firing process applied to green ceramics to increase its strength. Sintering is carried out below the melting temperature thus no liquid phase presents during sintering. However, for sintering to take place, the temperature must generally be maintained above one-half the absolute melting point of the material. During sintering, the green ceramic product shrinks and experiences a reduction in porosity. This leads to an improvement in its mechanical integrity. These changes involve different mass transport mechanisms that cause coalescence of powder particles into a more dense mass. With sintering, the grain boundary and bulk atomic diffusion contribute to densification, surface diffusion and evaporation condensation can cause grain growth, but do not cause densification. After pressing, ceramic particles touch one another. During initial stages of sintering, necks form along the contact regions between adjacent particles thus every interstice between particles becomes a pore. The pore channels in the compact grow in size, resulting in a significant increase in strength. With increase in sintering time, pores become smaller in size. The driving force for sintering process is the reduction in total particle surface area, and thus the reduction in total surface energy. During sintering, composition, impurity control and oxidation protection are provided by the use of vacuum conditions or inert gas atmospheres.

References

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, Introduction to Ceramics, Second Edition, Wiley, New York, 1976.
2. L. H. Van Vlack, Physical Ceramics for Engineers, Addison-Wesley Longman, Reading, MA, 1964.
3. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.

Chapter 11. Applications and Processing of Polymers

Polymers play a very important role in human life. In fact, our body is made of lot of polymers, e.g. Proteins, enzymes, etc. Other naturally occurring polymers like wood, rubber, leather and silk are serving the humankind for many centuries now. Modern scientific tools revolutionized the processing of polymers thus available synthetic polymers like useful plastics, rubbers and fiber materials. As with other engineering materials (metals and ceramics), the properties of polymers are related their constituent structural elements and their arrangement. The suffix in polymer 'mer' is originated from Greek word *meros* – which means part. The word polymer is thus coined to mean material consisting of many parts/mers. Most of the polymers are basically organic compounds, however they can be inorganic (e.g. silicones based on Si-O network). This chapter introduces classification of polymers, processing and synthesis of polymers, followed by mechanism of deformation and mechanical behavior of polymers.

11.1 Polymer types and Polymer synthesis & processing

Polymers are classified in several ways – by how the molecules are synthesized, by their molecular structure, or by their chemical family. For example, linear polymers consist of long molecular chains, while the branched polymers consist of primary long chains and secondary chains that stem from these main chains. However, linear does not mean straight lines. The better way to classify polymers is according to their mechanical and thermal behavior. Industrially polymers are classified into two main classes – *plastics* and *elastomers*.

Plastics are moldable organic resins. These are either natural or synthetic, and are processed by forming or molding into shapes. Plastics are important engineering materials for many reasons. They have a wide range of properties, some of which are unattainable from any other materials, and in most cases they are relatively low in cost. Following is the brief list of properties of plastics: light weight, wide range of colors, low thermal and electrical conductivity, less brittle, good toughness, good resistance to acids, bases and moisture, high dielectric strength (use in electrical insulation), etc. Plastics are again classified in two groups depending on their mechanical and thermal behavior as *thermoplasts* (thermoplastic polymers) and *thermosets* (thermosetting polymers).

Thermoplasts: These plastics soften when heated and harden when cooled – processes that are totally reversible and may be repeated. These materials are normally fabricated by the simultaneous application of heat and pressure. They are linear polymers without any cross-linking in structure where long molecular chains are bonded to each other by secondary bonds and/or inter-wined. They have the property of increasing plasticity with increasing temperature which breaks the secondary bonds between individual chains. Common thermoplasts are: acrylics, PVC, nylons, polypropylene, polystyrene, polymethyl methacrylate (plastic lenses or perspex), etc.

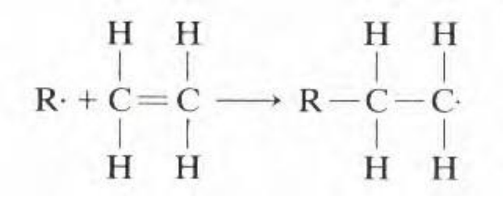
Thermosets: These plastics require heat and pressure to mold them into shape. They are formed into a permanent shape and cured or 'set' by chemical reactions such as extensive cross-linking. They cannot be re-melted or reformed into another shape but decompose upon being heated to too high a temperature. Thus thermosets cannot be recycled, whereas thermoplasts can be recycled. The term thermoset implies that heat is required to permanently set the plastic. Most thermosets composed of long chains that are strongly cross-linked (and/or covalently bonded) to one another to form 3-D network structures to form a rigid solid. Thermosets are generally stronger, but more brittle than thermoplasts. Advantages of thermosets for engineering design applications include one or more of the following: high thermal stability, high dimensional stability, high rigidity, light weight, high electrical and thermal insulating properties and resistance to creep and deformation under load. There are two methods whereby cross-linking reaction can be initiated – cross-linking can be accomplished by heating the resin in a suitable mold (e.g. bakelite), or resins such as epoxies (araldite) are cured at low temperature by the addition of a suitable cross-linking agent, an amine. Epoxies, vulcanized rubbers, phenolics, unsaturated polyester resins, and amino resins (ureas and melamines) are examples of thermosets.

Elastomers: Also known as rubbers, these are polymers which can undergo large elongations under load, at room temperature, and return to their original shape when the load is released. There are number of man-made elastomers in addition to natural rubber. These consist of coil-like polymer chains those can reversibly stretch by applying a force.

Processing of polymers mainly involves preparing a particular polymer by synthesis of available raw materials, followed by forming into various shapes. Raw materials for polymerization are usually derived from coal and petroleum products. The large molecules of many commercially useful polymers must be synthesized from substances having smaller molecules. The synthesis of the large molecule polymers is known as polymerization in which monomer units are joined over and over to become a large molecule. More upon, properties of a polymer can be enhanced or modified with the addition of special materials. This is followed by forming operation. Addition polymerization and condensation polymerization are the two main ways of polymerization.

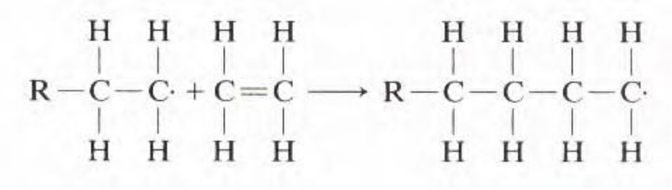
Addition polymerization, also known as chain reaction polymerization, is a process in which multi-functional monomer units are attached one at a time in chainlike fashion to form linear/3-D macro-molecules. The composition of the macro-molecule is an exact multiple of for that of the original reactant monomer. This kind of polymerization involves three distinct stages – initiation, propagation and termination. To initiate the process, an initiator is added to the monomer. This forms free radicals with a reactive site that attracts one of the carbon atoms of the monomer. When this occurs, the reactive site is transferred to the other carbon atom in the monomer and a chain begins to form in propagation stage. A common initiator is benzoyl peroxide. When polymerization is nearly complete, remaining monomers must diffuse a long distance to reach reactive site, thus the growth rate decreases.

The process for polyethylene is as follows



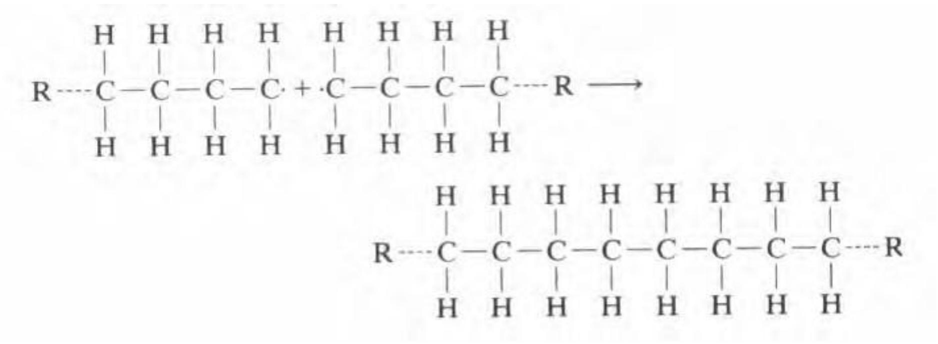
$$R\cdot + CH_2{=}CH_2 \longrightarrow R{-}CH_2{-}CH_2\cdot$$

Here R$^{\cdot}$ represents the active initiator. Propagation involves the linear growth of the molecule as monomer units become attached to one another in succession to produce the chain molecule, which is represented, again for polyethylene, as follows



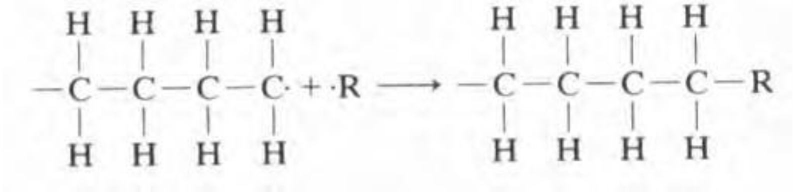
$$R{-}CH_2{-}CH_2\cdot + CH_2{=}CH_2 \longrightarrow R{-}CH_2{-}CH_2{-}CH_2{-}CH_2\cdot$$

As we need polymers with controlled molecular weight, polymerization needs to be terminated at some stage. Propagation may end or terminate in different ways. First, the active ends of two propagating chains may react or link together to form a non-reactive molecule, as follows:



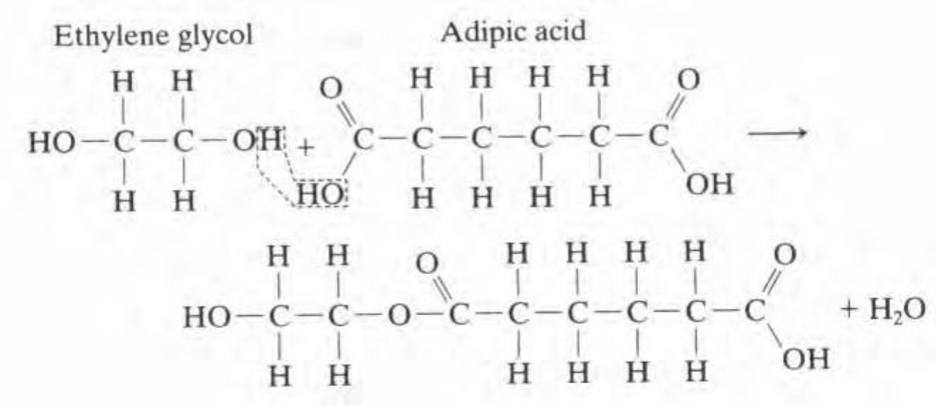
$$R\cdots(CH_2)_3{-}CH_2\cdot + \cdot CH_2{-}(CH_2)_3\cdots R \longrightarrow R\cdots(CH_2)_8\cdots R$$

thus terminating the growth of each chain or an active chain end may react with an initiator or other chemical species having a single active bond, as follows:



with the resultant cessation of chain growth. Polyethylene, polypropylene, PVC, and polystyrene are synthesized using addition polymerization.

Condensation polymerization, also known as step growth polymerization, involves more than one monomer species; and there is usually a small molecular weight by-product such as water, which is eliminated. The repeat unit here forms from original monomers, and no product has the chemical formula of mere one mer repeat unit. The polymerization of dimethyl terephthalate and ethylene glycol to produce polyester is an important example. The by-product, methyl alcohol, is condensed off and the two monomers combine to produce a larger molecule (mer repeat unit). Another example, consider the formation of a polyester from the reaction between ethylene glycol and adipic acid; the intermolecular reaction is as follows:



This stepwise process is successively repeated, producing, in this case, a linear molecule. The intermolecular reaction occurs every time a mer repeat unit is formed. Reaction times for condensation are generally longer than for addition polymerization. Polyesters, phenol-formaldehyde, nylons, polycarbonates etc are produced by condensation polymerization. Condensation polymerization reactions also occur in sol-gel processing of ceramic materials. Some polymers such as nylon may be polymerized by either technique.

Polymers, unlike organic/inorganic compounds, do not have a fixed molecular weight. It is specified in terms of *degree of polymerization* – number of repeat units in the chain or ration of average molecular weight of polymer to molecular weight of repeat unit. Average molecular weight is however defined in two ways. *Weight average molecular* *weight* is obtained by dividing the chains into size ranges and determining the fraction of chains having molecular weights within that range. *Number average molecular weight* is based on the number fraction, rather than the weight fraction, of the chains within each size range. It is always smaller than the weight average molecular weight.

Most of polymer properties are intrinsic i.e. characteristic of a specific polymer. Foreign substances called additives are intentionally introduced to enhance or modify these properties. These include – fillers, plasticizers, stabilizers, colorants, and flame retardants. Fillers are used to improve tensile and compressive strength, abrasion resistance, dimensional stability etc. wood flour, sand, clay, talc etc are example for fillers. Plasticizers aid in improving flexibility, ductility and toughness of polymers by lowering glass transition temperature of a polymer. These are generally liquids of low molecular weight. Stabilizers are additives which counteract deteriorative processes such as oxidation, radiation, and environmental deterioration. Colorants impart a specific color to a polymer, added in form of either dyes (dissolves) or pigments (remains as a separate phase). Flame retardants are used to enhance flammability resistance of combustible polymers. They serve the purpose by interfering with the combustion through the gas phase or chemical reaction.

Polymeric materials are formed by quite many different techniques depending on (a) whether the material is thermoplast or thermoset, (b) melting/degradation temperature, (c) atmospheric stability, and (d) shape and intricacy of the product. Polymers are often formed at elevated temperatures under pressure. Thermoplasts are formed above their glass transition temperatures while applied pressure ensures that the product retain its shape. Thermosets are formed in two stages – making liquid polymer, then molding it.

Different molding techniques are employed in fabrication of polymers. *Compression molding* involves placing appropriate amount of polymer with additives between heated male and female mold parts. After pouring polymer, mold is closed, and heat and pressure are applied, causing viscous plastic to attain the mold shape. *Figure*-11.1 shows a typical mould employed for compression molding.

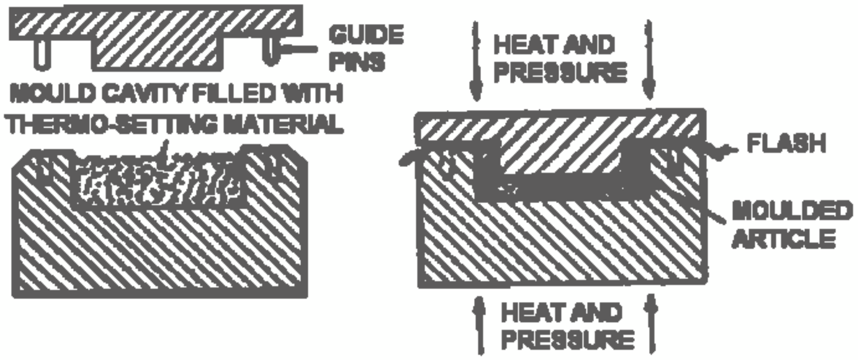


Figure 11.1: *Schematic diagram of a mould employed for compression molding*

Transfer molding differs from compression molding in how the materials is introduced into the mold cavities. In transfer molding the plastic resin is not fed directly into the mold cavity but into a chamber outside the mold cavities. When the mold is closed, a plunger forces the plastic resin into the mold cavities, where and molded material cures. In *injection molding,* palletized materials is fed with use of hopper into a cylinder where charge is pushed towards heating chamber where plastic material melts, and then molten plastic is impelled through nozzle into the enclosed mold cavity where product attains its shape. Most outstanding characteristic of this process is the cycle time which is very short. The schematic diagram of injection-molding machine is shown in *figure-11.2*

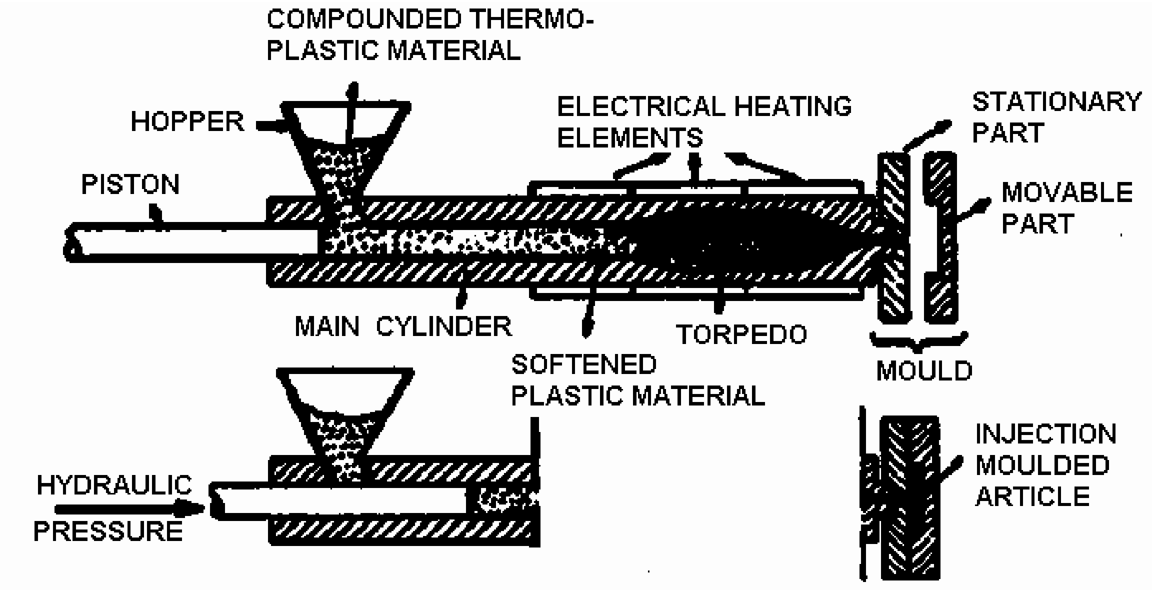


Figure 11.2: *Schematic diagram of injection-molding machine*

Extrusion is another kind of injection molding, in which a thermoplastic material is forced through a die orifice, similar to the extrusion of metals. This technique is especially adapted to produce continuous lengths with constant cross-section. The schematic diagram of a simple extrusion machine is shown in *figure- 11.3*

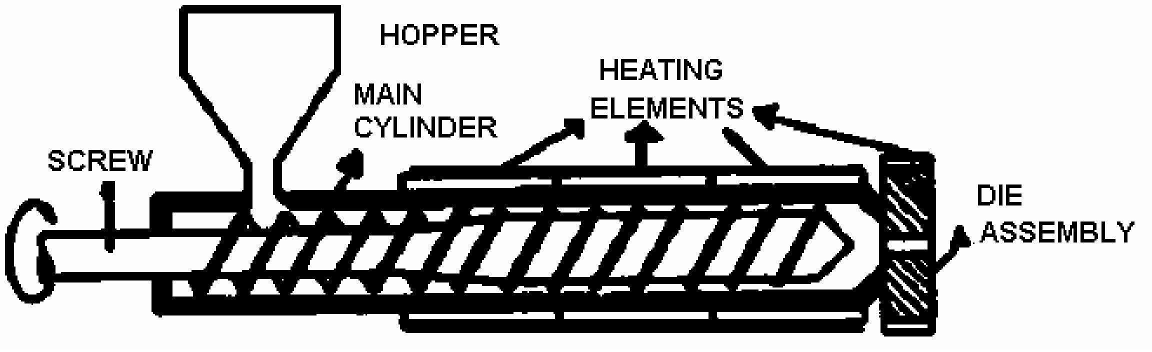


Figure 11.3: *Schematic diagram of a simple extrusion machine*

Blow molding of plastics is similar to blowing of glass bottles. Polymeric materials may be cast similar to metals and ceramics.

11.2 Crystallization, melting and glass transition

Polymers are known by their high sensitivity of mechanical and/or thermal properties. This section explains their thermal behavior. During processing of polymers, they are cooled with/without presence of presence from liquid state to form final product. During cooling, an ordered solid phase may be formed having a highly random molecular structure. This process is called crystallization. The melting occurs when a polymer is heated. If the polymer during cooling retains amorphous or non-crystalline state i.e. disordered molecular structure, rigid solid may be considered as frozen liquid resulting from glass transition. Thus, enhancement of either mechanical and/or thermal properties needs to consider crystallization, melting, and the glass transition.

Crystallization and the mechanism involved play an important role as it influences the properties of plastics. As in solidification of metals, polymer crystallization involves nucleation and growth. Near to solidification temperature at favorable places, nuclei forms, and then nuclei grow by the continued ordering and alignment of additional molecular segments. Extent of crystallization is measured by volume change as there will be a considerable change in volume during solidification of a polymer. Crystallization rate is dependent on crystallization temperature and also on the molecular weight of the polymer. Crystallization rate decreases with increasing molecular weight.

Melting of polymer involves transformation of solid polymer to viscous liquid upon heating at melting temperature, $\boldsymbol{T_m}$. Polymer melting is distinctive from that of metals in many respects – melting takes place over a temperature range; melting behavior depends on history of the polymer; melting behavior is a function of rate of heating, where increasing rate results in an elevation of melting temperature. During melting there occurs rearrangement of the molecules from ordered state to disordered state. This is influenced by molecular chemistry and structure (degree of branching) along with chain stiffness and molecular weight.

Glass transition occurs in amorphous and semi-crystalline polymers. Upon cooling, this transformation corresponds to gradual change of liquid to rubbery material, and then rigid solid. The temperature range at which the transition from rubbery to rigid state occurs is termed as glass transition temperature, $\boldsymbol{T_g}$. This temperature has its significance as abrupt changes in other physical properties occur at this temperature. Glass transition temperature is also influenced by molecular weight, with increase of which glass transition temperature increases. Degree of cross-linking also influences the glass transition such that polymers with very high degree of cross-linking do not experience a glass transition. The glass transition temperature is typically 0.5 to 0.75 times the absolute melting temperature. Above the glass transition, non-crystalline polymers show viscous behavior, and below the glass transition they show glass-brittle behavior (as chain motion is very restricted), hence the name glass transition.

Melting involves breaking of the inter-chain bonds, so the glass- and melting-temperatures depend on:

- chain stiffness (e.g., single vs. double bonds)
- size, shape of side groups
- size of molecule
- side branches, defects
- cross-linking

11.3 Mechanical behavior of polymers

Polymer mechanical properties can be specified with many of the same parameters that are used for metals such as modulus of elasticity, tensile/impact/fatigue strengths, etc. However, polymers are, in many respects, mechanically dissimilar to metals. To a much greater extent than either metals or ceramics, both thermal and mechanical properties of polymers show a marked dependence on parameters namely temperature, strain rate, and morphology. In addition, molecular weight and temperature relative to the glass transition play an important role that are absent for other type of materials.

A simple stress- strain curve can describe different mechanical behavior of various polymers. As shown in *figure* – 11.4, the stress-strain behavior can be brittle, plastic and highly elastic (elastomeric or rubber-like). Mechanical properties of polymers change dramatically with temperature, going from glass-like brittle behavior at low temperatures to a rubber-like behavior at high temperatures. Highly crystalline polymers behave in a brittle manner, whereas amorphous polymers can exhibit plastic deformation. These phenomena are highly temperature dependent, even more so with polymers than they are with metals and ceramics. Due to unique structures of cross-linked polymers, recoverable deformations up to very high strains / point of rupture are also observed with polymers (elastomers). Tensile modulus (modulus) and tensile strengths are orders of magnitude smaller than those of metals, but elongation can be up to 1000 % in some cases. The tensile strength is defined at the fracture point and can be lower than the yield strength.

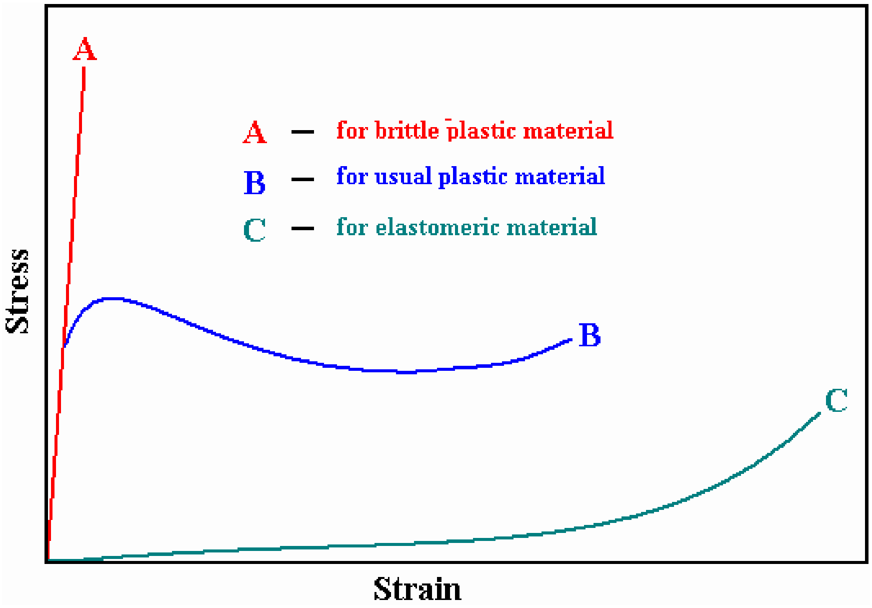


Figure-11.4: *Typical stress-strain curves for polymers.*

As the temperature increases, both the rigidity and the yield strength decrease, while the elongation increases. Thus, if high rigidity and toughness are the requirements, the temperature consideration is important. In general, decreasing the strain rate has the same influence on the strain-strength characteristics as increasing the temperature: the material becomes softer and more ductile. Despite the similarities in yield behavior with temperature and strain rate between polymers, metals, and ceramics, the mechanisms are quite different. Specifically, the necking of polymers is affected by two physical factors that are not significant in metals: dissipation of mechanical energy as heat, causing softening magnitude of which increases with strain rate; deformation resistance of the neck, resulting in strain-rate dependence of yield strength. The relative importance of these two factors depends on materials, specimen dimensions and strain rate. The effect of temperature relative to the glass transition is depicted in terms of decline in modulus values. Shallow decline of modulus is attributed to thermal expansion, whereas abrupt changes are attributable to viscoelastic relaxation processes.

Together molecular weight and crystallinity influence a great number of mechanical properties of polymers including hardness, fatigue resistance, elongation at neck, and even impact strength. The chance of brittle failure is reduced by raising molecular weight, which increases brittle strength, and by reducing crystallinity. As the degree of crystallinity decreases with temperature close to melting point, stiffness, hardness and yield strength decrease. These factors often set limits on the temperature at which a polymer is useful for mechanical purposes.

Elastomers, however, exhibit some unique mechanical behavior when compared to conventional plastics. The most notable characteristics are the low modulus and high deformations as elastomers exhibit large, reversible elongations under small applied stresses. Elastomers exhibit this behavior due to their unique, cross-linked structure.

Elastic modulus of elastomers (resistance to the uncoiling of randomly orientated chains) increases as with increase in temperature. Unlike non-cross-linked polymers, elastomers exhibit an increase inelastic modulus with cross-link density.

11.4 Mechanisms of deformation and strengthening of polymers

An understanding of deformation mechanisms of polymers is important in order to be able to manage the optimal use of these materials, a class of materials that continues to grow in terms of use in structural applications. Despite the similarities in ductile and brittle behavior with to metals and ceramics respectively, elastic and plastic deformation mechanisms in polymers are quite different. This is mainly due to (a) difference in structure they made of and (b) size of the entities responsible for deformation. Plastic deformation in metals and ceramics can be described in terms of dislocations and slip planes, whereas polymer chains must undergo deformation in polymers leading to different mechanism of permanent deformation. Unique to most of the polymers is the viscoelasticity – means when an external force is applied, both elastic and plastic (viscous) deformation occur. For viscoelastic materials, the rate of strain determines whether the deformation in elastic or viscous. The viscoelastic behavior of polymeric materials is dependent on both time and temperature.

Plastic polymers deform elastically by elongation of the chain molecules from their stable conformations in the direction of the applied stress by the bending and stretching of the strong covalent bonds. In addition, there is a possibility for slight displacement of adjacent molecules, which is resisted by weak secondary / van der Waals bonds. Plastic deformation in polymers is not a consequence of dislocation movement as in metals. Instead, chains rotate, stretch, slide and disentangle under load to cause permanent deformation. This permanent deformation in polymers might occur in several stages of interaction between lamellar and intervening amorphous regions. Initial stages involve elongation of amorphous tie chains, and eventual alignment in the loading direction. Continues deformation in second stage occurs by the tilting of the lamellar blocks. Next, crystalline block segments separate before blocks and tie chains become orientated in the direction of tensile axis in final stage. This leads to highly orientated structure in deformed polymers.

Elastomers, on the other hand, deform elastically by simple uncoiling, and straightening of molecular chains that are highly twisted, kinked, and coiled in unstressed state. The driving force for elastic deformation is change in entropy, which is a measure of degree of disorder in a system. When an elastomer is stretched, the system's order increases. If elastomer is released from the applied load, its entropy increases. This entropy effect results in a rise in temperature of an elastomer when stretched. It also causes the modulus of elasticity to increase with increasing temperature, which is opposite to the behavior of other materials.

Fracture of polymers is again dependent on morphology of a polymer. As a thumb rule, thermosets fracture in brittle mode. It involves formation of cracks at regions where there is a localized stress concentration. Covalent bonds are severed during the fracture.

However, both ductile and brittle modes are possible mode of fracture for thermoplasts. Many of thermoplasts can exhibit ductile-to-brittle transition assisted by reduction in temperature, increase in strain rate, presence of notch, increased specimen thickness and a modification of the polymer structure. Unique to polymer fracture is crazing – presence of regions of very localized yielding, which lead to formation of small and interconnected microvoids. Crazes form at highly stressed regions associated with scratches, flaws and molecular inhomogeneties; and they propagate perpendicular to the applied tensile stress and typically are 5 μm or less thick. A craze is different from a crack as it can support a load across its face.

The deformation of plastic materials can be primarily elastic, plastic, or a combination of both types. The deformation mode and resistance of deformation depends on many parameters for different plastics. The following factors influence the strength of a thermoplast: average molecular mass, degree of crystallization, presence of side groups, presence of polar and other specific atoms, presence of phenyl rings in main chains and addition of reinforcements. Effect of every one of these factor can be used to strengthen a thermoplast. Thermosets are, however, strengthened by reinforcement methods.

Strength of a thermoplast is directly dependent on its average molecular mass since polymerization up to a certain molecular-mass range is necessary to produce a stable solid. This method is not used so often as after a critical mass range, increasing the average molecular mass does not greatly increase its strength. In general, as the degree of crystallinity increases, the strength, modulus and density all increase for a thermoplast. Another method to increase the strength is to create more resistance to chain slippage. This can be achieved by addition of bulky side groups on main chains, which results in increase of strength but reduces the ductility. Increased resistance to chain slippage can be achieved by increasing the molecular bonding forces between the polymer chains. E.g.: introducing a chlorine atom on every other carbon atom of main chain to make polyvinylchloride (PVC). Introducing an ether linkage (i.e. introduction of oxygen atom) or amide linkage (i.e. introduction of oxygen and nitrogen atoms) into the main chain can increase the rigidity of thermoplasts. One of the most important strengthening methods for thermoplasts is the introduction of phenylene rings in the main chain. It is commonly used for high-strength engineering plastics. The phenylene rings cause steric hindrance to rotation within the polymer chain and electronic attraction of resonating electrons between adjacent molecules. Another method of strengthening is introduction of reinforcements like glass fibers. Glass content ranges from 20 to 40%, depending on trade-off between desired strength, ease of processing and economics.

Thermosets are strengthened by reinforcements again. Different reinforcements are in use according to the necessity. Glass fibers are most commonly used to form structural and molding plastic compounds. Two most important types of glass fibers are E (electrical)- and S (high strength)- glasses. *E-glass* (lime-aluminium-borosilicate glass with zero or low sodium and potassium levels) is often used for continuous fibers. *S-glass* (65%SiO_2, 25%Al_2O_3 and 10% MgO) has higher strength-to-weight ratio and is more expansive thus primary applications include military and aerospace applications. Carbon fiber reinforced plastics are also often used in aerospace applications. However they are very expansive.

The other classes of reinforcements include aramid (aromatic polyamide) fibers. They are popularly known as *Kevlar*. Presently two commercial variants of Kevlar are available – Kevlar29 and Kevlar49. Kevlar29 is a low-density, high strength aramid fiber designed for applications such as ballistic protection, ropes and cables. Kevlar49 is characterized by a low density and high strength/modulus; is used in aerospace, marine, automotive and other industrial applications. Thermosets without reinforcements are strengthened by creation of network of covalent bonds throughout the structure of the material. Covalent bonds can be developed during casting or pressing under heat and pressure.

11.5 Characteristics and typical applications of few plastic materials.

a) Thermo plastics

1. Acrylonitrile-butadiene-styrene (ABS):
Characteristics: Outstanding strength and toughness, resistance to heat distortion; good electrical properties; flammable and soluble in some organic solvents.
Application: Refrigerator lining, lawn and garden equipment, toys, highway safety devices.

2. Acrylics (poly-methyl-methacrylate)
Characteristics: Outstanding light transmission and resistance to weathering; only fair mechanical properties.
Application: Lenses, transparent aircraft enclosures, drafting equipment, outdoor signs

3. Fluorocarbons (PTFE or TFE)
Characteristics: Chemically inert in almost all environments, excellent electrical properties; low coefficient of friction; may be used to 260^{o} C; relatively weak and poor cold-flow properties.
Application: Anticorrosive seals, chemical pipes and valves, bearings, anti adhesive coatings, high temperature electronic parts.

4. Polyamides (nylons)
Characteristics: Good mechanical strength, abrasion resistance, and toughness; low coefficient of friction; absorbs water and some other liquids.
Application: Bearings, gears, cams, bushings, handles, and jacketing for wires and cables

5. Polycarbonates
Characteristics: Dimensionally stable: low water absorption; transparent; very good impact resistance and ductility.
Application: Safety helmets, lenses light globes, base for photographic film

6. Polyethylene
Characteristics: Chemically resistant and electrically insulating; tough and relatively low coefficient of friction; low strength and poor resistance to weathering.
Application: Flexible bottles, toys, tumblers, battery parts, ice trays, film wrapping materials.

7. Polypropylene
Characteristics: Resistant to heat distortion; excellent electrical properties and fatigue strength; chemically inert; relatively inexpensive; poor resistance to UV light.
Application: Sterilizable bottles, packaging film, TV cabinets, luggage

8. Polystyrene
Characteristics: Excellent electrical properties and optical clarity; good thermal and dimensional stability; relatively inexpensive
Application: Wall tile, battery cases, toys, indoor lighting panels, appliance housings.

9. Polyester (PET or PETE)
Characteristics: One of the toughest of plastic films; excellent fatigue and tear strength, and resistance to humidity acids, greases, oils and solvents
Application: Magnetic recording tapes, clothing, automotive tire cords, beverage containers.

b) Thermo setting polymers

1. Epoxies
Characteristics: Excellent combination of mechanical properties and corrosion resistance; dimensionally stable; good adhesion; relatively inexpensive; good electrical properties.
Application: Electrical moldings, sinks, adhesives, protective coatings, used with fiberglass laminates.

2. Phenolics
Characteristics: Excellent thermal stability to over 150^{o} C; may be compounded with a large number of resins, fillers, etc.; inexpensive.
Application: Motor housing, telephones, auto distributors, electrical fixtures.

References

1. V. R. Gowariker, N. V. Viswanathan, and Jayadev Sreedhar, Polymer Science, New Age International (P) Limited publishers, Bangalore, 2001
2. C. A. Harper, Handbook of Plastics Elastomers and Composites, Third Edition, McGrawHill Professional Book Group, New York, 1996.
3. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.

Chapter 12. Composites

There is a great need for materials with special properties with emergence of new technologies. However, conventional engineering materials are unable to meet this requirement of special properties like high strength and low density materials for aircraft applications. Thus, emerged new class of engineering materials – *composites*. Unfortunately, there is no widely accepted definition for a composite material. For the purpose of this module, the following definition is adopted: any multiphase material that is artificially made and exhibits a significant proportion of the properties of the constituent phases. The constituent phases of a composite are usually of macro sized portions, differ in form and chemical composition and essentially insoluble in each other.

Composites are, thus, made by combining two distinct engineering materials in most cases; one is called *matrix* that is continuous and surrounds the other phase – *dispersed phase*. The properties of composites are a function of the properties of the constituent phases, their relative amounts, and size-and-shape of dispersed phase.

Millions of combinations of materials are possible and thus so number of composite materials. For ease of recognition, composite materials are classified based on different criteria like: (1) type of matrix material – metal matrix composites, polymer matrix composites and ceramic matrix composites (2) size-and-shape of dispersed phase – particle-reinforced composites, fiber-reinforced composites and structural composites. It is understandable that properties of composite materials are nothing but improved version of properties of matrix materials due to presence of dispersed phase. However, engineers need to understand the mechanics involved in achieving the better properties. Hence the following sections highlight the mechanics of composites, which depend on size-and-shape of dispersed phase.

12.1 Particle-reinforced composites

This class of composites is most widely used composites mainly because they are widely available and cheap. They are again two kinds: dispersion-strengthened and particulate-reinforced composites. These two classes are distinguishable based upon strengthening mechanism – dispersion-strengthened composites and particulate composites.

In *dispersion-strengthened composites*, particles are comparatively smaller, and are of 0.01-0.1μm in size. Here the strengthening occurs at atomic/molecular level i.e. mechanism of strengthening is similar to that for precipitation hardening in metals where matrix bears the major portion of an applied load, while *dispersoids* hinder/impede the motion of dislocations. Examples: thoria (ThO_2) dispersed Ni-alloys (TD Ni-alloys) with high-temperature strength; SAP (sintered aluminium powder) – where aluminium matrix is dispersed with extremely small flakes of alumina (Al_2O_3).

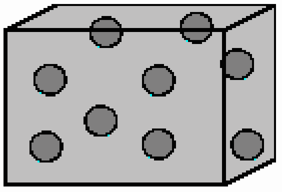


Figure 12.1. Particulate reinforced composite

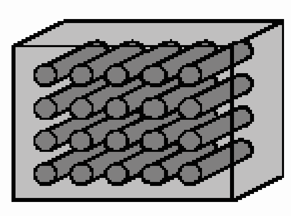


Figure 12.2. Particulate reinforced composite

Particulate composites are other class of particle-reinforced composites. These contain large amounts of comparatively coarse particles. These composites are designed to produce unusual combinations of properties rather than to improve the strength. Mechanical properties, such as elastic modulus, of particulate composites achievable are in the range defined by *rule of mixtures* as follows:

Upper bound is represented by: $E_c(u) = E_m V_m + E_p V_p$

And lower bound is represented by: $E_c(l) = \dfrac{E_m E_p}{E_p V_m + E_m V_p}$

where *E* and *V* denote elastic modulus and volume fractions respectively while *c*, *m*, and *p* represent composite, matrix and particulate phases. A schematic diagram of these bounds is shown in the *figure*-12.1.

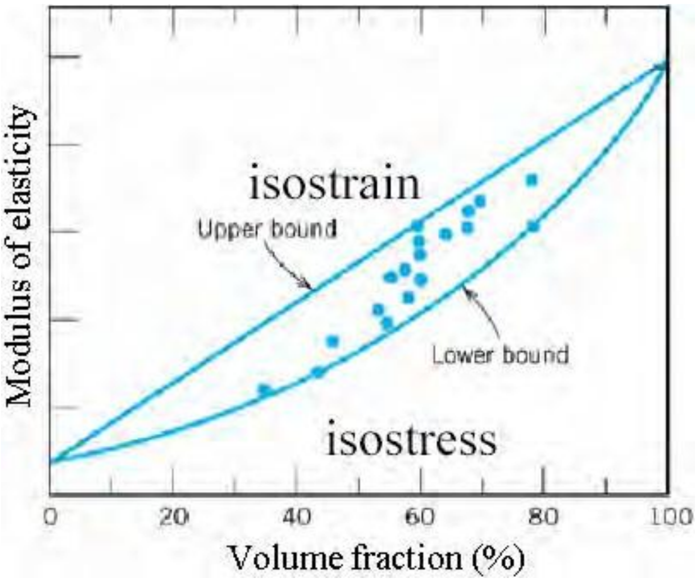


Figure-12.3: *Schematic presentation of rule-of-mixture bounds.*

Particulate composites are used with all three material types – metals, polymers and ceramics. Cermets contain hard ceramic particles dispersed in a metallic matrix. Eg.: tungsten carbide (WC) or titanium carbide (TiC) embedded cobalt or nickel used to make cutting tools. Polymers are frequently reinforced with various particulate materials such as carbon black. When added to vulcanized rubber, carbon black enhances toughness and abrasion resistance of the rubber. Aluminium alloy castings containing dispersed SiC particles are widely used for automotive applications including pistons and brake applications.

Concrete is most commonly used particulate composite. It consists of cement as binding medium and finely dispersed particulates of gravel in addition to fine aggregate (sand) and water. It is also known as Portland cement concrete. Its strength can be increased by additional reinforcement such as steel rods/mesh.

12.2 Fiber-reinforced composites

Most fiber-reinforced composites provide improved strength and other mechanical properties and strength-to-weight ratio by incorporating strong, stiff but brittle fibers into a softer, more ductile matrix. The matrix material acts as a medium to transfer the load to the fibers, which carry most off the applied load. The matrix also provides protection to fibers from external loads and atmosphere.

These composites are classified as either continuous or discontinuous. Generally, the highest strength and stiffness are obtained with continuous reinforcement. Discontinuous fibers are used only when manufacturing economics dictate the use of a process where the fibers must be in this form.

The mechanical properties of fiber-reinforced composites depend not only on the properties of the fiber but also on the degree of which an applied load is transmitted to the fibers by the matrix phase. Length of fibers, their orientation and volume fraction in addition to direction of external load application affects the mechanical properties of these composites.

Effect of fiber length: Some critical length (l_c) is necessary for effective strengthening and stiffening of the composite material, which is defined as:

$$l_c = \frac{\sigma^* d}{2\tau_c}$$

σ^*_f – ultimate/tensile strength of the fiber, d – diameter of the fiber, τ_c – interface bond strength. Fibers for which $l >> l_c$ (normally l >15 l_c) are termed as continuous, discontinuous or short fibers on the other hand.

Effect of fiber orientation and concentration: with respect to orientation, two extremes possibilities are – parallel alignment and random alignment. Continuous fibers are normally aligned, whereas discontinuous fibers are randomly or partially orientated. Two instants of loading are: longitudinal loading and transverse loading.

(a) Continuous fiber composites:

Under *longitudinal loading*, by assuming that deformation of both matrix and fiber is the same i.e. isostrain condition, rule-of-mixtures results in the following:

$$\sigma_c = \sigma_m \frac{A_m}{A_c} + \sigma_f \frac{A_f}{A_c}$$

where A_m/A_c and A_f/A_c are the area fractions of the matrix and fiber phases respectively. In the composite, if matrix and fiber are all of equal length, area fractions will be equal to volume fractions. Thus,

$$\sigma_c = \sigma_m V_m + \sigma_f V_f$$

When the isostrain assumption is taken into account, the above equation transforms into

$$E_{cl} = E_m V_m + E_f V_f = E_m (1 - V_f) + E_f V_f$$

The ratio of the load carried by the fibers to that carried by the matrix is given by

$$\frac{F_f}{F_m} = \frac{E_f V_f}{E_m V_m}$$

In case of *transverse loading*, it is assumed that both matrix and fiber will experience the equal stress i.e. isostress condition. Then the modulus of the composite is given by:

$$E_{ct} = \frac{E_m E_f}{E_f V_m + E_m V_f} = \frac{E_m E_f}{E_f (1 - V_f) + E_m V_f}$$

Longitudinal tensile strength: as mentioned earlier, matrix material is softer i.e. fibers strain less and fail before the matrix. And once the fibers have fractured, majority of the load that was borne by fibers is now transferred to the matrix. Based on this criterion the following equation can be developed for longitudinal strength of the composite:

$$\sigma^*_{cl} = \sigma'_m (1 - V_f) + \sigma^*_f V_f$$

where σ'_m – stress in the matrix at fiber failure, σ^*_f – fiber tensile strength.

Whereas longitudinal strength is dominated by fiber strength, a variety of factors will have a significant influence on the transverse strength. These include properties of both the fiber and matrix, interface bond strength, and the presence of voids.

(b) Discontinuous and aligned fiber composites:

Even though reinforcement efficiency is lower for discontinuous fiber composites than continuous fiber composites, discontinuous and aligned fiber composites are commercially gaining an important place. The longitudinal strength of these composites is given by:

$$\sigma^*_{cd} = \sigma^*_f V_f \left(1 - \frac{l_c}{2l}\right) + \sigma'_m (1 - V_f) \text{ when } l > l_c \text{ and}$$

$$\sigma^*_{cd'} = \frac{l \tau_c}{d} V_f + \sigma'_m (1 - V_f) \text{ when } l < l_c$$

where τ_c – smaller of either the fiber-matrix bond strength or the matrix shear yield strength.

(c) Discontinuous and randomly orientated fiber composites:

Reinforcement efficiency of these fiber composites is difficult to calculate, and is usually characterized by a parameter known as fiber efficiency parameter, K. K depends on V_f and the E_f/E_m ratio. If rule-of-mixtures can be applied, elastic modulus of these composites is given by:

$$E_{cl} = K(E_m V_m + E_f V_f)$$

12. 3 Structural composites

These are special class of composites, usually consists of both homogeneous and composite materials. Properties of these composites depend not only on the properties of the constituents but also on geometrical design of various structural elements. Two classes of these composites widely used are: laminar composites and sandwich structures.

Laminar composites: there are composed of two-dimensional sheets/layers that have a preferred strength direction. These layers are stacked and cemented together according to the requirement. Materials used in their fabrication include: metal sheets, cotton, paper, woven glass fibers embedded in plastic matrix, etc. Examples: thin coatings, thicker protective coatings, claddings, bimetallics, laminates. Many laminar composites are designed to increase corrosion resistance while retaining low cost, high strength or light weight.

Sandwich structures: these consist of thin layers of a facing material joined to a light weight filler material. Neither the filler material nor the facing material is strong or rigid, but the composite possesses both properties. Example: corrugated cardboard. The faces bear most of the in-plane loading and also any transverse bending stresses. Typical face materials include Al-alloys, fiber-reinforced plastics, titanium, steel and plywood. The core serves two functions – it separates the faces and resists deformations perpendicular to the face plane; provides a certain degree of shear rigidity along planes that are perpendicular to the faces. Typical materials for core are: foamed polymers, synthetic rubbers, inorganic cements, balsa wood. Sandwich structures are found in many applications like roofs, floors, walls of buildings, and in aircraft for wings, fuselage and tailplane skins.

References

1. K. K. Chawla, Composite Materials Science and Engineering, Second Edition, Springer-Verlag, New York, 1998.
2. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
3. D. Hull and T. W. Clyne, An Introduction to Composite Materials, Second Edition, Cambridge University Press, New York, 1996.

Chapter 13. Corrosion and Degradation of Materials

All engineering materials are subjected to numerous external mechanical and environmental factors during their service. Those factors include temperature, chemical attack, mechanical vibration, applied mechanical loads, etc. Under the influence of these factors, engineering materials gets degraded i.e. they loss their potential to perform the task. The degradation is inevitable, and may lead to catastrophes! Hence, degradation of materials not only be monitored but also be factored into the design of different components by engineers.

Mechanism of degradation is different for each engineering material group. Material loss either by dissolution or by formation of nonmetallic scale/film occurs in metals, while ceramics are relatively resistant to degradation. However, ceramics also gets deteriorated at extreme temperatures. Polymers, on the other hand, may dissolve/distort in presence of a liquid solvent or even when exposed to electromagnetic radiation. Degradation due to chemical/electro-chemical factors is termed as corrosion, while oxidation refers to formation of nonmetallic scales in metals.

13.1 Corrosion of Metals

Corrosion can be defined as unintentional deterioration of materials due to electrochemical factors. It usually starts at the external material surface. The corrosion resistance of metals and alloys is a basic property related to the easiness with which these materials react with a given environment. Corrosion is a natural process that seeks to reduce the binding energy in metals, and is of electrochemical in nature. The end result of corrosion involves a metal atom being oxidized, whereby it loses one or more electrons and leaves the bulk metal. The process is called oxidation, some times also called anodic reaction. The lost electrons are conducted through the bulk metal to another site where they are reduced, the process called reduction**.** The site where metal atoms lose electrons is called the anode, and the site where electrons are transferred to the reducing species is called the cathode. As a consequence of these electrochemical reactions, metal ions either go into the corroding solution as ions, or they may form an insoluble compound with nonmetallic elements.

Corrosion tendencies for different metals in measured in terms of electromotive force (EMF). A series of metals are arranged according to their potential to get corroded with respect to hydrogen electrode. Table13.1 presents EMF series, and those at top are noble, while those metals at bottom are more prone to corrosion than the metals above them in the series.

As corrosion is electrochemical in nature, temperature and concentration of chemical does have an important role to play. In general, with increase of either or both temperature and concentration, corrosion rate increases. The electrode potential depends on the concentration of the electrolyte. At 25°C, the following equation, known as Nernst equation, gives the electrode potential:

$$E = E_0 + \frac{0.0592}{n} \log(C_{ion})$$

where E is the electrode potential in a solution containing a concentration C_{ion} (in moles) of metals, n is the valence of the metallic ion, E_0 is the standard electrode potential in a 1-M solution.

Faraday's equation gives the amount (w, weight in gms.) of metal plated on the cathode, or removed from the metal by corrosion:

$$w = \frac{ItM}{nF}$$

where I is the current (A), M is the atomic mass of the metal, n is the valence of the metal ion, t is the time (secs), and F is the Faraday's constant (96,500 C).

Table 13.1: *Standard EMF series of metals.*

	Metal	Electrode potential, E_0 (V)
Increasingly inert ↑	$Au \rightarrow Au^{3+} + 3e^-$	+1.420
	$2H_2O \rightarrow O_2 + 4H^+ + 4e^-$	+1.229
	$Pt \rightarrow Pt^{2+} + 2e^-$	+1.200
	$Ag \rightarrow Ag^+ + e^-$	+0.800
	$Fe^{2+} \rightarrow Fe^{3+} + e^-$	+0.771
	$4(OH)^- \rightarrow O_2 + 2H_2O + 4e^-$	+0.401
	$Cu \rightarrow Cu^{2+} + 2e^-$	+0.340
	$H_2 \rightarrow 2H^+ + 2e^-$	0.000
	$Pb \rightarrow Pb^{2+} + 2e^-$	-0.126
	$Sn \rightarrow Sn^{2+} + 2e^-$	-0.136
	$Ni \rightarrow Ni^{2+} + 2e^-$	-0.250
	$Co \rightarrow Co^{2+} + 2e^-$	-0.277
	$Cd \rightarrow Cd^{2+} + 2e^-$	-0.403
	$Fe \rightarrow Fe^{2+} + 2e^-$	-0.440

Increasingly active ↓	$\mathrm{Cr} \rightarrow \mathrm{Cr}^{3+}+3e^-$	-0.744
	$\mathrm{Zn} \rightarrow \mathrm{Zn}^{2+}+2e^-$	-0.763
	$\mathrm{Al} \rightarrow \mathrm{Al}^{3+}+3e^-$	-1.662
	$\mathrm{Mg} \rightarrow \mathrm{Mg}^{2+}+2e^-$	-2.363
	$\mathrm{Na} \rightarrow \mathrm{Na}^{+}+e^-$	-2.714
	$\mathrm{K} \rightarrow \mathrm{K}^{+}+e^-$	-2.924
	$\mathrm{Li} \rightarrow \mathrm{Li}^{+}+e^-$	-3.050

Galvanic Series

Anodic or Least Noble end
Magnesium
Magnesium Alloys Zinc Aluminum 5052, 3004, 3003, 1100, 6053 Cadmium Aluminum 2117, 2017, 2024 Mild Steel (1018), Wrought Iron Cast Iron, Low Alloy High Strength Steel Chrome Iron (Active) Stainless Steel, 430 Series (Active) 302, 303, 321, 347, 410,416, Stainless Steel (Active) Ni - Resist 316, 317, Stainless Steel (Active) Carpenter 20cb-3 Stainless (Active) Aluminum Bronze (Ca 687) Hastelloy C (Active) Inconel 625 (Active) Titanium (Active) Lead - Tin Solders Lead Tin Inconel 600 (Active) Nickel (Active) 60 Ni-15 Cr (Active) 80 Ni-20 Cr (Active) Hastelloy B (Active) Brasses Copper (Ca102) Manganese Bronze (Ca 675), Tin Bronze (Ca903, 905) Silicone Bronze Nickel Silver Copper - Nickel Alloy 90-10 Copper - Nickel Alloy 80-20 430 Stainless Steel Nickel, Aluminum, Bronze (Ca 630, 632) Monel 400, K500

Silver Solder
Nickel (Passive)
60 Ni- 15 Cr (Passive)
Inconel 600 (Passive)
80 Ni- 20 Cr (Passive)
Chrome Iron (Passive)
302, 303, 304, 321, 347, Stainless Steel (Passive)
316, 317, Stainless Steel (Passive)
Carpenter 20 Cb-3 Stainless (Passive), Incoloy 825nickel - Molybdeum - Chromium - Iron Alloy (Passive)
Silver
Titanium (Pass.) Hastelloy C & C276 (Passive), Inconel 625(Pass.)
Graphite
Zirconium
Gold
Platinum
Cathode Or Most Noble end

Passivity: Under particular environments, some active metals and alloys may turn inert, known as passivity. It is usually observed in metals such as Cr, Fe, Ni, Ti and their alloys. Passivity is inherited as a result of formation of a thin and highly adherent oxide film on the metal surface. For example: stainless steels are highly resistant to corrosion in many environments, except in hard sea water. For steels, Cr forms a protective layer. Aluminium is also observed to passive in many environments. This is due to formation thin oxide layer.

Corrosion in metals is classified based on the manner in which it is manifest into eight forms, namely: uniform, galvanic, crevice, pitting, inter-granular, selective leaching, erosion-corrosion, and stress corrosion.

Uniform corrosion: As the name suggests, electrochemical corrosion occurs over entire exposed surface with equal intensity. It usually results in formation of a scale/film. This form of corrosion is most common, easy to monitor, and less damaging than the other forms. Painting the surface is best counter measure for it.

Galvanic corrosion: This occurs when two metals with different compositions i.e. EMF are electrically connected in presence of an electrolyte. Out-of-these two metals, less inert metal will corrode, while the other will be protected from the corrosion. Corrosion rate depends on the relative anode-to-cathode surface areas that are exposed to the electrolyte. the following measures are observed to reduce the damage due to this form of corrosion: when coupling of two different metals is unavoidable, choose that are close together in EMF series; tailor the anode-to-cathode area ratio; metals may be electrically insulated form each other; or use a third metal to protect the both, known as cathodic protection.

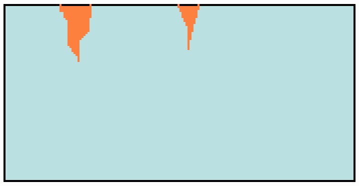
Crevice corrosion: This occurs as a result of concentration difference between two regions of same metal component. Corrosion takes place in the locale that is of lower concentration. Probable sites for this corrosion are: crevices and recesses, under scales of dirt/corrosion product. This form of corrosion may be avoided if enough care is taken to avoid concentration differences. E.g.: welding instead of riveting. Other measures include removing accumulated deposits; avoiding stagnant areas.

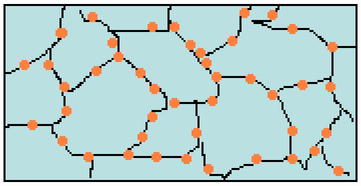
Pitting: It is another form of localized corrosion. As a consequence of this, small pits, and holes form, hence the name. This occurs in normal direction to the surface exposed, and so is difficult to monitor. Thus, it is extremely dangerous form of corrosion, as material loss is often undetectable till the component fails. Same measures as for crevice corrosion are part of care against this form of corrosion. In addition, polishing of surfaces could do good.

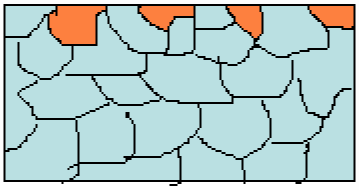

Uniform Corrosion

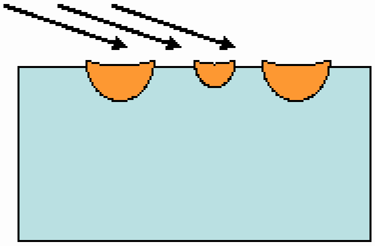

Galvanic corrosion

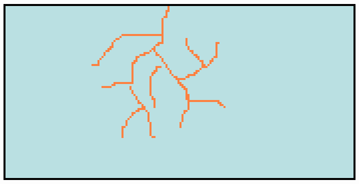

Crevice corrosion


Pitting Corrosion


Inter-granular corrosion


Selective leaching


Erosion-corrosion


Stress corrosion

Inter-granular corrosion: This occurs again due to concentration difference, but now along grain boundaries, which are usually of different chemical composition compared with rest of the grains. Thus grain boundaries are highly susceptible to corrosion. This form of corrosion is very severe in stainless steels. These materials may be protected by suitable heat treatment; lowering carbon content; addition of alloying elements those readily forms of carbides.

Selective leaching: It is the preferential removal of a solid alloy by corrosion process. Best example can be dezincification of brass. It also occurs as the loss of Ni, Sn and Cr from Cu-alloys; Fe from cast iron; Ni from steel alloys, and Co from satellite. As a result of this corrosion, mechanical properties of the components are impaired as the component becomes porous. Measures of protection include change if composition; change of environment; use of cathodic protection.

Erosion-corrosion: It is defined as acceleration of corrosion attack in a metal due to relative motion of a corrosive fluid and a metal surface i.e. is result of combined action of chemical attack and mechanical abrasion/wear. Practically, all metals and alloys are prone to this corrosion form. It has more damaging influence on metals which passivate by forming a protective film and also on soft metals like Cu and Pb. Counter measures include: change of design; choosing different metal for the component; removal of particulates and bubbles form the fluid to lessen the erosive effects.

Stress corrosion: This form of corrosion, also known as stress corrosion cracking (SCC), occurs as a result of combined action of applied tensile stresses and corrosive environment. During SCC, the metal's surface is usually attacked very little while highly localized cracks propagate through the metal section. The stresses which cause SCC can be residual or applied. Only certain combinations of alloys and environments cause SCC. E.g.: Stainless steels crack in chloride environments but not in ammonia-containing ones. Best counter measure is elimination of external stress. Other measures are: to eliminate detrimental environment; change of the alloy if neither environment nor stress level can be changed; cathodic protection: addition of inhibitors.

Hydrogen embrittlement is more failure than a form of corrosion, but it is often results from the hydrogen, produced from corrosion. Atomic hydrogen produced during corrosion diffuses interstitially through crystal lattice, and interferes with dislocation motion, leading to failure. It is similar to stress corrosion in the sense that ductile materials experience brittle failures as a result. Counter measures to hydrogen embrittlement include: heat treatment to reduce strength of the alloy; removal of source of hydrogen; baking the component to drive out any dissolved hydrogen.

13.2 Corrosion of Ceramics

Ceramics, being compounds between metallic and nonmetallic elements, may be thought of as having already been corroded. It is often said that one of the biggest advantages which ceramics have over other engineering materials is their corrosion resistance, that is, their chemical inertness in corrosive environments.

Corrosion is generally understood as property degradation due to environmental attack. Ceramics corrode by simple chemical dissolution, in contrast to the electrochemical processes involved with corrosion of metals. There are a number of environments in which ceramics can degrade at a rapid rate. There exists a tremendous need for reliable and corrosion resistant structural ceramic or partly ceramic materials which can be used in aggressive environments such as:

- high energy battery systems (such as sodium-sulphur): beta-alumina is being investigated

- gas turbines: silicon nitride and/or carbide are being investigated

- heat exchangers: SiC, composites are being investigated

Ceramics are indeed much more environmentally stable, as compared to any other group of engineering materials, e.g. metals or plastics. Still, the potential for ceramics as corrosion resistant engineering structural materials are far from being fully realized, because of:

- mechanical non-reliability of structural ceramic components

- difficult design with brittle materials

- a shortage of information and standardization of ceramics

- human reluctance to use non-ductile materials

13.3 Degradation of polymers

While the plastics industry searches for solutions to the problem of plastics waste, there is, surprisingly, a growing band of people trying to save plastics. The crucial fact is that plastics are organic and can are prone to physiochemical attacks. Polymeric materials degrade by non-corrosive processes.

As all other engineering materials, plastics do indeed degrade during their service. However, degradation of plastics is not termed as corrosion, as it is of physiochemical in nature in contrast to electrochemical corrosion of metals. Degradation of polymers, thus, involves a wide variety of reactions and results like absorption and swelling, dissolution, bond rupture due to heat, chemical effects, or radiation; weathering, etc. Degradation of polymers might be due to exposure to light (especially UV), humidity, oxygen, heat, bacteria or external loads/stress. However, due to complex nature of chemical bonds in polymers, their degradation mechanisms are not well understood.

Swelling and dissolution: Polymers, when exposed to liquids, they get swelling as a result of solute diffusion and absorption of solute. Because of swelling i.e. separation of chains, secondary bonds become weaker. As a consequence, the material becomes softer and more ductile. Swelling is considered as partial dissolution, while the dissolution or complete solubility is worse than the swelling. Swelling and dissolution effects are influenced by temperature because of their physiochemical nature.

Bond rupture: Bond rupture in polymers due to degradation is known as scission. Polymers bonds may get rupture due to many effects like radiation, heat energy, or chemical reactions. When polymers are exposed to certain types of radiation, which may result in broken bonds and rearrangement of atoms leads to degradation of polymers. At elevated temperatures, bonds in polymers may get weakened, leading to deterioration of polymers. Some chemical elements like oxygen, ozone can alter the chain scission rate as a result of chemical reactions. This is especially pronounced in vulcanized rubbers.

Weathering: Polymers are exposed to outdoor conditions in many instants of their service. Any degradation of polymer under these circumstances is termed as weathering, which in fact is a combination of several different processes. It primarily involves oxidation, which may be initiated by ultraviolet radiation from the Sun. polymer's resistance to weathering to very difficult to quantify and is very diverse.

Some examples of polymer degradation include:

- deterioration of acrylic paintings and pieces of art

- decomposition of photographic films

- decolorization of plastics pieces preserved in museums

Solutions to polymer degradation: As the recognition of polymer degradation improves, conservation guidelines are beginning to emerge. High-tech solutions which could help in theory are prohibitively expensive, but tailor made scavengers such as activated charcoal or Ageless help to create a low oxygen environment. Ageless is a reactive powdered iron and is normally used to prolong the shelf-life of dry foods by absorbing oxygen. Epoxidised soya bean oil (ESBO), has also been tested with encouraging results as an acid absorbing coating on degrading cellulose nitrate.

References

1. M. G. Fontana, Corrosion Engineering, Third Edition, McGraw-Hill, New York, 1986.
2. ASM Handbook, Corrosion, Vol. 13, ASM Internation, Materials Park, OH, 1987.
3. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.

Chapter 14. Electrical properties

Engineering materials are important in everyday life because of their versatile structural properties. Other than these properties, they do play an important role because of their physical properties. Prime physical properties of materials include: electrical properties; thermal properties; magnetic properties; and optical properties. The electrical behaviors of engineering materials are diverse, and so are their uses in electrical applications. Few examples of electrical applications: copper and aluminium wires used in power transmission; amorphous silicon in solar cells; LEDs; gallium-arsenide chips used in modern day computers, cell phones, digital assistants; etc.

This chapter shall describe electrical properties and behavior of engineering materials in electrical applications. First, basic electrical property, electrical conductivity, will be described followed by different types of electrical conduction including semi- and super-conductivities. Some peculiar phenomena like ferro-electricity and peizo-electricity are explained in addition to electrical insulation.

14.1 Electrical conduction

Electrical conductivity of a material is defined in terms of ease with which a material transmits an electrical current. Electrical current (I) is flow of electrons, and driving force for the flow of electrons is called voltage (V). Ohm's law relates these parameters as follows:

$V \alpha I$

$V = IR$……………………………….………..(1)

where R – is the materials resistance to flow of electrons through it.

V, I, and R respectively have units as volts, amperes, and ohms (Ω).

Electrical resistance of a material is influenced by its geometric configuration; hence a new parameter called *electrical resistivity* (ρ) is defined such as it is independent of the geometry.

$$\rho = \frac{RA}{l} \text{ ………………………………..……….(2)}$$

where A – cross-sectional area perpendicular to the direction of the current, and l – the distance between points between which the voltage is applied. Units for ρ are ohm-meters (Ω-m).

Reciprocal of the electrical resistivity, known as *electrical conductivity* (***σ***), is used to express the electrical behavior of a material, which is indicative of the ease with which a material allows of flow of electrons.

$$\sigma = \frac{1}{\rho} = \frac{l}{RA} \text{ …………………………..……..(3)}$$

Electrical conductivity has the following units: $(\Omega\text{-m})^{-1}$ or mho/meter.

Resistivity or *conductivity* does not depend on the dimensions of the material. These properties are microstructure-sensitive instead, like many other intrinsic properties, for example yield strength. Thus either of it (usually *conductivity*) allows us to compare different materials. Solid engineering materials exhibit very wide range of electrical conductivity (about 27 orders of variation). Hence the materials for electrical applications are classified according to their electrical conductivity as: *conductors*, *semiconductors* and *insulators/dielectrics*. Metals are conductors having conductivities in range of 10^7 $(\Omega\text{-m})^{-1}$, while semiconductors have conductivities in range from 10^{-6} to 10^4 $(\Omega\text{-m})^{-1}$, and materials with conductivity lower than $10^{-10}(\Omega\text{-m})^{-1}$ are termed as insulators.

In components designed to conduct electrical energy, minimizing losses is important to conserve energy as well as minimize heating. The electric power (P) lost when a current flows through a material is given by:

$$P = VI = I^2R \text{……………………………………..(4)}$$

Electrical losses are also known as Joule heating losses.

Second form of Ohm’s law can be obtained by combining the equations (3) and (4) to give:

$$\frac{I}{A} = \sigma \frac{V}{l} \Rightarrow J = \sigma E \text{ …………………………(5)}$$

where J – current density (A/cm^2), E – electric field strength (V/cm).

Current density can be defined in different terms as

$J = n.q.\bar{v}$ ……………………………….……….(6)

where n – number of charge carriers ($1/cm^3$), q – charge on each carrier ($1.6x10^{-19}$C), $\bar{v}$ - average drift velocity (cm/s), which occurs as a result of an applied electric or magnetic field.

The term '$\bar{v}$/E' is called mobility, μ (cm^2/V.s) of the carriers, thus

$\sigma = n.q.\mu$ ……………………………………(7)

From the above equation, the following can be concluded: electrical conductivity of a material can be controlled by (i) controlling number of charge carriers, n or (ii) controlling the mobility of the carriers, μ. Mobility is important for metals or electrical conductors, whereas number of carriers is important for semi-conductors and insulators.

Electrons are charge carriers in metals. In ionic materials (for example ionic ceramics), conduction is result of net movement of charged particles (cations and anions) in addition to any electron motion. However, both electrons and holes are charge carriers in semiconductors and in doped conductive polymers. Thus equation 7 can be written as follows:

for metals : $\sigma = n_e.q_e.\mu_e$ …………………………(8)

for semiconductors : $\sigma = n.q.\mu_n + n.q.\mu_p$ …………………..(9)

for ionic materials : $\sigma_{total} = \sigma_{electronic} + \sigma_{ionic}$ ……………(10)

In all conductors, semiconductors, and many insulating materials, only electronic conduction exists i.e. electrons are the only charge carriers. Ionic conduction results because of net motion of charged ions. Movement different particles in various materials depend on more than one parameter. These include: atomic bonding, imperfections, microstructure, ionic compounds, diffusion rates and temperature. Because of these, electrical conductivity of materials varies tremendously.

In metals not all electrons will accelerate in the presence of an electric field. Instead the number of charge carriers is dependent on the arrangement of electron states/levels with respect to energy, and the manner in which these states are occupied by electrons. Electrons in metals are arranged into shells and sub-shells in accordance with Pauli exclusion principle. Each of these shells / sub-shells is distinguished from every other by their energy state/level. Basic existence of metals is dependent on metallic bond where large groups of atoms come together and share valence electrons in what is called *metallic bond*. Thus there are N atoms in the lattice with N atomic orbitals and N

molecular orbitals, many of which are of same energy. This leads to bands of electrons. The electrons not involved in bonding remain in what is called the core band, whereas the valence electrons that form the electron gas enter into the valence band. The remaining un-filled orbitals form higher-energy bands, called conduction band. It is the conduction band that gives metals and alloys the ability to freely conduct electrons.

Four different types of band structures are possible at 0 K: (1) Where one outermost band is only partially filled with electrons. The energy corresponding to the highest filled state at 0 K is called the Fermi energy, E_f. (2) There is overlap of an empty band and a filled band. (3) and (4) One filled band (valence band) and an empty band (conduction band) are separated by an energy band gap which lies between them. These are schematically shown in the *figure 14-1*.

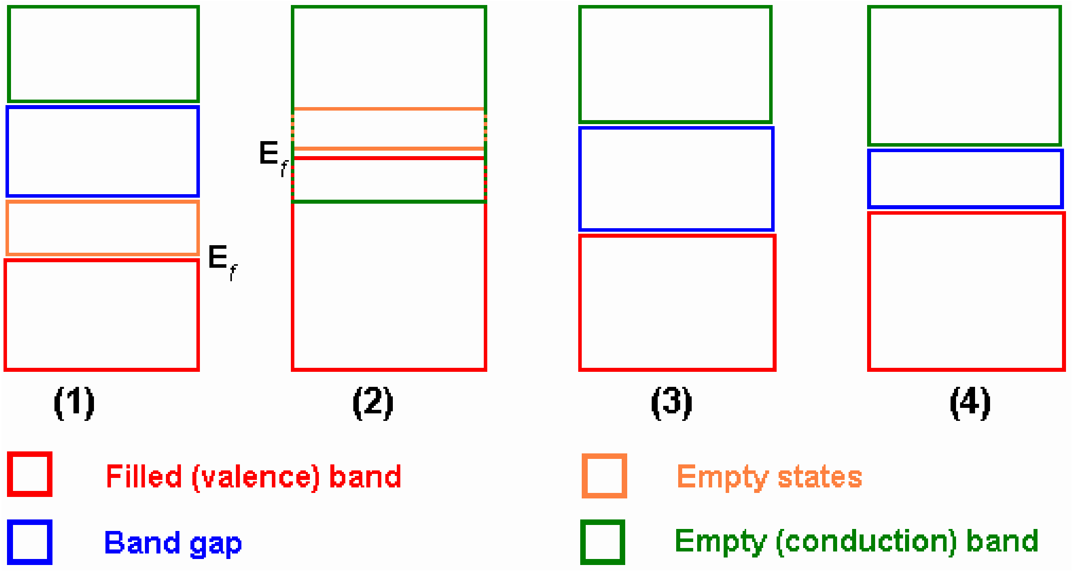


Figure 14-1: Possible electron band structures.

Band structures (1) and (2) are found in metals / conductors. Band structures (3) and (4) are distinguished by the size of energy band gap. Narrower energy band gap i.e. size < 2 eV, is found in semiconductors, while the broader energy band gap i.e. size > 4 eV, is found in insulators. Fermi energy for these two band structures lies within the band gap near its center.

Metals have high conductivities because of the large number of free electrons that have been excited into empty states above the Fermi energy. If the metal crystal lattices were perfect and there are no lattice vibrations, the electrons would pass through the lattice un-scattered, encountering no resistance. However, lattice vibrations and phonon scattering play a role in disrupting the mean free path of electrons. In addition, crystalline defects and impurity atom affect the conductivity. These scattering mechanisms act independently on one another. Thus the effective resistivity of metals can be represented as follows:

$\rho = \rho_{thermal} + \rho_{residual}$ ………………….………..(11)

Equation 11 is some times known as Mathiessen's rule. With increase of temperature, thermal vibrations increase so the resistivity, and vice versa. In the same manner, with increase of either defects or impurities, resistivity increases. Thus, in general, for pure metals, the resistivity approaches zero at absolute zero temperature.

Some metals, however, such as Pb, lose all resistivity abruptly and completely at some low temperatures. This phenomenon is called *superconductivity*, and the materials that exhibit it are called *superconductors*. The temperature at which the resistivity vanishes is called the *critical transition temperature*, T_c. Many elemental metals, solid-solution alloys, some ceramics and intermetallic compounds exhibit superconductivity. For example: Ti, V, Zn, W, Al, Hg, NbTi, Nb_3Sn, MgB_2, La-Sr-Cu oxide, $YBa_2Cu_3O_{7-x}$, carbon nanotubes, etc. BCS (Bardeen, Cooper and Schrieffer) theory explains that the super conductivity is caused by electron-lattice interaction and that the superconducting electrons consist of paired ordinary electrons called a *Cooper pair*.

14.2 Semiconductivity

Conductivity of semiconductors is not as high as metals / conductors. However, their electrical properties are unique, in the sense that their electrical properties are extremely sensitive to even minute concentrations of impurities. For intrinsic semiconductors, their electrical behavior is based on inherent electronic structure of the pure material. On the other hand, if the electrical properties are dominated by impurities, they are called extrinsic semiconductors.

Intrinsic semiconduction: Energy band gap between the valence and conduction bands can be used to classify materials as conductors, insulators, and semiconductors. The magnitude of the band gap is characteristic of the lattice alone, and varies widely for different crystals. In semiconductors, the valence and conduction bands do not overlap as in metals, but they possess enough electrons in the valence band those can be promoted to the conduction band at a certain temperature. Promotion of electrons leaves behind positively charged holes in the valence band which maintain the charge neutrality. These holes can be regarded as moving positive charges through the crystal with a charge magnitude and ass as that of an electron. Conduction that arises from thermally / optically excited electrons is called intrinsic semiconduction. It usually takes place at elevated temperatures since sufficient thermal agitation is necessary to transfer a reasonable number of electrons from the valence band to the conduction band. At still higher temperatures, the concentration of thermally excited electrons in the conduction band becomes so high that the semiconductor behaves more like a metal. The elements capable of intrinsic conduction are limited, and include: Si, Ge, Sn, Pb, etc.

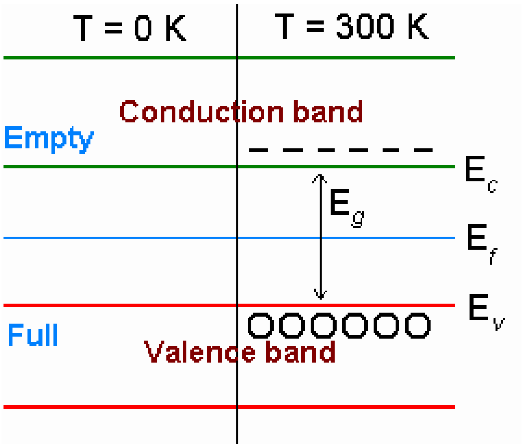


Figure 14-2: Band structure and intrinsic conductivity.

In intrinsic semiconductors, the number of holes equals the number of mobile electrons. The resulting conductivity can be expressed as

$$\sigma = n_e.q_e.\mu_e + n_h.q_h.\mu_h = n_e.q_e.(\mu_e + \mu_h) \ldots..(12)$$

as $n_e = n_h$, and both electron and hole possess equal magnitude of charge.

The mobility of holes is always less than those of electrons i.e. $\mu_h < \mu_e$. In Si and Ge, the ratio μ_e / μ_h is approximately three and two respectively.

Extrinsic semiconduction: The charge carrier density can also be increased by adding impurities of either higher or lower valence. This addition of impurities is known as doping, and impure atoms in the element are called donor atoms.

When a pentavalent substitutional atoms like P, As, or Sb are placed into a covalently bonded tetravalent materials such as Si or Ge, only four of their five valence electrons are required to participate in covalent bonding, while the fifth electron remains weakly bound to the donor atom. However, this binding energy is in order of 0.01 eV, much less than that of a covalently bonded electron. The energy state of this electron is indicated as E_d. Energy gap between the donor and conducting bands is much smaller than that between the valence band and conduction band, E_g. Thermal agitation, even at room temperature, is good enough to excite this electron to the conduction band, leaving behind a positively charged hole in the donor band. Conductivity is due to motion of negatively charged electrons in the conduction band, and so the material is called *n-type semiconductor*. Fermi level (E_f) of *n*-type extrinsic semiconductor is higher than that of intrinsic semiconductor i.e. it shifted upward in the band gap, to within the vicinity of the donor state; and its exact position is a function of both temperature and donor concentration.

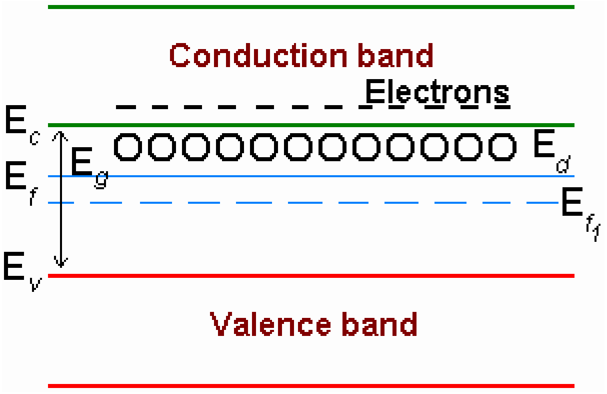


Figure 14-3: Band structure and *n*-type semiconductivity.

Other possibility of susbstituion involves impurity with lower valence. For example trivalent elements like B, Al, Ga, In, Th doped to Si or Ge. Since the donor atoms are deficient in bonding electrons, one of their bonding orbitals will contain a hole that is capable of accepting an electron from elsewhere in the crystal. Binding energy is small, and promotion of an electron from the valence band to the acceptor band, E_a, leaves a positively charged hole in the valence band that can act as a charge carrier. Materials with this *p*-type extrinsic semiconduction are called *p-type semiconductors*. Fermi level (E_f) for *p*-type semiconductors is positioned within the band gap and near to the acceptor level.

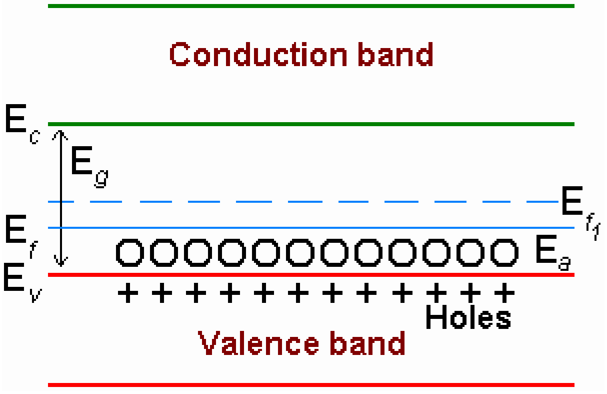


Figure 14-4: Band structure and *p*-type semiconductivity.

For an *n*-type semiconductor, electrons in the conduction band are charge carriers and holes participate in bonding, whereas for a *p*-type semiconductor, the electrons participate in bonding, and the holes are the charge carriers. Conductivity equations for these extrinsic semiconductors are written as:

n-type semiconductor: $\sigma \cong n.q.\mu_e$ …….(13)

p-type semiconductor: $\sigma \cong n.q.\mu_h$ ……..(14)

In both types of extrinsic semiconductors, the doping elements are chosen so that both acceptor and donor levels are located closer to the corresponding energy bands and only a small energy gap in involved when exciting electrons. Intrinsic semiconductors are pure to the order greater than 10^7, while extrinsic semiconductors are made with impurity levels in order of 1-1000 *ppm*.

In intrinsic semiconductors, the conductivity is dominated by the exponential relation between temperature and band-gap. However, the conductivity of extrinsic semiconductors is governed by competing forces: charge carrier density and charge carrier mobility. At low temperatures, the number of charge carriers initially increases with temperature. At intermediate temperatures, for *n*-type extrinsic semiconductor, most of the donor electrons have been promoted and the number of charge carriers is nearly independent of temperature in what is known as the exhaustion range. At higher temperatures, the number of valence electrons excited to the conduction band greatly exceeds the total number of electrons from the donor atoms, and extrinsic semiconductor behaves like an intrinsic semiconductor. That is, at low temperatures the conductivity varies with temperature as the charge carrier concentration, at higher temperatures the charge carrier mobility dominates and the conductivity decreases with temperature, and at still higher temperatures the extrinsic semiconductor behaves like an intrinsic semiconductor.

14.3 Electrical conduction in ionic ceramics and in polymers

In addition to the conduction of charge via electrons, charge can be conducted via ions. Ions are present in most crystalline ceramic materials such as oxides and halides. This process is called ionic conduction, and may occur either in conjunction with or separately from electronic conduction.

Several types of compounds show exceptionally high ionic conductivity. Such phases fall into three broad categories: halide and chalcogenides of silver and copper; oxides with β-alumina structure; and oxides of fluorite structure. Conductivity of halides and chalcogenides is result of larger number of available sites than the number of cations. In oxides with β-alumina structure ($AM_{11}O_{17}$) monovalent cation A (such as Na, K, Rb, Li) is mobile. High dopant levels in the fluorite-type solid solutions leads to large defect concentrations and vacancy ordering which results in rapid oxygen migrations. This is the base for development of high-temperature superconductors. For example: La_2CuO_4 (T_c = 30 K), YBC compounds – yttrium doped perovskite structure, $YBa_2Cu_3O_7$ (T_c = 92 K).

Although ceramics are insulators, by properly engineering the point defects in ceramics, it is possible to convert them into semiconductors. For example: Indium tin oxide (ITO) is used as a transparent conductive coating on plate glass, and also for tough screen displays; Ytrria-stabilized zirconia (YSZ) is used as solid electrolyte in solid oxide fuel cells; Lithium cobalt oxide is used as solid electrolyte in lithium ion batteries.

Polymers have a band structure with a large energy gap leading to low conductivity because their valence electrons are involved in covalent boding. Hence polymers are used frequently in applications that require electrical insulation. In some cases however, low conductivity is undesirable. For example; static electricity can accumulate on housing for electronic equipment making the polymer vulnerable to electromagnetic radiation that damages the internal devices. These problems can be solved in two approaches: (1) introducing an additive to the polymer to improve conductivity, and (2) creating polymers with inherent conductivity.

Adding ionic compound to the polymer can reduce resistivity because the ions migrate to the polymer surface and attract moisture, which in turn, dissipates static charges. Introducing conductive fillers such as carbon black can also dissipate static charges. Some polymers inherently have good conductivity as result of doping or processing techniques. When acetal polymers are doped with agents such as arsenic pentafluoride, electrons or holes are able to jump freely from one atom to another along the backbone of the chain, increasing the conductivity to near that of metals. However, unlike semiconductors, the dopant atoms or molecules do not substitute for or replace the polymer atoms. Other polymers include polyparaphynylene, polypyrole, and polyaniline. Some other polymers such as polyphthaocyanine can be cross-linked by special curing processes to raise its conductivity.

14.4 Dielectric behavior

A dielectric is a material separating two charged bodies. For a material to be a good dielectric, it must be an electrical insulator. Thus, any insulator is also called a dielectric. Dielectric materials are used in capacitors, devices used to store the electric energy.

When a voltage is applied between two parallel plates in a vacuum, one plate becomes positively charged and the other negatively charged, with corresponding field directed from the positive to the negative. Then the energy stored, Q, is given by:

$$Q = CV \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (15)$$

where C is the capacitance (units: coulombs per volt *or* farads [F]). It is dependent on the geometry of plates as follows:

$$C = \varepsilon_0 \frac{A}{l} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (16)$$

where A – area of the plates, l – distance between them, and ε_0 – permittivity of vacuum (8.85×10^{-12} F/m). If a dielectric material is inserted between the plates,

$$C = \varepsilon \frac{A}{l} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (17)$$

ε – permittivity of the medium, which will be greater in magnitude than ε_0. The *relative permittivity* ε_r, also called *dielectric constant*, is defined as

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

which is greater than unity, and presents the increase in charge storing capacity by insertion of the dielectric medium between the plates. This is result of polarization – orientation of permanent or induced dipoles under externally applied electric field. Polarization causes positive charge to accumulate on the bottom surface next to the negatively charged plate, and causes negative charge to accumulate toward the positively charged plate on the top. This tends to decrease the effective surface charge density on either plate. The expected decrease in effective charge corresponds to the polarization, of the material which equals the induced dipole moment per unit volume of polarizable material. The magnitude of the charge per unit area on either plate, called the dielectric displacement *or* flux density *or* surface charge density, *D*, is defined as:

$$D = \varepsilon\xi \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (19)$$

where ξ – electric field strength,

$$\xi = V / l \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (20)$$

In the presence of the dielectric, surface charge density can also be presented as

$$D = \varepsilon\xi = \varepsilon_0\xi + P \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

where polarization is defined as,

$$P = \varepsilon_0(\varepsilon_r - 1)\xi \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (22)$$

the increase in charge density above that for a vacuum because of the presence of the dielectric.

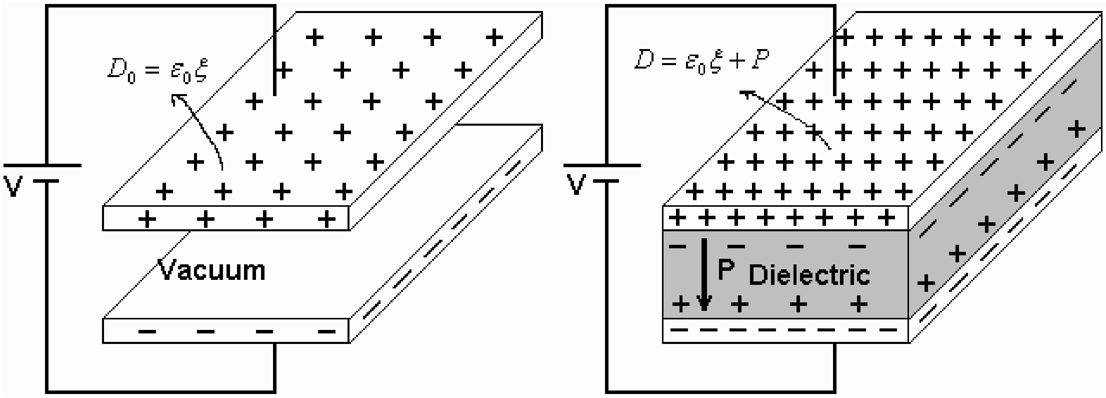


Figure 14-5: Schematic representation of capacitor and polarization phenomenon.

The polarization may also be thought of as the total dipole moment per unit volume of the dielectric material or as a polarization electric field within the dielectric that results from the mutual alignment of the many a atomic or molecular dipoles. Polarization of a dielectric material may be caused by at least one of four major types of polarization: electronic, ionic (atomic), orientation, and space charge (interfacial).

Electronic polarization arises because the center of the electron cloud around a nucleus is displaced under an applied electric field. Ionic polarization occurs in ionic materials because an applied field acts to displace cations in the direction of the applied field while displacing anions in a direction opposite to the applied field. Orientation polarization can occur in materials that possess permanent electric dipole moments. These permanent dipoles tend to become aligned with the applied electric field, but entropy and thermal effects tend to counter this alignment, thus it is highly temperature dependent. Space charge polarization results from the accumulation of charge at structural interfaces in heterogeneous materials. Such polarization occurs when one of the phases has a much higher resistivity than the other.

The total polarization subjected to an alternating electric field depends upon the ease with which the permanent or induced dipoles can reverse their alignment. The time required for dipole reversal is called the *relaxation time*, and its reciprocal is called the *relexation frequency*. As the frequency of the applied electric field approached the relaxation frequency, the polarization response increasingly lags behind the applied field. The reorientation of each type of dipole is opposed by internal friction, which leads to heating in the sample and power loss. The power loss depends on the degree to which the polarization lags behind the electric field. This is also called *dielectric loss*, and a low dielectric loss is desired at the frequency of utilization.

At high enough frequencies, the dielectric will experience electrical breakdown initiated by the field-induced excitation of a number of electrons into the conduction band, and the insulator become a conductor. The magnitude of the electric field required to cause dielectric breakdown is called the *dielectric strength* or *breakdown strength*.

Many ceramics and polymers are utilized as insulators and in capacitors. For example: glass, porcelain, stealite, mica. These have dielectric constants within the range of 6-10. Typical applications: electrical insulation, switch bases, light receptacles. Dielectric constant for most polymers lies in the range of 2-5, less than that for ceramics, since the later exhibit greater dipole moments. Typical applications: insulation for wires, cables, motors, generators, some capacitors.

14.5 Ferro-electricity and Piezoelectricity

Two peculiar phenomena related to electric dipoles are *ferro-electricity* and *piezo-electricity*. Ferro-electricity is defined as the spontaneous alignment of electric dipoles by their mutual interaction in the absence of an applied electric field. This arises from the fact that the local field increases in proportion to the polarization. Thus, ferro-electric materials must possess permanent dipoles. Ex.: $BaTiO_3$, Rochelle salt

($NaKC_4H_4O_6.4H_2O$), potassium dihydrogen phosphate (KH_2PO_4), potassium niobate ($KNbO_3$). These materials have extremely high dielectric constants at relatively low applied field frequencies. Thus, capacitors made from ferro-electric materials are smaller than capacitors made of other dielectric materials.

Piezo-electricity, or pressure electricity, is defined as polarization induced by the application of external force. Hence, by reversing the direction of external force, direction of the field can be reversed i.e. the application of an external electric field alters he net dipole length and causes a dimensional change. This property is characteristic of materials having complicated structures with a low degree of symmetry. Thus, piezo-electric materials can be used as transducers – devices that convert mechanical stress into electrical energy and vice versa. Application for these materials includes microphones, ultrasonic generators, sonar detectors, and mechanical strain gauges. Ex.: Barium titanate, lead titanate, lead zirconate ($PbZrO_3$), ammoinium dihydrogen phosphate ($NH_4H_2PO_4$), and quartz.

References

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, Introduction to Ceramics, Second Edition, Wiley, New York, 1976.
2. R. E. Hummel, Electronic Properties of Materials, Third Edition, Springer-Verlag, New York, 2000.
3. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.

Chapter 15. Thermal properties

Engineering materials are important in everyday life because of their versatile structural properties. Other than these properties, they do play an important role because of their physical properties. Prime physical properties of materials include: electrical properties; thermal properties; magnetic properties; and optical properties. Selection of materials for use at elevated temperatures and/or temperature changes require an engineer to know and understand their thermal properties. This chapter deals with the thermal properties of materials. Physical property of a solid body related to application of heat energy is defined as a thermal property.

This chapter shall describe thermal properties like heat capacity, thermal expansion, thermal conductivity, and thermal stresses. It is important to know and understand the concept of thermal expansion which is the root cause for thermal stresses. Thermal stresses are stresses leading to failure of engineering structures at elevated temperatures.

15.1 Heat capacity

Many engineering solids when exposed to heat experiences an increase in temperature i.e. it absorbs heat energy. This property of a material i.e. material's ability to absorb heat energy is called its *heat capacity*, C. It is defined as the energy required to change a material's temperature by one degree. Mathematically, it is expressed as:

$$C = \frac{dQ}{dT}, \qquad \text{J/mol-K or Cal/mol-K}$$

where dQ is the energy required to produce a temperature change equal to dT.

Heat capacity is not an intrinsic property i.e. total heat a material can absorb depends on its volume / mass. Hence another parameter called *specific heat*, c, it defined as heat capacity per unit mass (J/kg-K, Cal/kg-K).

With increase of internal energy, geometrical changes may occur. Accordingly, heat capacity is measured either at constant volume, C_v, or at constant external pressure, C_p. The magnitude of C_p is always greater than C_v but only marginally.

Heat energy absorption of a (solid, liquid or gaseous) material exists in mode of thermal energy vibrations of constituent atoms or molecules, apart from other mechanisms of heat absorption such as electronic contribution. With increase of energy, atoms vibrate at higher frequencies. However, the vibrations of adjacent atoms are coupled through atomic bonding, which may lead to movement of lattices. This may be represented as elastic waves (phonon) *or* sound waves. Vibrational contribution of heat capacity of solids varies with temperature according to the following relation at low temperatures:

$$C_v = AT^3$$

C_v is equal to zero at 0 K, but increase rapidly with temperature. It represents increased ability of atomic vibrations or enhanced energy of lattice waves with ascending temperature. Above a temperature called Debye temperature, θ_d, dependence of volumetric heat capacity value reaches saturation. This saturation values if approximately equal to 3R (≈ 6 cal/mol-K), R being the universal gas constant. For many solids, value of θ_d is below room temperature.

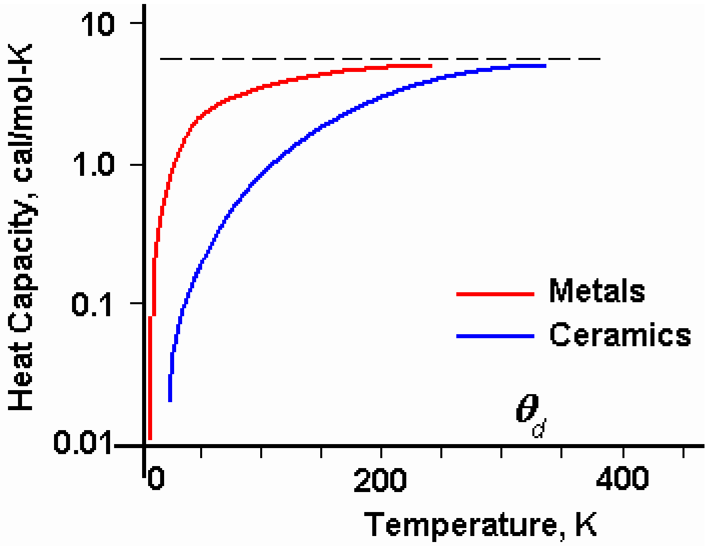


Figure 15-1: Heat capacity as a function of temperature.

15.2 Thermal expansion

After heat absorption, vibrating atoms behaves as though they have larger atomic radius, which leads to increase in materials dimensions. The phenomenon is called thermal expansion. It is quantified in terms of thermal expansion coefficient. Linear coefficient of thermal expansion (α) defined as the change in the dimensions of the material per unit length, and is expressed as:

$$\alpha = \frac{l_f - l_0}{l_0(T_f - T_0)} = \frac{\Delta l}{l_0 \Delta T} = \frac{\varepsilon}{\Delta T}$$

where T_0 and T_f are the initial and final temperatures (in K), l_0 and l_f are the initial and final dimensions of the material and ε is the strain. α has units as $(^\circ C)^{-1}$. For most metals α is the range of 5-25x10^{-6}, for ceramics 0.5-15x10^{-6}, and for polymers 50-400x10^{-6}.

A volume coefficient of thermal expansion, α_v (=3α) is used to describe the volume change with temperature.

$$\alpha_v = \frac{\Delta v}{v_0 \Delta T}$$

An instrument known as dilatometer is used to measure the thermal expansion coefficient. It is also possible to trace thermal expansion using XRD.

At microscopic level, thermal expansion can be attributed to the increase in the average distance between the atoms. Thus, the coefficient of thermal expansion of a material is related to the strength of the atomic bonds. The relation between inter-atomic distance and potential energy is shown in the *figure 15.2*. As shown in the figure, the potential energy curve has trough and is characterized by a minimum. This minimum corresponds to the equilibrium inter-atomic distance ay 0 K. With increase in temperature of the material, vibrational amplitude increase, and the mean of it represents the average inter-atomic distance. Because of asymetric nature of the potential energy trough, rather than the increase in atomic vibration amplitude, the average inter-atomic distance increases with the temperature of a material.

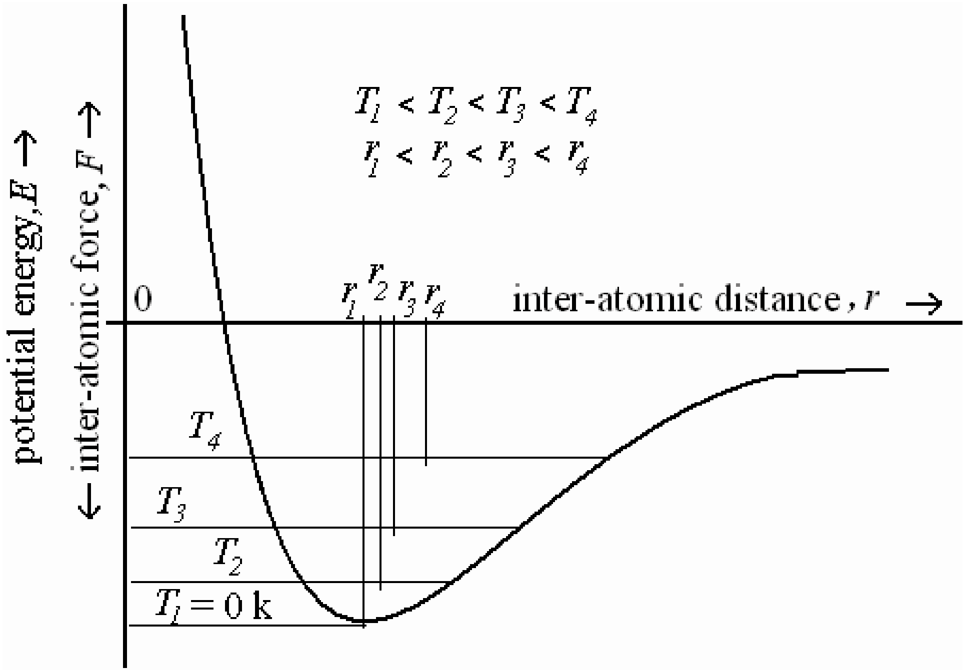


Figure 15.2: *Change if inter-atomic distance with temperature.*

If a very deep energy trough caused by strong atomic bonding is characteristic of the material, the atoms separate to a lesser and the material has low linear coefficient of thermal expansion. This relationship also suggests that materials having a high melting temperature – also due to strong atomic bonds – have low thermal expansion coefficients. Most ceramics having strong atomic bonds, thus, have low thermal expansion coefficients compared with metals and polymers. However, as a consequence of non-uniform dimensional changes, brittle materials like ceramics may experience fracture, known as *thermal shock*. The capacity of material to withstand this failure is known as thermal shock resistance. (TSR).

$$TSR \cong \frac{\sigma_f k}{E\alpha}$$

where σ_f – fracture strength.

Thermal shock behavior is affected by several factors: thermal expansion coefficient – a low value is desired; thermal conductivity – a high value is desired; elastic modulus – low value is desired; fracture strength – high value is desired; phase transformations. Thermal shock may be prevented by altering the external conditions to the degree that cooling or heating rates are reduced and temperature gradients across the material are minimized. Thermal shock is usually not a problem in most metals because metals normally have sufficient ductility to permit deformation rather than fracture. However, it is more of a problem in ceramics and glass materials. It is often necessary to remove thermal stresses in ceramics to improve their mechanical strength. This is usually accomplished by an annealing treatment.

15.3 Thermal conductivity

The ability of a material to transport heat energy from high temperature region to low temperature region is defined as thermal conductivity. Similar to diffusion coefficient, thermal conductivity is a microstructure sensitive property. The heat energy, Q, transported across a plane of area A in presence of a temperature gradient $\Delta T/\Delta l$ is given by

$$Q = kA \frac{\Delta T}{\Delta l}$$

where k is the thermal conductivity of the material. It has units as W/m.K. Metals have k values in the range 20-400, ceramics 2-50, while polymers have in order of 0.3.

Heat energy in solids in transported by two mechanisms: lattice vibrations (phonons) and free electrons. However, usually only one or other predominates the proceedings. Valence electrons gain energy, move toward the colder areas of the material, and transfer their energy to other atoms. The amount of energy transported depends on number of excited electrons, their mobility i.e. type of material, lattice imperfections, and temperature. The thermal energy associated with phonons is transported in the direction of their motion.

In metals, since the valence band is not completely filled, with little thermal excitation number of electron move and contribute to the transfer of heat energy. Thus thermal conduction I metals is primarily due to movement of electrons. It is the same for electrical conduction. Both conductivities are related through the following relation:

$$\frac{k}{\sigma T} = L$$

where L – Lorentz constant, $5.5\text{x}10^{-9}$ cal.ohm/sec.K^2 *or* $2.44\text{x}10^{-8}$ W.ohm/ K^2. The relation is termed as *Wiedemann-Franz law*. Lorentz constant is supposed to be independent of temperature and the same for all metals if the heat energy is transported entirely by free electrons. Thus, the relationship is followed to a limited extension in many metals.

With increase in temperature, both number of carrier electrons and contribution of lattice vibrations increase. Thus thermal conductivity of a metal is expected to increase. However, because of greater lattice vibrations, electron mobility decreases. The combined effect of these factors leads to very different behavior for different metals. For example: thermal conductivity of iron initially decreases then increases slightly; thermal conductivity decreases with increase in temperature for aluminium; while it increases for platinum.

In ceramics, phonons are primarily responsible for thermal conduction. Since the electronic contribution is absent, the thermal conductivity of most ceramics is lower than that of metals. However, main reason for experimentally observed low conductivity of ceramics is the level of porosity, as phonons are effectively scattered by imperfections. The scattering of phonons becomes more pronounced with rising temperature. Hence, the thermal conductivity of ceramic materials normally diminishes with increasing temperature. Advanced ceramic materials like AlN, SiC are good thermal conductors, they are also electrical insulators. Therefore these materials are useful as electronic packaging substrates where heat dissipation is needed.

Thermal conductivity of polymers is even low, compared with ceramic materials. Vibration and movement/rotation of molecular chains transfer heat energy. In these materials thermal conductivity depends on degree of crystallinity; a polymer with highly crystalline and ordered structure will have higher conductivity than amorphous polymer.

15.4 Thermal stresses

Apart from thermal shock, another instance of problem exists with thermal expansion of a material where there is no scope of dimensional changes. Thus due to temperature changes, material may experience thermal stresses ($\sigma_{thermal}$).

$$\sigma_{thermal} = \alpha E \Delta T$$

where E – elastic modulus of the material. Thermal stresses in a constrained body will be of compressive nature if it is heated, and vice versa.

Another source for thermal stresses is thermal gradient within the body when a solid body is heated or cooled. It is because temperature distribution will depend on its size and shape. These thermal stresses may be established as a result of temperature gradients across a body, which are frequently caused by rapid heating or cooling.

Engineering materials can be tailored using multi-phase constituents so that the overall material can show a zero thermal expansion coefficient. For example: Zerodur – a glass-ceramic material that consists of 70-80% crystalline quartz, and the remaining as glassy phase. Negative thermal expansion coefficient of glassy phase compensates for the positive thermal expansion coefficient of the crystalline base, leading to a zero thermal expansion. This material has many applications like parts of telescopes in astronomy. Many ceramic materials for critical thermal applications are developed based on sodium-zirconium-phosphate (NZP) that have a near-zero thermal expansion coefficient.

References

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, Introduction to Ceramics, Second Edition, Wiley, New York, 1976.
2. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.

Chapter 16. Magnetic properties

Engineering materials are important in everyday life because of their versatile structural properties. Other than these properties, they do play an important role because of their physical properties. Prime physical properties of materials include: electrical properties; thermal properties; magnetic properties; and optical properties. Magnetic properties play an important role in daily life. Magnetic materials are used in electric motors, generators, transformers. Modern-day devices use data storage technology that is based on magnetic particles. Magnetic materials are used in devices like telephones, televisions, supercomputers, etc. they are also used in medical technology, for example DNA sequencing.

This chapter shall describe magnetic properties like dia-, para-, and ferro-magnetism along with anti-ferro- and ferri- magnetism. It also includes discussion about temperature effects on magnetic behavior, followed by magnetic domains and hysteresis.

Magnetism is a phenomenon by which a material exerts either attractive or repulsive force on another. Basic source of magnetic force is movement of electrically charged particles. Magnetic behavior of a material can be traced to the structure of atoms. Electrons in atoms have a planetary motion in that they go around the nucleus. This orbital motion and its own spin cause separate magnetic moments, which contribute to the magnetic behavior of materials. Thus every material can respond to a magnetic field. However, the manner in which a material responds depend much on its atomic structure, and determines whether a material will be strongly or weakly magnetic.

Magnetic moment due to spin of an electron is known as *Bohr magneton,*M_B, is the most fundamental magnetic moment.

$$M_B = \frac{qh}{4\pi m_e} = 9.274 X 10^{-24} A.m^2$$

where q is the charge on the electron, h – Planck's constant, m_e – mass of electron. This moment is directed along the axis of electron spin. Magnetic moment resulted from particles in nucleus is much smaller than that result from spin of electron, so it is ignored. If magnetic moment of electrons could sum up, then the world would be a magnetic place. Fortunately, two reasons are found to explain this phenomenon:- First: according to *Pauli exclusion rule*, two electrons with same energy level must have opposite spins – thus so are their magnetic moments, which cancel out each other. Second: orbital moments of electrons also cancel out each other – thus no net magnetic moments if there is no unpaired electron(s). Some elements such as transition elements, lanthanides, and actinides have a net magnetic moment since some of their energy levels have an unpaired electron.

Magnetic dipoles are found to exist in magnetic materials, analogous to electric dipoles. A magnetic dipole is a small magnet composed of north and south poles instead of positive and negative charges. Within a magnetic field, the force of field exerts a torque that tends to orient the dipoles with the filed. Magnetic forces are generated by moving electrically charged particles. These forces are in addition to any electrostatic forces that may already exist. It is convenient to think magnetic forces in terms of distributed field, which is represented by imaginary lines. These lines also indicate the direction of the force. If a magnetic field is generated by passing current I through a coil of length l and number of turns n, then the magnetic field strength is given by:

$$H = \frac{nI}{l}$$

Units for magnetic field strength, thus, are A/m.

Magnetic flux density (induction) is the measure of lines within a medium. It has units as weber (Wb) /m^2 *or* tesla.

$$B = \mu H$$

where μ – permeability. It is a specific property of the medium, and has units as Wb/A.m or henries (H) /m.

Flux density, B, is determined by the manner in which induced and permanent dipoles interact with the applied field. If the magnetic moments reinforce the applied field i.e. greater number of lines of flux are created, and the field is magnified. This is represented as

$$\mu > \mu_0$$

where μ_0 – magnetic permeability of vacuum.

Several parameters may be used to describe magnetic properties of solids. One of them is relative permeability (μ_r). It is a measure of the degree to which the material can be magnetized.

$$\mu_r = \frac{\mu}{\mu_0}$$

Another field quantity called magnetization, *M*, is defined as

$$B = \mu_0 H + \mu_0 M = \mu_0 \mu_r H$$

$$M = \chi_m H$$

χ_m is called the magnetic susceptibility.

$$\chi_m = \mu_r - 1$$

There are many ways a material can be magnetized i.e. many types of magnetism. Three basic magnetisms are: dia-magnetism, para-magnetism and ferro-magnetism. Anti-ferro-magnetism and ferri-magnetisms are considered as subclasses of ferro-magnetism. A material exhibits one of these magnetisms.

16.1 Dia-, Para-, and Ferro-magnetism

Dia-magnetism is very weak form of magnetism. It exists only when an external field is applied, and is non-permanent. The applied external field acts on atoms of a material, slightly unbalancing their orbiting electrons, and creates small magnetic dipoles within atoms which oppose the applied field. This action produces a negative magnetic effect known as diamagnetism. The induced magnetic moment is small, and the magnetization (*M*) direction is opposite to the direction of applied field (*H*). Thus the relative permeability is less than unity i.e. magnetic susceptibility is negative, and is in order of -10^{-5}. Dia-magnetism is virtually found in all materials; however it is observable only in absence of other magnetisms. This form of magnetism is of no practical importance. Materials such as Cu, Ag, Si, Ag and alumina are diamagnetic at room temperature. Superconductors are perfect dia-magnets (χ_m=-1); they lose their superconductivity at higher temperatures or in the presence of a magnetic field.

Materials which exhibit a small positive magnetic susceptibility in the presence of a magnetic field are called para-magnetic, and the effect is termed as *para-magnetism*. When materials have unpaired electrons, a net magnetic moment due to electron spin is associated with each atom. In the absence of an external field, the orientations of these atomic magnetic moments are random leading to no net magnetization. When an external field is applied dipoles line-up with the field, resulting in a positive magnetization. However, because the dipoles do not interact, extremely large magnetic fields are required to align all of the dipoles. In addition, the effect is lost as soon as the magnetic field is removed. Since thermal agitation randomizes the directions of the magnetic dipoles, an increase in temperature decreases the paramagnetic effect. Para-magnetism is produced in many materials like aluminium, calcium, titanium, alloys of copper.

Magnetic susceptibility of these materials is slightly positive, and lies in the range 10^{-5} to 10^{-2}.

Both dia- and para- magnetic materials are considered as non-magnetic because they exhibit magnetization only in presence of an external field. Certain materials possess permanent magnetic moments even in the absence of an external field. This is result of permanent unpaired dipoles formed from unfilled energy levels. These dipoles can easily line-up with the imposed magnetic field due to the exchange interaction or mutual reinforcement of the dipoles. These are chrematistics of *ferro-magnetism*. Materials with ferro-magnetism (Examples: Fe, Co, Ni, Gd) possess magnetic susceptibilities approaching 10^6. Consequently,

$$H << M \text{ , and } B \cong \mu_0 M$$

The mutual spin alignment in ferro-magnetic materials exists over relatively large volume regions of the crystal called *domains*. The susceptibility of ferro-magnetic materials depends upon the intensity of the applied field. Above the Curie temperature, ferro-magnetic materials behave as para-magnetic materials and their susceptibility is given by the Curie-Weiss law, defined as follows:

$$\chi_m = \frac{C}{T - T_c}$$

where C – material constant, T – temperature, T_c – Curie temperature.

In some materials such as Mn, Cr, MnO, NiO, CoO, $MnCl_2$ the magnetic moments produced in neighboring dipoles line up in opposition to one another in the magnetic field, even though the strength of each dipole is very high. This will result in zero magnetization, and the effect is called *anti-ferro-magnetism*. Exchange interaction which is responsible for parallel alignment of spins is extremely sensitive to inter-atomic spacing and to the atomic positions. This sensitivity causes anti-parallel alignment of spins. When the strength of anti-parallel spin magnetic moments is equal, no net spin moment exists, and resulting susceptibilities are quite small. One noticeable characteristic of anti-ferro-magnets is they attain maximum susceptibility at a critical temperature called *Neel temperature*. At temperatures above this, anti-ferro-magnets become para-magnetic.

On the other hand, some ceramic materials exhibit net magnetization. This is because of either the strength or number of opposing dipoles is not equal. In a magnetic field, the dipoles of a cation may line up with the field, while dipoles of other cation may not. These ceramics are called ferrites, and the effect is known as *ferri-magnetism*. Ferri-magnetism is similar to anti-ferro-magnetism in that the spins of different atoms or ions line up anti-parallel. However, the spins do not cancel each other out, and a net spin moment exists. Below the Neel temperature, therefore, ferromagnetic materials behave very much like ferromagnetic materials and are paramagnetic above the Neel temperature. These materials exhibit a large but field dependent magnetic susceptibility similar to ferro-magnets. They also show Curie-Weiss behavior. As these ceramics are good insulators, electrical losses are minimal, and hence ferrites have lot of applications in devices such as high frequency transformers. Examples: Fe_3O_4, $NiFe_2O_4$, $(Mn.Mg)Fe_2O_4$, $PbFe_{12}O_{19}$, Ba $Fe_{12}O_{19}$, YIG – yttrium iron garnet $Y3Fe_5O_{12}$.

The table 16.1 compares different magnetism.

Table 16-1: *Various types of magnetisms.*

Magnetism	Magnetic susceptibility		Examples
	sign	magnitude	
Dia	-	Small, Constant	Organic materials, superconducting materials, metals like Bi
Para	+	Small, Constant	Alkali and transition metals, rare earth elements
Ferro	+	Large, Function of *H*	Transition metals (Fe, Ni, Co), rare earth elements (Gd)
Anti-Ferro	+	Small, Constant	Salts of transition elements (MnO)
Ferri	+	Large, Function of *H*	Ferrites ($MnFe_2O_4$, $ZnFe2O_4$) and chromites

16.2 Influence of temperature on magnetic behavior

Temperature does influence the magnetic characteristics of materials, as it influences electrical properties. With rising temperature, magnitude of the atom thermal vibrations increases. This may lead to more randomization of atomic magnetic moments as they are free to rotate. Usually, atomic thermal vibrations counteract forces between the adjacent atomic dipole moments, resulting in dipole misalignment up to some extent both in presence and absence of external field. As a consequence of it, saturation magnetization initially decreases gradually, then suddenly drops to zero at a temperature called *Curie temperature*, T_c. The magnitude of the Curie temperatue is dependent on the material. For example: for cobalt – 1120 ˚C, for nickel – 335 ˚C, for iron – 768 ˚C, and for Fe_3O_4 – 585 ˚C.

16.3 Domains and Hysteresis

In addition to susceptibility differences, the different types of magnetism can be distinguished by the structure of the magnetic dipoles in regions called domains. Each domain consists of magnetic moments that are aligned, giving rise to a permanent net magnetic moment per domain. Each of these domains is separated from the rest by domain boundaries / domain walls. Boundaries, also called *Bolch walls*, are narrow zones in which the direction of the magnetic moment gradually and continuously changes from that of one domain to that of the next. The domains are typically very small about 50 μm or less, while the Bloch walls are about 100 nm thick. For a polycrystalline specimen, each grain may have more than one microscopic sized domain.

Domains exist even in the absence of an external field. In a material that has never been exposed to a magnetic field, the individual domains have a random orientation. This type of arrangement represents the lowest free energy. The domain structure of material is determined by many types of energies, with most stable structure being attained when the overall potential energy of the material is a minimum. Total magnetic energy of a material is the sum of the contributions of the following: exchange energy, magneto-static energy, magneto-crystalline anisotropy energy, domain wall energy and magneto-strictive energy. When the bulk material is un-magnetized, the net magnetization of these domains is zero, because adjacent domains may be orientated randomly in any number of directions, effectively canceling each other out.

The average magnetic induction of a ferro-magnetic material is intimately related to the domain structure. When a magnetic field is imposed on the material, domains that are nearly lined up with the field grow at the expense of unaligned domains. This process continues until only the most favorably oriented domains remain. In order for the domains to grow, the Bloch walls must move, the external field provides the force required for this moment. When the domain growth is completed, a further increase in the magnetic field causes the domains to rotate and align parallel to the applied field. At this instant material reaches saturation magnetization and no further increase will take place on increasing the strength of the external field. Saturation magnetization is the greatest amount of magnetization tat the material can obtain. Under these conditions the permeability of these materials becomes quite small. *Figure 16-1* in the following presents the relation between the applied field strength and magnetization of the material.

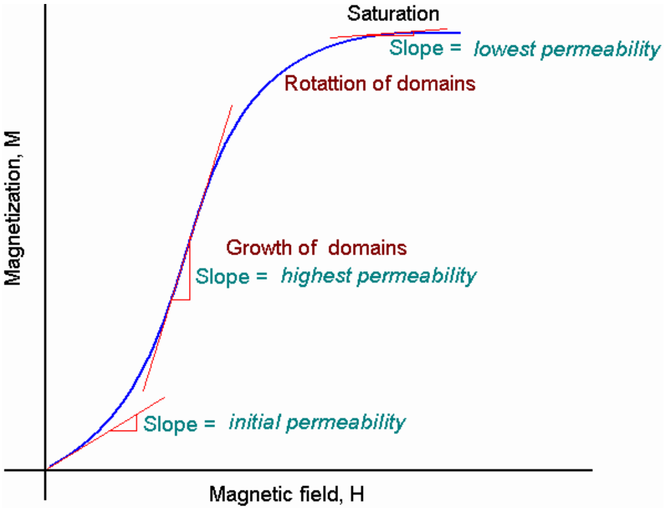


Figure 16-1: *Magnetization saturation of a material with applied field.*

As shown in *figure 16-1*, magnetization increases with applied filed, and reaches a saturation value in a sufficiently stronger filed. Once magnetic saturation has been achieved, a decrease in the applied field back to zero results in a macroscopically permanent or residual magnetization, known as *remanance*, M_r. The corresponding induction, B_r, is called *retentivity* or *remanent induction* of the magnetic material. This effect of retardation by material is called *hysteresis*. The material acts as a permanent magnet, even at zero applied field. At this point, spin orientations within domains have readily rotated back to their favorable position, but the original random domain arrangement is not achieved. This is because the resistance offered by the domain walls prevents re-growth of the domains i.e. domain growth process is not entirely reversible, and domain wall motion is limited. The magnetic field strength needed to bring the induced magnetization to zero is termed as *coercivity*, H_c. This must be applied anti-parallel to the original field.

A further increase in the field in the opposite direction results in a maximum induction i.e. saturation magnetization, but in the opposite direction. The field can once again be reversed, and the field-magnetization loop can be closed, and so the corresponding field-induction loop. This loop is known as *hysteresis loop* or *B-H plot* or *M- H plot*. The area within the hysteresis loop represents the energy loss per unit volume of material for one cycle. The B-vs-H curve in *figure 16-2* represents a hysteresis loop taken to saturation. However, it is not necessary to increase the field strength to saturation to generate the loop.

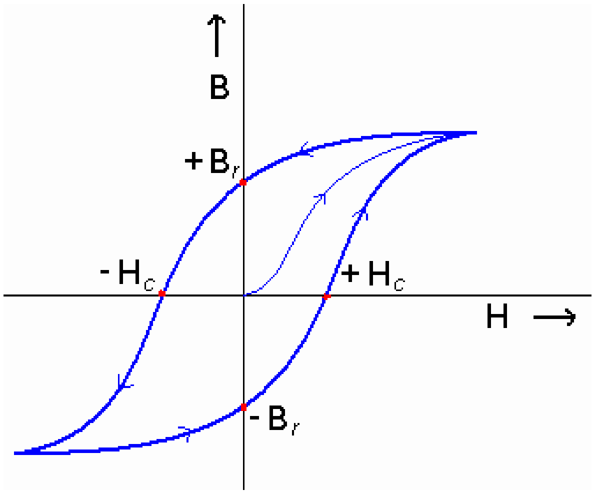


Figure 16-2: *Schematic presentation of a typical hysteresis loop.*

The coercivity of the material is a micro-structure sensitive property. This dependence is known as *magnetic shape anisotropy*. The coercivity of recording materials needs to be smaller than that for others since data written onto a data storage medium should be erasable. On the other hand, the coercivity values should be higher since the data need to be retained. Thus such materials are called magnetically semi-hard. Examples: Hard ferrites based on Ba, CrO_2, γ-Fe_2O_3; alloys based on Co-Pt-Ta-Cr, Fe-Pt and Fe-Pd, etc.

Soft magnets are characterized by low coercive forces and high magnetic permeabilities; and are easily magnetized and de-magnetized. They generally exhibit small hysteresis losses. Application of soft magnets include: cores for electro-magnets, electric motors, transformers, generators, and other electrical equipment. Examples: ingot iron, low-carbon steel, Silicon iron, superalloy (80% Ni-5% Mo-Fe), 45 Permalloy (55%Fe-45%Ni), 2-79 Permalloy (79% Ni-4% Mo-Fe), MnZn ferrite / Ferroxcube A (48% $MnFe_2O_4$-52%$ZnFe_2O_4$), NiZn ferrite / Ferroxcube B (36% $NiFe_2O_4$-64% $ZnFe_2O_4$), etc.

Hard magnets are characterized by high remanent inductions and high coercivities. These are also called *permanent magnets* or *hard magnets*. They generally exhibit large hysteresis losses. Examples: Co-steel, Tungsten steel, $SmCo_5$, $Nd_2Fe_{14}B$, ferrite Bao.6Fe_2O_3, Cunife (60% Cu 20% Ni-20% Fe), Alnico (alloy of Al, Ni, Co and Fe), etc. Applications include fractional horse-power motors, automobiles, audio- and video-recorders, earphones, computer peripherals, and clocks.

References

1. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.
2. V. Raghavan, Materials Science and Engineering, third edition, Prentice Hall of India Private Limited, New Delhi, 1990.
3. D. Jiles, Introduction to Magnetism and Magnetic Materials, Nelson Thornes, Cheltenham, UK 1998.

Chapter 17. Optical properties

Engineering materials are important in everyday life because of their versatile structural properties. Other than these properties, they do play an important role because of their physical properties. Prime physical properties of materials include: electrical properties; thermal properties; magnetic properties; and optical properties. Optical properties play an important role in daily life. They have had a significant impact on the development of the communications infrastructure and the information technology. They are also useful in fields like medicine, manufacturing, astronomy, etc.

The goal of this chapter is to present basic concepts about optical properties, optical properties of metals and non-metals, and finally applications of optical phenomena.

17.1 Basic concepts

Optical property of a material is related to the interaction of it with electromagnetic radiation. This radiation may have characteristics that fall in the visible light spectrum, or may be even out of it. Electromagnetic spectrum of radiation spans the wide range from γ-rays with wavelength as 10^{-12} m, through x-rays, ultraviolet, visible, infrared, and finally radio waves with wavelengths as along as 10^{5} m. Visible light is one form of electromagnetic radiation with wavelengths ranging from 0.39 to 0.77 µm. It contains color bands from violet through red, as shown in the *figure 17.1*. White light is simply a mixture of all colors. The ultraviolet region covers the range from about 0.01 to about 0.40 µm, and the infrared region extends from about 0.75 to 1000 µm.

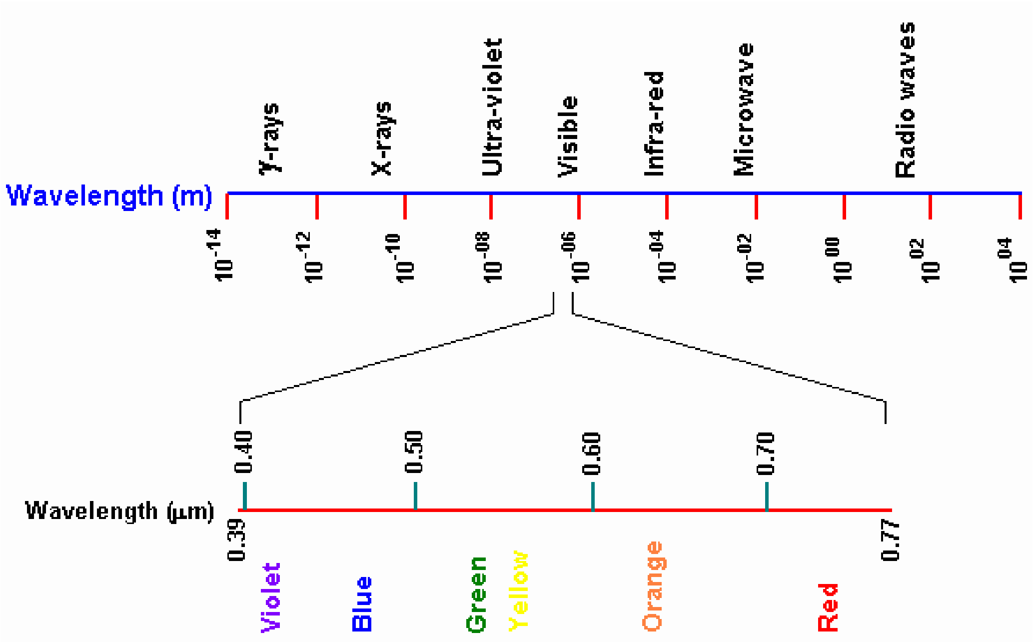


Figure-17.1: *Wave length spectrum of electro-magnetic waves.*

The true nature of the light will probably never be known. However, light can be considered as having waves and consisting of particles called photons. The important characteristics of the photons are related by the following equation. This equation allows us to consider the photon as a particle of energy, or as a wave with a characteristic wavelength and frequency.

$$E = h\nu = \frac{hc_0}{\lambda}$$

where E – energy, h – Planck's constant ($6.62\text{x}10^{-34}$ J.sec), ν – frequency, c_0 – speed of light in vacuum ($3\text{x}10^{8}$ m/sec), and λ – wavelength.

All materials interact in some way with light. Interaction of photons with the electronic or crystal structure of a material leads to a number of phenomena. The photons may give their energy to the material (absorption); photons give their energy, but photons of identical energy are immediately emitted by the material (reflection); photons may not interact with the material structure (transmission); *or* during transmission photons are changes in velocity (refraction).

At any instance of light interaction with a material, the total intensity of the incident light striking a surface is equal to sum of the absorbed, reflected, and transmitted intensities i.e.

$$I_0 = I_A + I_R + I_T$$

The intensity is defined as the number of photons impinging on a surface per unit area per unit time. Materials that are capable of transmitting light with relatively little absorption and reflection are called *transparent materials* i.e. we can see through them. *Translucent materials* are those through which light is transmitted diffusely i.e. objects are not clearly distinguishable when viewed through. Those materials that are impervious to the transmission of visible light are termed as *opaque materials*.

17.2 Optical properties of metals and alloys

Typical characteristic of metals with respect to crystal structure is that they possess a high-energy band that is only partially filled with electrons. When visible light in directed on a metal surface, the energy is used to excite electrons into unoccupied energy states above the Fermi level, thus making metals behave as opaque materials i.e. light is absorbed. Except of thin sections, metals strongly reflect and/or absorb incident radiation for long wavelengths to the middle of the ultraviolet range i.e. metals are opaque to all electromagnetic radiation on the low end of the frequency spectrum, from radio waves, through infrared, visible, into middle of the ultraviolet radiation. However, metals are transparent to high end frequencies, ex. x-ray and γ–ray radiation. Total absorption by metals is 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.

Most of the absorbed radiation is emitted from the metallic surface in the form of visible light of the same wavelength as reflected light. The reflectivity of metals is about 0.95, while the rest of impinged energy is dissipated as heat. The amount of energy absorbed by metals depends on the electronic structure of each particular metal. For example: with copper and gold there is greater absorption of the short wavelength colors such as green and blue and a greater reflection of yellow, orange and red wavelengths. Other metals such as silver and aluminium strongly reflect all parts of the visible spectrum and show a white silvery color.

17.3 Optical properties of non-metallic materials

By virtue of their electron structure with characteristic energy band structures, non-metallic materials may be transparent to visible lights. Thus, all four optical phenomena such as absorption, reflection, transmission and refraction are important for these materials.

Refraction: When light photons are transmitted through a material, they causes polarization of the electrons in the material and by interacting with the polarized materials, photons lose some of their energy. As a result of this, the speed of light is reduced and the beam of light changes direction.

The relative velocity of light passing through a medium is expressed by the optical property called the index of refraction (n), and is defined as:

$$n = \frac{c_0}{c}$$

where c_0 – speed of light in vacuum, c – speed of light in the concerned material. If the angle of incidence from a normal to the surface is θ_i, and the angle of refraction is θ_r, the refractive index of the medium, n, is given by

$$n = \frac{\sin\theta_i}{\sin\theta_r}$$

provided that the incident light is coming from a phase of low refractive index such as vacuum or air. Reflection and refraction of light is shown in *figure 17.2*.

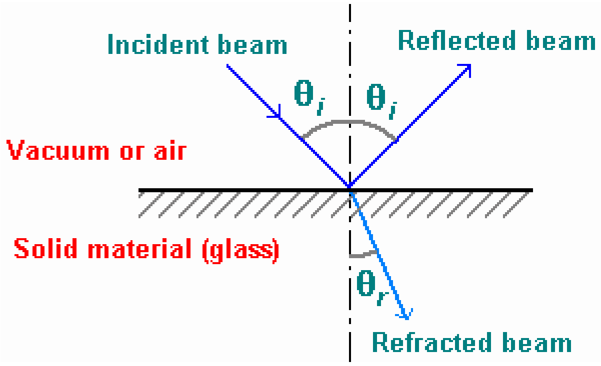


Figure-17.2: *Reflection and refraction of light.*

The speed of light in a material can be related to its electrical and magnetic properties as:

$$c = \frac{1}{\sqrt{\mu.\varepsilon}}$$

where ε – electrical permittivity, and μ – magnetic permeability. Thus,

$$n = \frac{c_0}{c} = \frac{\sqrt{\mu.\varepsilon}}{\sqrt{\mu_0.\varepsilon_0}} = \sqrt{\mu_r.\varepsilon_r}$$

Since most substances are only slightly magnetic i.e. $\mu_r \approx 1$, then

$$n \cong \sqrt{\varepsilon_r}$$

Thus, for transparent materials, index of refraction and dielectric constant are related. Refractive indices of some materials are listed in the *table 17.1*.

Table-17.1: *Index of refraction for typical materials.*

Material	***Refractive index***	***Material***	***Refractive index***
Air	1.00	Epoxy	1.58
Ice	1.309	Polystyrene	1.60
Water	1.33	Spinel, $MgAl_2O_3$	1.72
Teflon	1.35	Sapphire, Al_2O_3	1.76
Silica glass	1.458	Rutile, TiO_2	2.68
Polymethyl methacrylate	1.49	Diamond	2.417
Silicate glass	1.50	Silicon	3.29
Polyethylene	1.52	Gallium arsenide	3.35
NaCl	1.54	Germanium	4.00

Snell's law of light refraction – refractive indices for light passing through from one medium with refractive index n through another of refractive index n' is related to the incident angle, θ, and refractive angle, θ', by the following relation

$$\frac{n}{n'} = \frac{\sin\theta'}{\sin\theta}$$

If light passes from a medium with a high refractive index to one with a low refractive index, there is a critical angle of incidence, θ_c, which if increased will result in total internal reflection of the light. This angle is defined as θ' (refraction) = 90˚.

<u>*Reflection*</u>: When a beam of photons strikes a material, some of the light is scattered at the interface between the two media even if both are transparent. Reflectivity, R, is a measure of fraction of incident light which is reflected at the interface, and is given by

$$R = \frac{I_R}{I_0}$$

Where I_0 and I_R are the incident and reflected bean intensities respectively. If the material is in a vacuum or in air:

$$R = \left(\frac{n-1}{n+1}\right)^2$$

If the material is in some other medium with an index of refraction of n_i, then:

$$R = \left(\frac{n-n_i}{n+n_i}\right)^2$$

The above equations apply to the reflection from a single surface and assume normal incidence. The value of R depends upon the angle of incidence. Materials with a high index of refraction have a higher reflectivity than materials with a low index. Because the index of refraction varies with the wavelength of the photons, so does the reflectivity. In metals, the reflectivity is typically on the order of 0.90-0.95, whereas for glasses it is close to 0.05. The high reflectivity of metals is one reason that they are opaque. High reflectivity is desired in many applications including mirrors, coatings on glasses, etc.

Absorption: When a light beam in impinged on a material surface, portion of the incident beam that is not reflected by the material is either absorbed or transmitted through the material. The fraction of beam that is absorbed is related to the thickness of the materials and the manner in which the photons interact with the material's structure. Thus, according to *Bouguer's law*:

$$I = I_0 \exp(-\alpha . x)$$

where I – intensity of the beam coming out of the material, I_0 – intensity of the incident beam, x – path through which the photons move, and α – linear absorption coefficient, which is characteristic of a particular material.

Absorption in materials occurs mainly by two mechanisms (1) Rayleigh scattering – where photon interacts with the electrons orbiting an atom and is deflected without any change in photon energy. This is significant for high atomic number atoms and low photon energies. Ex.: Blue color in the sunlight gets scattered more than other colors in the visible spectrum and thus making sky look blue. *Tyndall effect* is where scattering occurs from particles much larger than the wavelength of light. Ex.: Clouds look white. (2) Compton scattering – here incident photon knocks out an electron from the atom loosing some of its energy during the process. This is also significant for high atomic number atoms and low photon energies. *Photoelectric effect* occurs when photon energy is consumed to release an electron from atom nucleus. This effect arises from the fact that the potential energy barrier for electrons is finite at the surface of the metal. Absorption occurs at particular levels of photon energies, which are equal to that of binding energies. The energy at which this occurs is called the absorption edge.

Transmission: the fraction of beam that is not reflected or absorbed is transmitted through the material. Thus the fraction of light that is transmitted through a transparent material depends on the losses incurred by absorption and reflection. Thus,

$$R + A + T = 1$$

where R – reflectivity, A – aborptivity, and T – transmitivity. Each of these parameters are characteristic of material, and they also depend on light wavelength.

If the incident light is of intensity I_0, then the loss due to reflection at the front end of the material is RI_0. Thus the fraction of beam intensity entering the material is

$$I_{after\ .reflection} = (1 - R) I_0$$

Once the beam enters the material, a portion of it is absorbed. Thus

$$I_{after.absorption} = (1 - R)I_0 \exp(-\alpha .x)$$

Before the beam exits at the back surface, a portion of it will be reflected again. Thus

$$I_{after.reflection.at.back.surface} = R(1 - R)I_0 \exp(-\alpha .x)$$

Thus, the fraction of beam that is actually transmitted through the material is given by

$$I_{transmitted} = I_{after.absorption} - I_{after.reflection.at.back.surface} = (1 - R)I_0 \exp(-\alpha .x) - R(1 - R)I_0 \exp(-\alpha .x)$$

Thus, $I_t = I_0(1 - R)^2 \exp(-\alpha .x)$

The process of light transmission is shown schematically in the *figure 17.3*.

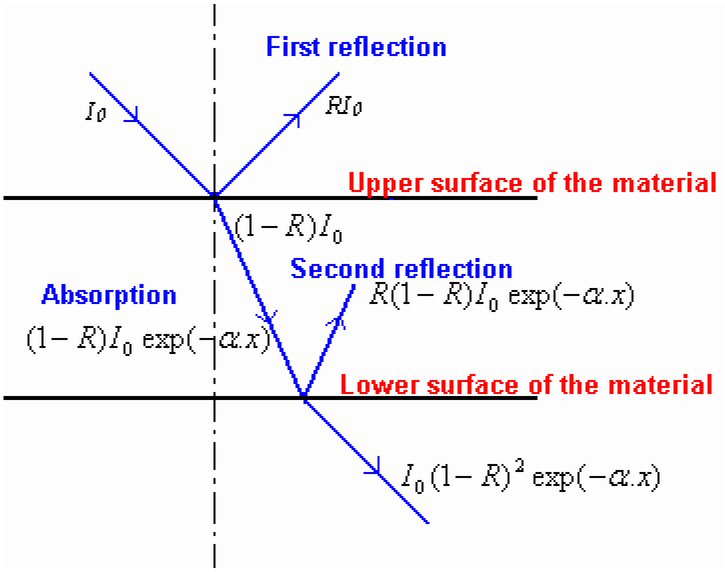


Figure-17.3: *Light transmission.*

17.4 Optical phenomena

As mentioned in earlier sections, light interacts with a material in many ways. Depending on the material, its crystal-/micro-structure, and also on the characteristics of incident light, there are many peculiar phenomena occurs, which are known as optical phenomena. These include: luminescence, lasers, thermal emission, photo-conductivity, and optical fibers. All these find quite many applications in technology for every day life.

Luminescence: Luminescence is defined as the process in which a material absorbs energy and then immediately emits visible or near-visible radiation. This occurs as a result of excitation of electrons of a material from the valence band into the conduction band. The source of input energy may be high energy electrons *or* light photons. During luminescence, the excited electrons drop back to lower energy levels. If the emission takes place within 10^{-8} sec.s after excitation, the luminescence is called *fluorescence*, and if it takes longer than 10^{-8} sec.s, it is known as *phosphorescence*.

Luminescence takes place in outer valence- and conduction- bands, while x-rays are produced during electron transitions in the inner-energy levels of an atom. Luminescence does not occur in metals. In certain ceramics and semi-conductors, however, the energy gap between the valence and conduction bands is such that an electron dropping through this gap produces a photon in the visible range.

Ordinarily pure materials do not display this phenomenon. Special materials called *phosphors* have the capability of absorbing high-energy radiation and spontaneously emitting lower-energy radiation. Ex.: some sulfides, oxides, tungstates, and few organic materials. The emission spectra of these are controlled by added impurities referred as *activators* which provide discrete energy levels in the gap between valence and conduction bands. The excited electrons first drop to donor level, and get trapped. When the source is removed, they gradually escape the trap and emit light over some additional period of time. The intensity of the luminescence is given by

$$I = I_0 \exp(-\frac{t}{\tau})$$

where I_0 – initial intensity of luminescence, I – fraction of luminescence after time, t, τ - relaxation time, constant for a material.

Luminescence process is classified based on the energy source for electron excitation as *photo-luminescence, cathode-luminescence*, and *electro-luminescence*.

Photo-luminescence occurs in fluorescent lamps. Here ultra-violet radiation from low-pressure mercury arc is converted to visible light by calcium halo-phosphate phosphor ($Ca_{10}F_2P_6O_{24}$). In commercial lamps, about 20% of F^- ions are replaced with Cl^- ions. Antimony, Sb^{3+}, ions provide a blue emission while manganese, Mn^{2+}, ions provide an orange-red emission band.

Cathode-luminescence is produced by an energized cathode which generates a beam of high-energy bombarding electrons. Applications of these include: electron microscope; cathode-ray oscilloscope; color television screens. Here, however, relaxation time must not be too long which makes the picture blur or they may even overlap. The modern televisions have very narrow, about 0.25 mm wide, vertical stripes of red-, green-, and blue- emitting phosphors deposited on the inner surface of the screens. With help of steel shadow mask with small elongated holes, the incoming signal is scanned over the entire screen at 30 times per second. The small and large number of phosphors consecutively exposed in the rapid scan of 15750 horizontal lines per second. The persistence of human eye makes possible a clear visible picture with good resolution. Commercial phosphors for different colors are: red – yttrium oxy-sulfide (Y_2O_2S) with 3% europium (Eu); green – (Zn,Cd)S with a Cu^+ acceptor and Al^{3+} donor; blue – zinc sulfide (ZnS) with Ag^+ acceptor and Cl^- donor.

Electro-luminescence occurs in devices with p-n rectifying junctions which are stimulated by an externally applied voltage. When a forward biased voltage is applied across the device, electrons and holes recombine at the junction and emit photons in the visible range (mono-chromatic light i.e. singe color). These diodes are called *light emitting diodes* (LEDs). The characteristic color of an LED depends on the particular semi-conducting material that is used. GaAs, GaP, GaAlAs, and GaAsP are typical materials for LEDs. LEDs emit light of many colors, from red to violet, depending on the composition of the semiconductor material used. Some even emit light outside of the visible spectrum, i.e., infrared and ultraviolet. The following *table 17.2* lists materials used in different colored LEDs.

Table-17.2: *Materials for colored LEDs.*

Wave length (nm)	Color	Material
-	Infra-red	GaAs
660	Red	$GaP_{0.40}As_{0.60}$ *or* $Al_{0.25}Ga_{0.75}As$
635	Orange	$GaP_{0.65}As_{0.35}$
578	Yellow	$GaP_{0.85}As_{0.15}$
556	Green	GaP ($GaP_{1.00}As_{0.00}$)
-	Blue	$Ga_{0.94}NIn_{0.06}$

Lasers: Laser is an acronym for *light amplification by stimulated emission of radiation*. It is in fact special application of luminescence. Unlike most radiation processes, such as luminescence, which produce incoherent light, the light produced by laser emission is coherent i.e. light waves are all in phase with each other. Consequently, laser light waves are does not spread out i.e. parallel, directional, and monochromatic i.e. entirely of one wavelength.

In certain materials, electrons excited by a stimulus produce photons which in turn excite additional photons of identical wavelength. Thus a large amplification of the photons emitted in the material occurs. By selecting stimulant and material properly, laser beam can be in the visible range. Lasers are useful in many applications such as welding, metal cutting, heat treatment, surgery, mapping, reading compact disks, etc. A variety of materials are used to produce lasers. Ex.: Ruby, single crystal of Al_2O_3 doped with little amount of Cr_2O_3; yttrium aluminium garnet ($Y_3Al_5O_{12}$ – YAG) doped with neodymium, Nd; CO_2 gas; He-Ne gas; some semi-conductors like GaAs and InGaAsP. Gas lasers generally produce lower intensities and powers, but are more suitable for continuous operation since solid-state lasers generate appreciable amounts of heat.

Laser operation: when the laser material is exposed to stimulant, for example flash lamp, electrons that initially fills the lowest-energy levels gets excited into higher energy-levels. These electrons can decay back by two paths: one in which they fall directly back –

associated photon emissions are not part of the laser beam; others decay into a intermediate meta-stable state where they reside for about 3 ms before spontaneous emission. This initial spontaneous emission acts as stimulus and triggers an avalanche of emissions from remaining electrons in the meta-stable state. The photons are of the same energy and are in phase. The beam is collimated through the use of a tube with silvered mirrors at each end. As stimulated emission occurs, only those photons traveling nearly parallel to the log axis of the material are reflected. These reflected photons stimulate the emission of more photons. Reflected photons traveling up and down the length of the crystal produce ever-increasing number of stimulated photons. Finally, high-energy, highly-collimated, monochromatic beam of coherent light is emitted from the laser device.

Thermal emission: When a material is heated, electrons are excited to higher energy levels, particularly in the outer energy levels where the electrons are less strongly bound to the nucleus. These excited electrons, upon dropping back to the ground state, release photons in process what is called *thermal emission*.

During thermal emission a continuous spectrum of radiation is emitted with a minimum wavelength and the intensity distribution is dependent on the temperature. Higher the temperature, wider will be the range of wavelengths emitted. By measuring the intensity of a narrow band of the emitted wavelengths with a pyrometer, material’s temperature can be estimated.

Photo-conductivity: As mentioned in earlier section upon absorption of photons at surface, electron may be released from its atom nucleus. Thus electricity can be generated from the surface of a metal when it is bombarded with photons. Similarly, bombardment of semiconductors by photons, with energy equal to greater than the band gap, may result in creation of electron-hole pairs that can be used to generate current. This process is called photo-conductivity, and is different from photo-electric effect in the sense that an electron-hole pair is generated whose energy is related to the band gap energy instead of free electron alone whose energy is related to the Fermi level. The current produced in photo-conductivity is directly related to the incident light intensity.

This phenomenon is utilized in photographic light meters. Cadmium sulfide (CdS) is commonly used for the detection of visible light, as in light meters. Photo-conductivity is also the underlying principle of the photo-voltaic cell, known to common man as solar cell, used for conversion of solar energy into electricity.

Optical fibers: Recently the buzz word in the communications sector is the optical fiber, using in place of metallic copper wires. Signal transmission through a metallic wire conductor is electronic, whereas in fibers it is photonic i.e. by photons. This enables faster transmission at higher densities to longer distances with reduction in error rate. These systems consists of transmitter (a semiconductor laser) to convert electrical signals to light signals, optical fiber to transmit the light signals, and a photodiode to convert light signals back to electrical signals.

Optical fiber is the heart of the communication system. It must have extremely low loss of light, must be able to guide the light pulses over long distances without significant loss and/or distortion. It primarily consists of core, cladding and coating. The core transmits the signals, while the cladding constrains the light beam to the core; outer coating protects the core and cladding from the external environment. Optical fiber operates on the principle of total internal reflectance. Typically both the core and cladding are made of special types of glass with carefully controlled indices of refraction. The indices of refraction are selected such that

$$n_{cladding} < n_{core}$$

Once the light enters the core from the source, it is reflected internally and propagates along the length of the fiber. Internal reflection is accomplished by varying the index of refraction of the core and cladding glass materials. Usually two designs are employed in this regard. In step-index optical fiber, there is a sharp change in refractive index between the core and cladding. In this design output pulse will be broader than the input one. It is because light rays traveling in different trajectories have a variety of path lengths. It is possible to avoid pulse broadening by using graded-index fiber. This results in a helical path for the light rays, as opposed to zig-zag path in a step-index fiber. Here impurities such as boron oxide (B_2O_3) or germanium dioxide /GeO_2) are added to the silica glass such that the index of refraction varied gradually in parabolic manner across the cross section. This enables light to travel faster while close to the periphery than at the center. This avoids pulse broadening i.e. light rays arrive at output at approximately same time. Both step- and graded- index fibers are termed as multi-mode fibers. Third type optical fiber is called single-mode fiber in which light travels largely parallel to the fiber axis with little distortion of the digital light pulse. These are used for long transmission lines.

Core and cladding materials are selected not only on the basis of their refractive indices, but also on basis of ease of manufacturability, light loss, mechanical strength properties and dispersion properties. However, density (ρ) and refractive index (n) are critical. These two parameters are related approximately as

$$n = \frac{\rho + 10.4}{8.6}$$

High-purity silica-based glasses are used as fiber material, with fiber diameter ranging from 5 to 100 μm. The fibers are carefully fabricated to be virtually free from flaws, as a result, are extremely strong and flexible. Uniformity of fiber cross-sectional dimensions and core roundness are critical; allowable tolerances of these parameters are within a micro-meter over 1 km of length.

References

1. William D. Callister, Jr, Materials Science and Engineering – An introduction, sixth edition, John Wiley & Sons, Inc. 2004.

2. K. M. Ralls, T. H. Courtney, and J. Wulff, Introudction to Materials Science and Engineering, Wiley, New York, 1976.
1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, Introduction to Ceramics, Second Edition, Wiley, New York, 1976.

Chapter 18. Economic, Environmental and Social issues of material usage

Engineering materials are important in everyday life because of their versatile structural and physical properties. Selection of the engineering materials is dependent on these properties. However, choice of material has implications throughout the life-cycle of a product, influencing many aspects of economic and environmental performance. In other words, different material issues are needed to be dealt with during material selection if the product is to be commercially successful and competitive in the market. These special considerations are related to their cost, production cost, environmental and social issues like pollution, disposal, recycling, etc. This chapter deals with these special issues of material selection.

18.1 Economic considerations

Engineering profession deals with utilization of scientific and technological advances to design and manufacture components and systems that perform reliably and satisfactorily. However, there are economics as the driving force behind it. Economics of engineering a component / system depends on three factors: component design, material usage, and manufacturing costs. These three factors are inter-related in the sense that one or two might influence the choice of others.

Initially any component / system need to go through conception and then design stage. This includes generation of concept about the component / system. Later on design stage takes care of its size, shape, and configuration which will influence the performance of it during the service. Usually engineers deal with not a single component, but with complex assembly of components / a system. Thus, each component needs to be designed for greater efficiency of the system. This may some times act as constraint to optimal design of a component. Hence, design of a component is usually an iterative process. Less the number of iterations, lower will be the cost of the component / system.

Design stage is followed by material selection. Material is selected depending on its properties, which are suitable to serve the purpose during the service. Other than the properties, cheaper materials are preferred, if choice is available. Thus, usually a family of materials is selected that satisfy the design constraints, then comparisons are made on the basis of cost per unit. This cost also includes unavoidable material wastage during manufacturing stage.

After design and material selection, it is up to manufacturing method to reduce the product cost. Usually manufacturing includes both primary stage and secondary stage. The cost considerations include the capital on tooling, maintenance cost, labor, repair costs, and material wastage. More the number of manufacturing stages, higher will be the product cost. Inspection, assembly, and final packaging will add-on to the product cost.

18.2 Environmental and Social considerations

Manufacturing of a product does have impact on environment and also on society in many ways. This is because resources required to produce a product comes from different parts of the world, and the industrial wealth and prosperity are shared by more the one region. Along with these, detrimental effects of industrialization also spread its wings to various parts of the world.

A material used to produce a product goes through number of stages / phases. These include extraction of raw materials from natural resources through production, use during the service, and finally its disposal. It is some times known as cradle-to-grave life cycle of a material. A schematic life cycle model has been presented in the *figure 18.1*.

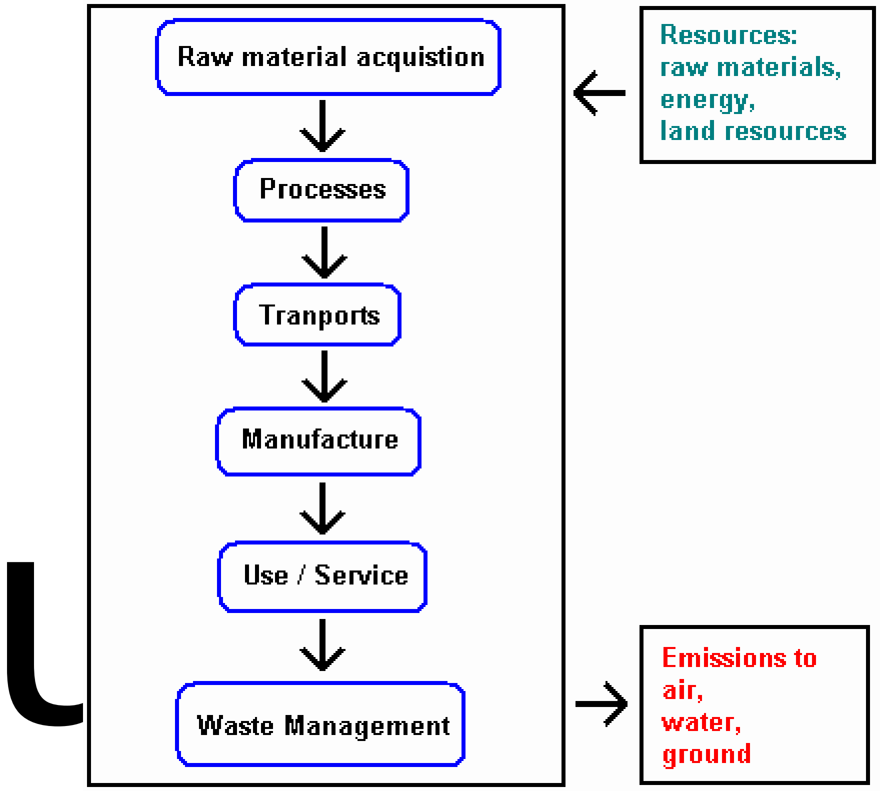


Figure-18.1: *Life cycle model.*

As depicted in the above figure, raw materials are first extracted from natural earthy resources through drilling, mining, etc. later-on these are subjected to purification, refining to convert them into metals, ceramics, rubber, fuel, etc. These primary products are further processed to obtain engineered materials like metallic alloy, glass, plastics, semi-conductors, etc. Now the engineered materials are shaped, heat treated to make components which are assembled into products, devices that are ready for use by society. During the service, products become old, out fashioned, break down, or may not serve the purpose efficiently. So they are discarded. Upon this, they go as waste into land fills, or some times recycled and some components may be reused. This competes the life cycle.

Earthy resources are, unfortunately, limited. However, some are renewable while majority are not! As society and thus population grows, resources per head become lesser and lesser. Hence, there is a great and in fact immediate attention must be paid to the better and effective use of earthy resources. It will be even better if we stop using the resources, which is of course impractical. In addition to this, energy is used at every stage of material cycle, which it self is a resource. That means that energy supply is limited, and energy need to be used conservatively, and effectively. Moreover, during extraction and production of materials / products, interaction with environment is usually detrimental to the society in many ways. It includes – water pollution, air pollution, and landscape spoilage. Thus great care needs to be taken during extraction, manufacturing, and use of a material / product. At end of its service, product may be recycled instead of disposal. There are many benefits from recycling those include: to conserve raw material resources and to conserve energy required to extract and refine. Recycling also avoids environmental pollution. Hence, final products are needed to be designed and manufactured such that they are environmentally friendly, and easy to recycle with minimum usage of raw materials and energy. In case of disposal into environment, products need to be bio-degradable. This shows that material life cycle involves interactions and exchanges among materials, energy and the environment including the society.

There is a need for better solutions to the environmental problems. Thus there is additional cost to the product if it is required to be produced under conditions such that detrimental environmental effects are kept to minimum i.e. green product cost more than usual products. Industrial approach to assess the environmental performance of products is termed as *life cycle analysis / assessment* (LCA). LCA is explained in detail in forth coming sections.

18.3 Recycling issues

As stated in earlier section, recycling of a product is important for many reasons. Recycling and disposal plays an important role in material selection and design of modern products. Recycling can be defined as unflawed entry of material into its cycle after completion of life as a product for infinite number of times. Environmentally friendly means product after disposal into the environment, must interact with it and get deteriorated so that material returns to the same state from where it is extracted and put into life cycle. For ease of disposal and recycling, product design should ease the dismantling of system. Different engineering materials are recycled and/or bio-degradable up to different extents.

Metals and alloys tend to get corroded up to some extent i.e. bio-degradable. However, some of them are toxic. On the other hand, most metals and alloys are recyclable. Ceramics / glasses are, however, are hardly recycled. It is because their raw materials are inexpensive, and recycling process is time consuming and expensive. Plastics and rubber are highly popular among public because of their inertness to the environment and moisture. This same inertness makes them hard to bio-degrade. Thus, plastics are mostly recycled, and just disposed through land-fills. Thermo-plastic polymers are easily recycled up on heating to higher temperatures. On the other hand recycling of thermo-set plastics is much more difficult. Hence these are usually disposed. Thus, there is a trend to use alternative materials which are recyclable. Ex.: thermo-plastic elastomers in place of traditions rubber.

18.4 Life cycle analysis (LCA) and its use in design

Materials extracted from the earth are needed to be taken care of for one prime reason - to conserve material and energy. However, materials, energy and environment effects are interrelated i.e. each one has impact on another either directly / indirectly. Industrial approach to assess the environmental performance of products is termed as *life cycle analysis / assessment* (LCA). In other words, LCA is the assessment of environmental impact of a product across its life cycle. The complex interaction between a product and the environment is dealt with in the Life Cycle Assessment (LCA) method. It is also known *Ecobalance*. LCA systematically describes and assesses all flows to and from nature, from a cradle to grave perspective.

Environmental life cycle of a product consists of all stages from raw material extraction to its waste management. In environmental life cycle assessment, natural resource use and pollutant emission are described in quantitative terms. One important reason for undertaking an LCA study is that there are growing concerns about a variety of environmental issues as expressed by public opinion, political bodies, and industry. Environmental concern may be related to the long-term resource base of human societies or may be more health related or it may be a concern for the natural environment as such. The strength of LCA is that it studies a whole product system. Since a whole life cycle is studied, LCA is not site specific. Thus, environmental impact cannot be modeled at a very detailed level. Economical and social aspects are not included in LCA other than when used as a basis for comparison. Risk is another aspect not dealt with in LCA.

LCA is a technique for assessing the environmental aspects and potential impacts associated with a product by

- complaining an inventory of relevant inputs and outputs of a product system;
- evaluating the potential environmental impacts associated with those inputs and outputs;

- interpreting the results of the inventory analysis and impact assessment phases in relation to the objectives of the study.

LCA is not only product-orientated; it is also quantitative and thus seemingly objective. Thus, it was no longer necessary to reply on simple rules of thumb. With LCA came the notion that it was possible to possible to quantitatively compare alternatives in order to identify the environmentally preferred option. Moreover, it could deal with environmental issues in a structured way and it could handle several environmental issues at a time. Quantitative description of the environmental impacts of emissions and resource using the three categories – resources use, human health, and ecological consequences – is not an uncomplicated affair. Environmental problems are complex and still not fully understood. As a consequence, different attempts at developing models to describe environmental impacts have led to different results.

LCA can be used in many different ways in product development. Life cycle assessment is a technique for assessing the environmental aspects associated with a product over its life cycle. The most important applications are: (1) Analysis of the contribution of the life cycle stages to the overall environmental load, usually with the aim to prioritise improvements on products or processes, (2) Comparison between products for internal or internal communications.

Product design and development has been seen as the principal area of application of LCA since the early days of LCA. The focal point of LCA, the product, coincides with that of the product design and development process. This explains in part why LCA has been brought forward as a tool for environmental adoption of product designs. The other part is that product development is seen as a decisive activity for achieving sustainability in industrial society – most of the environmental attributes are determined and built into the product during the design stage when materials are selected and constructions designed.

With respect to product design, there is a need to understand how a product impacts on the environment. To develop truly sustainable products, it must be possible to assess which design solution is environmentally preferable. LCA tools can help in this difficult area of *eco-design*. Ten simple guide lines for eco-design are:

1. Do not design products, but life cycles;
2. Natural materials are not always better;
3. Energy consumption often underestimated;
4. Increase product life time;
5. Do not design products, but services;
6. Use a minimum of material;
7. Use recycled materials;
8. Make sure product is recyclable;
9. Ask stupid questions i.e. be suspicious about the established result;
10. Become member O2: an international network for sustainable design.

As the whole product lifecycle should be regarded at once, representatives from design, production, marketing, purchasing, and project management should work together on the eco-design of a new product as they have together the best chance to predict the holistic effects of changes of the product and their environmental impact. Environmental aspects which ought to be analyzed for every stage of the lifecycle can be seen as:

1. Consumptions of resources (energy, materials, water or ground)
2. Emissions to air, water or the ground
3. Waste and waste products
4. Miscellaneous (noise and vibration)

Having made up a list on which phase of the lifecycle has which particular environmental aspect, these aspects are evaluated according to their environmental impact on the basis of a number parameters such as extend of environmental impact potential for improvement or potential of change. The evaluations can also take marketing and marketing issues into account. According to this ranking the recommended changes are carried out and are reviewed after a certain time. It must be done regularly to keep up with annual review of environmental legislation by national and international watchdogs.