SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF MECHANICAL ENGINEERING

DEPARTMENT OF AERONAUTICAL ENGINEERING

UNIT – I - MECHANICAL BEHAVIOUR OF AIRCRAFT MATERIALS – SAE1616

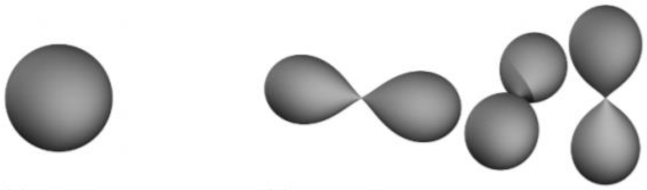
UNIT 1 Mechanical Behaviour of Engineering Materials

Introduction:

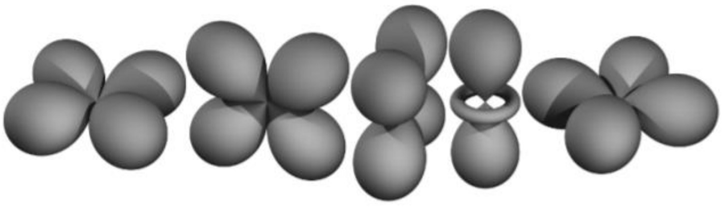
The mechanical properties of materials are determined by their atomic structure. To understand these properties, some knowledge of the structure of materials is therefore required.

1.1 Atomic structure and the chemical bond

Atoms consist of a positively charged nucleus surrounded by negatively charged electrons. Almost the complete mass of the atom is concentrated in the nucleus because it comprises heavy elementary particles, the protons and neutrons. The number of positively charged protons within the nucleus determines the atomic number and thus the chemical element. Thus hydrogen, containing one proton in the nucleus, has an atomic number of 1, oxygen an atomic number of 8, and iron of 26. The nucleus is not involved in chemical reactions which are governed by the electrons surrounding it. The electrons of an atom are not arranged in an arbitrary configuration. Instead, they are confined to so-called electron shells that are arranged in increasing distance around the nucleus and that can only contain a limited number of electrons. The further away an electron shell is from the nucleus, the higher is the energy of the electrons in this shell so that electrons on the outer shells are more weakly bound to the nucleus than those on the inner ones.



(a) *s* orbital (b) *p* orbitals



(c) *d* orbitals

Fig. 1.1. Sketch of selected electron orbitals

The basic structure of all electron shells is the same in all atoms. The innermost shell, called K shell, can contain at most two electrons because there is only one, spherically symmetric, orbital (the s orbital) in it. The next shell, the L shell, can be occupied by up to eight electrons. Two of these are situated in a spherically symmetric s orbital, whereas the other six occupy directed orbitals, the three p orbitals. The subsequent M shell offers space to 18 electrons in s, p, and d orbitals. As nature tends to states of lowest energy, these shells will be filled in the atoms starting with the innermost one, until the number of electrons equals the atomic number so that the atom is electrically neutral. As the energy of the electrons is higher on the outer shells than on the inner ones, it is only the electrons on these shells that are involved in chemical reactions. The binding energy of the weakest bound electron is called the ionisation energy because when the electron is removed a positively charged ion remains. Thus, the ionisation energy is a measure of the binding strength of an electron in the outermost shell.

A chemical bond between atoms is formed by several atoms ‘sharing’ their electrons, or by one atom completely transferring electrons to another to achieve a favourable electron configuration. Hydrogen, for instance, with only one electron on the K shell needs another electron to fill this shell. Therefore, two hydrogen atoms can bond with each other and share their electrons. A hydrogen molecule H_2 is formed. In this, rather simplified, picture of the chemical bond, each atom can form as many bonds as there are electrons missing on the outermost shell. This type of bond is called covalent. The number of bonds formed by an atom is called its valency. So fluorine has a valency of 1, oxygen of 2, and carbon of 4. The chemical bond between two atoms causes an attraction between them. If they get too close, the electrostatic repulsion of the electron shells causes a repulsive force. Another repulsive effect comes about because the size of the orbitals reduces when they approach, which, as explained, is energetically unfavourable. An equilibrium distance is reached where the energy is minimised and there is no net force on the. Typically, atomic distances of covalent bonds are between 0.1 nm and 0.3 nm. Depending on the elements forming the bond, different types of bonds with distinctly different properties can be formed.

1.2 Metals:

Metals are an especially important class of materials. They are distinguished by several special properties, namely their high thermal and electrical conductivity, their ductility, and the characteristic lustre of their surfaces. Their ductility, together with the high strength that can be achieved by alloying, renders metals particularly attractive as engineering materials. In nature, metals occur only seldom as they possess a high tendency for oxidation.

If one looks at the pure elements, more than two thirds of them are in a metallic state. Many elements are soluble in metals in the solid state and thus allow to form a metallic alloy. For instance, steels can be produced by alloying iron with carbon. The large number of metallic elements offers a broad range of possible alloys. Of most technical importance are alloys based on iron (steels and cast irons), aluminium, copper (bronzes and brasses), nickel, titanium, and magnesium.

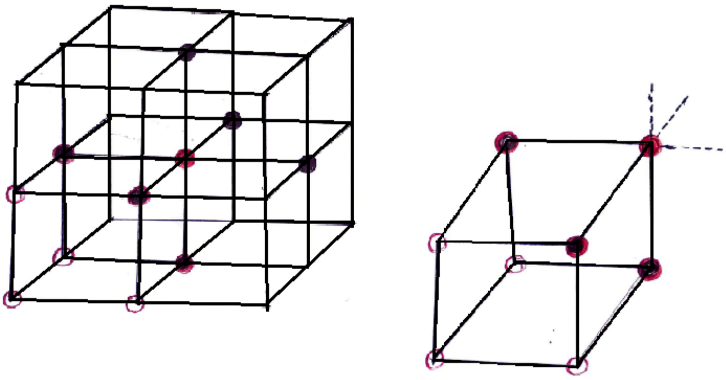
1.2.1 Metallic bond:

Metals are distinguished by possessing rather few electrons on their outer shell and thus would need a large number of electrons to fill this shell. On the other hand, they have the possibility to achieve a fully occupied outer shell by dispensing with their outer electrons. The ionisation energy of metals is, therefore, rather small. Due to the small number of outer electrons, the metallic bond cannot be based on several atoms sharing their electrons to achieve a full outer shell. That, nevertheless, a bond forms is due to the property of electrons to tend to spread over as large a region as possible. For example, Lithium is an alkali metal with only one electron on the outermost shell, thus offering seven unoccupied sites for other electrons. If two lithium atoms approach, both outer electrons, the valence electrons, of the atoms can occupy the space around both atoms and can thus reduce their energy. If a third lithium atom is added, this atom can also spread out its electron over all three atoms, thus forming a Li_3 molecule. A further lithium atom can also add its electron to the mix. Finally, a structure is formed in which each lithium atom is surrounded by eight nearest neighbours and shares its electrons with them. Each bond between two lithium atoms contains on average one quarter of an electron. The bond between the electrons is caused by the spreading of the electrons. This spreading of the electrons makes it impossible to assign the electrons to the atoms they originally belonged to. The electrons spread over the whole material so that on average one electron is always close to any lithium atom, but this electron is not stationary and can move about freely. This is the reason why it is often said that the atoms release their electrons to a common electron gas, resulting in positively charged metallic ions surrounded by a ‘gas’ of negatively charged electrons.

The mobility of the electrons within the electron gas explains many of the physical properties of metals because the excellent electrical and thermal conductivity are based on it. The shininess of metals is also caused by it, for the electrons can easily vibrate in an oscillating electromagnetical field (e. g., light) and thus bar it from entering the metal. As the metallic bond does not result in a fully occupied shell of the single atoms, it is weaker than other types of bond. The binding energy of a metallic bond between any two atoms takes values between approximately 0.1 eV and 0.3 eV. On the other hand, each atom in a metal has a relatively large number of nearest neighbours so that in total relatively large binding energies result, for example 1.1 eV for sodium and 3.5 eV for copper. As the binding energies are lower than in ceramics, which possess fully occupied outer shells, the melting temperature of metals is usually lower as well. The distribution of the electrons over a large region leads to a slow decrease of the interatomic force with the distance of the atoms compared to other types of bonds. Because it is thus possible to displace single atoms with a rather small amount of energy, metals can be easily deformed plastically. If some metal atoms are replaced by those of another metallic element, the metallic bond is usually not destroyed because, for the bond, it is mainly relevant that electrons are released to the electron gas.

1.2.2 Crystal structures

Atoms in a metallic solid arrange themselves so that their electrons can spread over many atoms. This spreading is most easy if the atoms are arranged in a dense and regular manner. Therefore, metals form crystals which are distinguished by their well-ordered structure.



Simple cubic crystal lattice

unit cell

Fig. 1.2. Simple Cubic Crystal Structure

Mathematically, a crystal can be considered as a three-dimensional arrangement of points (i. e., a lattice of points) that looks identical from each of the points. In a real-world crystal each of these points will be occupied by an atom. The crystal has a regular, periodic structure that repeats itself exactly. It thus not only possesses a short-range order, but also a long-range order, for the structure of even a remotely distant region can be predicted exactly from each point. The crystal can be visualised as consisting of cubes that all look alike. These cubes are the ‘building blocks’ from which the crystal can be constructed by putting them together. These building blocks are called unit cells. Unit cells cannot have arbitrary shapes. As the crystal has to be built from them without gaps, only such unit cells can form a crystal that can completely fill space. Altogether, there are 14 different possibilities to arrange atoms on a lattice so that the lattice looks the same from each lattice point. These are called Bravais lattices, named for their discoverer, Auguste Bravais.

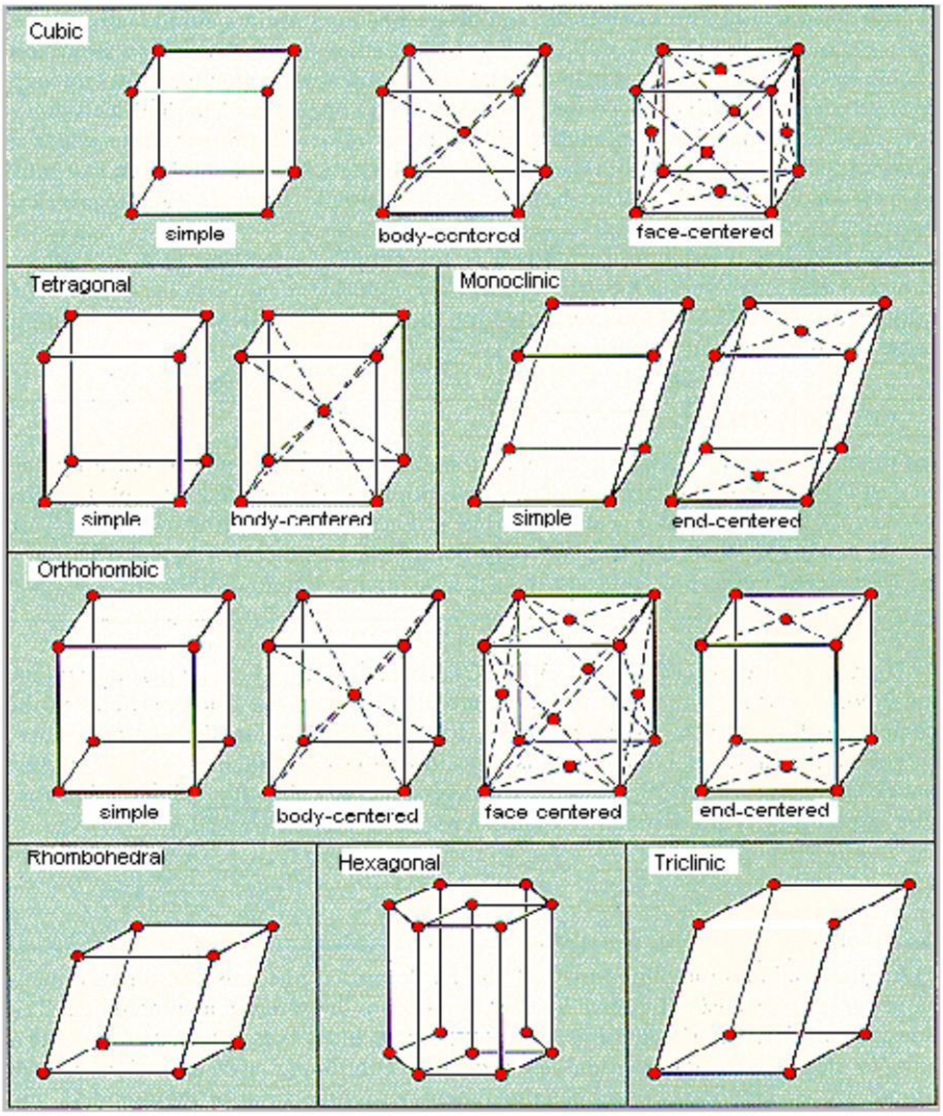


Fig. 1.4. Fourteen different Bravais lattices

For instance, the simple orthorhombic and the simple cubic lattice differ in the orthorhombic unit cell being a quadrangular prism with differing edge lengths, whereas the unit cell of the cubic lattice is a cube. The geometry of the different crystal types will be explained in more detail below. Some of the 14 Bravais lattices are very similar. The simple cubic and the body-centred cubic lattice differ only in the additional atom that is situated in the centre of the unit cell. Such similarities can be described using the symmetries of a crystal. A symmetry of an object is defined as an operation that leaves the object unchanged. The simple cubic crystal structure shown in figure, for example, remains unchanged when it is rotated by 90° along one of its edges, by 120° along the cube diagonal, or if it is reflected using any of the mid-planes of the cube as mirror plane. All crystal types possessing the same symmetries with respect to rotations and reflections as this cubic crystal are grouped into the same crystal system, the cubic crystal system. Although the body-centred cubic, the face-centred cubic and the simple cubic lattice differ in the arrangement of their atoms, they all possess the same cubic symmetry. The 14 Bravais lattices can be grouped into seven crystal systems according to their symmetry as listed. Generally, each crystal system is characterised by six numbers: three lattice constants, indicating the edge lengths of the three axes making up the unit cell, and the three angles between these axes. Typical values of the lattice constant in metals are between 0.2 nm and 0.6 nm. The symmetry of a crystal type is relevant because frequently it is reflected in its material properties. A cubic crystal, for instance, has the corresponding symmetries in its mechanical properties. The lower the symmetry of a crystal, the more complicated is the anisotropy of its properties.

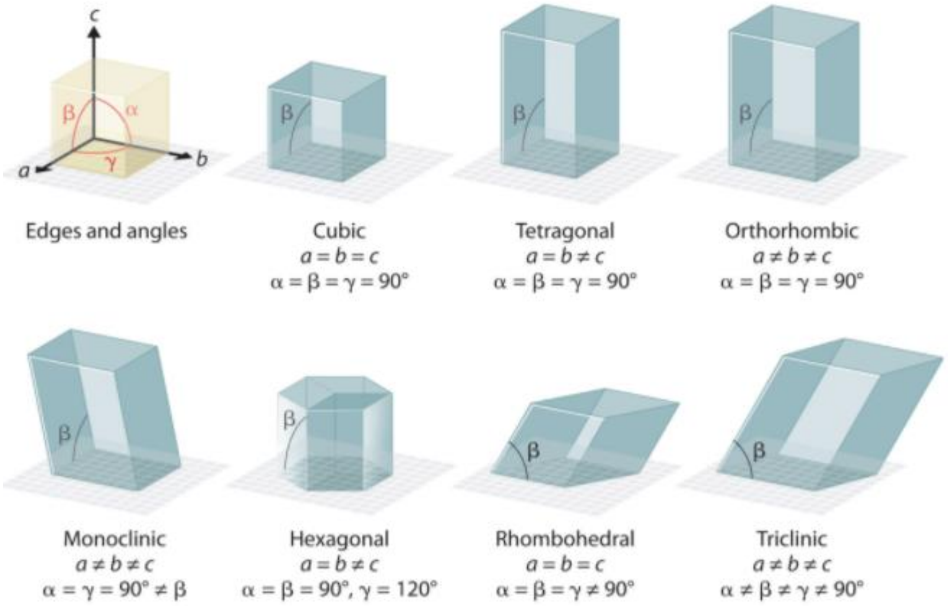


Fig. 1.5. Seven crystal systems

In metals, three lattice structures are especially frequent. Two of these are Bravais lattices with cubic symmetry:

- face-centred cubic
- body-centred cubic

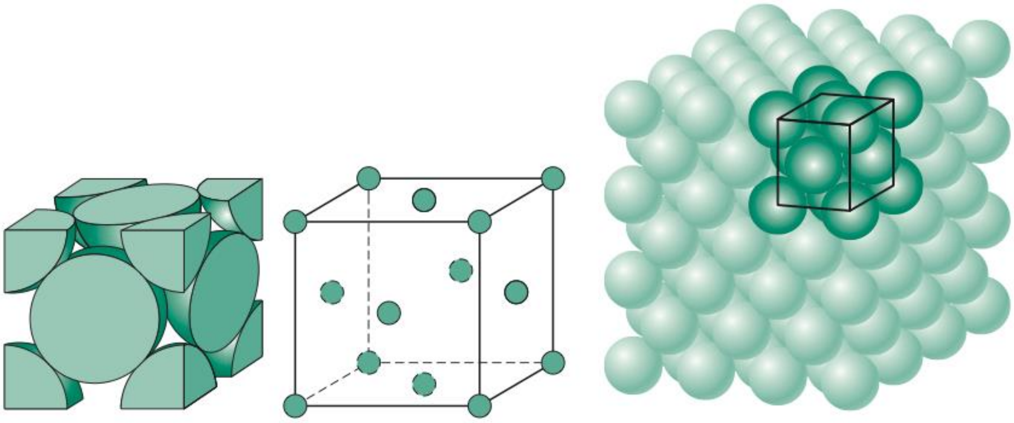


Fig. 1.6. Face-centred cubic

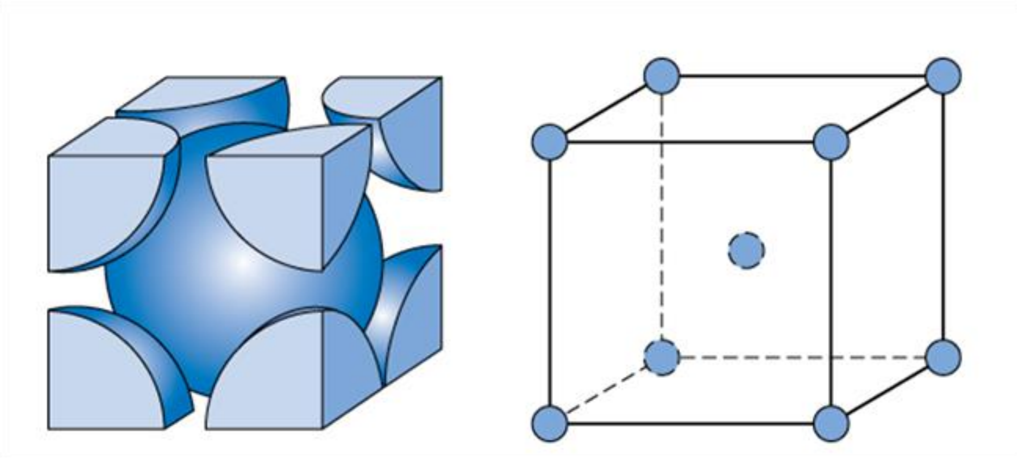
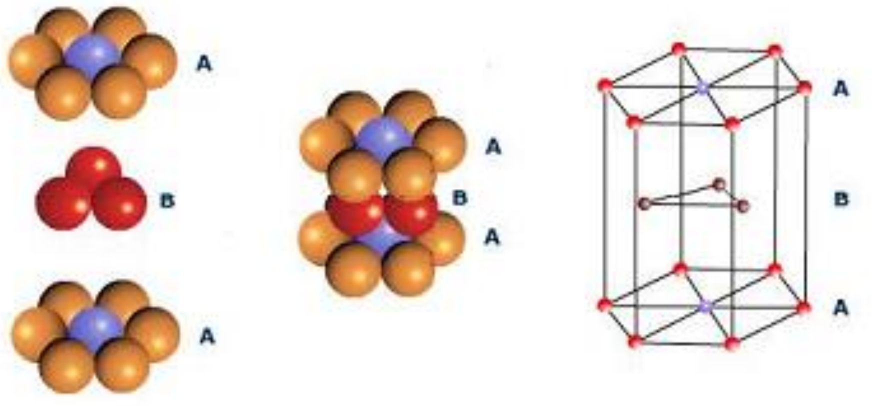


Fig. 1.7. Body-centred cubic

The third important crystal structure of metals is the hexagonal close packed structure, abbreviated hcp. This structure is not a Bravais lattice as not all atoms occupy identical positions.



(Hard sphere model) (Lattice point representation)

Fig. 1.8. Hexagonal close-packed structure

Looking at figure 1.8, it can be seen that the atom at the front right edge of the cell has a neighbour that can be reached by moving up by $c/2$ and to the left and back by $a/\sqrt{3}$. If this step is repeated from the atom reached in this way, there is no atom at the new position. The hexagonal close-packed lattice can be constructed by stacking two simple hexagonal lattices into each other. Such lattices are called lattices with a basis.

1.3 Polycrystalline metals

If a metal is cooled down from a melt and solidifies, it starts to crystallise. Depending on the cooling rate, many small nuclei of crystallisation form, small solidified regions with crystalline structure. These nuclei then grow and coalesce. As the initial nuclei develop independently, they possess no long-range order between them. Therefore, a metal does not usually consist of one single crystal with long-range order, but rather of several crystalline regions called crystallites or grains. They have a diameter of the order of a few micrometres up to a fraction of a millimetre, but can also be much larger in special cases. Grains can be made visible by polishing the surface of the metal and then etching it because the acid attacks differently oriented grains differently.

The structure of the grains of a metal is usually termed its microstructure. The grain boundaries i. e., the interfaces between the grains, do not have a perfectly crystalline order as differently oriented regions adjoin here. Therefore, they can be considered as lattice imperfections. Frequently, they strongly influence the properties of a material because, for example, they may be preferred diffusion paths for corroding media. This kind of weakening of grain boundaries may then lead to failure of the material. This is called inter-crystalline fracture. Technical alloys frequently consist of different phases i. e., regions with differing chemical composition or crystal structure. The particles of a second phase that are enclosed by a matrix of a first phase are especially important to influence mechanical properties. One example for this is iron carbide (cementite, Fe_3C) that increases the strength of steels when precipitated as fine particles.

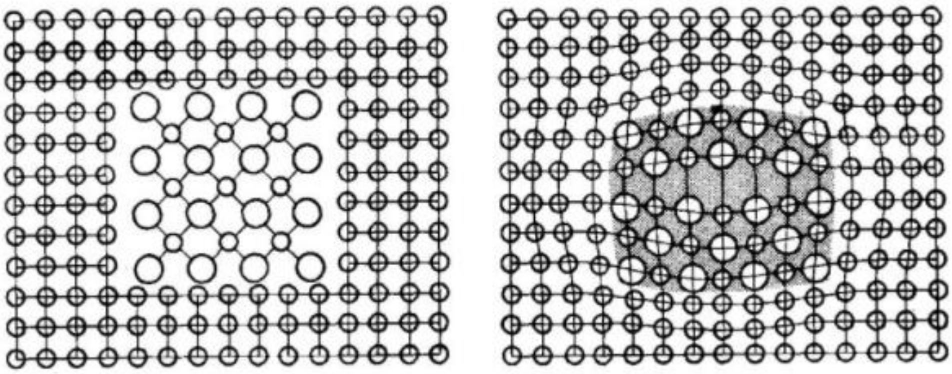


Fig. 1.9. Coherent and non-coherent crystal structure

Depending in the crystal structure of the two phases, the interface between them may adopt different structures: If the crystal structures and the crystal orientation of both phases are identical and the lattice constants do not differ too much, the particles of the second phase will be coherent i. e., the lattice planes of the matrix continue within the particle. If the lattice structure and orientation are identical, but the lattice constants differ strongly, the particles will be semi-coherent because some lattice planes of the matrix continue inside the particle but others do not. Generally, the crystal lattice is distorted near to the coherent or semi-coherent particle. If the lattice structure of both phases or the lattice orientation differ, the particles are incoherent; the lattice planes of particle and matrix have no relation at all.

1.4 Ceramics:

All non-metallic, non-organic materials are called ceramics. Physically, the distinction between ceramics and metals can be based on their bond type – ceramics do not possess a metallic bond, but bond types that result in a completely filled outer shell. Ceramics can be elementary i. e., they may consist of only one element (carbon, for example, can exist in two different ceramic forms, as diamond or graphite), or they can be compounds of different elements. Of technical importance are silicate ceramics, containing silicon oxide (for example, porcelain or mullite), oxide ceramics i. e., compounds of metallic elements with oxygen (for example, aluminium oxide Al_2O_3, zirconium oxide ZrO_2, or magnesium oxide MgO), and non-oxide ceramics i. e., oxygen-free compounds like silicon carbide and silicon nitride. Ceramics can be chemically bound in different ways. Rather strong bond types are the covalent and ionic bonds, weaker ones are van der Waals, dipole, and hydrogen bonds.

1.5 Covalent bond:

Atoms that lack only a few electrons to achieve a fully occupied outer shell share some of their electrons. To form a solid with strong bonds between the atoms, it is insufficient if each electron lacks only one electron because in this case a two-atomic molecule will form only. An atom with a valency of four, like carbon, can form large units in which each atom has four bonded neighbours. In this ceramic silicon oxide, each oxygen atom is linked to two silicon atoms which serve as the nodes in the three-dimensional network. In contrast to the metallic bond, the covalent bond is directed. Thus, the electrons do not spread evenly over a wide region of the crystal, but are concentrated on the connecting line between two atoms. Therefore, it is much more difficult to move atoms in a covalent crystal against each other, resulting in brittleness and poor deformability of these ceramics. The binding energy of the covalent bond is typically about 1 eV per bond, but reaches a value of 1.85 eV in diamond. Due to the smaller number of nearest neighbours, the difference between the overall binding energy of ceramics and metals is smaller – even in diamond the binding energy of an atom is 7.4 eV, only twice that of copper, a metal with a rather high binding energy. In other covalent crystals, typical values are between 3 eV and 5 eV, again approximately twice that of typical metals.

1.6 Ionic bond:

Many ceramics are compounds of a metal and a non-metal. Common salt, for instance, consists of sodium and chlorine (NaCl). From this formula and the fact that common salt forms a crystal, it can be deduced that the bond cannot be covalent, for as chlorine has a valency of only one, only a diatomic molecule could form, but not a crystal. Instead, an ionic bond is formed. The ionic bond is based on the high electron affinity (also known as electronegativity) of the non-metal (the chlorine in the example of common salt), whereas the metal (the sodium in the example) has only a small ionisation energy. If the outer electron of the metal is transferred to the non-metal, only a comparably small amount of energy is needed. Additional energy can be gained because the two resulting ions are electrically charged and attract each other. A diatomic molecule forms, held together by the electrostatic attraction of its ions. In an ionic crystal, the binding energy is even higher than in a diatomic molecule because each ion is surrounded by several oppositely charged ions. Similar to the covalent bond, the ionic bond is directed. Shifting the atoms would strongly increase the electrostatic repulsion of the ions. Therefore, ionic crystals are also brittle. There is a smooth transition between covalent and ionic bonds. The ionisation energy of metals increases with increasing number of outer electrons, whereas the electron affinity of the non-metals decreases with an increasing valency. In between the purely covalent and the purely ionic bond, there are also intermediate states where the electron is preferentially located at one atom, but can also be found at the other. One example of this is carbon dioxide (CO_2) in which the oxygen atoms have a higher electron affinity than the carbon atom. The electrons therefore have a higher tendency of being close to the oxygen atoms so that these are partially negatively charged, whereas the carbon atom has a partially positive charge. The molecule is electrically polar and can be considered as consisting of two electric dipoles. This kind of bond is called a polar bond.

1.7 Dipole bond:

If carbon dioxide (CO_2) is cooled down to −78°C, it forms dry ice, a solid. As the atoms of each CO_2 molecule have fully occupied shells, none of the binding mechanisms discussed so far can be responsible for the cohesion between molecules. The bond between the carbon dioxide molecules is due to the polarity of the molecules in which electrical charges are distributed inhomogeneously. Because the molecules form electric dipoles, this type of bond is called dipole bond. As the atoms in the molecules do not carry complete elementary charges, but are charged rather weakly, the attractive force between the molecules is correspondingly small. Typical binding energies lie in the range of 0.2 eV–0.4 eV per bond. Solids like dry ice are, according to the definition, ceramics, but due to the small binding forces they are not used as engineering materials. However, the dipole bond plays an important role in binding polymers.

1.8 Van der Waals bond:

Even completely nonpolar molecules like oxygen or the noble gases finally solidify if cooled down sufficiently. The attraction between such molecules is even smaller than that between molecular dipoles, but it is nevertheless present. This attractive force is called van der Waals force or, sometimes, dispersion force. The van der Waals force originates in charge fluctuations in the electron shell of the atoms. Slightly simplified, it can be imagined that the charge distribution of an atom is not static because the outer electrons move about. At any instant in time, the atom therefore forms a weak dipole, although on the average it is still electrically neutral. Neighbouring atoms possessing such dipole moments attract each other, for proximity is energetically favourable if the movement of the electrons is correlated. A van der Waals force acts between all molecules. Because it is the weakest of all bond types, it can only play a role if no other binding mechanism is present. The strength of the van der Waals force is between 0.01 eV and 0.1 eV per bond. In addition, it is very short-ranged and decreases rapidly with growing distance of the molecules.13 The van der Waals force is stronger in large atoms than in small ones because, due to their larger radius, they can produce larger dipole moments.

1.9 Hydrogen bond:

Water has very special properties. If we compare the boiling temperature of hydrogen compounds of elements of the sixth group of the periodic table (tellurium, selenium, sulfur, and oxygen), these values are −2°C for H_2Te, −42°C for H_2Se, and −60°C for H_2S. The decrease is due to the decreasing dipole moments with decreasing atomic radius. Therefore, we would expect water to have a very low boiling temperature. Instead, H_2O boils at +100°C. The binding force between the water molecules is thus much higher than expected from the comparison with other molecules. Water is a polar molecule and as oxygen has a slightly higher electron affinity than, for example, sulfur, the larger boiling temperature may at least partially be due to this, but a detailed calculation shows that the dipole bond is far too weak to explain the large boiling temperature. The special property of water is based on the formation of so-called hydrogen bonds.

1.10 Polymers:

Polymers (plastics) consist of macromolecules, frequently in the form of large molecular chains in which the atoms are held together by covalent bonds, whereas the bonds between the different chains are much weaker. For this reason, chain molecules can be considered as the basic building units of a polymer. Contrary to metals and ceramics, polymers are thus composed not of pointlike particles (atoms), but of linear components. Therefore, their structure is more complicated than that of the other classes of materials.

1.11 Mechanical Properties:

Often materials are subject to forces (loads) when they are used. Mechanical engineers calculate those forces and material scientists how materials deform (elongate, compress, twist) or break as a function of applied load, time, temperature, and other conditions. Materials scientists learn about these mechanical properties by testing materials. Results from the tests depend on the size and shape of material to be tested (specimen), how it is held, and the way of performing the test. That is why we use common procedures, or standards. The engineering tension test is widely used to provide basic design information on the strength of materials and as an acceptance test for the specification of materials.

In the tension test a specimen is subjected to a continually increasing uniaxial tensile force while simultaneous observations are made of the elongation of the specimen. The parameters, which are used to describe the stress-strain curve of a metal, are the tensile strength, yield strength or yield point, percent elongation, and reduction of area. The first two are strength parameters; the last two indicate ductility. In the tension test a specimen is subjected to a continually increasing uniaxial tensile force while simultaneous observations are made of the elongation of the specimen. An engineering stress-strain curve is constructed from the load elongation measurements.

The tensile test is probably the simplest and most widely used test to characterize the mechanical properties of a material. The test is performed using a loading apparatus such as the Tinius Olsen machine. The capacity of this machine is 10,000 pounds (tension and compression). The specimen of a given material (i.e. steel, aluminum, cast iron) takes a cylindrical shape that is 2.0 in. long and 0.5 in. in diameter in its undeformed (with no permanent strain or residual stress), or original shape.

The results from the tensile test have direct design implications. Many common engineering structural components are designed to perform under tension. The truss is probably the most common example of a structure whose members are designed to be in tension (and compression).

1.11.1 Concepts of Stress and Strain

Stress can be defined by ratio of the perpendicular force applied to a specimen divided by its original cross-sectional area, formally called engineering stress. To compare specimens of different sizes, the load is calculated per unit area, also called normalization to the area. Force divided by area is called stress. In tension and compression tests, the relevant area is that perpendicular to the force. In shear or torsion tests, the area is perpendicular to the axis of rotation. The stress is obtained by dividing the load (F) by the original area of the cross section of the specimen (A_O).

$$\sigma = \frac{F}{A_0}$$

The unit is the Megapascal = 10^6 Newtons/m^2.

There is a change in dimensions, or deformation elongation, ΔL as a result of a tensile or compressive stress. To enable comparison with specimens of different length, the elongation is also normalized, this time to the length l_o. This is called strain. So, Strain is the ratio of change in length due to deformation to the original length of the specimen, formally called engineering strain. strain is unitless, but often units of m/m (or mm/mm) are used.

The strain used for the engineering stress-strain curve is the average linear strain, which is obtained by dividing the elongation of the gage length of the specimen, by its original length.

$$\varepsilon = \frac{l_i - l_o}{l_o} = \frac{\Delta l}{l_o}$$

Since both the stress and the strain are obtained by dividing the load and elongation by constant factors, the load-elongation curve will have the same shape as the engineering stress-strain curve. The two curves are frequently used interchangeably. The shape and magnitude of the stress-strain curve of a metal will depend on its composition, heat treatment, prior history of plastic deformation, and the strain rate, temperature, and state of stress imposed during the testing. The parameters used to to describe stress-strain curve are tensile strength, yield strength or yield point, percent elongation, and reduction of area. The first two are strength parameters; the last two indicate ductility.

The general shape of the engineering stress-strain curve requires further explanation. In the elastic region stress is linearly proportional to strain. When the load exceeds a value corresponding to the yield strength, the specimen undergoes gross plastic deformation. It is permanently deformed if the load is released to zero. The stress to produce continued plastic deformation increases with increasing plastic strain, i.e., the metal strain-hardens. The volume of the specimen remains constant during plastic deformation, $A \cdot L = A_0 \cdot L_0$ and as the specimen elongates, it decreases uniformly along the gage length in cross-sectional area.

Initially the strain hardening more than compensates for this decrease in area and the engineering stress (proportional to load P) continues to rise with increasing strain. Eventually a point is reached where the decrease in specimen cross-sectional area is greater than the increase in deformation load arising from strain hardening. This condition will be reached first at some point in the specimen that is slightly weaker than the rest. All further plastic deformation is concentrated in this region, and the specimen begins to neck or thin down locally. Because the cross-sectional area now is decreasing far more rapidly than strain hardening increases the deformation load, the actual load required to deform the specimen falls off and the engineering stress likewise continues to decrease until fracture occurs.

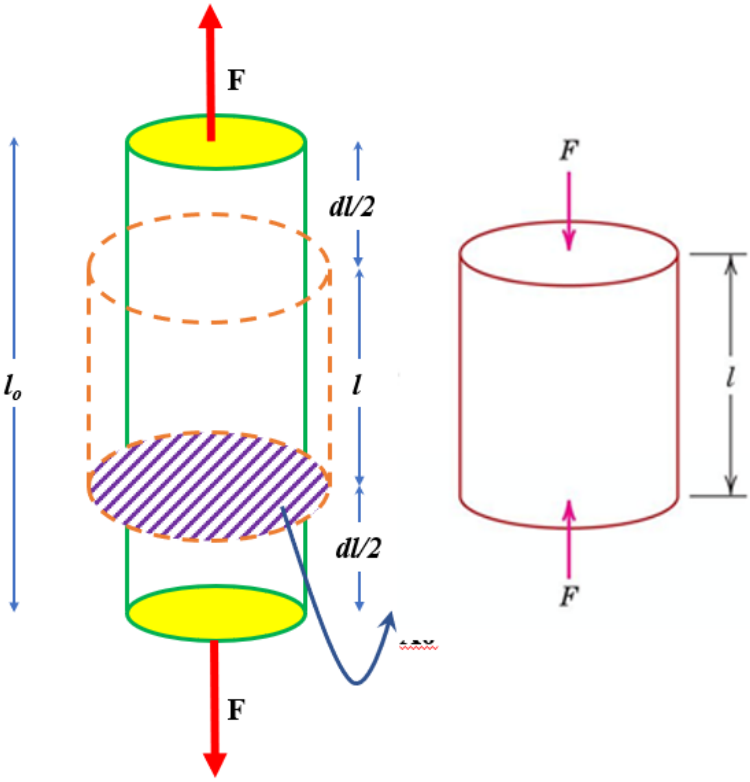


Fig.1.10. Tensile and compressional stress can be defined in terms of forces applied to a uniform rod.

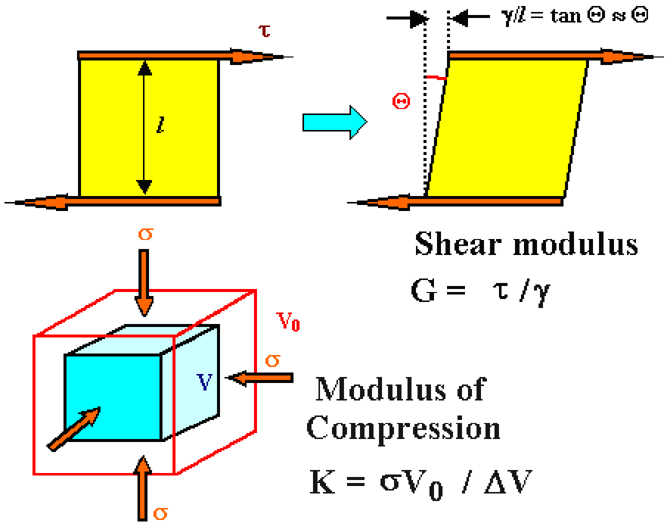


Fig.1.11. Shear stress to deform a joining member

Shear stress is defined in terms of a couple that tends to deform a joining member.

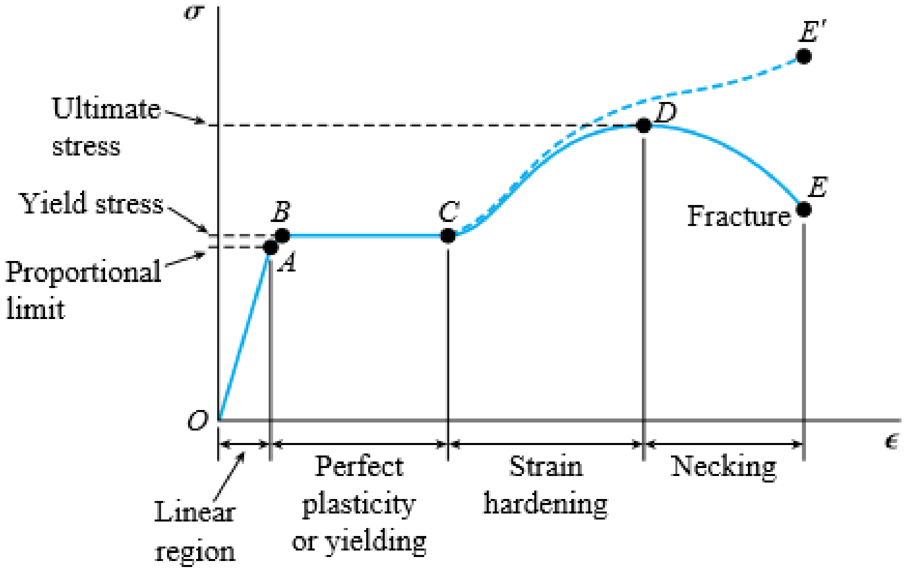


Fig.1.12. A typical stress-strain curve showing the linear region, necking and eventual break.

Shear strain is defined as the tangent of the angle theta, and, in essence, determines to what extent the plane was displaced. In this case, the force is applied as a couple (that is, not along the same line), tending to shear off the solid object that separates the force arms.

$$\text{shear strain} = \frac{\Delta x}{l}$$

where, Δx = deformation in m

l = width of a sample in m

In this case, the force is applied as a couple (that is, not along the same line), tending to shear off the solid object that separates the force arms. The strain in this case is defined as the fractional change in dimension of the sheared member.

1.12 Stress—Strain Behaviour:

1.12.1 Hooke’s Law

For materials stressed in tension, at relatively low levels, stress and strain are proportional through:

$$\sigma = \mathbf{E}\varepsilon$$

constant E is known as the modulus of elasticity, or young’s modulus measured in MPa and can range in values from ~$4.5\text{x}10^4$ - $40\text{x}10^7$ MPa.

The engineering stress strain graph shows that the relationship between stress and strain is linear over some range of stress. If the stress is kept within the linear region, the material is essentially elastic in that if the stress is removed, the deformation is also gone. But if the elastic limit is exceeded, permanent deformation results. The material may begin to "neck" at some location and finally break. Within the linear region, a specific type of material will always follow the same curves despite different physical dimensions. Thus, it can say that the linearity and slope are a constant of the type of material only. In tensile and compressional stress, this constant is called the modulus of elasticity or Young's modulus (E)

$$E = \frac{F/A}{\Delta l/l}$$

Where,

stress = F/A in N/m^2

strain = Δl/l unitless

E = Modulus of elasticity in N/m^2

The modulus of elasticity has units of stress, that is, N/m^2, The following table gives the modulus of elasticity for several materials. In an exactly similar fashion, the shear modulus is defined for shear stress-strain as modulus of elasticity.

Table 1.1 The modulus of elasticity

Material	**Modulus of elasticity (N/m^2)**
Aluminium	$6.89x10^{10}$
Copper	$11.73x10^{11}$
Steel	$2.1x10^{8}$

1.12.2 Stress-strain curve

If a ductile bar of uniform cross-sectional area is subjected to gradually increasing axial tensile force (generally is done in Universal Testing Machine) till failure of the bar occurs, when the stress-strain curve plots the curve may be divided into following parts.

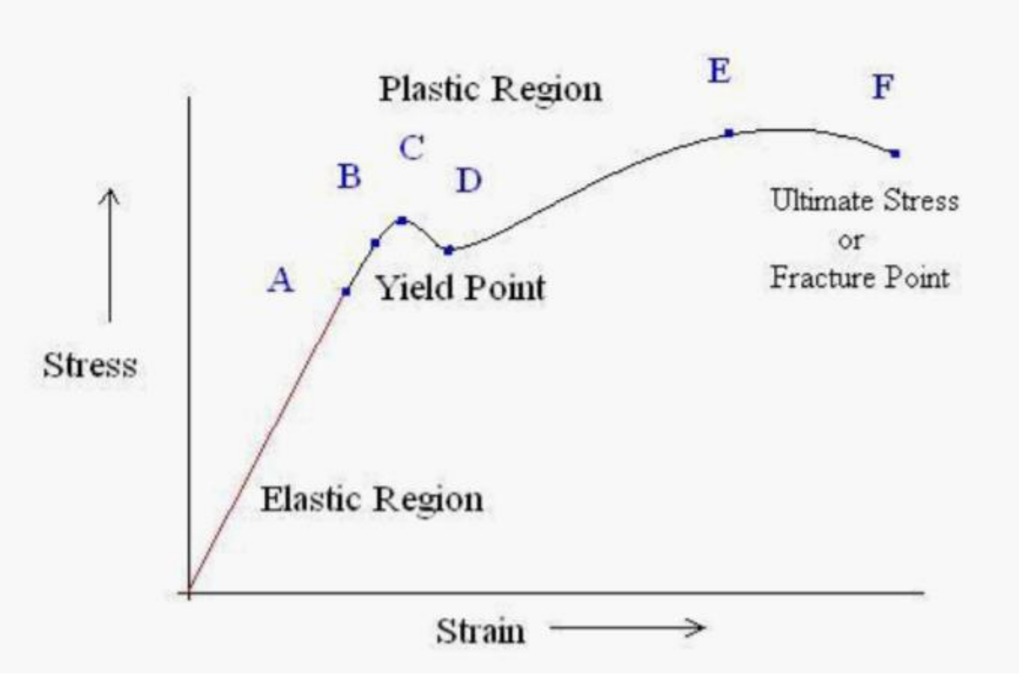


Fig.1.13. stress-strain curve.

Portion OA: This portion is absolutely straight, where the stress is proportional to strain and the material obeys Hooke's law ($\sigma = E \epsilon$). The value of stress at point A is called proportional limit.

Portion AB: In this portion, Hook’s law is not obeyed, although the material may still be elastic. The point B indicates the elastic limit.

Portion BC: In this portion, the metal shows a strain even without increase in stress and the strain is not fully return when load is removed.

Portion CD: Yielding start in this portion and there is a drop of stress at the point D directly after yielding begins at C. The point D is termed as lower yield point and C is called upper yield point.

Portion DE: After yielding has taken place at D, further straining takes place at this portion by increasing the stress and the stress–strain curve continues to rise up to the point E. Strain in this portion is about 100 times that of portion O-A. At the point E, the bar begins to form a local neck. The point E is termed as ultimate tensile stress point.

Portion EF: In this portion, the load is falling off from the maximum and fracture at F takes place. The point F is termed as fracture or breaking point and the identical stress is called breaking stress.

A stress-strain curve with each region identified is shown below. The curve has been sketched using the assumption that the strain in the specimen is monotonically increasing - no unloading occurs. It should also be emphasized that a lot of variation from what's shown is possible with real materials, and each of the above regions will not always be so clearly delineated. It should be emphasized that the extent of each region in stress-strain space is material dependent, and that not all materials exhibit all of the above regions.

A stress-strain curve is a graph derived from measuring load (stress - σ) versus extension (strain - ε) for a sample of a material. The nature of the curve varies from material to material. The following diagrams illustrate the stress-strain behaviour of typical materials in terms of the engineering stress and engineering strain where the stress and strain are calculated based on the original dimensions of the sample and not the instantaneous values. In each case the samples are loaded in tension although in many cases similar behaviour is observed in compression.

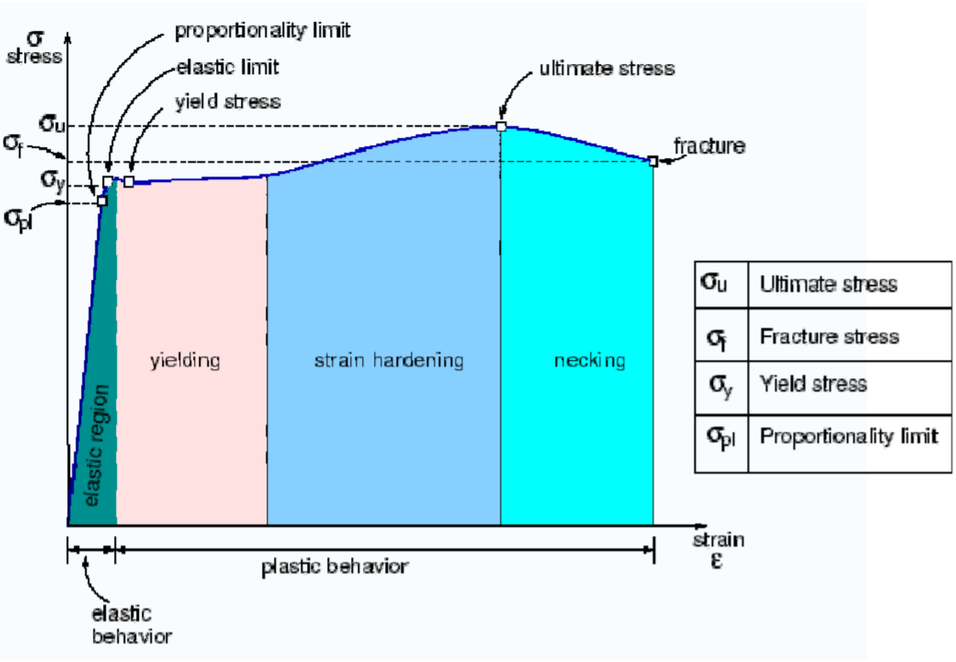


Fig.1.14. Various regions in stress-strain curve.

1.12.3 Brittle and Ductile Behaviour:

The behaviour of materials can be broadly classified into two categories; brittle and ductile. Steel and aluminium usually fall in the class of ductile materials. Glass, ceramics, plain concrete and cast-iron fall in the class of brittle materials. The two categories can be distinguished by comparing the stress-strain curves, such as the ones shown in Figure

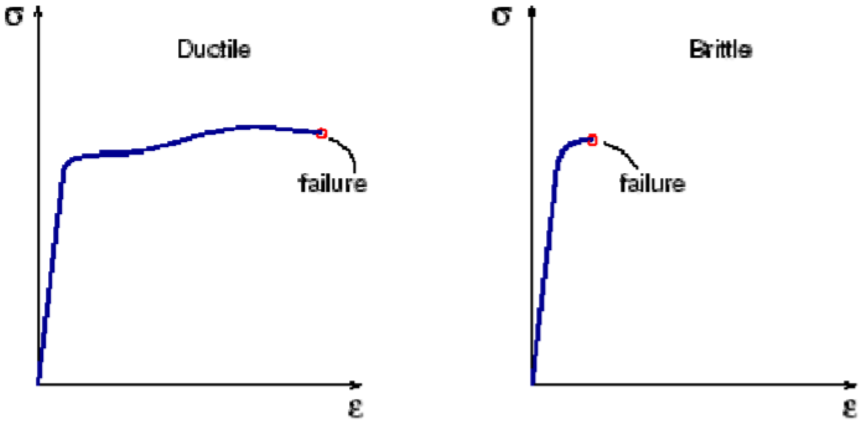


Fig.1.15. Ductile and brittle material behaviour

The material response for ductile and brittle materials are exhibited by both qualitative and quantitative differences in their respective stress-strain curves. Ductile materials will withstand large strains before the specimen ruptures; brittle materials fracture at much lower strains. The yielding region for ductile materials often takes up the majority of the stress-strain curve, whereas for brittle materials it is nearly non-existent. Brittle materials often have relatively large Young's moduli and ultimate stresses in comparison to ductile materials.

These differences are a major consideration for design. Ductile materials exhibit large strains and yielding before they fail. On the contrary, brittle materials fail suddenly and without much warning. Thus ductile materials such as steel are a natural choice for structural members in buildings as we desire considerable warning to be provided before a building fails. The energy absorbed (per unit volume) in the tensile test is simply the area under the stress strain curve. Clearly, by comparing the curves in Figure, It can be observed that ductile materials are capable of absorbing much larger quantities of energy before failure.

Finally, it should be emphasized that not all materials can be easily classified as either ductile or brittle. Material response also depends on the operating environment; many ductile materials become brittle as the temperature is decreased. With advances in metallurgy and composite technology, other materials are advanced combinations of ductile and brittle constituents.

Often in structural design, structural members are designed to be in service below the yield stress. The reason being that once the load exceeds the yield limit, the structural members will exhibit large deformations (imagine for instance a roof sagging) that are undesirable. Thus, materials with larger yield strength are preferable.

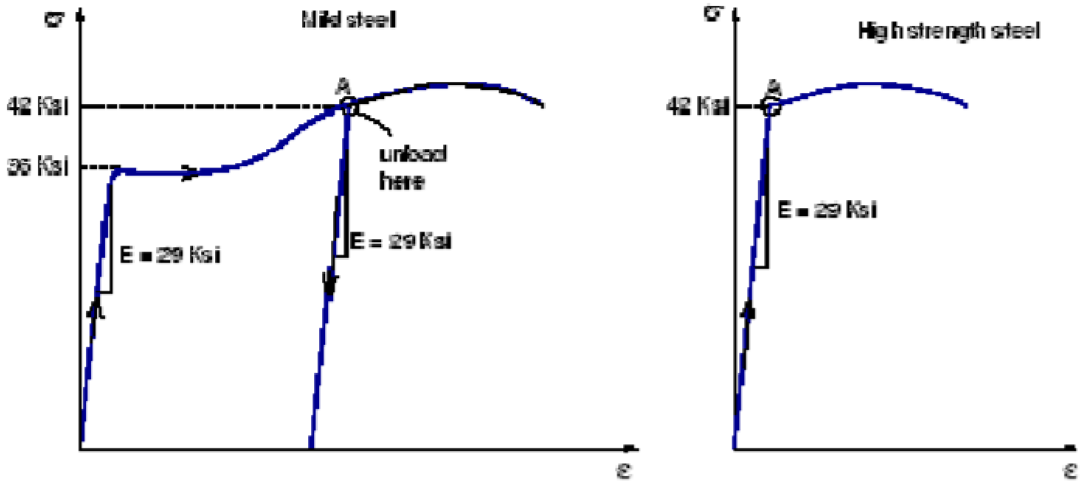


Fig. 1.16. Stress-strain curve of a mild steel (left) resembles that of high-strength steel (right).

Mild steels have a yield strength somewhere between 240 and 360 N/mm^2. When work hardened, the yield strength of this steel increases. Work hardening is the process of loading mild steel beyond its yield point and unloading as shown in Figure. When the material is loaded again, the linear elastic behaviour now extends up to point A as shown. The negative aspect of work hardening is some loss in ductility of the material. It is noteworthy that mild steel is usually recycled. Because of this, the yield strength may be a little higher than expected for the mild steel specimens tested in the laboratory. Often in structural design, structural members are designed to be in service below the yield stress. The reason being that once the load exceeds the yield limit, the structural members will exhibit large deformations (imagine for instance a roof sagging) that are undesirable. Thus materials with larger yield strength are preferable. Generally, the stress strain distribution varies from a material to another and could be in different forms as follows. Consequently, the type of material and fracture pattern can be defined and determined according to its stress-strain distribution diagram.

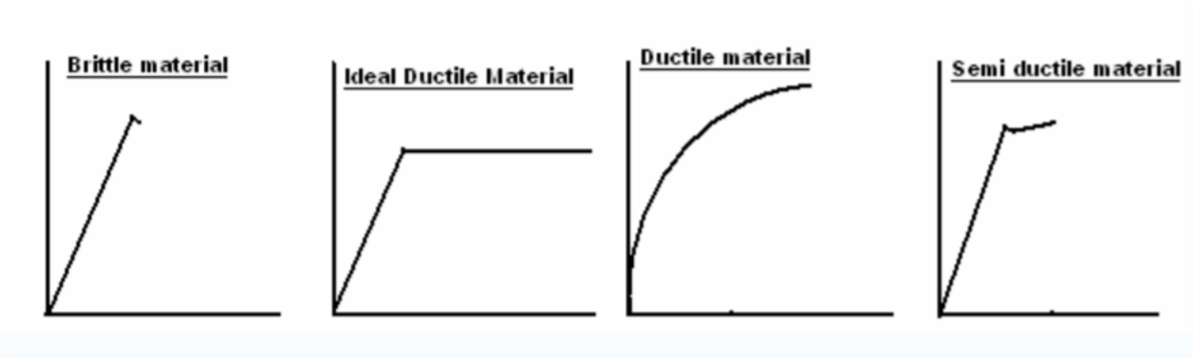


Fig. 1.17. Various stress-strain diagrams for different engineering materials

1.13 Yield strength

The yield point, is defined in engineering and materials science as the stress at which a material begins to plastically deform. Prior to the yield point the material will deform elastically and will return to its original shape when the applied stress is removed. Once the yield point is passed some fraction of the deformation will be permanent and non-reversible. Knowledge of the yield point is vital when designing a component since it generally represents an upper limit to the load that can be applied. It is also important for the control of many materials production techniques such as forging, rolling, or pressing. In structural engineering, yield is the permanent plastic deformation of a structural member under stress. This is a soft failure mode which does not normally cause catastrophic failure unless it accelerates buckling. It is often difficult to precisely define yield due to the wide variety of stress-strain behaviours exhibited by real materials. In addition there are several possible ways to define the yield point in a given material. Yield occurs when dislocations first begin to move. Given that dislocations begin to move at very low stresses, and the difficulty in detecting such movement, this definition is rarely used.

1.14 Elastic Limit

The lowest stress at which permanent deformation can be measured. This requires a complex iteractive load-unload procedure and is critically dependent on the accuracy of the equipment and the skill of the operator.

1.15 Proportional Limit

The point at which the stress-strain curve becomes nonlinear. In most metallic materials the elastic limit and proportional limit are essentially the same.

1.16 Offset Yield Point (proof stress)

Due to the lack of a clear border between the elastic and plastic regions in many materials, the yield point is often defined as the stress at some arbitrary plastic strain (typically 0.2%). This is determined by the intersection of a line offset from the linear region by the required strain. In some materials there is essentially no linear region and so a certain value of plastic strain is defined instead. Although somewhat arbitrary this method does allow for a consistent comparison of materials and is the most common.

1.17 Yield point.

If the stress is too large, the strain deviates from being proportional to the stress. The point at which this happens is the yield point because there the material yields, deforming permanently (plastically).

1.18 Yield stress.

Hooke's law is not valid beyond the yield point. The stress at the yield point is called yield stress, and is an important measure of the mechanical properties of materials. In practice, the yield stress is chosen as that causing a permanent strain of 0.002, which called as proof stress. The yield stress measures the resistance to plastic deformation.

The yield strength is the stress required to produce a small-specified amount of plastic deformation. The usual definition of this property is the offset yield strength determined by the stress corresponding to the intersection of the stress strain curve and a line parallel to the elastic part of the curve offset by a specified strain. In the United States the offset is usually specified as a strain of 0.2 or 0.1 percent ($e = 0.002$ or 0.001).

1.19 Elastic Properties of Materials (Deformation is reversible, non-permanent)

When the stress is removed, the material returns to the dimension it had before the load was applied. Valid for small strains (except the case of rubbers).

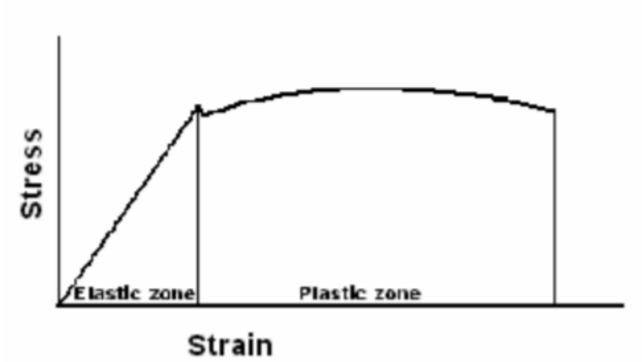


Fig. 1.18. zones in S-n Curves

Materials subject to tension shrink laterally. Those subject to compression, bulge. The ratio of lateral and axial strains is called the Poisson's ratio. When a material is placed under a tensile stress, an accompanying strain is created in the same direction.

Poisson's ratio is the ratio of the lateral to axial strains.

$$\nu = -\frac{\varepsilon_x}{\varepsilon_z} = \frac{\varepsilon_y}{\varepsilon_z}$$

The elastic modulus, shear modulus and Poisson's ratio are related by E = 2G(1+ ν)

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

- Theoretically, isotropic materials will have a value for Poisson's ratio of 0.25.
- The maximum value of ν is 0.5
- Most metals exhibit values between 0.25 and 0.35

1.20 Plastic deformation.

When the stress is removed, the material does not return to its previous dimension but there is a permanent, irreversible deformation. For metallic materials, elastic deformation only occurs to strains of about 0.005. After this point, plastic (non-recoverable) deformation occurs, and Hooke's Law is no longer valid. On an atomic level, plastic deformation is caused by slip, where atomic bonds are broken by dislocation motion, and new bonds are formed.

1.21 Anelasticity

Here the behavior is elastic but not the stress-strain curve is not immediately reversible. It takes a while for the strain to return to zero. The effect is normally small for metals but can be significant for polymers.

1.22 Tensile strength.

When stress continues in the plastic regime, the stress-strain passes through a maximum, called the tensile strength, and then falls as the material starts to develop a neck and it finally breaks at the fracture point. Note that it is called strength, not stress, but the units are the same, MPa. For structural applications, the yield stress is usually a more important property than the tensile strength, since once the it is passed, the structure has deformed beyond acceptable limits. The tensile strength, or ultimate tensile strength (UTS), is the maximum load divided by the original cross-sectional area of the specimen.

The tensile strength is the value most often quoted from the results of a tension test; yet in reality it is a value of little fundamental significance with regard to the strength of a metal. For ductile metals the tensile strength should be regarded as a measure of the maximum load, which a metal can withstand under the very restrictive conditions of uniaxial loading. It will be shown that this value bears little relation to the useful strength of the metal under the more complex conditions of stress, which are usually encountered.

Further, because the tensile strength is easy to determine and is a quite reproducible property, it is useful for the purposes of specifications and for quality control of a product. Extensive empirical correlations between tensile strength and properties such as hardness and fatigue strength are often quite useful. For brittle materials, the tensile strength is a valid criterion for design.

1.23 Ductility

The ability to deform before braking. It is the opposite of brittleness. Ductility can be given either as percent maximum elongation emax or maximum area reduction. At our present degree of understanding, ductility is a qualitative, subjective property of a material. In general, measurements of ductility are of interest in three ways:

1. To indicate the extent to which a metal can be deformed without fracture in metal working operations such as rolling and extrusion.
2. To indicate to the designer, in a general way, the ability of the metal to flow plastically before fracture. A high ductility indicates that the material is "forgiving" and likely to deform locally without fracture should the designer err in the stress calculation or the prediction of severe loads.

3. To serve as an indicator of changes in impurity level or processing conditions. Ductility measurements may be specified to assess material quality even though no direct relationship exists between the ductility measurement and performance in service.

The conventional measures of ductility that are obtained from the tension test are the engineering strain at fracture ef (usually called the elongation) and the reduction of area at fracture q. Both of these properties are obtained after fracture by putting the specimen back together and taking measurements of L_f and A_f .

1.24 Resilience

The resilience of the material is the triangular area underneath the elastic region of the curve. Resilience generally means the ability to recover from (or to resist being affected by) some shock, insult, or disturbance. However, it is used quite differently in different fields. In physics and engineering, resilience is defined as the capacity of a material to absorb energy when it is deformed elastically and then, upon unloading to have this energy recovered. In other words, it is the maximum energy per volume that can be elastically stored. It is represented by the area under the curve in the elastic region in the Stress-Strain diagram.

Modulus of Resilience, Ur , can be calculated using the following formula:

$$U_r = \frac{\sigma^2}{2E} = 0.5\sigma\epsilon = 0.5\sigma(\frac{\sigma}{E}),$$

where σ is yield stress, E is Young's modulus, and ε is strain

The ability of a material to absorb energy when deformed elastically and to return it when unloaded is called resilience. This is usually measured by the modulus of resilience, which is the strain energy per unit volume required to stress the material from, zero stress to the yield stress. The ability of a material to absorb energy when deformed elastically and to return it when unloaded is called resilience. This is usually measured by the modulus of resilience, which is the strain energy per unit volume required to stress the material from, zero stress to the yield stress s. The strain energy per unit volume for uniaxial tension is

$$U_0 = \frac{1}{2}\sigma_x e_x$$

1.25 Toughness

The area underneath the stress-strain curve is the toughness of the material- i.e. the energy the material can absorb prior to rupture. It also can be defined as the resistance of a material to crack propogation. In materials science and metallurgy, toughness is the resistance to fracture of a material when stressed. It is defined as the amount of energy that a material can absorb before rupturing, and can be found by finding the area (i.e., by taking the integral) underneath the stress-strain curve.

The ability of a metal to deform plastically and to absorb energy in the process before fracture is termed toughness. The emphasis of this definition should be placed on the ability to absorb energy before fracture. Recall that ductility is a measure of how much something deforms plastically before fracture, but just because a material is ductile does not make it tough. The key to toughness is a good combination of strength and ductility. A material with high strength and high ductility will have more toughness than a material with low strength and high ductility. Therefore, one way to measure toughness is by calculating the area under the stress strain curve from a tensile test. This value is simply called “material toughness” and it has units of energy per volume. Material toughness equates to a slow absorption of energy by the material.

The toughness of a material is its ability to absorb energy in the plastic range. The ability to withstand occasional, stresses above the yield stress without fracturing is particularly desirable in parts such as freight-car couplings, gears, chains, and crane hooks. Toughness is a commonly used concept, which is difficult to pin down and define. One way of looking at toughness is to consider that it is the total area under the stress-strain curve. This area is an indication of the amount of work per unit volume, which can be done, on the material without causing it to rupture. The following Figure shows the stress-strain curves for high- and low-toughness materials. The high-carbon spring steel has a higher yield strength and tensile strength than the medium-carbon structural steel. However, the structural steel is more ductile and has a greater total elongation. The total area under the stresstrain curve is greater for the structural steel, and therefore it is a tougher material. This illustrates that toughness is a parameter that comprises both strength and ductility. The crosshatched regions in Figure indicate the modulus of resilience for each steel. Because of its higher yield strength, the spring steel has the greater resilience.

Several mathematical approximations for the area under the stress-strain curve have been suggested. For ductile metals that have a stress-strain curve like that of the structural steel, the area under the curve can be approximated by either of the following equations:

$$U_T \approx S_u e_f$$

$$U_T \approx \frac{s_0 + s_u}{2} \cdot e_f$$

For brittle materials the stress-strain curve is sometimes assumed to be a parabola, and the area under the curve is given by

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

All these relations are only approximations to the area under the stress-strain curves. Further, the curves do not represent the true behaviour in the plastic range, since they are all based on the original area of the specimen.

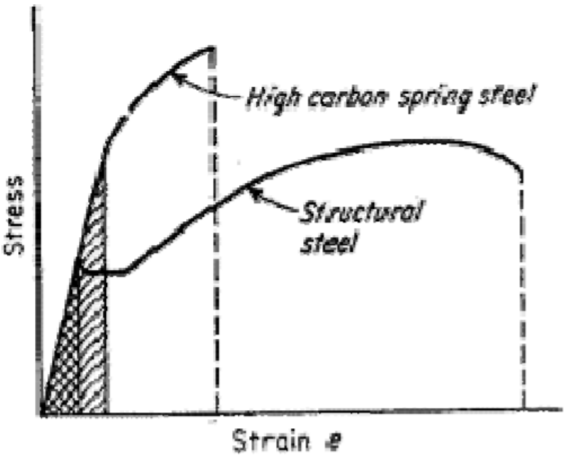


Fig. 1.18. Toughness of materials

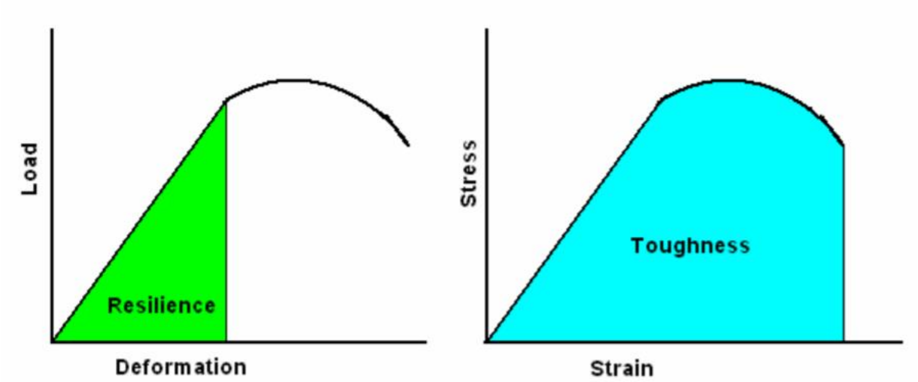


Fig. 1.19. Comparison between resilience and toughness of metals

1.26 Impact Toughness

The impact toughness (AKA Impact strength) of a material can be determined with a Charpy or Izod test. These tests are named after their inventors and were developed in the early 1900's before fracture mechanics theory was available. Impact properties are not directly used in fracture mechanics calculations, but the economical impact tests continue to be used as a quality control method to assess notch sensitivity and for comparing the relative toughness of engineering materials.

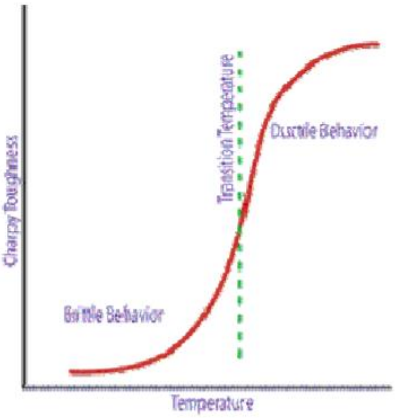


Fig. 1.20. Impact Toughness

The two tests use different specimens and methods of holding the specimens, but both tests make use of a pendulum-testing machine. For both tests, the specimen is broken by a single overload event due to the impact of the pendulum. A stop pointer is used to record how far the pendulum swings back up after fracturing the specimen. The impact toughness of a metal is determined by measuring the energy absorbed in the fracture of the specimen. This is simply obtained by noting the height at which the pendulum is released and the height to which the pendulum swings after it has struck the specimen. The height of the pendulum times the weight of the pendulum produces the potential energy and the difference in potential energy of the pendulum at the start and the end of the test is equal to the absorbed energy.

Since toughness is greatly affected by temperature, a Charpy or Izod test is often repeated numerous times with each specimen tested at a different temperature. This produces a graph of impact toughness for the material as a function of temperature. An impact toughness versus temperature graph for a steel is shown in the image. It can be seen that at low temperatures the material is more brittle and impact toughness is low. At high temperatures the material is more ductile and impact toughness is higher. The transition temperature is the boundary between brittle and ductile behavior and this temperature is often an extremely important consideration in the selection of a material.

1.27 Notch-Toughness

Notch toughness is the ability that a material possesses to absorb energy in the presence of a flaw. As mentioned previously, in the presence of a flaw, such as a notch or crack, a material will likely exhibit a lower level of toughness. When a flaw is present in a material, loading induces a triaxial tension stress state adjacent to the flaw. The material develops plastic strains as the yield stress is exceeded in the region near the crack tip. However, the amount of plastic deformation is restricted by the surrounding material, which remains elastic. When a material is prevented from deforming plastically, it fails in a brittle manner.

Notch-toughness is measured with a variety of specimens such as the Charpy Vnotch impact specimen or the dynamic tear test specimen. As with regular impact testing the tests are often repeated numerous times with specimens tested at a different temperature. With these specimens and by varying the loading speed and the temperature, it is possible to generate curves such as those shown in the graph. Typically only static and impact testing is conducted but it should be recognized that many components in service see intermediate loading rates in the range of the dashed red line.

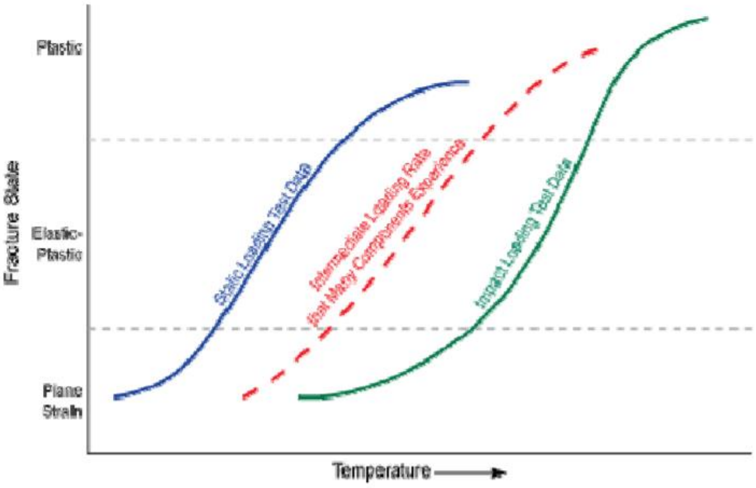


Fig. 1.20. Notch Toughness

1.28 Fracture Toughness

In materials science, fracture toughness is a property which describes the ability of a material containing a crack to resist fracture, and is one of the most important properties of any material for virtually all design applications.

Fracture toughness is a quantitative way of expressing a material's resistance to brittle fracture when a crack is present. If a material has a large value of fracture toughness it will probably undergo ductile fracture. Brittle fracture is very characteristic of materials with a low fracture toughness value.

Fracture mechanics, which leads to the concept of fracture toughness, was largely based on the work of A. A. Griffith who, amongst other things, studied the behaviour of cracks in brittle materials. Fracture toughness is an indication of the amount of stress required to propagate a preexisting flaw. It is a very important material property since the occurrence of flaws is not completely avoidable in the processing, fabrication, or service of a material/component. Flaws may appear as cracks, voids, metallurgical inclusions, weld defects, design discontinuities, or some combination thereof. Since engineers can never be totally sure that a material is flaw free, it is common practice to assume that a flaw of some chosen size will be present in some number of component This approach uses the flaw size and features, component geometry, loading conditions and the material property called fracture toughness to evaluate the ability of a component containing a flaw to resist fracture.

A parameter called the stress-intensity factor (K) is used to determine the fracture toughness of most materials. A Roman numeral subscript indicates the mode of fracture and the three modes of fracture are illustrated in the image to the right. Mode I fracture is the condition in which the crack plane is normal to the direction of largest tensile loading. This is the most commonly encountered mode and, therefore, for the remainder of the material we will consider K_I

The stress intensity factor is a function of loading, crack size, and structural geometry. The stress intensity factor may be represented by the following equation:

$$K_I = \sigma\sqrt{\pi a \beta}$$

Where:

K_I is the fracture toughness

σ is the applied stress in MPa or psi

a is the crack length in meters or inches

B is a crack length and component geometry factor that is different for each specimen and is dimensionless

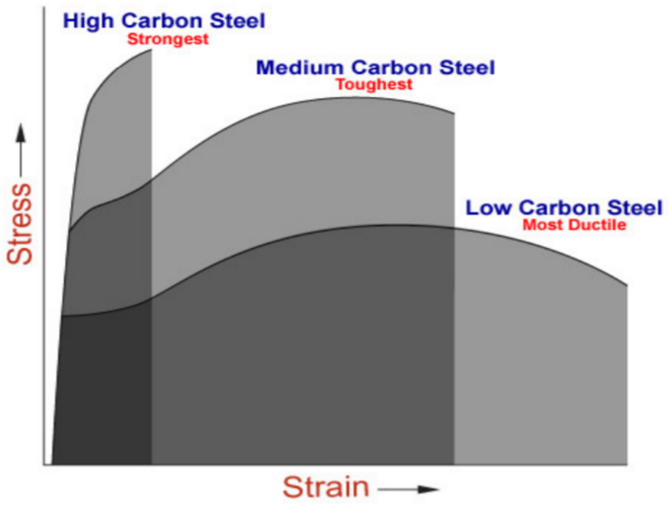


Fig. 1.21. Fracture Toughness

There are several variables that have a profound influence on the toughness of a material. These variables are:

- Strain rate (rate of loading)
- Temperature
- Notch effect

A metal may possess satisfactory toughness under static loads but may fail under dynamic loads or impact. As a rule ductility and, therefore, toughness decrease as the rate of loading increases. Temperature is the second variable to have a major influence on its toughness. As temperature is lowered, the ductility and toughness also decrease. The third variable is termed notch effect, has to due with the distribution of stress. A material might display good toughness when the applied stress is uniaxial; but when a multiaxial stress state is produced due to the presence of a notch, the material might not withstand the simultaneous elastic and plastic deformation in the various directions. There are several standard types of toughness test that generate data for specific loading conditions and/or component design approaches.

1.29 True Stress and Strain

When one applies a constant tensile force the material will break after reaching the tensile strength. The material starts necking (the transverse area decreases) but the stress cannot increase beyond σ_{TS}. The ratio of the force to the initial area, what we normally do, is called the engineering stress. If the ratio is to the actual area (that changes with stress) one obtains the true stress.

Stress has units of a force measure divided by the square of a length measure, and the average stress on a cross-section in the tensile test is the applied force divided by the cross-sectional area. Similarly, we may approximate the strain component along the long axis of the specimen as the change in length divided by the original reference length.

Engineering stress and strain measures are distinguished by the use of fixed reference quantities, typically the original cross-sectional area or original length. More precisely,

$$\sigma_E = \frac{P}{A_0}, \; \epsilon_E = \frac{\Delta l}{l_0}.$$

In most engineering applications, these definitions are accurate enough, because the cross-sectional area and length of the specimen do not change substantially while loads are applied. In other situations (such as the tensile test), the cross-sectional area and the length of the specimen can change substantially. In such cases, the engineering stress calculated using the above definition (as the ratio of the applied load to the undeformed cross-sectional area) ceases to be an accurate measure. To overcome this issue alternative stress and strain measures are available. Below we discuss true stress and true strain.

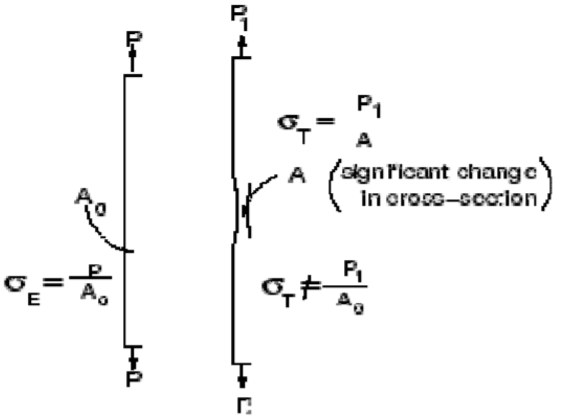


Fig. 1.22 Engineering stress measures vs. true stress measures.

True Stress: The true stress is defined as the ratio of the applied load (P) to the instantaneous cross-sectional area (A):

True stress can be related to the engineering stress if we assume that there is no volume change in the specimen. Under this assumption, stress can be related to the engineering stress if we assume that there is no volume change in the specimen.

$$A \cdot l = A_0 \cdot l_0,$$

Which leads to,

$$\sigma_T = \frac{P}{A} = \frac{P}{A_0} \cdot \frac{l}{l_0} = \sigma_E(1+\epsilon_E).$$

True Strain:

The true strain is defined as the sum of all the instantaneous engineering strains. Letting

$$d\epsilon = \frac{dl}{l},$$

the true strain is then,

$$\epsilon_T = \int d\epsilon = \int_{l_0}^{l_f} \frac{dl}{l} = \epsilon_T = \int d\epsilon = \int_{l_0}^{l_f} \frac{dl}{l} = \ln \frac{l_f}{l_0}. \frac{l_f}{l_0}.$$

In closing, you should note that the true stress and strain are practically indistinguishable from the engineering stress and strain at small deformations, as shown below in Figure 4. You should also note that as the strain becomes large and the cross-sectional area of the specimen decreases, the true stress can be much larger than the engineering stress.

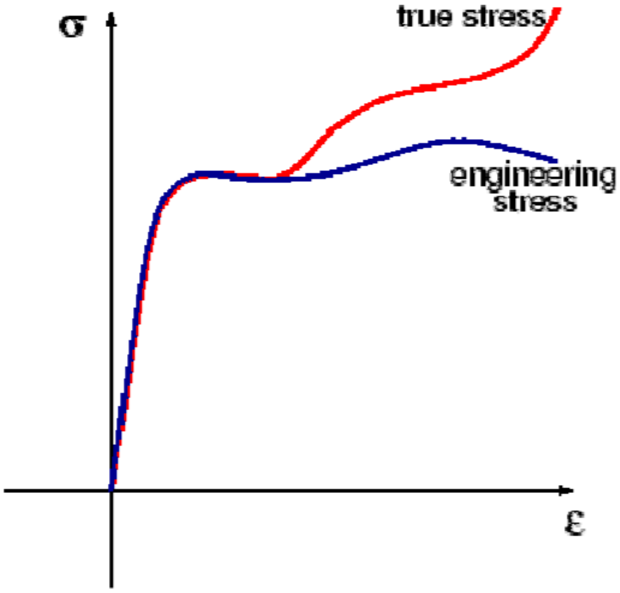


Fig. 1.23 Engineering stress-strain curve vs. a true stress-strain curve

1.3 Non-Destructive Testing

Introduction

Non-destructive Testing is one part of the function of Quality Control and is complementary to other long-established methods.

By definition non-destructive testing is the testing of materials, for surface or internal flaws or metallurgical condition, without interfering in any way with the integrity of the material or its suitability for service. The technique can be applied on a sampling basis for individual investigation or may be used for 100% checking of material in a production quality control system.

Whilst being a high technology concept, evolution of the equipment has made it robust enough for application in any industrial environment at any stage of manufacture - from steelmaking to site inspection of components already in service. A certain degree of skill is required to apply the techniques properly in order to obtain the maximum amount of information concerning the product, with consequent feed back to the production facility.

Non-destructive Testing is not just a method for rejecting substandard material; it is also an assurance that the supposedly good is good. The technique uses a variety of principles; there is no single method around which a black box may be built to satisfy all requirements in all circumstances.

What follows is a brief description of the methods most commonly used in industry, together with details of typical applications, functions and advantages. The methods covered are:

- Radiography
- Magnetic Particle Crack Detection
- Dye Penetrant Testing
- Ultrasonic Flaw Detection
- Eddy Current and Electro-magnetic Testing

However, these are by no means the total of the principles available to the N.D.T. Engineer. Electrical potential drop, sonics, infra-red, acoustic emission and spectrography, to name but a few, have been used to provide information that the above techniques have been unable to yield, and development across the board continues.

1.3.1 Radiography - X And Gamma

This technique is suitable for the detection of internal defects in ferrous and nonferrous metals and other materials.

X-rays, generated electrically, and Gamma rays emitted from radio-active isotopes, are penetrating radiation which is differentially absorbed by the material through which it passes; the greater the thickness, the greater the absorbtion. Furthermore, the denser the material the greater the absorbtion. X and Gamma rays also have the property, like light, of partially converting silver halide crystals in a photographic film to metallic silver, in proportion to the intensity of the radiation reaching the film, and therefore forming a latent image. This can be developed and fixed in a similar way to normal photographic film. Material with internal voids is tested by placing the subject between the source of radiation and the film. The voids show as darkened areas, where more radiation has reached the film, on a clear background. The principles are the same for both X and Gamma radiography.

In X-radiography the penetrating power is determined by the number of volts applied to the X-Ray tube - in steel approximately 1000 volts per inch thickness is necessary. In Gamma radiography the isotope governs the penetrating power and is unalterable in each isotope. Thus Iridium 192 is used for 1/2" to 1" steel and Caesium 134 is used for 3/4" to 21/2" steel.

In X-radiography the intensity, and therefore the exposure time, is governed by the amperage of the cathode in the tube. Exposure time is usually expressed in terms of milliampere minutes. With Gamma rays the intensity of the radiation is set at the time of supply of the isotope. The intensity of radiation from isotopes is measured in Becquerel’s and reduces over a period of time. The time taken to decay to half the amount of curies is the half life and is characteristic of each isotope. For example, the half life of Iridium 192 is 74 days, and Caesium 134 is 2.1 years. The exposure factor is a product of the number of curies and time, usually expressed in curie hours.

The time of exposure must be increased as the isotope decays - when the exposure period becomes uneconomical the isotope must be renewed. As the isotope is continuously emitting radiation it must be housed in a container of deleted uranium or similar dense shielding material, whilst not exposed to protect the environment and personnel.

To produce an X or Gamma radiograph, the film package (comprising film and intensifying screens - the latter being required to reduce the exposure time – enclosed in a light tight cassette) is placed close to the surface of the subject. The source of radiation is positioned on the other side of the subject some distance away, so that the radiation passes through the subject and on to the film. After the exposure period the film is removed, processed, dried, and then viewed by transmitted light on a special viewer.

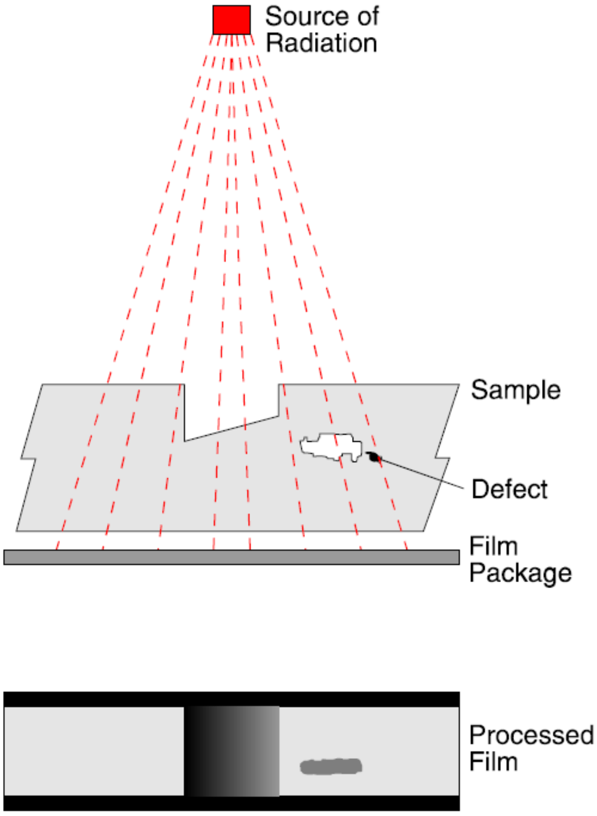


Fig. 1.24 Radiography

Various radiographic and photographic accessories are necessary, including such items as radiation monitors, film markers, image quality indicators, darkroom equipment, etc. Where the last is concerned there are many degrees of sophistication, including fully automatic processing units. These accessories are the same for both X and Gamma radiography systems. Also required are such consumable items as radiographic film and processing chemicals.

Advantages of Radiography

- Information is presented pictorially.
- A permanent record is provided which may be viewed at a time and place
- distant from the test.
- Useful for thin sections.
- Sensitivity declared on each film.

- Suitable for any material.

Disadvantages of Radiography

- Generally an inability to cope with thick sections.
- Possible health hazard.
- Need to direct the beam accurately for two-dimensional defects.
- Film processing and viewing facilities are necessary, as is an exposure
- compound.
- Not suitable for automation, unless the system incorporates fluoroscopy with
- an image intensifier or other electronic aids
- Not suitable for surface defects.
- No indication of depth of a defect below the surface

1.3.2 Magnetic Particle Inspection

This method is suitable for the detection of surface and near surface discontinuities in magnetic material, mainly ferritic steel and iron.

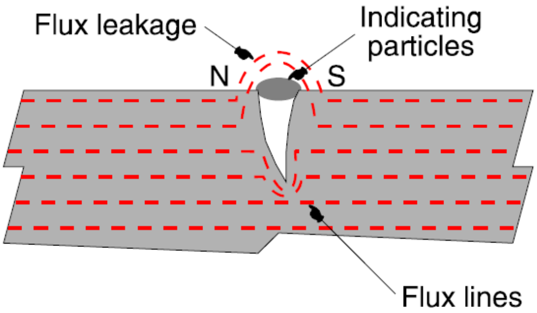


Fig. 1.24 Illustration of the Principle of Magnetic Particle Inspection

The principle is to generate magnetic flux in the article to be examined, with the flux lines running along the surface at right angles to the suspected defect. Where the flux lines approach a discontinuity, they will stray out in to the air at the mouth of the crack. The crack edge becomes magnetic attractive poles North and South. These have the power to attract finely divided particles of magnetic material such as iron fillings. Usually, these particles are of an oxide of iron in the size range 20 to 30 microns, and are suspended in a liquid which provides mobility for the particles on the surface of the test piece, assisting their migration to the crack edges. However, in some instances they can be applied in a dry powder form.

The particles can be red or black oxide, or they can be coated with a substance, which fluoresces brilliantly under ultra-violet illumination (black light). The object is to present as great a contrast as possible between the crack indication and the material background. The technique not only detects those defects which are not normally visible to the unaided eye, but also renders easily visible those defects which would otherwise require close scrutiny of the surface.

There are many methods of generating magnetic flux in the test piece, the simplest one being the application of a permanent magnet to the surface, but this method cannot be controlled accurately because of indifferent surface contact and deterioration in magnetic strength. Modern equipments generate the magnetic field electrically either directly or indirectly.

In the direct method a high amperage current is passed through the subject and magnetic flux is generated at right angles to the current flow. Therefore, the current flow should be in the same line as the suspected defect. If it is not possible to carry out this method because of the orientation of the defect, then the indirect method must be used. This can be one of two forms:

1. Passing a high current through a coil that encircles the subject.
2. Making the test piece form part of a yoke, which is wound with a current carrying coil. The effect is to pass magnetic flux along the part to reveal transverse and circumferential defects.

If a bar with a length much greater than its diameter is considered, then longitudinal defects would be detected by current flow and transverse and circumferential defects by the indirect method of an encircling coil or magnetic flux flow.

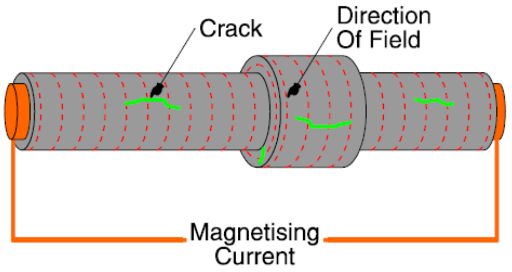
Subjects in which cracks radiating from a hole are suspected can be tested by means of the threading bar technique, whereby a current carrying conductor is passed through the hole and the field induced is cut by any defects. Detection of longitudinal defects in hollow shafts is a typical application of the threader bar technique.

The electricity used to generate the magnetic flux in any of these methods can be alternating current, half wave rectified direct current or full wave rectified direct current. A.C. generated magnetic flux, because of the skin effect, preferentially follows the contours of the surface and does not penetrate deeply into the material. H.W.D.C. penetrates more deeply but is inclined not to follow sharp changes in section. H.W.D.C. is useful for the detection of slightly subsurface defects. The pulsing effect of A.C. and H.W.D.C. gives additional mobility to the indicating particles. D.C. penetrates even more deeply but does not have this facility. Furthermore, demagnetising of the material after D.C. magnetising is far more difficult than after A.C. magnetising.

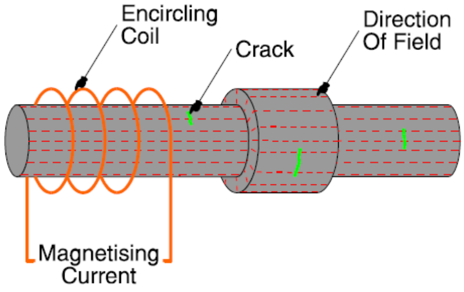
Normally, to ensure that a test piece has no cracks, it is necessary to magnetise it in at least two directions and after each magnetising - and ink application – visually examine the piece for crack indications.

Since this double process, which would include adjustment of the magnetising equipment controls in between each magnetising takes time it is obviously advantageous to have the facility to reduce the time required. The recent development of the Swinging Field method of multi-directional magnetising will indicate all defects, regardless of their orientation on the surface, with one magnetising shot and therefore requires only one inspection.

Basically magnetic crack detection equipment takes two forms. Firstly, for test pieces which are part of a large structure, or pipes, heavy castings, etc. which cannot be moved easily, the equipment takes the form of just a power pack to generate a high current. This current is applied to the subject either by contact prods on flexible cables or by an encircling coil of cable. These power packs can have variable amperages up to a maximum of 2000 Amps for portable units, and up to 10,000 Amps for mobile equipments. Both A.C. and H.W.D.C. magnetising current is available. The indicating material is applied by means of a spray and generally the surplus runs to waste.



Schematic arrangement for detecting longitudinal defects using current flow through the subject. The source of the current can be a portable / mobile power pack or, in the case of a bench unit, from the built-in power pack.



Schematic arrangement of an encircling coil for the detection of circumferential and transverse defects. The coil may be cable wrapped round loosely or wound on a former, as in a bench unit.

Fig. 1.25 Process of Magnetic Particle Inspection

Advantages

- Quantitative.
- Can be automated, apart from viewing. (Though modern developments in automatic defect recognition can be used in parts of simple geometry e.g. billets and bars. In this case a special camera captures the defect indication image and processes it for further display and action)

Disadvantages

- Restricted to ferromagnetic materials.
- Restricted to surface or near surface flaws.
- Not fail safe in that lack of indication could mean no defects or process not carried out properly.

1.3.3 Dye Penetrant Testing

This method is frequently used for the detection of surface breaking flaws in nonferromagnetic materials.

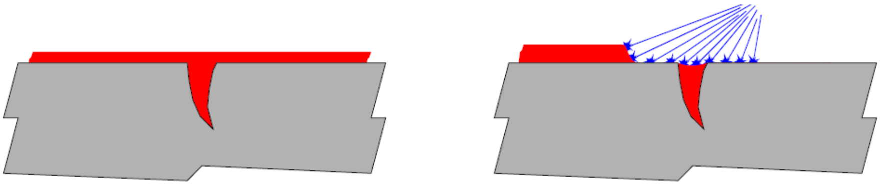
The subject to be examined is first of all chemically cleaned, usually by vapour phase, to remove all traces of foreign material, grease, dirt, etc. from the surface generally, and also from within the cracks.

Next the penetrant (which is a very fine thin oil usually dyed bright red or ultra-violet fluorescent) is applied and allowed to remain in contact with the surface for approximately fifteen minutes. Capillary action draws the penetrant into the crack during this period.

The surplus penetrant on the surface is then removed completely and thin coating of powdered chalk is applied.

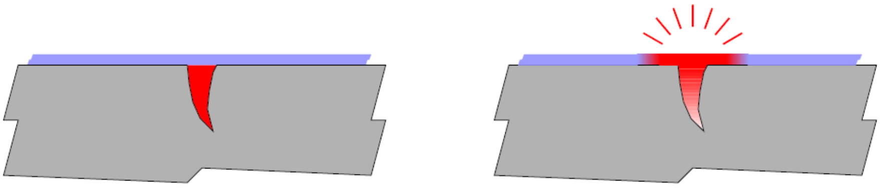
After a further period (development time) the chalk draws the dye out of the crack, rather like blotting paper, to form a visual, magnified in width, indication in good contrast to the background.

The process is purely a mechanical/chemical one and the various substances used may be applied in a large variety of ways, from aerosol spray cans at the simplest end to dipping in large tanks on an automatic basis at the other end. The latter system requires sophisticated tanks, spraying and drying equipment but the principle remains the same.



A. Penetrant applied to the surface and enters defect

B. Excess penetrant removed from surface



C. Developer powder applied to draw penetrant out of crack.

D. Accentuated indication of crack as penetrant spreads around the opening.

Fig. 1.26 Illustration of the Dye Penetrant Testing

Advantages of Dye Penetrant Testing

- Simplicity of operation.
- Best method for surface breaking cracks in non-ferrous metals.
- Suitable for automatic testing, with reservation concerning viewing
- Quantative.

Disadvantages of Dye Penetrant Testing

- Restricted to surface breaking defects only.
- Decreased sensitivity.
- Uses a considerable amount of consumables.

1.3.4 Ultrasonic Flaw Detection

This technique is used for the detection of internal and surface (particularly distant surface) defects in sound conducting materials.

The principle is in some respects similar to echo sounding. A short pulse of ultrasound is generated by means of an electric charge applied to a piezo electric crystal, which vibrates for a very short period at a frequency related to the thickness of the crystal. In flaw detection this frequency is usually in the range of one million to six million times per second (1 MHz to 6 MHz). Vibrations or sound waves at this frequency have the ability to travel a considerable distance in homogeneous elastic material, such as many metals with little attenuation. The velocity at which these waves propagate is related to the Young's

Modulus for the material and is characteristic of that material. For example the velocity in steel is 5900 metres per second, and in water 1400 metres per second.

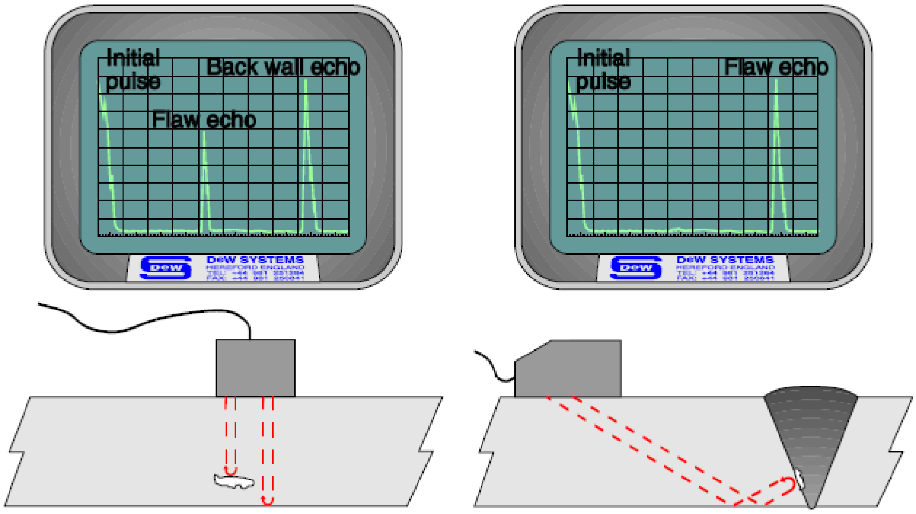
Ultrasonic energy is considerably attenuated in air, and a beam propagated through a solid will, on reaching an interface (e.g. a defect, or intended hole, or the backwall) between that material and air reflect a considerable amount of energy in the direction equal to the angle of incidence.

For contact testing the oscillating crystal is incorporated in a hand-held probe, which is applied to the surface of the material to be tested. To facilitate the transfer of energy across the small air gap between the crystal and the test piece, a layer of liquid (referred to as ‘couplant’), usually oil, water or grease, is applied to the surface. As mentioned previously, the crystal does not oscillate continuously but in short pulses, between each of which it is quiescent. Piezo electric materials not only convert electrical pulses to mechanical oscillations, but will also transduce mechanical oscillations into electrical pulses; thus we have not only a generator of sound waves but also a detector of returned pulses. The crystal is in a state to detect returned pulses when it is quiescent. The pulse takes a finite time to travel through the material to the interface and to be reflected back to the probe.

The standard method of presenting information in ultrasonic testing is by means of a cathode ray tube, in which horizontal movement of the spot from left to right represents time elapsed. The principle is not greatly different in digitised instruments that have a LCD flat screen. The rate at which the spot moves is such that it gives the appearance of a horizontal line on the screen. The system is synchronised electronically so that at the instant the probe receives its electrical pulse the spot begins to traverse the screen. An upward deflection (peak) of the line on the left-hand side of the screen is an indication of this occurrence. This peak is usually termed the initial pulse.

Whilst the base line is perfectly level the crystal is quiescent. Any peaks to the right of the initial pulse indicate that the crystal has received an incoming pulse reflected from one or more interfaces in the material. Since the spot moves at a very even speed across the tube face, and the pulse of ultrasonic waves moves at a very even velocity through the material, it is possible to calibrate the horizontal line on the screen in terms of absolute measurement. The use of a calibration block, which produces a reflection from the back wall a known distance away from the crystal together with variable controls on the flaw detector, allows the screen to be calibrated in units of distance, and therefore determination of origins of returned pulses obtained from a test piece.

It is therefore possible not only to discover a defect between the surface and the back wall, but also to measure its distance below the surface. It is important that the equipment is properly calibrated and, since it is in itself not able to discriminate between intended boundaries of the object under test and unintended discontinuities, the operator must be able to identify the origin of each peak. Further as the pulses form a beam it is also possible to determine the plan position of a flaw. The height of the peak (echo) is roughly proportional to the area of the reflector, though there is on all instruments a control, which can reduce or increase the size of an indication - variable sensitivity in fact. Not only is party of the beam reflected at a material/air interface but also at any junction where there is a velocity change, for example steel/slag interface in a weld.



Schematic diagram of ultrasonic detection of slag in steel section using a normal probe.

Schematic diagram of the use of an angle probe to detect defects not directly under the probe. Such as in weld inspection.

Fig. 1.27 Illustration of the Ultrasonic Flaw Detection

Probing all faces of a test piece not only discovers the three-dimensional defect and measures its depth, but can also determine its size. Two-dimensional (planar) defects can also be found but, unlike radiography, it is best that the incident beam impinges on the defect as near to right angles to the plane as possible. To achieve this some probes introduce the beam at an angle to the surface. In this manner longitudinal defects in tubes (inner or outer surface) are detected. Interpretation of the indications on the screen requires a certain amount of skill, particularly when testing with hand held probes. The technique is, however, admirably suited to automatic testing of regular shapes by means of a monitor – an electronic device that fits into the main equipment to provide an electrical signal when an echo occurs in a particular position on the trace. The trigger level of this signal is variable and it can be made to operate a variety of mechanical gates and flaw warnings. Furthermore, improvements in computer technology allow test data and results to be displayed and out-putted in a wide variety of formats.

Modern ultrasonic flaw detectors are fully solid state and can be battery powered, and are robustly built to withstand site conditions. Since the velocity of sound in any material is characteristic of that material, it follows that some materials can be identified by the determination of the velocity. This can be applied, for example in S.G. cast irons to determine the percentage of graphite nodularity.

Advantages

- Thickness and lengths up to 30 ft can be tested.
- Position, size and type of defect can be determined.
- Instant test results.
- Portable.
- Extremely sensitive if required.
- Capable of being fully automated.
- Access to only one side necessary.
- No consumables.

Disadvantages

- No permanent record available unless one of the more sophisticated test results and data collection systems is used.
- The operator can decide whether the test piece is defective or not whilst the test is in progress.
- Indications require interpretation (except for digital wall thickness gauges).
- Considerable degree of skill necessary to obtain the fullest information from the test.
- Very thin sections can prove difficult.

1.3.5 Eddy Current and Electro-Magnetic Methods

The main applications of the eddy current technique are for the detection of surface or subsurface flaws, conductivity measurement and coating thickness measurement. The technique is sensitive to the material conductivity, permeability and dimensions of a product.

Eddy currents can be produced in any electrically conducting material that is subjected to an alternating magnetic field (typically 10Hz to 10MHz). The alternating magnetic field is normally generated by passing an alternating current through a coil. The coil can have many shapes and can between 10 and 500 turns of wire. The magnitude of the eddy currents generated in the product is dependent on conductivity, permeability and the set up geometry. Any change in the material or geometry can be detected by the excitation coil as a change in the coil impedance. The simplest coil comprises a ferrite rod with several turns of wire wound at one end and which is positioned close to the surface of the product to be tested. When a crack, for example, occurs in the product surface the eddy currents must travel farther around the crack and this is detected by the impedance change.

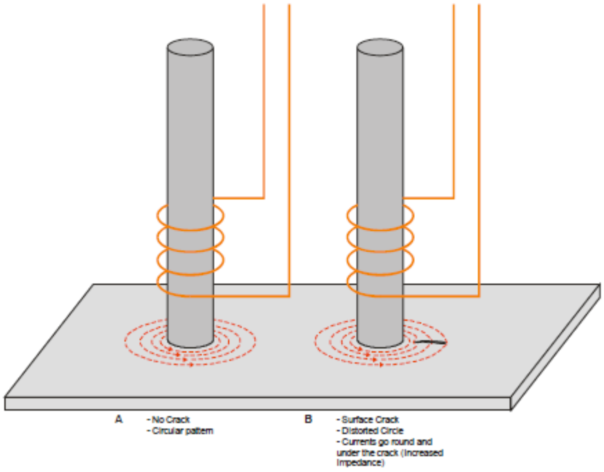


Fig. 1.28 Coil with single winding

Coils can also be used in pairs, generally called a driven pair, and this arrangement can be used with the coils connected differentially. In this way ‘lift off’ (distance of the probe from the surface) signals can be enhanced.

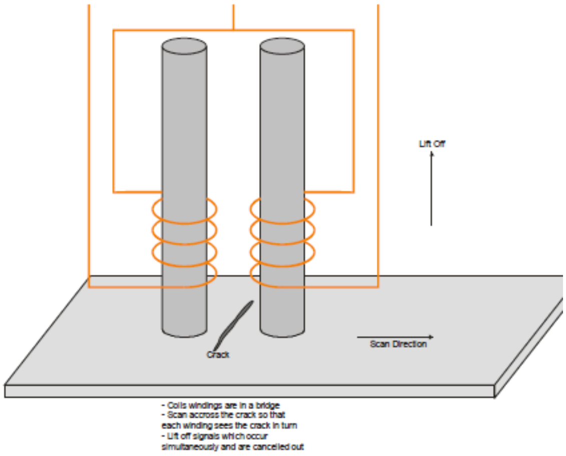


Fig. 1.29 Coil with two windings, known as a driver pair or differential probe

Coils can also be used in a transformer type configuration where one coil winding is a primary and one (or two) coil windings are used for the secondaries.

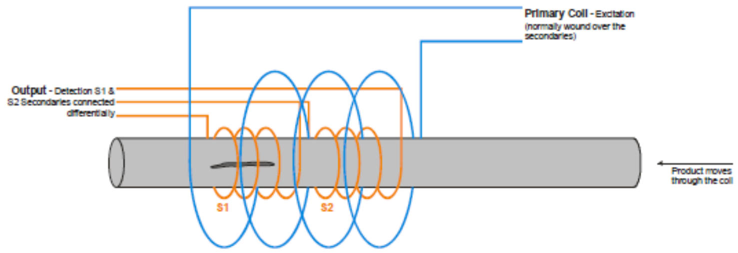


Fig. 1.30 Transformer type coil with 3 windings

The detected eddy current signals contain amplitude and phase information and which can be displayed on CRT type displays – non digital displays. Signals can be displayed as the actual, i.e. absolute signal, or with appropriate electronics, only a signal change is displayed. The best results are obtained where only one product parameter is changes, e.g. the presence of a crack. In practice changes in eddy current signals are caused by differences in composition, hardness, texture, shape, conductivity, permeability and geometry. In some cases the effects of the crack can be hidden by changes in other parameters and unnecessary rejection can occur. However, the coils can be selected for configuration, size and test frequency in order to enhance detection of cracks, conductivity, metal loss etc. as required.

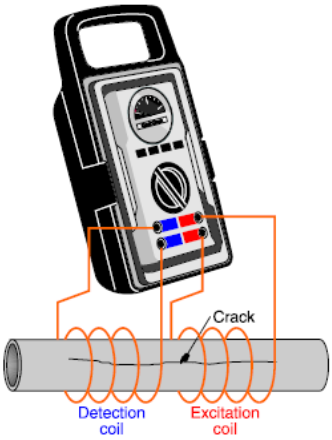
The depth to which the eddy currents penetrate a material can be changed by adjusting the test frequency – the higher the frequency, the lower the penetration; however, the lower the frequency, the lower sensitivity to small defects. Larger coils are less sensitive to surface roughness and vice versa. The latest electronic units are able to operate a wide range of coil configurations in absolute or differential modes and at a wide range of frequencies.

For surface testing for cracks in single or complex shaped components, coils with a single ferrite cored winding are normally used. The probe is placed on the component and ‘balanced’ by use of the electronic unit controls. As the probe is scanned across the surface of the component the cracks can be detected.

Where surfaces are to be scanned automatically the single coil windings are suitable only if the lift off distance is accurately maintained. Generally differential coil configurations are used with higher speed scanning systems where lift off effects, vibration effects, etc. can be cancelled out to an acceptable extent.

Tubes, bar and wire can be inspected using an encircling coil and these usually have a coil configuration with one primary and two secondaries connected differentially. Most eddy current electronics have a phase display and this gives an operator the ability to identify defect conditions. In many cases signals from cracks, lift off and other parameters can be clearly identified. Units are also available which can inspect a product simultaneously at two or more different test frequencies. These units allow specific unwanted effects to be electronically cancelled in order to give improved defect detection.

The eddy current test is purely electrical. The coil units do not need to contact the product surface and thus the technique can be easily automated. Most automated systems are for components of simple geometry where mechanical handling is simplified.



Schematic illustration of eddy current testing of drawn tube for longitudinal defects.

Fig. 1.31 Illustration of Eddy Current Testing Equipment

Advantages

- Suitable for the determination of a wide range of conditions of conducting material, such as defect detection, composition, hardness, conductivity, permeability etc. in a wide variety of engineering metals.
- Information can be provided in simple terms: often go/no go. Phase displayelectronic units can be used to obtain much greater product information.
- Extremely compact and portable units are available.
- No consumables (except probes – which can sometimes be repaired).
- Flexibility in selection of probes and test frequencies to suit different applications.
- Suitable for total automation.

Disadvantages

- The wide range of parameters which affect the eddy current responses means that the signal from a desired material characteristic, e.g. a crack, can be masked by an unwanted parameter, e.g. hardness change. Careful selection of probe and electronics will be needed in some applications.
- Generally tests restricted to surface breaking conditions and slightly subsurface flaws.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF MECHANICAL ENGINEERING

DEPARTMENT OF AERONAUTICAL ENGINEERING

UNIT – II - MECHANICAL BEHAVIOUR OF AIRCRAFT MATERIALS

– SAE1616

UNIT 2 SELECTION OF MATERIALS FOR AIRCRAFT AND ROCKETS

Aluminum alloy

An aluminum alloy is a chemical composition where other elements are added to pure aluminum in order to enhance its properties, primarily to increase its strength. These other elements include iron, silicon, copper, magnesium, manganese and zinc at levels that combined may make up as much as 15 percent of the alloy by weight. Alloying requires the thorough mixing of aluminum with these other elements while the aluminum is in molten – liquid – form.

Cast aluminium alloys

Cast aluminium contains larger percentages of alloying elements when compared to wrought aluminium. Cast aluminium also has a generally lower tensile strength than wrought aluminium due to the difficulty in eliminating casting defects.

Properties of cast aluminium

Cast aluminium alloys make use of a four-digit numbering system and include a decimal point after the third digit. This system was developed and maintained by the Aluminium Association and is the most widely used naming convention for aluminium alloys. The aluminium Association works in accordance with ANSI regulations. The first digit will indicate the principal alloying elements and is thus the most important.

Cast alloy production

Cast aluminium alloy is produced from bauxite. This is a naturally occurring mineral containing 15-20 % aluminium and is the only ore still used for commercial aluminium extraction. The process to extract pure aluminium from bauxite is highly complex and energy-intensive.

The process consists of dissolving the bauxite into caustic soda at high temperatures, once dissolved the temperature of the mixture is lowered and the alumina crystallizes while the remaining elements either settle out or recrystallise separately. This is known as the Bayer process. The alumina is then broken down in an electrolytic cell whereby an electric current separates the aluminium from the oxygen in the presence of molten cryolite. The aluminium separates out to the bottom of the cell and is removed on a regular basis and sent to the cast house where impurities are separated out. The pure aluminium is cast into billets for further processing, these billets are then smelted along with the necessary alloying elements to produce the desired grade. This smelted alloy is then either cast into billet form or into the final product shape. The casting can be done via sand casting, die casting or investment casting.

Cast alloy applications

Cast aluminium alloys are not often used for structural components due to their comparatively low tensile strength. This can be overcome by specialised processing techniques but, in general, cast alloys are used for the following applications:

- Machine tools
- Engine cylinder heads
- Gearbox housings
- Axle housings
- Cast wheels
- Window fittings
- Farm equipment
- Garden tools

Cast alloy advantages

Cast aluminium alloys offer various advantages compared to wrought alloys as indicated in the list below:

- Lower price per kilogram when compared to wrought aluminium

- Wide range of shapes can be achieved due to casting flexibility
- Some specialised alloys are only available as castings due to their low ductility
- Parts can be produced that require limited post-cast machining

Wrought aluminium

Wrought aluminium has exceptional mechanical properties and can be formed into various standard and nonstandard shapes.

Wrought aluminium properties

A wrought aluminium alloy can be identified by a four-digit number. The first digit will indicate the principal alloying elements and is thus the most important. The second digit, unless it is a 0, indicates a modification of the alloy and the third and fourth digits are identification numbers for the specific alloy.

Wrought alloy production

Wrought aluminium is produced by smelting pure aluminium ingots with the specific alloying elements required to make a given grade of aluminium. The smelted alloy is then cast into billets or large slabs. This material is then either rolled, forged or extruded into its final shape. In some cases, the alloys are heat treated to further enhance their properties.

Wrought alloy applications

Wrought aluminium tends to have better tensile strength when compared to cast alloys, as can be seen in the two tables above. Their typical uses are listed below:

- Extrusions
- Electrical conductors & bus bars
- Aircraft Airframes
- Cooking utensils
- Welding rods
- Pressure vessels

- Motorcycle frames
- Sheeting

Wrought alloy advantages

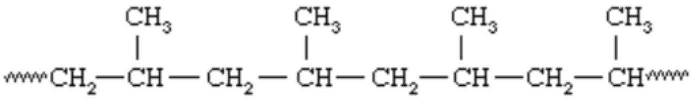
Wrought aluminium alloys offer numerous advantages when compared to cast aluminium as indicated in the list below;

- Excellent mechanical properties
- Structural integrity i.e. no defects from casting
- Better surface finishes
- Ease of manufacture for example welding and machining
- Ease of forming, for example, aluminium can be extruded into an almost endless range of cross sections that can be custom designed to suit a specific application

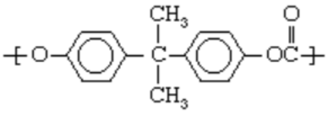
plastics

plastic, polymeric material that has the capability of being molded or shaped, usually by the application of heat and pressure. This property of plasticity, often found in combination with other special properties such as low density, low electrical conductivity, transparency, and toughness, allows plastics to be made into a great variety of products. These include tough and lightweight beverage bottles made of polyethylene terephthalate (PET), flexible garden hoses made of polyvinyl chloride (PVC), insulating food containers made of foamed polystyrene, and shatterproof windows made of polymethyl methacrylate.

Plastics also can be divided into two distinct categories on the basis of their chemical composition. One category is plastics that are made up of polymers having only aliphatic (linear) carbon atoms in their backbone chains. All the commodity plastics listed above fall into this category. The structure of polypropylene can serve as an example; here attached to every other carbon atom is a pendant methyl group (CH_3):



The other category of plastics is made up of heterochain polymers. These compounds contain atoms such as oxygen, nitrogen, or sulfur in their backbone chains, in addition to carbon. Most of the engineering plastics listed above are composed of heterochain polymers. An example would be polycarbonate, whose molecules contain two aromatic (benzene) rings:



plastics are primarily defined not on the basis of their chemical composition but on the basis of their engineering behaviour. More specifically, they are defined as either thermoplastic resins or thermosetting resins.

Thermoplastics such as polyethylene and polystyrene are capable of being molded and remolded repeatedly. Thus, a foamed-polystyrene cup can be heated and reshaped into a new form—for instance, a dish. The polymer structure associated with thermoplastics is that of individual molecules that are separate from one another and flow past one another. The molecules may have low or extremely high molecular weight, and they may be branched or linear in structure, but the essential feature is that of separability and consequent mobility.

Thermosets, on the other hand, cannot be reprocessed upon reheating. During their initial processing, thermosetting resins undergo a chemical reaction that results in an infusible, insoluble network. Essentially, the entire heated, finished article becomes one large molecule. For example, the epoxy polymer used in making a fibre-reinforced laminate for a golf club undergoes a cross-linking reaction when it is molded at a high temperature. Subsequent application of heat does not soften the material to the point where it can be reworked and indeed may serve only to break it down.

Rubber:

Rubber is an elastic material that can be produced naturally from various plant sources or synthetically through a variety of chemical processes. It has been in use for thousands of years, during which time it has been produced in numerous variations with distinct characteristics that make them suitable for different applications.

It serves as an essential raw material in the production of everything from automobile tires to surgical gloves. However, successfully producing these components necessitates choosing the right type of rubber for the given part design and application environment. For this reason, we’ve provided an overview of some of the most commonly available rubber types, outlining what they are, their key properties, and their typical uses.

Basic Properties of Rubber

As indicated above, rubber comes in several variations, each of which has unique properties. However, most—if not all—rubbers also share a few common characteristics, such as:

- **Elasticity:** The molecular structure of rubber materials allows them to return to their normal shape after being compressed or stretched. This characteristic is apparent in rubber bands. Stretching or compressing a rubber band temporarily pulls or pushes the individual molecules out of alignment with each other. As the molecules are attached to one another, they return to their original position once the stretching or compressing force is removed.
- **Thermal contraction:** While most materials expand when heated, rubber contracts. This unusual phenomenon results from how the rubber molecules respond to heat. When heat is applied, the already-tangled molecules become more tangled and twisted. When heat is removed, the molecules return to their resting state and the rubber regains its original shape.

- **Durability:** Most rubbers are highly durable, resisting damage and degradation from abrasive and tearing forces, impact, low temperatures, and water. They also exhibit a relatively slow heat buildup rate.

Types of Rubber

Each type of rubber material—whether natural or rubber—demonstrates distinct properties that make it suitable for certain applications. Some of the most common types of rubber and their properties include:

Natural Rubber

Natural rubber—also known as India or gum rubber—is derived from the milky liquid (i.e., latex) present in the *Hevea brasiliensis* tree. Some of the material's key characteristics are high tear and tensile strength, resilience, and resistance to abrasion, friction, extreme temperatures, and water swell. Typical uses include adhesives, flooring and roofing, gloves, insulation, and tires.

Neoprene Rubber

Neoprene rubber—also called chloroprene—is one of the oldest types of synthetic rubber. Compared with natural rubber and other synthetic rubbers, it exhibits an exceptionally low susceptibility to burning, corrosion, and degradation. This quality makes it the perfect base material for adhesives and corrosion-resistant coatings. Its ability to maintain good mechanical properties over a wide temperature range also suits it for use in high-pressure gaskets, belts, and window and door seals.

Silicone Rubber

Silicone rubber—also referred to as polysiloxane—is known for its malleability, biocompatibility, and resistance to extreme temperatures, fire, ozone, and ultraviolet (UV) radiation. It is available in both solid and liquid forms in a variety of colors. Its chemically inert nature makes it ideal for use in parts and products that demand biocompatibility (such as gloves, respiratory masks, implants, and other medical products) and chemical resistance (such as baby care items, cosmetic applicators, and food containers and instruments).

Nitrile Rubber

Nitrile rubber—also known as Buna-N rubber or nitrile butadiene rubber (NBR)—demonstrates several desirable mechanical and chemical properties, such as resistance to compression set, heat, oil and gas, and wear. These traits make it suitable for use in automotive gaskets and seals, O-rings, and engine hoses. It is also used in medical products (e.g., surgical gloves) since it lacks the allergenic proteins of latex-based rubbers and maintains its structural integrity better than silicone rubber.

EPDM Rubber

Ethylene propylene diene monomer (EPDM) rubber is a synthetic rubber that exhibits excellent durability, resisting damage and degradation from extreme temperatures and weather conditions. These qualities make it suitable for use in outdoor parts and products, such as roofing sealants, hoses, and seals. Its superior noise and thermal insulation properties also suit it for use in automobile systems.

Styrene-Butadiene Rubber (SBR)

Styrene-butadiene rubber (SBR) is a copolymer of styrene and butadiene characterized by its superior hardness and durability. It demonstrates better abrasion, alcohol, compression set, and water swell resistance than some of the more expensive synthetic rubbers, which makes it ideal for use in seals integrated into hydraulic brake systems. Other common uses include cutting boards, gaskets, and shoe soles.

Butyl Rubber

Butyl rubber—also known as isobutylene isoprene—offers one of the highest levels of gas impermeability. This quality, combined with the material's excellent flexibility, makes it suitable for the manufacture of airtight components, such as inner tubes, sport balls, and sealants. As a liquid compound, it is also often used as an additive in diesel and petroleum fuels and chewing gums.

Fluorosilicone Rubber

Fluorosilicone rubber—also called FVMQ—is highly resistant to extreme temperatures (-100–350 degrees Fahrenheit), transmission fluids, petroleum oils and fuels, synthetic lubricants, fire, and ozone. These properties make it an ideal material for aircraft fuel systems and other highly specialized industrial applications.

Fiber Reinforced Composites/Fibre Reinforced Polymer (FRP) Composites

Fibers are the important class of reinforcements, as they satisfy the desired conditions and transfer strength to the matrix constituent influencing and enhancing their properties as desired. Glass fibers are the earliest known fibers used to reinforce materials. Ceramic and metal fibers were subsequently found out and put to extensive use, to render composites stiffer more resistant to heat. Fibers fall short of ideal performance due to several factors. The performance of a fiber composite is judged by its length, shape, orientation, and composition of the fibers and the mechanical properties of the matrix. The orientation of the fiber in the matrix is an indication of the strength of the composite and the strength is greatest along the longitudinal directional of fiber. This doesn't mean the longitudinal fibers can take the same quantum of load irrespective of the direction in which it is applied. Optimum performance from longitudinal fibers can be obtained if the load is applied along its direction. The slightest shift in the angle of loading may drastically reduce the strength of the composite. Unidirectional loading is found in few structures and hence it is prudent to give a mix of orientations for fibers in composites particularly where the load is expected to be the heaviest. Monolayer tapes consisting of continuous or discontinuous fibers can be oriented unidirectional stacked into plies containing layers of filaments also oriented in the same direction. More complicated orientations are possible too and nowadays, computers are used to make projections of such variations to suit specific needs. In short, in planar composites, strength can be changed from unidirectional fiber oriented composites that result in composites with nearly isotropic properties. Properties of angle-plied composites which are not quasi-isotropic may vary with the number of plies and their orientations. Composite variables in such composites are assumed to have a constant ratio and the matrices are considered relatively weaker than the fibers. The strength of the fiber in any one of the three axes would, therefore be one-third the unidirectional fiber composite, assuming that the volume percentage is equal in all three axes. However, orientation of short fibers by different methods is also possible like random orientations by sprinkling on to given plane or addition of matrix in liquid or solid state before or after the fiber deposition. Even three-dimensional orientations can achieve in this way. There are several methods of random fiber orientations, which in a two-dimensional one, yield composites with one-third the strength of a unidirectional fiber-stressed composite, in the direction of fibers. In a 3-dimension, it would result in a composite with a comparable ratio, about less than one-fifth. In very strong matrices, moduli and strengths have not been observed. Application of the strength of the composites with such matrices and several orientations is also possible. The longitudinal strength can be calculated on the basis of the assumption that fibers have been reduced to their effective strength on approximation value in composites with strong matrices and non-longitudinally orientated fibers. It goes without saying that fiber composites may be constructed with either continuous or short fibers. Experience has shown that continuous fibers (or filaments) exhibit better orientation, although it does not reflect in their performance. Fibers have a high aspect ratio, i.e., their lengths being several times greater than their effective diameters. This is the reason why filaments are manufactured using continuous process. This finished filaments. Mass production of filaments is well known and they match with several matrices in different ways like winding, twisting, weaving and knitting, which exhibit the characteristics of a fabric. Since they have low densities and high strengths, the fiber lengths in filaments or other fibers yield considerable influence on the mechanical properties as well as the response of composites to processing and procedures. Shorter fibers with proper orientation composites that use glass, ceramic or multi-purpose fibers can be endowed with considerably higher strength than those that use continuous fibers. Short fibers are also known to their theoretical strength. The continuous fiber constituent of a composite is often joined by the filament winding process in which the matrix impregnated fiber wrapped around a mandrel shaped like the part over which the composite is to be placed, and equitable load distribution and favorable orientation of the fiber is possible in the finished product. However, winding is mostly confined to fabrication of bodies of revolution and the occasional irregular, flat surface. Short-length fibers incorporated by the open- or close-mould process are found to be less efficient, although the input costs are considerably lower than filament winding. Most fibers in use currently are solids which are easy to produce and handle, having a circular cross-section, although a few non-conventional shaped and hollow fibers show signs of capabilities that can improve the mechanical qualities of the composites. Given the fact that the vast difference in length and effective diameter of the fiber are assets to a fiber composite, it follows that greater strength in the fiber can be achieved by smaller diameters due to minimization or total elimination of surface of surface defects. After flat-thin filaments came into vogue, fibers rectangular cross sections have provided new options for applications in high strength structures. Owing to their shapes, these fibers provide perfect packing, while hollow fibers show better structural efficiency in composites that are desired for their stiffness and compressive strengths. In hollow fibers, the transverse compressive strength is lower than that of a solid fiber composite whenever the hollow portion is more than half the total fiber diameter. However, they are not easy to handle and fabricate.

Types of Fiber Reinforcement

The fiber reinforcement provides the structural performance required of the final part. The fibers or filaments come in many chemical types and forms and are the primary contributor to the stiffness, strength and other properties of the composite. The dominant chemical types of commercially available fibers are: fiberglass, aramid, carbon, polyester and vectran. Other fiber types may be suitable for special applications. The dominant forms that fibers are sold include: Strands (or roving or tow or yarn. Many fibers or filaments stranded together in a bundle, wound in a spool or reel,) woven fabrics (flattened strands of filaments woven in a variety of weaves to a type of fabric or cloth,) unidirectional (strands laid side by side and stitched or held together by other means, forming a kind of fabric that bares reinforcement only in the fill direction,) multiaxials (unidirectional woven fabrics stitched together in a combination of orientations,) and chopped strand mat (chopped strands held together with some kind of glue or “binder” in the form of a non-woven fabric.) All fibers designated for use in composites undergo chemical treatments and are coated with some kind of “sizing”. Sizing is a chemical that binds the filaments together, reduces abrasion, facilitates impregnation and acts like a coupling agent that enhances compatibility with one or more types of resin. Selection of reinforcement type greatly influences costs. This is not only because of the quality and cost of the material itself, but also because higher cost material usually requires highly skilled personnel, more sophisticated production facilities and often more labour hours.

Fiberglass is an attractive, lightweight (when compared to non composite products), durable material. When fiberglass is impregnated with epoxy resin, the resulting composite has high tensile strength and is stronger than steel on an equal weight basis. This high strength for its relatively low weight is the primary reason why fiberglass composites became so popular, and are widely used for many decades now. Glass fiber can be combined with other fiber material such as Kevlar/Aramid creating a hybrid fabric. This way the properties of the composing fibers are enhanced by synergy.

Fiberglass is made of natural minerals and some manufactured chemicals. The basic raw materials are silica sand, limestone and soda ash. Fiberglass is often made of recycled glass as well. The raw materials are mixed together, after being carefully weighted in the exact quantities, in a process that is called batching, and then melt into glass. Following thereafter, the molten glass is transferred to the forming equipment, passes through superfine holes, where very thin filaments are created. Fiberglass is the material of choice for making relatively inexpensive shaped shell structures, like large covers with round corners or complex shapes, as it offers flexibility in design. Manufacturing with fiberglass is cost-effective and affordable not only because it saves maintenance costs due to its durability, but also because the final product is lightweight, which results in less storage and shipping costs.

Fiberglass reinforcements are classified according to their properties. There are six major types of glass used to make fibers.

1. E-glass is the most common type used for fiberglass production today (more than 50% of the fibers made are from E-glass.) ‘E’ implies that it is an electrical insulator. It is inexpensive and appropriate for general purposes.

2. S-glass, S2-glass, (‘S’ comes from Strength,) is (15%-25%) stronger than E-glass, has higher modulus, improved mechanical properties, higher melt temperature and is considerably more expensive.
3. C-glass or T-glass is very resistant to chemicals and corrosion.
4. A-glass is a high-alkali glass. It offers good chemical resistance, but has lower electrical properties.
5. D-glass has a low dielectric constant and is used in circuit boards.
6. AR-glass is resistant to alkali environment.

Fiberglass’ main properties are:

- Dimensional stability
- Moisture resistance
- High strength
- Fire and heat resistance
- Chemical resistance
- Electrical insulation
- Freedom of design
- Good acoustic and vibration insulation
- Good fatigue resistance
- Very good resin adhesion

Fiberglass is widely used in applications such as:

- Automotive: Body parts for trucks and buses, roofs, air fairings, side fairings, doors.
- Aerospace: Interior panels, wall, separators, toilet compartments and furniture.
- Marine: Boat hulls, decks, internal and external components.
- Windmill blades.

- Military – defense: Protective equipment e.g. armor, helmets, clothing, vehicles, radomes.
- Construction: Structural reinforcement, roofing laminate, canopies, swimming pools and spas, hot tubs.
- Electrical and telecommunications industry: printed circuit board, insulation components.
- Sports: Surfboards, kayaks, beach rackets, bicycles, masts.
- Tubes, pipes and other profiles.
- Machinery and equipment enclosures.
- Storage tanks.

Carbon fibers

Carbon fibers have high tensile strength, and are very light and very stable. Carbon crystals are bonded together in a chain, creating a very strong material, which compared to steel is 5 times stronger on an equal weight basis. Carbon fiber diameter is very small, it ranges from 5-10 microns. Production and consumption of carbon fibers has grown recently because of their great mechanical properties. High manufacturing cost is balanced by its high strength in both tension and compression, and high resistance to corrosion, creep and fatigue, low weight and high performance.

Woven Carbon fabric is used in various applications like marine, sporting goods, defense and many others. The two most common weaving styles are «plain» and «twill». Both have an equal amount of carbon fiber going each direction, and their strengths are quite similar. Other styles are satin, unidirectional, biaxial. Alternatively, fabrics can be added to a resin system (like epoxy) that hardens, and this way structural composite parts are formed. Since resin systems are strong low density materials, the composite part is also very strong and light-weight at the same time.

The dominant raw material for carbon fiber manufacturing is polyacrylonitrile (PAN), pitch follows, and a very small amount of carbon fibers are derived from rayon.

Carbon fibers are usually grouped according to the modulus or strength band in which their properties belong. These bands are commonly referred as: high strength, intermediate modulus, high modulus and ultra high modulus etc. These references for carbon fiber quality are not very clear, as different companies that produce different qualities may consider or evaluate certain quality differently. PAN fiber density ranges from 1.75gr/cm3 to 1.90gr/cm3. PAN tensile strength can be as high as 1000Ksi.

Carbon was first invented near Cliveland, Ohio in 1958, but this process was inefficient as only 20% of the final fiber was carbon. Later another process using PAN as precursor was developed, with a resulting fiber of 55% carbon. But with the advent of a new manufacturing process in 1963, developed at a UK research center, carbon fiber’s high potential strength was realized, and the carbon manufacturing industry began to grow. In the 1970s research on alternative raw materials led to carbon fibers containing 85% of carbon, this time made from pitch.

Modern carbon fiber content is above 90% and is approaching 100%. The leading companies that make the highest qualities in our days are from Japan, and have many carbon yarn manufacturing sites around the world. In the early years, carbon composites have been very expensive, so were rarely used, and most exclusively for aerospace applications. However, as time passed, carbon fiber became affordable for more applications, manufacturing techniques have improved, and all this resulted in increased (and growing) consumption.

The atomic structure of carbon fiber is similar to that of graphite, consisting of carbon atom layers arranged in a hexagonal pattern. Depending on precursors and manufacturing process, the layers might be turbostratic, graphitic or in a hybrid structure. In graphitic structure the sheets are stacked parallel in a regular fashion. Bonding between planes is weak, giving graphite its soft characteristics. Carbon fibers made of PAN are turbostratic and can provide higher strength, while pitch can provide higher modulus.

PAN carbon fiber is produced by pyrolysis of a precursor fiber in an inert atmosphere at temperatures above 982°C.

The most common precursor material used to manufacture carbon is polyacrylonitrile (or PAN), that is 90% of all carbon fiber production. The process consists of the five following steps:

- Spinning (and Polymerization): PAN is mixed with other ingredients and spun into fibers, which are washed and stretched. The quality of finished fiber depends a lot on the quality of the precursor.
- Stabilizing (or Oxidation): Chemical alteration to stabilize bonding at about 200°C-400°C.
- Carbonizing: Stabilized fibers are heated to very high temperature (~1000°C) to remove hydrogen, oxygen, nitrogen and other non carbon elements, forming tightly bonded carbon crystals. In order to manufacture stiffer (high modulus) fibers, the pyrolizing process continues for longer, and at higher temperature, (up to 3000°C,) creating a smaller diameter round fiber. Higher modulus fibers are more expensive and more brittle. For this reason, they should be further processed (when weaving, winding, etc) with greater care.
- Surface treatment: Surface of oxidized fibers is treated to improve bonding properties.
- Sizing: Fibers are coated and wound onto bobbins.

Over other materials, carbon fiber offers many advantages. The main advantageous characteristics are:

- High strength
- Light in weight
- Corrosion resistance
- Excellent creep resistance
- Good thermal and electrical conductivity
- Compatible with most resin systems
- Very high dimensional stability
- Low thermal expansion coefficient
- X-ray permeability

Resins in FRP

Plastic resins are the formable matrix within the composite. Resins, broadly defined, encompass a large class of synthetic products that have some of the physical properties of natural resins but are different chemically and are used chiefly in the manufacturing of plastics, fiberglass, and other composites.

Virtually all of the physical properties (chemical, electrical, and thermal) are strongly influenced by the resin. Most resins used in composites are the thermosetting type – once formed they cannot be reformed, and they do not melt or flow. Thermoplastics are also used in lower temperature applications. Furthermore, the use of organic or inorganic additives may be used to gain functionality of the finished FRP product.

Resin Types

- Polyester
- Vinyl Ester
- Epoxy

- Polyurethane
- Pehnolic

Polyester Resins are thermoset polyesters. They are versatile, offer good dimensional stability, and have good mechanical, chemical-resistance, and electrical properties.

Vinyl Ester Resins are stronger than polyesters and more resilient than epoxy resins. They are flexible (double bonded vinyl group) in nature and are useful when creating products that are designed to withstand flexing, impacts or compression.

Epoxy Resins have an extended range of properties when compared to polyester and vinyl ester resins. They demonstrate extremely low shrinkage, good dimensional stability, high temperature resistance, as well as good fatigue and adherence to reinforcements. In addition, they have excellent resistance to basic (alkali) environments/solutions.

Polyurethane Resins are known throughout the fiberglass industry for their durability and robustness. They are flexible in nature and can be used in a wide variety of applications.

Pehnolic Resins posses many desirable attributes in the fiberglass world. They offer formability to complex contours, as well as flexibility. They are heat and chemical resistant and demonstrate flame retardance. They are ideal for high temperature applications where parts/components must meet fire safety, smoke emission, and combustion and toxicity requirements. In addition, they also have electrical non-conductivity characteristics.

Superalloys

Superalloys are complex, high-performance alloys, which have a high tolerance of oxidising environments and high temperatures. They are typically classified according to their predominant matrix element; nickel, cobalt, or iron, and they contain multiple alloying elements including the refractory metals (Nb, Mo, W, Ta), chromium, and titanium. They exhibit high mechanical strength, creep resistance and corrosion resistance, especially at high temperatures. These properties make them more challenging to produce and costlier than other alloys, but they are also critical for components in industries such as aerospace.

Properties of superalloys

Superalloys are intended for use in high-temperature applications, which means they need to maintain their shape at elevated temperatures close to their melting points (above 650 °C or 1200°F). When alloyed with certain elements, at extreme temperatures superalloys can maintain high strength, stability, and corrosion and oxidation resistance.

Superalloys are classified into three main categories :

Nickel-based superalloys

- High strength
- High thermal resistance
- High corrosion resistance
- Machinability
- Shape memory
- Low coefficient of thermal expansion

Cobalt-based superalloys

- Higher melting point compared to nickel- or iron-based alloys
- Superior hot corrosion resistance compared to nickel- or iron-based alloys
- Higher thermal fatigue resistance and weldability compared to nickel-based alloys

Iron-based superalloys

- High strength at room temperature
- High resistance to creep, oxidation, corrosion and wear

Of these three categories, nickel-based alloys have the widest range of applications, particularly in the aerospace industry. The essential solutes in the nickel-based alloys are aluminium (Al) and titanium, with concentrations of less than 10 wt. %. This allows the generation of a two-phase equilibrium microstructure that consists of the phases known as gamma (γ) and gamma-prime (γ’). The matrix of superalloys is composed of the γ-phase, while their primary hardening is a result of the γ’-phase. The high-temperature strength, as well as other mechanical properties of superalloys, are also a result of the presence of the γ’-phase.

Examples of superalloys

The high-temperature properties of superalloys are provided by alloying the matrix element (Ni, Co or Fe) with various other elements such as chromium (Cr), titanium (Ti), aluminium (Al), boron (B), and iron (Fe). In some cases, refractory metals are added, such as molybdenum (Mo), cobalt (Co), niobium (Nb), zirconium (Zr), amongst others.

Superalloys have many applications. These mainly include aircraft components, chemical plant equipment and petrochemical equipment.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF MECHANICAL ENGINEERING

DEPARTMENT OF AERONAUTICAL ENGINEERING

UNIT – III - MECHANICAL BEHAVIOUR OF AIRCRAFT MATERIALS

– SAE1616

UNIT 3 Heat Treatment & Corrosion.

Heat treatment

Heating and cooling a solid metal or alloy in a way so as to obtain specific conditions and/or properties

Objective of Heat treatment

- To increase strength, hardness and wear resistance (bulk hardening, surface hardening)
- To increase ductility and softness (Tempering, Recrystallization Annealing)
- To increase toughness (Tempering, Recrystallization annealing)
- To obtain fine grain size (Recrystallization annealing, Full annealing, Normalizing)
- To remove internal stresses induced by differential deformation by cold working, nonuniform
- cooling from high temperature during casting and welding (Stress relief annealing)
- To improve machinability (Full annealing and Normalizing)
- To improve cutting properties of tool steels (Hardening and Tempering)
- To improve surface properties (surface hardening, high temperature resistance-resistance precipitation hardening, surface treatment)
- To improve electrical properties (Recrystallization, Tempering, Age hardening)
- To improve magnetic properties (Hardening, Phase transformation)

Fundamentals of the Heat Treating of Steel

The common dictionary definition is “a hard, tough metal composed of iron, alloyed with various small percentage of carbon and often variously with other metals such as nickel, chromium, manganese, etc.” Although this definition is not untrue, it is hardly adequate. In the glossary of this book (see Appendix A, “Glossary of Heat-Treating Terms”) the principal portion of the definition for steel is “an iron base alloy, malleable in some temperature range as initially cast, containing manganese, usually carbon, and often other alloying elements. In carbon steel and low-alloy steel, the maximum carbon is about 2.0%; in high-alloy steel, about 2.5%. The dividing line between low-alloy and high-alloy steels is generally regarded as being at about 5% metallic alloying elements”. Fundamentally, all steels are mixtures, or more properly, alloys of iron and carbon. However, even the so-called plain-carbon steels have small, but specified, amounts of manganese and silicon plus small and generally unavoidable amounts of phosphorus and sulphur. The carbon content of plain-carbon steels may be as high as 2.0%, but such an alloy is rarely found. Carbon content of commercial steels usually ranges from 0.05 to about 1.0%. The alloying mechanism for iron and carbon is different from the more common and numerous other alloy systems in that the alloying of iron and carbon occurs as a two-step process. In the initial step, iron combines with 6.67% C, forming iron carbide, which is called cementite. Thus, at room temperature, conventional steels consist of a mixture of cementite and ferrite (essentially iron). Each of these is known as a phase (defined as a physically homogeneous and distinct portion of a material system). When a steel is heated above 725°C (1340 F), cementite dissolves in the matrix, and a new phase is formed, which is called austenite. Note that phases of steel should not be confused with structures. There are only three phases involved in any steel—ferrite, carbide (cementite), and austenite, whereas there are several structures or mixtures of structures.

Classification of Steels

It is impossible to determine the precise number of steel compositions and other variations that presently exist, although the total number probably exceeds 1000; thus, any rigid classification is impossible. However, steels are arbitrarily divided into five groups, which has proved generally satisfactory to the metalworking community. These five classes are:

- Carbon steels

- Alloy steels (sometimes referred to as low-alloy steels)
- Stainless steels
- Tools steels
- Special-purpose steels

The first four of these groups are well defined by designation systems developed by the Society of Automotive Engineers (SAE) and the American Iron and Steel Institute (AISI). Each general class is subdivided into numerous groups, with each grade indentified. The fifth group comprises several hundred different compositions; most of them are proprietary. Many of these special steels are similar to specific steels in the first four groups but vary sufficiently to be marked as separate compositions. For example, the SAE-AISI designation system lists nearly 60 stainless steels in four different general subdivisions. In addition to these steels (generally referred to as “standard grades”), there are well over 100 other compositions that are nonstandard. Each steel was developed for a specific application.

Metallurgical Phenomena

The broad possibilities provided by the use of steel are attributed mainly to two all-important metallurgical phenomena: iron is an allotropic element; that is, it can exist in more than one crystalline form; and the carbon atom is only 1 ⁄30 the size of the iron atom. These phenomena are thus the underlying principles that permit the achievements that are possible through heat treatment. In entering the following discussion of constitution, however, it must be emphasized that a maximum of technical description is unavoidable. This portion of the subject is inherently technical. To avoid that would result in the discussion becoming uninformative and generally useless. The purpose of this chapter is, therefore, to reduce the prominent technical features toward their broadest generalizations and to present those generalizations and underlying principles in a manner that should instruct the reader interested in the metallurgical principles of steel. This is done at the risk of some oversimplification.

Constitution of Iron

It should first be made clear to the reader that any mention of molten metal is purely academic; this book deals exclusively with the heat treating range that is well below the melting temperature. The objective of this section is to begin with a generalized discussion of the constitution of commercially pure iron, subsequently leading to discussion of the ironcarbon alloy system that is the basis for all steels and their heat treatment. All pure metals, as well as alloys, have individual constitutional or phase diagrams. As a rule, percentages of two principal elements are shown on the horizontal axis of a figure, while temperature variation is shown on the vertical axis. However, the constitutional diagram of a pure metal is a simple vertical line. The constitutional diagram for commercially pure iron is presented in Fig. 1. This specific diagram is a straight line as far as any changes are concerned, although time is indicated on the horizontal. As pure iron, in this case, cools, it changes from one phase to another at constant temperature. No attempt is made, however, to quantify time, but merely to indicate as a matter of interest that as temperature increases, reaction time decreases, which is true in almost any solidsolution reaction. Pure iron solidifies from the liquid at 1538 °C (2800 F) (top of Fig. 1). A crystalline structure, known as ferrite, or delta iron, is formed (point a, Fig. 1). This structure, in terms of atom arrangement, is known as a body-centered cubic lattice (bcc), shown in Fig. 2(a). This lattice has nine atoms—one at each corner and one in the center. As cooling proceeds further and point b (Fig. 1) is reached (1395 °C, or 2540 F), the atoms rearrange into a 14-atom lattice a shown in Fig. 2(b). The lattice now has an atom at each corner and one at the center of each face. This is known as a face-centered cubic lattice (fcc), and this structure is called gamma iron.

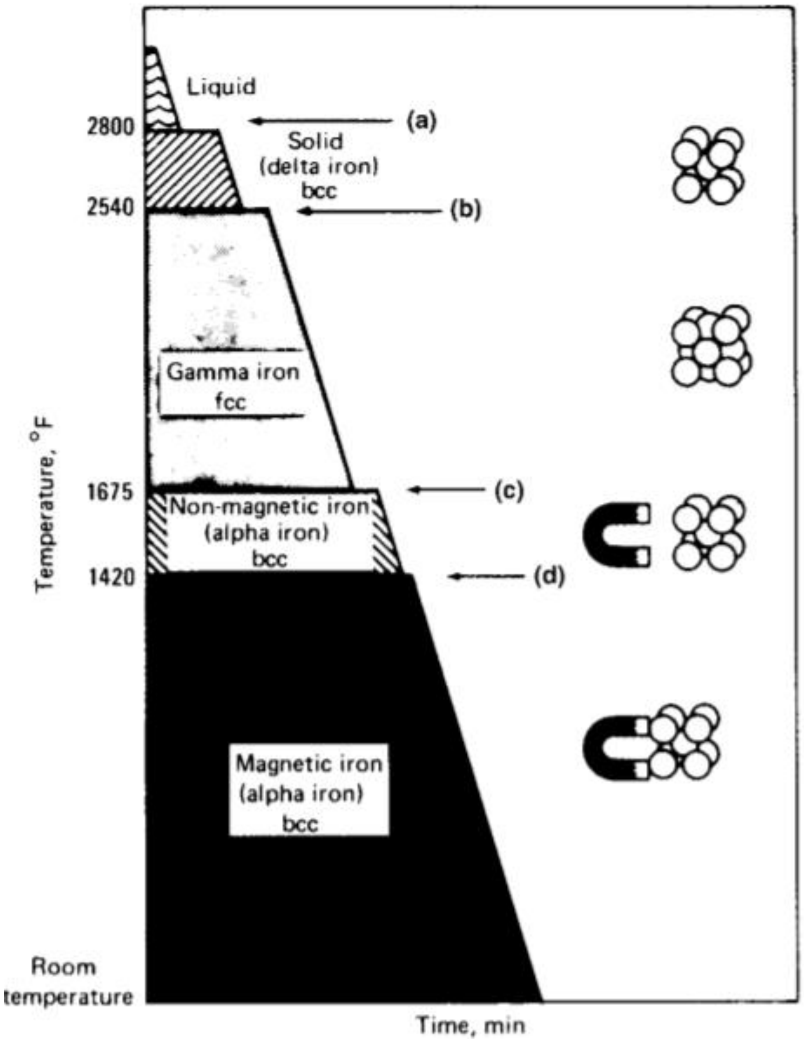


Fig. 1. Changes in pure iron as it cools from the molten state to room temperature.

As cooling further proceeds to 910 C (1675 F) (point c, Fig. 1), the structure reverts to the nine-atom lattice or alpha iron. The change at point d on Fig. 1 (770 C, or 1420 F) merely denotes a change from nonmagnetic to magnetic iron and does not represent a phase change. The entire field below 910 C (1675 F) is composed of alpha ferrite, which continues on down to room temperature and below. The ferrite forming above the temperature range of austenite is often referred to as delta ferrite; that forming below A3 as alpha ferrite, though both are structurally similar. In this Greek-letter sequence, austenite is gamma iron, and the interchangeability of these terms should not confuse the fact that only two structurally distinct forms of iron exist. Figures 1 and 2 thus illustrate the allotropy of iron. In the following sections of this chapter, the mechanism of allotropy as the all-important phenomenon relating to the heat treatment of iron-carbon alloys is discussed.

Alloying Mechanisms

Metal alloys are usually formed by mixing together two or more metals in their molten state. The two most common methods of alloying are by atom exchange and by the interstitial mechanism. The mechanism by which two metals alloy is greatly influenced by the relative atom size. The exchange mechanism simply involves trading of atoms from one lattice system to another. An example of alloying by exchange is the copper nickel system wherein atoms are exchanged back and forth.

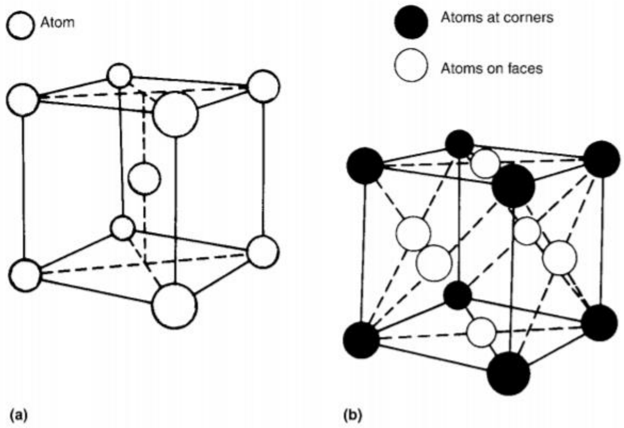


Fig. 2 Arrangement of atoms in the two crystalline structures of pure iron. (a) Body-centered cubic lattice. (b) Face-centered cubic lattice

Interstitial alloying requires that there be a large variation in atom sizes between the elements involved. Because the small carbon atom is 1⁄30 the size of the iron atom, interstitial alloying is easily facilitated. Under certain conditions, the tiny carbon atoms enter the lattice (the interstices) of the iron crystal (Fig. 2). A description of this basic mechanism follows.

Effect of Carbon on the Constitution of Iron

As an elemental metal, pure iron has only limited engineering usefulness despite its allotropy. Carbon is the main alloying addition that capitalizes on the allotropic phenomenon and lifts iron from mediocrity into the position of a unique structural material, broadly known as steel. Even in the highly alloyed stainless steels, it is the quite minor constituent carbon that virtually controls the engineering properties. Furthermore, due to the manufacturing processes, carbon in effective quantities persists in all irons and steels unless special methods are used to minimize it.

Carbon is almost insoluble in iron, which is in the alpha or ferritic phase (910 C, or 1675 F). However, it is quite soluble in gamma iron. Carbon actually dissolves; that is, the individual atoms of carbon lose themselves in the interstices among the iron atoms. Certain interstices within the fcc structure (austenite) are considerably more accommodating to carbon than are those of ferrite, the other allotrope. This preference exists not only on the mechanical basis of size of opening, however, for it is also a fundamental matter involving electron bonding and the balance of those attractive and repulsive forces that underlie the allotrope phenomenon.

The effects of carbon on certain characteristics of pure iron are shown in Fig. 3 (Ref 3). Figure 3(a) is a simplified version of Fig. 1; that is, a straight line constitutional diagram of commercially pure iron. In Fig. 3(b), the diagram is expanded horizontally to depict the initial effects of carbon on the principal thermal points of pure iron. Thus, each vertical dashed line, like the solid line in Fig. 3(a), is a constitutional diagram, but now for iron containing that particular percentage of carbon. Note that carbon lowers the freezing point of iron and that it broadens the temperature range of austenite by raising the temperature A4 at which (delta) ferrite changes to austenite and by lowering the temperature A3 at which the austenite reverts to (alpha) ferrite. Hence, carbon is said to be an austenitizing element. The spread of arrows at A3 covers a two-phase region, which signifies that austenite is retained fully down to the temperatures of the heavy arrow, and only in part down through the zone of the lesser arrows.

In a practical approach, however, it should be emphasized that Fig. 1, as well as Fig. 3, represents changes that occur during very slow cooling, as would be possible during laboratory-controlled experiments, rather than under conditions in commercial practice. Furthermore, in slow heating of iron, these transformations take place in a reverse manner. Transformations occurring at such slow rates of cooling and heating are known as equilibrium transformations, due to the fact that temperatures indicated in Fig. 1. Therefore, the process by which iron changes from one atomic arrangement to another when heated through 910 °C (1675 F) is called a transformation. Transformations of this type occur not only in pure iron but also in many of its alloys; each alloy composition transforms at its own characteristic temperature. It is this transformation that makes possible the variety of properties that can be achieved to a high degree of reproducibility through use of carefully selected heat treatments.

Iron-Cementite Phase Diagram

When carbon atoms are present, two changes occur (see Fig. 3). First, transformation temperatures are lowered, and second, transformation takes place over a range of temperatures rather than at a single temperature. These data are shown in the well-known iron-cementite phase diagram (Fig. 4). However, a word of explanation is offered to clarify the distinction between phases and phase diagrams.

A phase is a portion of an alloy, physically, chemically, or crystallographically homogeneous throughout, which is separated from the rest of the alloy by distinct bounding surfaces. Phases that occur in iron-carbon alloys are molten alloy, austenite (gamma phase), ferrite (alpha phase), cementite, and graphite. These phases are also called constituents. Not all constituents (such as pearlite or bainite) are phases—these are microstructures. A phase diagram is a graphical representation of the equilibrium temperature and composition limits of phase fields and phase reactions in an alloy system. In the iron-cementite system, temperature is plotted vertically, and composition is plotted horizontally. The iron-cementite diagram (Fig. 4), deals only with the constitution of the iron-iron carbide system, i.e., what phases are present at each temperature and the composition limits of each phase. Any point on the diagram, therefore, represents a definite composition and temperature, each value being found by projecting to the proper reference axis. Although this diagram extends from a temperature of 1870 °C (3400 F) down to room temperature, note that part of the diagram lies below 1040 °C (1900 F). Steel heat treating practice rarely involves the use of temperatures above 1040 °C (1900 F). In metal systems, pressure is usually considered as constant.

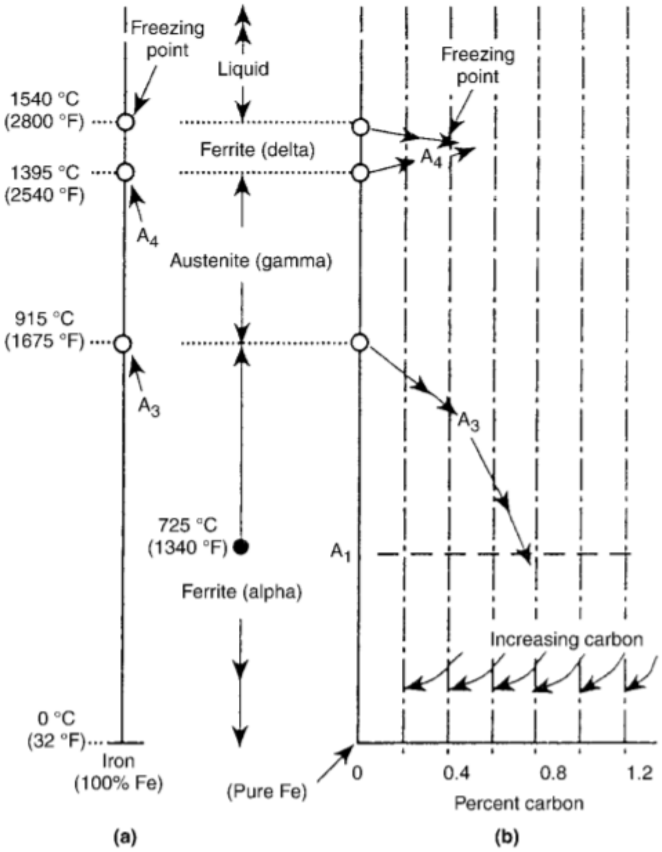


Fig. 3 Effects of carbon on the characteristics of commercially pure iron. (a) Constitutional diagram for pure iron. (b) Initial effects of carbon on the principal thermal points of pure iron.

Solubility of Carbon in Iron

In Fig. 4, the area denoted as austenite is actually an area within which iron can retain much dissolved carbon. In fact, most heat treating operations (notably annealing, normalizing, and heating for hardening) begin with heating the alloy into the austenitic range to dissolve the carbide in the iron. At no time during such heating operations are the iron, carbon, or austenite in the molten state. A solid solution of carbon in iron can be visualized as a pyramidal stack of basketballs with golf balls between the spaces in the pile. In this analogy, the basketballs would be the iron atoms, while the golf balls interspersed between would be the smaller carbon atoms. Austenite is the term applied to the solid solution of carbon in gamma iron, and, like other constituents in the diagram, austenite has a certain definite solubility for carbon, which depends on the temperature (shaded area in Fig. 4 bounded by AGFED). As indicated by the austenite area in Fig. 4, the carbon content of austenite can range from 0 to 2%. Under normal conditions, austenite cannot exist at room temperature in plain carbon steels; it can exist only at elevated temperatures bounded by the lines AGFED in Fig. 4. Although austenite does not ordinarily exist at room temperature in carbon steels, the rate at which steels are cooled from the austenitic range has a profound influence on the room temperature microstructure and properties of carbon steels. Thus, the phase known as austenite is fcc iron, capable of containing up to 2% dissolved carbon. The solubility limit for carbon in the bcc structure of iron-carbon alloys is shown by the line ABC in Fig. 4. This area of the diagram is labeled alp**ha (α),** and the phase is called ferrite. The maximum solubility of carbon in alpha iron (ferrite) is 0.025% and occurs at 725 °C (1340 F). At room temperature, ferrite can dissolve only 0.008% C, as shown in Fig. 4. This is the narrow area at the extreme left of Fig. 4 below approximately 910 °C (1675 F). For all practical purposes, this area has no effect on heat treatment and shall not be discussed further. Further discussion of Fig. 4 is necessary, although as previously stated, the area of interest for heat treatment extends vertically to only about 1040 °C (1900F) and horizontally to a carbon content of 2%. The large area extending vertically from zero to the line BGH (725 °C, or 1340 F) and horizontally to 2% C is denoted as a **two-phase area—α + Cm, or** alpha (ferrite) plus cementite (carbide). The line BGH is known as the lower transformation temperature (A1). The line AGH is the upper transformation temperature (A3). The triangular area ABG is also a two-phase area, but the phases are alpha and gamma, or ferrite plus austenite. As carbon content increases, the A3 temperature decreases until the eutectoid is reached— 725 °C (1340 F) and 0.80% C (point G). This is considered a saturation point; it indicates the amount of carbon that can be dissolved at 725 °C (1340 F). A1 and A3 intersect and remain as one line to point H as indicated. The area above 725 °C (1340 F) and to the right of the austenite region is another two-phase field—gamma plus cementite (austenite plus carbide).

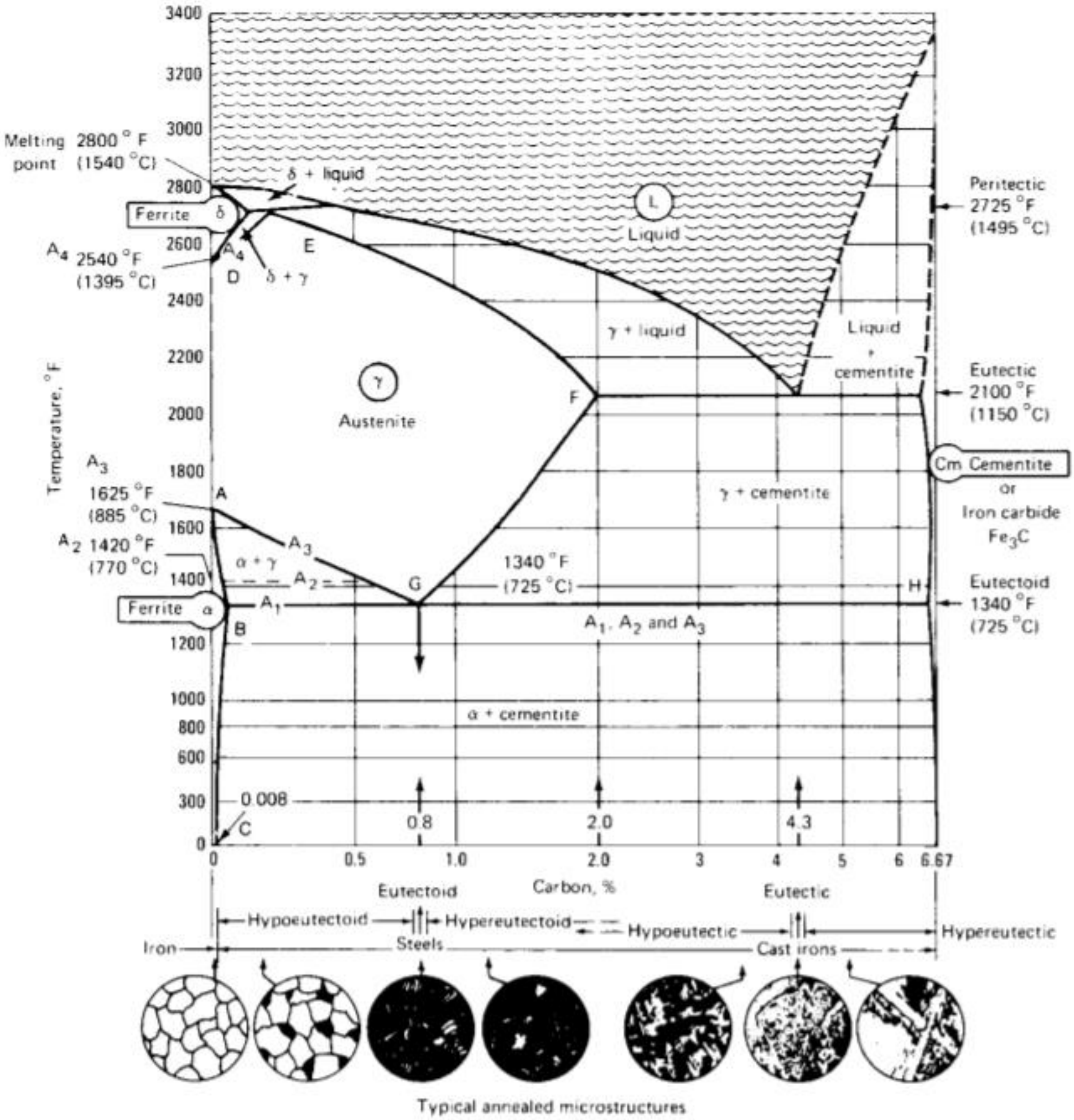


Fig. 4 Iron-cementite phase diagram

Now as an example, when a 0.40% carbon steel is heated to 725°C (1340F), its crystalline structure begins to transform to austenite; transformation is not complete however until a temperature of approximately 815°C (1500F) is reached. In contrast, as shown in Fig. 4, a steel containing 0.80% C transforms completely to austenite when heated to 725°C (1340 F).

Now assume that a steel containing 1.0% C is heated to 725°C (1340F) or just above. At this temperature, austenite is formed, but because only 0.80% C can be completely dissolved in the austenite, 0.20% C remains as cementite, unless the temperature is increased. However, if the temperature of a 1.0% carbon steel is increased above about 790°C (1450 F), the line GF is intersected, and all of the carbon is thus dissolved. Increasing temperature gradually increases the amount of carbon that can be taken into solid solution. For instance, at 1040°C (1900F), approximately 1.6% C can be dissolved (Fig.4).

Transformation of Austenite

Thus far the discussion has been confined to heating of the steel and the phases that result from various combinations of temperature and carbon content. Now what happens when the alloy is cooled? Referring to Fig. 4, assume that a steel containing 0.50% C is heated to 815 °C (1500 F). All of the carbon will be dissolved (assuming, of course, that holding time is sufficient). Under these conditions, all of the carbon atoms will dissolve in the interstices of the fcc crystal (Fig. 2b). If the alloy is cooled slowly, transformation to the bcc (Fig. 2a) or alpha phase begins when the temperature drops below approximately 790 °C (1450 F). As the temperature continues to decrease, the transformation is essentially complete at 725 °C (1340 F). During this transformation, the carbon atoms escape from the lattice because they are essentially insoluble in the alpha crystal (bcc). Thus, in slow cooling, the alloy for all practical purposes, returns to the same state (in terms of phase) that it was before heating to form austenite. The same mechanism occurs with higher carbon steels, except that the austenite-to-ferrite transformation does not go through a two-phase zone (Fig. 4). In addition to the entry and exit of the carbon atoms through the interstices of the iron atoms, other changes occur that affect the practical aspects of heat treating.

First, a magnetic change occurs at 770 °C (1420 F) as shown in Fig. 1. The heat of transformation effects may chemical changes, such as the heat that is evolved when water freezes into ice and the heat that is absorbed when ice melts. When an iron-carbon alloy is converted to austenite by heat, a large absorption of heat occurs at the transformation temperature. Likewise, when the alloy changes from gamma to alpha (austenite to ferrite), heat evolves. What happens when the alloy is cooled rapidly? When the alloy is cooled suddenly, the carbon atoms cannot make an orderly escape from the iron lattice. This cause "atomic bedlam" and results in distortion of the lattice, which manifests itself in the form of hardness and/or strength. If cooling is fast enough, a new structure known as martensite is formed, although this new structure (an aggregate of iron and cementite) is in the alpha phase.

Classification of Steels by Carbon Content

It must be remembered that there are only three phases in steels, but there are many different structures. A precise definition of eutectoid carbon is unavoidable; it varies from 0.77% to slightly over 0.80%, depending on the reference used. However, for the objectives of this book, the precise amount of carbon denoted as eutectoid is of no particular significance.

Hypoeutectoid Steels. Carbon steels containing less than 0.80% C are known as hypoeutectoid steels. The area bounded by AGB on the iron cementite diagram (Fig. 4) is of significance to the room temperature microstructures of these steels; within the area, ferrite and austenite each having different carbon contents, can exist simultaneously.

Assume that a 0.40% carbon steel has been slowly heated until its temperature throughout the piece is 870 °C (1600 F), thereby ensuring a fully austenitic structure. Upon slow cooling, free ferrite begins to form from the austenite when the temperature drops across the line AG, into the area AGB, with increasing amounts of ferrite forming as the temperature continues to decline while in this area. Ideally, under very slow cooling conditions, all of the free ferrite separates from austenite by the time the temperature of the steel reaches A1 (the line BG) at 725 °C (1340 F). The austenite islands, which remain at about 725 °C (1340 F), now have the same amount of carbon as the eutectoid steel, or about 0.80%. At or slightly below 725 °C (1340 F) the remaining untransformed austenite transforms—it becomes pearlite, which is so named because of its resemblance to mother of pearl. Upon further cooling to room temperature, the microstructure remains unchanged, resulting in a final room temperature microstructure of an intimate mixture of free ferrite grains and pearlite islands. A typical microstructure of a 0.40% carbon steel is shown in Fig. 5(a). The pure white areas are the islands of free ferrite grains described previously. Grains that are white but contain dark platelets are typical lamellar pearlite. These platelets are cementite or carbide interspersed through the ferrite, thus conforming to the typical two-phase structure indicated below the BH line in Fig. 4.

Eutectoid Steels. A carbon steel containing approximately 0.77% C becomes a solid solution at any temperature in the austenite temperature range, i.e., between 725 and 1370 °C (1340 and 2500 F). All of the carbon is dissolved in the austenite. When this solid solution is slowly cooled, several changes occur at 725 °C (1340 F). This temperature is a transformation temperature or critical temperature of the iron-cementite system. At this temperature, a 0.77% (0.80%) carbon steels transforms from a single homogeneous solid solution into two distinct new solid phases. This change occurs at constant temperature and with the evolution of heat. The new phases are ferrite an cementite, formed simultaneously; however, it is only at composition point G in Fig. 4 (0.77% carbon steel) that this phenomenon of the simultaneous formation of ferrite and cementite can occur.

The microstructure of a typical eutectoid steel is shown in Fig. 5(b). The white matrix is alpha ferrite and the dark platelets are cementite. All grains are pearlite—no free ferrite grains are present under these conditions. Cooling conditions (rate and temperature) govern the final condition of the particles of cementite that precipitate from the austenite at 725 °C (1340 F). Under specific cooling conditions, the particles become spheroidal instead of elongated platelets as shown in Fig. 5(b). Figure 5(c) shows a similar two-phase structure resulting from slowly cooling a eutectoid carbon steel just below A1. This structure is commonly known as spheroidite but is still a despersion of cementite particles in alpha ferrite. There is no indication of grain boundaries in Fig. 5(c). The spheroidized structure is often preferred over the pearlitic structure because spheroidite has superior machinability and formability. Combination structures (that is, partly lamellar and partly spheroidal cementite in a ferrite matrix) are also common. As noted previously, a eutectoid steel theoretically contains a precise amount of carbon. In practice, steels that contain carbon within the range of approximately 0.75 to 0.85% are commonly referred to as eutectoid carbon steels.

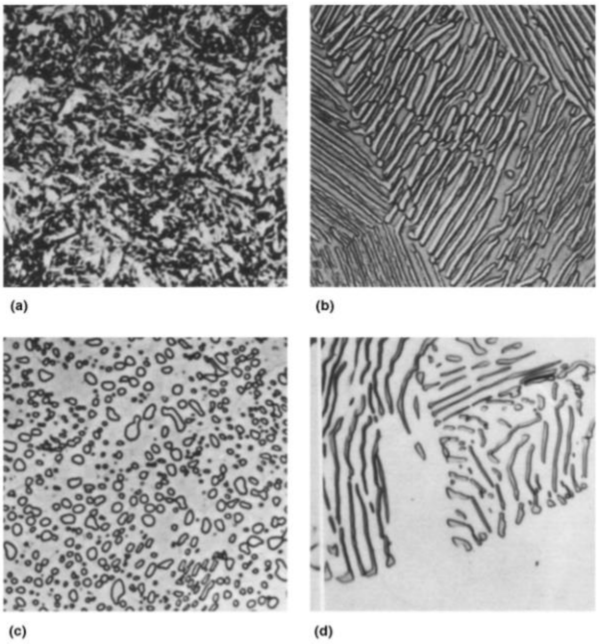


Fig. 5 Effects of carbon content on the microstructures of plain-carbon steels. (a) Ferrite grains (white) and pearlite (gray streaks) in a white matrix of a hypoeutectoid steel containing 0.4% C. 1000X. (b) Microstructure (all pearlite grains) of a eutectoid steel containing 0.77% C. 2000X. (c) Microstructure of a eutectoid steel containing 0.77% C with all cementite in the spheroidal form. 1000X. (d) Microstructure of a hypereutectoid steel containing 1.0% C containing pearlite with excess cementite bounding the grains. 1000X.

Hypereutectoid steels contain carbon contents of approximately 0.80 to 2.0%. Assume that a steel containing 1.0% C has been heated to 845 °C (1550 F), thereby ensuring a 100%

austenitic structure. When cooled, no change occurs until the line GF (Fig. 4), known as Acm or cementite solubility line, is reached. At this point, cementite begins to separate out from the austenite, and increasing amounts of cementite separate out as the temperature of the 1% carbon steels descends below the A line. The composition of austenite changes from 1% C toward 0.77% C. At a temperature slightly below 725 °C (1340 F), the remaining austenite changes to pearline. No further changes occur as cooling proceeds toward room temperature, so that the room temperature microstructure consists of pearline and free cementite. In this case, the free cementite exists as a network around the pearline grains (Fig. 5d). Upon heating hypereutectoid steels, reverse changes occur. At 725 °C (1340 F), pearlite changes to austenite. As the temperature increases above 725 °C (1340 F), free cementite dissolves in the austenite, so that when the temperature reaches the Acm line, all the cementite dissolves to form 100% austenite.

Hysteresis in Heating and Cooling

The critical temperatures (A_1, A_6, and A_{cm}) are “arrests” in heating or cooling and have been symbolized with the letter A, from the French word arret meaning arrest or a delay point, in curves plotted to show heating or cooling of samples. Such changes occur at transformation temperatures in the iron-cementite diagram if sufficient time is given and cab be plotted for steels showing lags at transformation temperatures, as shown for iron in Fig. 4. However, because heating rates in commercial practice usually exceed those in controlled laboratory experiments, changes on heating usually occur at temperatures a few degrees above the transformation temperatures shown in Fig. 4 and are known as Ac temperatures, such as Ac_1 or Ac_3. The “c” is from the French word chauffage, meaning heating. Thus, Ac_1 is a few degrees above the ideal A_1 temperature. Likewise, on slow cooling in commercial practice, transformation changes occur at temperatures a few degrees below those in Fig. 4. These are known as Ar, or Ar_3, the “r” originating from the French word refroidissement, meaning cooling.

This difference between the heating and cooling varies with the rate of heating or cooling. The faster the heating, the higher the Ac point; the faster the cooling , the lower the Ar point. Also, the faster the heating and cooling rate, the greater the gap between the Ac and Ar points of the reversible (equilibrium) point A. Going one step further, in cooling a piece of steel, it is of utmost importance to note that the cooling rate may be so rapid (as in quenching steel in water) as to suppress the transformation for several hundreds of degrees. This is due to the decrease in reaction rate with decrease in temperature. As discussed subsequently, time is an important factor in transformation, especially in cooling.

Effect of Time on Transformation

The foregoing discussion has been confined principally to phases that are formed by various combinations of composition and temperature; little reference has been made to the effects of time. In order to convey to the reader the effects of time on transformation, the simplest approach is by means of a time-temperature-transformation (TTT) curve for some constant iron-carbon composition. Such a curve is presented in Fig. 6 for a 0.77% (eutectoid) carbon steel. TTT curves are also known as “S” curves because the principal part of the curve is shaped like the letter “S.” In Fig. 6, temperature is plotted on the vertical axis, and time is plotted on a logarithmic scale along the horizontal axis. The reason for plotting time on a logarithmic scale is merely to keep the width of the chart within a manageable dimension. In analyzing Fig. 6, begin with line Ae1 (725 °C or 1340 F). Above this temperature, austenite exists only for a eutectoid steel (refer also to Fig. 4). When the steel is cooled and held at a temperature just below Ae_1 (705 °C or 1300 F), transformation begins (follow line (Ps Bs), but very slowly at this temperature; 1 h of cooling is required before any significant amount of transformation occurs, although eventually complete transformation occurs isothermally (meaning at a constant temperature), and the transformation product is spheroidite (Fig. 5c). Now assume a lower temperature (650 °C or 1200 F) on line Ps Bs (the line of beginning transformation); transformation begins in less than 1 min, and the transfor mation product is coarse pearlite (near the right side of Fig. 6). Next, assume a temperature of 540 °C (1000 F); transformation begins in approximately 1 s and is completely transformed to fine pearlite in a matter of a few minutes. The line Pf Bf represents the completion of transformation and is generally parallel with Ps Bs. However, if the steel is cooled very rapidly (such as by immersing in water) so that there is not sufficient time for transformation to begin in the 540 °C (1000 F) temperature vicinity, then the beginning of transformation time is substantially extended. For example, if the steel is cooled to and held at 315 °C (600 F), transformation does not begin for well over 1 min. It must be remembered that all of the white area to the left of line Ps Bs represents the austentic phase, although it is highly unstable.

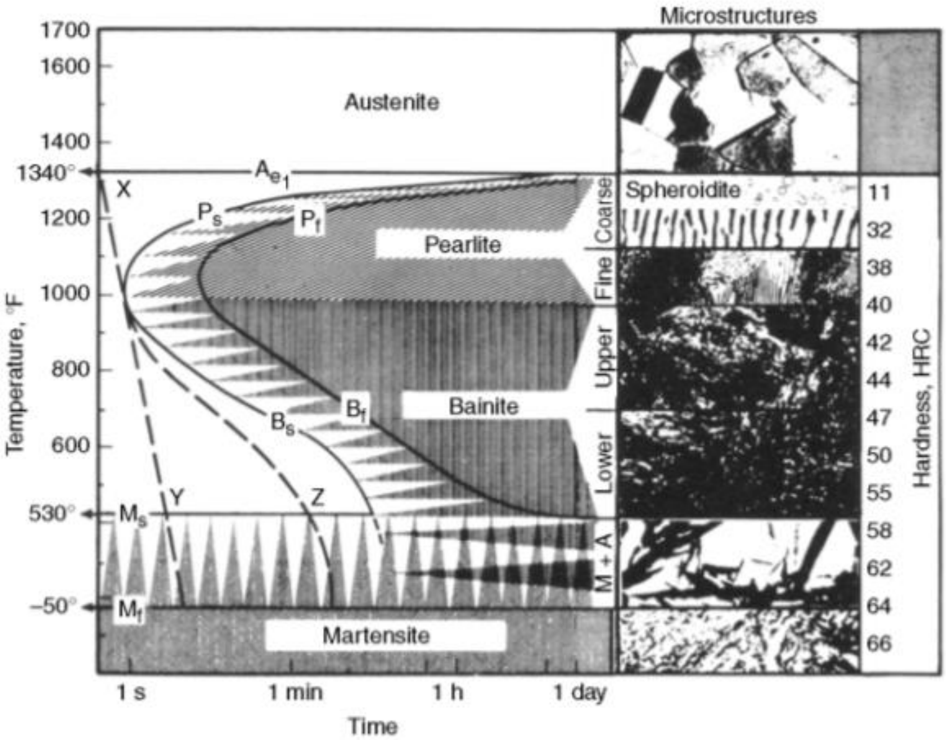


Fig. 6 Time-temperature-transformation (TTT) diagram for a eutectoid (0.77%) carbon steel.

When transformation takes place isothermally within the temperature range of approximately 290 to 425 °C (550 to 800 F), the transformation product is a microstructure called bainite (upper or lower as indicated toward the right of Fig. 6). A bainitic microstructure is shown in Fig. 7(b). In another example, steel is cooled so rapidly that no transformation takes place in the 540 °C (1000 F) region and rapid cooling is continued (note line XY in Fig. 6) to and below 275 °C (530 F) or Ms. Under these conditions, martensite is formed. Point Ms is the temperature at which martensite begins to form, and Mf indicates the complete finish of transformation. It must be remembered that martensite is not a phase but is a specific microstructure in the ferritic (alpha) phase. Martensite is formed from the carbon atoms jamming the lattice of the austenitic atomic arrangement. Thus, martensite can be considered as an aggregate of iron and cementite (Fig. 7a).

In Fig. 6, the microstructure of austenite (as it apparently appears at elevated temperature) is shown on the right. It is also evident that the lower the temperature at which transformation takes place, the higher the hardness.

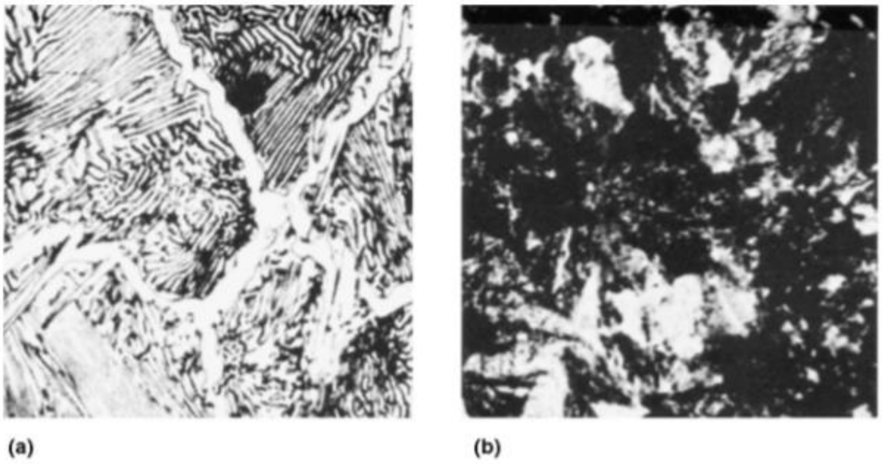


Fig. 7 (a) Microstructure of quenched 0.95% carbon steel. 1000X. Structure is martensitic. (b) Bainitic structure in a quenched 0.95% carbon steel. 550X.

It is also evident that all structures from the top to the region where martensite forms (Ae_1) are time-dependant, but the formation of martensite is not time-dependent. Each different steel composition has its own TTT curve; Fig. 6 is presented only as an example. However, patterns are much the same for all steels as far as shape of the curves is concerned. The most outstanding difference in the curves among different steels is the distance between the vertical axis and the nose of the S curve. This occurs at about 540 °C (1000 F) for the steel in Fig. 6. This distance in terms of time is about 1 s for a eutectoid carbon steel, but could be an hour or more for certain high-alloy steels, which are extremely sluggish in transformation. The distance between the vertical axis and the nose of the S curve is often called the “gap” and has a profound effect on how rapidly the steel must be cooled to form the hardened structure—martensite. Width of this gap for any steel is directly related to the critical cooling rate for that specific steel. Critical cooling rate is defined as the rate at which a given steel must be cooled from the austenite to prevent the formation of nonmartensitic products. In Fig. 6, it is irrelevant whether the cooling rate follows the lines X to Y or X to Z because they are both at the left of the beginning transformation line Ps Bs. Practical heat treating procedures are based on the fact that once the steel has been cooled below approximately 425 °C (800 F), the rate of cooling may be decreased. These conditions are all closely related to hardenability.

Heat Treating of Aluminum Alloys

Heat treating in its broadest sense, refers to any of the heating and cooling operations that are performed for the purpose of changing the mechanical properties, the metallurgical structure, or the residual stress state of a metal product. When the term is applied to aluminum alloys, however, its use frequently is restricted to the specific operations employed to increase strength and hardness of the precipitation hardenable wrought and cast alloys. These usually are referred to as the "heat-treatable" alloys to distinguish them from those alloys in which no significant strengthening can be achieved by heating and cooling. The latter, generally referred to as "non-heat treatable" alloys, depend primarily on cold work to increase strength.

Heating to decrease strength and increase ductility (annealing) is used with alloys of both types; metallurgical reactions may vary with type of alloy and with degree of softening desired. Except for the low-temperature stabilization treatment sometimes given for 5xxx series alloys (which is a mill treatment and not discussed in this article), complete or partial annealing treatments are the only ones used for non-heat-treatable alloys. A general overview of these heat treatments is covered in the article "Principles of Heat Treating of Nonferrous Alloys" in this Volume.

Precipitation from Solid Solution

One essential attribute of a precipitation hardening alloy system is a temperature dependent equilibrium solid solubility characterized by increasing solubility with increasing temperature (see, for example, the phase diagrams in Fig 8 and 9). Although this condition is met by most of the binary aluminum alloy systems, many exhibit very little precipitation hardening, and these alloys ordinarily are not considered heat treatable. Alloys of the binary aluminum-silicon and aluminum-manganese systems, for example, exhibit relatively insignificant changes in mechanical properties as a result of heat treatments that produce considerable precipitation. The major aluminum alloy systems with precipitation hardening include:

- Aluminum-copper systems with strengthening from $CuAI_2$
- Aluminum-copper-magnesium systems (magnesium intensifies precipitation)
- Aluminum-magnesium-silicon systems with strengthening from Mg_2Si
- Aluminum-zinc-magnesium systems with strengthening from $MgZn_2$
- Aluminum-zinc-magnesium-copper systems

The general requirement for precipitation strengthening of supersaturated solid solutions involves the formation of finely dispersed precipitates during aging heat treatments (which may include either natural aging or artificial aging). The aging must be accomplished not only below the equilibrium solvus temperature, but below a metastable miscibility gap called the Guinier-Preston (GP) zone solvus line. The supersaturation of vacancies allows diffusion, and thus zone formation, to occur much faster than expected from equilibrium diffusion coefficients. In the precipitation process, the saturated solid solution first develops solute clusters, which then become involved in the formation of transitional (nonequilibrium) precipitates.

The mechanism of strengthening from precipitation involves the formation of coherent clusters of solute atoms (that is, the solute atoms have collected into a cluster but still have the same crystal structure as the solvent phase). This causes a great deal of strain because of mismatch in size between the solvent and solute atoms. Consequently, the presence of the precipitate particles, and even more importantly the strain fields in the matrix surrounding the coherent particles, provide higher strength by obstructing and retarding the movement of dislocations. The characteristic that determines whether a precipitate phase is coherent or noncoherent is the closeness of match or degree of disregistry between atomic spacings on the lattice of the matrix and on that of the precipitate. These changes in properties result from the formation of solute-rich microstructural domains, or GP zones.

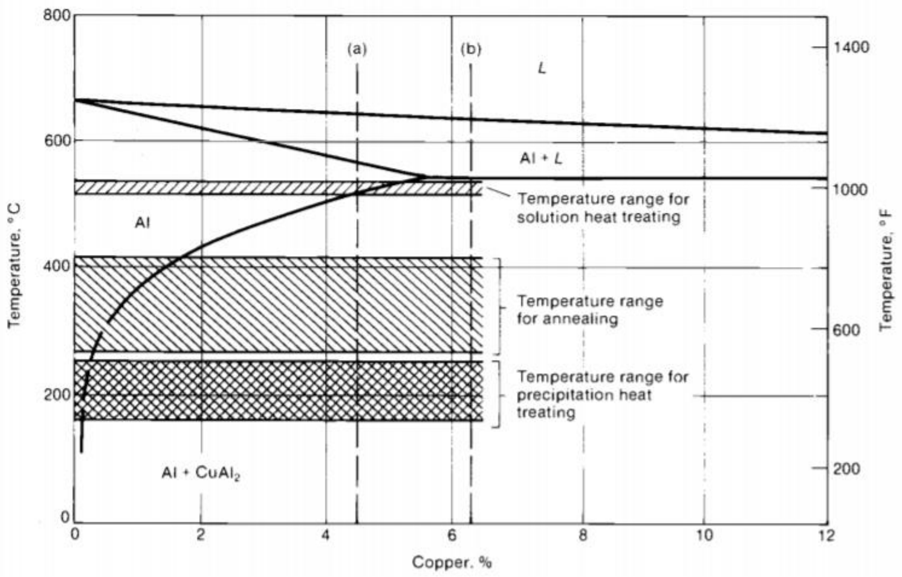


Fig 8 Portion of aluminum-copper binary phase diagram. Temperature ranges for annealing, precipitation heat treating, and solution heat treating are indicated. The

range for solution treating is below the eutectic melting point of 548 °C (1018 °F) at 5.65 wt% Cu.

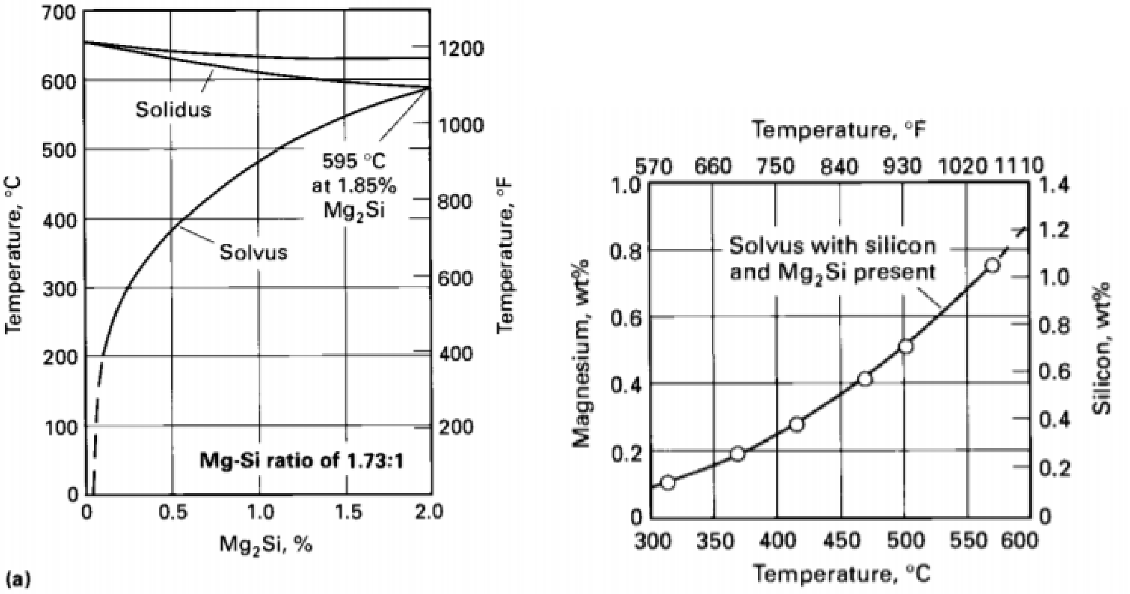


Fig 8 Equilibrium solubility as function of temperature for (a) Mg_2Si in aluminum with an Mg-Si ratio of 1.73-to-1 and (b) magnesium and silicon in solid aluminum when both Mg_2Si and silicon are present

The exact size, shape, and distribution of GP zones depend on the alloy in which they form and on the thermal and mechanical history of the specimen. Their shape can sometimes be deduced by refined studies of diffuse x-ray scattering. Under favorable conditions, GP zones can be seen in transmission electron micrographs. Spherical solute-rich zones usually form when the sizes of the solvent and solute atoms are nearly equal, as in the aluminum-silver and aluminum-zinc systems. If there is a large difference in atom sizes, as in the aluminum-copper system, the GP zones usually form as disks whose planes lie parallel with some low-index plane of the matrix lattice. Sometimes, the solute atoms occupy preferred lattice sites within the GP zone, and thus form a small region of lattice order.

The GP zones are of the size range of tens of angstroms in diameter. They are essentially distorted regions of the matrix lattice, rather than discrete particles of a new phase having a different lattice. As such, they are completely coherent with the matrix, imposing local but often large strains on it. These mechanical strains, as well as the presence of a locally solute-rich, sometimes ordered lattice, can account for large changes in mechanical properties of the alloy before any long-range microstructural changes occur.

The GP zones are characteristically metastable and thus dissolve in the presence of a more stable precipitate. This dissolution causes a precipitate-free, visibly denuded region to form around the stable precipitate particles. The final structure consists of equilibrium precipitates, which do not contribute as significantly to hardening.

Precipitation in Aluminum-Copper Alloys.

Figure 8, which illustrates the required solubility-temperature relationship needed in precipitation strengthening, shows the temperature ranges required for solution treatment and subsequent precipitate hardening in the aluminum-copper system. The equilibrium solid solubility of copper in aluminum increases as temperature increases-- from about 0.20% at 250 °C (480 °F) to a maximum of 5.65% at the eutectic melting temperature of 548 °C (1018 °F). (It is considerably lower than 0.20% at temperatures below 250 °C.) For aluminum-copper alloys containing from 0.2 to 5.6% Cu, two distinct equilibrium solid states are possible. At temperatures above the lower curve in Fig 8 (solvus), the copper is completely soluble, and when the alloy is held at such temperatures for sufficient time to permit needed diffusion, the copper will be taken completely into solid solution. At temperatures below the solvus, the equilibrium state consists of two solid phases: solid solution, ct, plus an intermetallic-compound phase 0 (AIECU). When such an alloy is converted to all solid solution by holding above the solvus temperature and then the temperature is decreased to below the solvus, the solid solution becomes supersaturated and the alloy seeks the equilibrium two-phase condition; the second phase tends to form by solid-state precipitation.

The preceding description is a gross oversimplification of the actual changes that occur under different conditions even in simple binary aluminum-copper alloys. A variety of different nonequilibrium precipitate structures is formed at temperatures below solvus. In alloys of the aluminum copper system, a succession of precipitates is developed from a rapidly cooled supersaturated solid solution (SSS). These precipitates develop sequentially either with increasing temperature or with increasing time at temperature between room temperature and the solvus.

At temperatures in the natural aging range (about -20 to 60 °C, or 0 to 140 °F), the distribution of copper atoms changes with time from random to the disklike planar aggregates (GP zones), which form on particular crystallographic planes of the aluminum matrix. These aggregates create coherency strain fields that increase resistance to deformation, and their formation is responsible for the changes in mechanical properties that occur during natural aging. At higher temperatures, transition forms of approximate composition AI2Cu develop and further increase strength. In the highest strength condition, both the 0" and 0' transition precipitates may be present. When time and temperature are increased sufficiently to form high proportions of the equilibrium 0, the alloy softens and is said to be "overaged."

The commercial heat-treatable aluminum alloys are, with few exceptions, based on ternary or quaternary systems with respect to the solutes involved in developing strength by precipitation. Commercial alloys whose strength and hardness can be significantly increased by heat treatment include 2xxx, 6xxx, and 7xxx series wrought alloys (except 7072) and 2xx.0, 3xx.0, and 7xx.O series casting alloys. Some of these contain only copper, or copper and silicon, as the primary strengthening alloy addition(s). Most of the heat-treatable alloys, however, contain combinations of magnesium with one or more of the elements copper, silicon, and zinc. Characteristically, even small amounts of magnesium in concert with these elements accelerate and accentuate precipitation hardening, while alloys in the 6xxx series contain silicon and magnesium approximately in the proportions required for formulation of magnesium silicide (MgESi). Although not as strong as most 2xxx and 7xxx alloys, 6xxx series alloys have good formability, weldability, machinability, and corrosion resistance, with medium strength.

In the heat-treatable wrought alloys, with some notable exceptions (2024, 2219, and 7178), such solute elements are present in amounts that are within the limits of mutual solid solubility at temperatures below the eutectic temperature (lowest melting temperature). In contrast, some of the casting alloys of the 2xx.O series and all of the 3xx.O series alloys contain amounts of soluble elements that far exceed solid-solubility limits. In these alloys, the phase formed by combination of the excess soluble elements with the aluminum will never be dissolved, although the shapes of the undissolved particles may be changed by partial solution. Most of the heat-treatable aluminum alloy systems exhibit multistage precipitation and undergo accompanying strength changes analogous to those of the aluminum-copper system. Multiple alloying additions of both major solute elements and supplementary elements employed in commercial alloys are strictly functional and serve with different heat treatments to provide the many different combinations of properties---physical, mechanical, and electrochemical--that are required for different applications. Some alloys, particularly those for foundry production of castings, contain amounts of silicon far in excess of the amount that is soluble or needed for strengthening alone. The function here is chiefly to improve casting soundness and freedom from cracking, but the excess silicon also serves to increase wear resistance, as do other microstructural constituents formed by manganese, nickel, and iron. Parts made of such alloys are commonly used in gasoline and diesel engines (pistons, cylinder blocks, and so forth).

Alloys containing the elements silver, lithium, and germanium are also capable of providing high strength with heat treatment, and in the case of lithium, both increased elastic modulus and lower density, which are highly advantageous--particularly for aerospace applications. Commercial use of alloys containing these elements has been restricted either by cost or by difficulties encountered in producing them. Such alloys are used to some extent, however, and research is being directed toward overcoming their disadvantages. In the case of alloys having copper as the principal alloying ingredient and no magnesium, strengthening by precipitation can be greatly increased by adding small fractional percentages of tin, cadmium, or indium, or combinations of these elements. Alloys based on these effects have been produced commercially but not in large volumes because of costly special practices and limitations required in processing, and in the case of cadmium, the need for special facilities to avoid health hazards from formation and release of cadmium vapor during alloying. Such alloys, as well as those containing silver, lithium, or other particle-forming elements, may be used on a selective basis in the future.

Aluminum-Lithium Alloys. Like other agehardened aluminum alloys, aluminum-lithium alloys achieve precipitation strengthening by thermal aging after a solution heat treatment. The precipitate structure is sensitive to a number of processing variables, including, but not limited to, the quenching rate following the solution heat treatment, the degree of cold deformation prior to aging, and the aging time and temperature. Minor alloying elements can also have a significant effect on the aging process by changing the interface energy of the precipitate, by increasing the vacancy concentration, and/or by raising the critical temperature for homogeneous precipitation. Like some other age-hardened 2xxx aluminum alloys, aluminum-lithium-base alloys also gain increased strength and toughness from deformation prior to aging. This unusual phenomenon has given rise to a number of thermomechanical processing steps for aluminum-lithium alloys aimed at optimizing mechanical properties after artificial aging. The age hardening of aluminum-lithium alloys involves the continuous precipitation of 8' (AlaLi) from a supersaturated solid solution. The aluminum and lithium in the 8' precipitates are positioned at specific locations. The eight shared corner sites are occupied by lithium, and the six shared faces are occupied by aluminum. This gives rise to the aluminum-lithium composition of 8' precipitates. The geometrical similarity between the lattice of the precipitates and the face-centered cubic lattices of the solid solution facilitates the observed cube/cube orientation. The lattice parameters of the precipitate are also closely matched to those of the matrix. Consequently, the microstructure of an aluminum-lithium alloy solution heat treated and aged for short times below the 8' solvus is characterized by a homogeneous distribution of coherent, spherical 8' precipitates.

pherical 8' precipitates. Aluminum-lithium-base alloys are microstructurally unique. They differ from most of the aluminum alloys in that once the major strengthening precipitate (8') is homogeneously precipitated, it remains coherent even after extensive aging. In addition, extensive aging at high temperatures (> 190 °C, or 375 °F) can result in the precipitation of icosahedral grain-boundary precipitates. with five-fold symmetry. Although the quasi-crystalline structure and the composition of these grain-boundary precipitates are not yet exactly known, it has been suggested that both the precipitates and the precipitate-free zones (PFZs) near the grain boundaries might play a major role in the fracture process.

Aluminum-lithium alloys can be traced, at least in part, to the inhomogeneous nature of their slip, resulting from coherent-particle hardening of spherical 8' precipitates. The presence of equilibrium ~ (aluminum lithium) precipitates at grain boundaries can also cause PFZs, which can induce further strain localization and promote intergranular failure. Consequently, for the development of commercial alloys, slip has been homogenized by introducing dispersoids (manganese, zirconium) and semicoherent/ incoherent precipitates, such as T 1 (Al_2CuLi), 0' (Al_2Cu), or S (AI_2LiMg), through copper or magnesium additions. Magnesium and copper improve the strength of aluminum-lithium alloys through solid-solution and precipitate strengthening, and they can minimize the formation of PFZs near grain boundaries. Zirconium, which forms the cubic Al_2Zr coherent dispersoid, stabilizes the subgrain structure and suppresses recrystallization.

nd suppresses recrystallization. Development of commercially available aluminum-lithium-base alloys was started by adding lithium to aluminum-copper, aluminum-magnesium, and aluminum-copper magnesium alloys. These alloys were chosen to superimpose the precipitationhardening characteristics of aluminumcopper-, aluminum-copper-magnesium-, and aluminum-magnesium-base precipitates to the hardening of lithium-containing precipitates. Proceeding in this manner, alloys 2020 (A1-Cu-Li-Cd), 01429 (AI-Mg-Li), 2090 (AI-Cu-Li), and 2091 and 8090 (A1-Cu-MgLi) evolved. Besides these registered alloys, other commercial aluminum-lithium alloys include Weldalite 049 and CP276. Properties and applications of these alloys are discussed in the article "AluminumLithium Alloys" in Volume 2 of the 10th Edition of Metals Handbook. In terms of 8' precipitation, the only effect of magnesium appears to be a reduction in the solubility of lithium. The microstructure of an aluminum-magnesium-lithium alloy in the early stages of aging is similar to that of an aluminum-lithium alloy. Precipitation in the aluminum-copper-lithium system is more complicated than that in either the aluminum-lithium or aluminum magnesium-lithium systems.

Strengthening by Heat Treatment

Heat treatment to increase strength of aluminum alloys is a three-step process:

- Solution heat treatment: dissolution of soluble phases
- Quenching: development of supersaturation
- Age hardening: precipitation of solute atoms either at room temperature (natural aging) or elevated temperature (artificial aging or precipitation heat treatment)

Solution Heat Treating

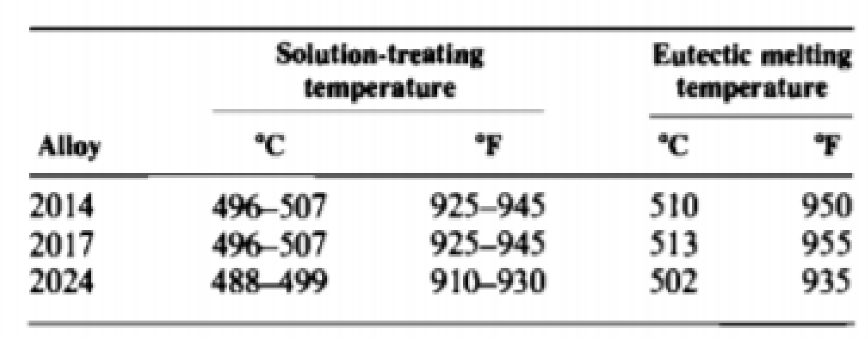
To take advantage of the precipitation hardening reaction, it is necessary first to produce a solid solution. The process by which this is accomplished is called solution heat treating, and its objective is to take into solid solution the maximum practical amounts of the soluble hardening elements in the alloy. The process consists of soaking the alloy at a temperature sufficiently high and for a time long enough to achieve a nearly homogeneous solid solution. Nominal commercial solution heat-treating temperature is determined by the composition limits of the alloy and an allowance for unintentional temperature variations. Although ranges normally listed allow variations of ±6 °C (± l0 °F) from the nominal, some highly alloyed, controlled-toughness, high-strength alloys require that temperature be controlled within more restrictive limits. Broader ranges may be allowable for alloys with greater intervals of temperature between their solvus and eutectic melting temperatures.

Overheating. Care must be exercised to avoid exceeding the initial eutectic melting temperature. If appreciable eutectic melting occurs as a result of overheating, properties such as tensile strength, ductility, and fracture toughness may be degraded. Materials that exhibit microstructural evidence of overheating are generally categorized as unacceptable by specification. Evidence of grain-boundary melting that occurs above the eutectic melting temperature of the alloy usually is not detectable by either visual examination or nondestructive testing.

Although maximum temperature must be restricted to avoid melting, the lower limit should, when possible, be above the temperature at which complete solution occurs (solvus). In the alloy represented by line (a) in Fig 1, these temperatures would be about 575 and 515 °C (1065 and 960 °F), respectively. However, under production conditions, the temperature interval for solution treatment (shown in Fig 8 for typical 2xxx or 2xx.x) alloys provides a margin to safeguard against eutectic melting and a cushion on the low side for increased solution and diffusion rates. For alloys containing more than 5.65% Cu, complete solution can never occur. For these alloys, such as alloy 2219 (which has 5.8 to 6.8% Cu), the minimum solution heattreating temperature is established so that it is as close as practical to the eutectic temperature while providing a margin of safety commensurate with the capability of the equipment. Line (b) in Fig 8 is another example of a composition above 5.65% Cu that does not allow complete dissolution of aluminum-copper precipitates.

For more complex ternary and quaternary systems, solution treatments are modified according to the effect of new elements on the solid solubility and/or the eutectic melting points of the basic binary system. In aluminum-lithium alloys, for example, magnesium reduces the solubility of lithium in aluminum. In the aluminum-copper system, magnesium also lowers the eutectic melting point. The proximity of typical solution treating temperature ranges to eutectic melting temperatures for three common aluminum-copper-magnesium alloys is shown in the following table:

Table.1. Temperature ranges to eutectic melting temperatures



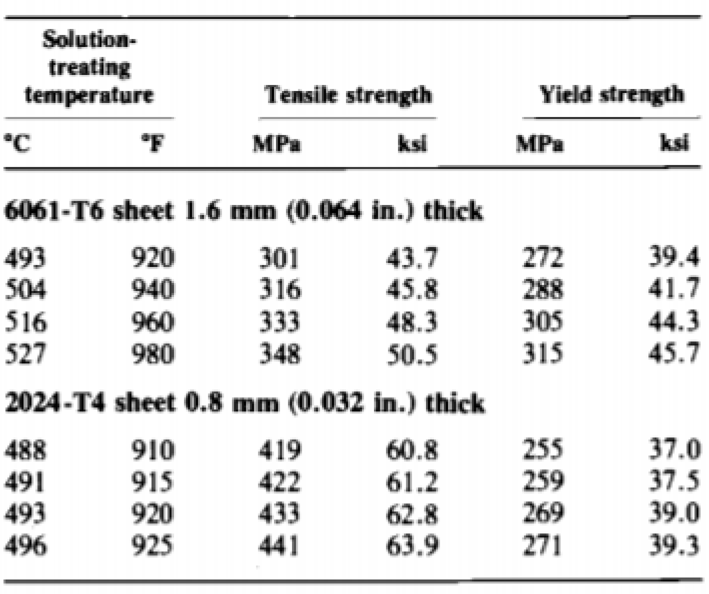
	Solution-treating temperature		Eutectic melting temperature	
Alloy	°C	°F	°C	°F
2014	496–507	925–945	510	950
2017	496–507	925–945	513	955
2024	488–499	910–930	502	935

Similar considerations apply to other agehardenable alloy systems such as aluminummagnesium-silicon alloys. For example, according to Fig 9(a), a 1.08% Mg_2Si alloy would be soaked at a temperature in excess of 500 °C (930 °F) but below the solidus of 595 °C (1100 °F) to avoid incipient melting. However, because some alloy constituents may form complex eutectics that melt at temperatures below the equilibrium eutectic temperature, the upper limit for solution treatment of aluminum-magnesium-silicon alloys is in the range of 515 to 540 °C (960 to 1000 °F). At 540 °C (1000 °F), about 0.6% Mg can be placed in solution (Fig 9b).

NonequUibrium Melting. When high heating rates are employed, the phenomenon of nonequilibrium melting must be considered. This phenomenon can also be explained with the help of the aluminum-copper phase diagram (Fig 8). The room-temperature microstructure of an F-temper product containing 4% Cu consists of a solid solution of copper in aluminum and particles of AI_2Cu. When this product is heated slowly, the AI_2Cu begins to dissolve, and if heating is slow enough, all of the Al_2Cu is dissolved when temperatures above the solvus (500 °C, or 932 °F) are reached. When the heating rate is high, however, much of the $A1_2Cu$ remains undissolved. If a material with this microstructure is heated at or above the eutectic temperature of 548 °C (1018 °F), melting will begin at the interface between the AI_2Cu and the matrix. With sufficient time above the eutectic temperature, this metastable liquid will dissolve to form a solid solution and will leave no trace provided that hydrogen gas has not condensed at the interface to form a void. If the product is quenched before the liquid has time to equilibrate, however, it will solidify and form fine eutectic rosettes. This nonequilibrium melting should not be confused with true equilibrium melting, which would occur in any alloy containing more than 5.65% Cu. In such an alloy, eutectic melting is equilibrium melting. No matter how long such an alloy is held above the eutectic temperature, the liquid will never solidify. In commercial alloys, which usually are temaries or quaternaries of the major alloying elements, the situation is more complex. Different phases have different solvus temperatures, and nonequilibrium melting may occur at different temperatures depending on composition, size of precipitates, and rate of heating. When new solution heat-treating equipment (which provides higher heating rates) is employed, careful examination of alloy microstructures should be included as part of the certification process.

Underheating. When the temperatures attained by the parts or pieces being heat treated are appreciably below the normal range, solution is incomplete, and strength somewhat lower than normal is expected. In the aluminum-copper system (Fig 8), the shallow slope of the solvus at its intersection with the composition line indicates that a slight decrease in temperature will result in a large reduction in the concentration of the solid solution and a correspondingly significant decrease in final strength. The effect of solution treating temperature on the strength of two aluminum alloys is illustrated by the following data:

Table.2. The effect of solution treating



Solution-treating temperature		Tensile strength		Yield strength	
°C	°F	MPa	ksi	MPa	ksi
6061-T6 sheet 1.6 mm (0.064 in.) thick					
493	920	301	43.7	272	39.4
504	940	316	45.8	288	41.7
516	960	333	48.3	305	44.3
527	980	348	50.5	315	45.7
2024-T4 sheet 0.8 mm (0.032 in.) thick					
488	910	419	60.8	255	37.0
491	915	422	61.2	259	37.5
493	920	433	62.8	269	39.0
496	925	441	63.9	271	39.3

Solution-Treating Time.

The time at the nominal solution heat-treating temperature (soak time) required to effect a satisfactory degree of solution of the undissolved or precipitated soluble phase constituents and to achieve good homogeneity of the solid solution is a function of microstructure before heat treatment. This time requirement can vary from less than a minute for thin sheet to as much as 20 h for large sand or plaster-mold castings. The time required to heat a load to the treatment temperature in furnace heat treatment also increases with section thickness and furnace loading, and thus total cycle time increases with these factors.

Soak time for alclad sheet and for parts made from alclad sheet must be held to a minimum, because excessive diffusion of alloying elements from the core into the cladding reduces corrosion protection. For the same reason, reheat treatment of alclad sheet less than 0.75 mm (0.030 in.) thick generally is prohibited, and the number of reheat treatments permitted for thicker alclad sheet is limited. The soak times for wrought alloys take into account the normal thermal lag between furnace and part and the difference between surface and center temperatures for commercial equipment qualified to the standards of MIL-H-6088. The rapid heating rates of salt baths permit all immersion time to be counted as soak time unless the bath temperature drops below the minimum of the range. Even then, soak time begins as soon as the bath temperature returns to the minimum. In air furnaces, soak time does not begin until all furnace instruments return to their original set temperature--that is, the temperature reading before insertion of the load.

In air furnaces, thermocouples may also be attracted to, or buried in, parts located in the load in such a manner as to represent the hottest and coldest temperatures in each zone. In this way, it is possible to ensure that adequate soaking is obtained. Special consideration is given also to establishing soak times for hand and die forgings; soak time in some specifications is extended to complete solution and homogenization in areas that received marginal reduction during forging. Considerable variation exists in the amount of soak time added; some specifications call for an arbitrary addition, such as one hour, and others require one hour per inch of thickness of the original forging.

In air furnaces, careful attention should be given to arrangement of the load. Air flow and natural temperature distribution within the furnace should be arranged to:

- Offer minimum resistance to air flow
- Produce the least disturbance in the natural temperature distribution
- Afford constant replenishment of the envelope of air around each part

High-Temperature Oxidation.

There is a condition, commonly but erroneously known as HTO or high-temperature oxidation, which can lead to deterioration of properties in aluminum alloys. High-temperature oxidation is a misnamed condition of hydrogen diffusion that affects surface layers during elevated-temperature treatment. This condition can result from moisture contamination in the furnace atmosphere and is sometimes aggravated by sulfur (as in heat-treatment furnaces also used for magnesium alloy castings) or other furnace refractory contamination. Moisture in contact with aluminum at high temperatures serves as a source of nascent hydrogen, which diffuses into the metal. Foreign materials, such as sulfur compounds, function as decomposers of the natural oxide surface film, eliminating it as a barrier either between the moisture and the aluminum or between the nascent hydrogen and the aluminum. The most common manifestation of high-temperature oxidation is surface blistering, but occasionally the only manifestations are internal discontinuities or voids, which can be detected only by careful ultrasonic inspection or by metallographic techniques.

It is important to recognize that the symptoms of high-temperature oxidation are identical to those of unsoundness or high gas content in the original ingot or of other improper mill practice. Blisters resulting from ingot defects, improper extrusion or improper rolling may be lined up in the direction of working. However, it usually is impossible to distinguish among defect sources, and therefore the possibility that a contaminated atmosphere is the cause of the defects must be checked. Not all alloys and product forms are equally vulnerable to this type of attack. The 7xxx series alloys are most susceptible, followed by the 2xxx alloys. Extrusions undoubtedly are the most susceptible form; forgings are probably second. Low-strength alloys and alclad sheet and plate are relatively immune to high-temperature oxidation. (Blistering of alclad material as a result of inadequate bonding is not the same as the blistering caused by high-temperature oxidation.) If the protective oxide film formed during mill operations is removed from the mill product by a subsequent mechanical conditioning operation such as sanding, the conditioned surface will be more susceptible to high-temperature oxidation than those from which the film was not removed.

Moisture can be minimized by thoroughly drying parts and racks before they are charged. Drain holes often are needed in racks of tubular construction to avoid en trapment of water. Another common requirement is adjustment of the position of the quench tank with respect to furnace doors and air intake. Because it is unlikely that all moisture can be eliminated from the atmosphere in a production heat-treating furnace, it is extremely important to eliminate all traces of other contaminants from both the parts and the furnace atmosphere. The most virulent contaminants in attacking aluminum are sulfur compounds. Residues from forming or machining lubricants, or from a sulfur dioxide protective atmosphere used in prior heat treatment of magnesium, are potential sources of sulfur contamination. In one plant, surface contamination resulted from sulfur-containing materials in tote boxes used to transport parts. In another, an epidemic of blistering was cured by rectifying a "sour" degreaser. In a third instance, it was found that a vapor-degreasing operation was not completely removing a thin, hard waxy residue, and an alkaline cleaning operation was added.

Quenching

Quenching is in many ways the most critical step in the sequence of heat-treating operations. The objective of quenching is to preserve the solid solution formed at the solution heat-treating temperature, by rapidly cooling to some lower temperature, usually near room temperature. From the preceding general discussion, this statement applies not only to retaining solute atoms in solution, but also to maintaining a certain minimum number of vacant lattice sites to assist in promoting the low-temperature diffusion required for zone formation. The solute atoms that precipitate either on grain boundaries, dispersoids, or other particles, as well as the vacancies that migrate (with extreme rapidity) to disordered regions, are irretrievably lost for practical purposes and fail to contribute to the subsequent strengthening.

In most instances, to avoid those types of precipitation that are detrimental to mechanical properties or to corrosion resistance, the solid solution formed during solution heat treatment must be quenched rapidly enough (and without interruption) to produce a supersaturated solution at room temperature---the optimum condition for precipitation hardening. The resistance to stress-corrosion cracking of certain copperfree aluminum-zinc-magnesium alloys, however, is improved by slow quenching. Most frequently, parts are quenched by immersion in cold water or, in continuous heat treating of sheet, plate, or extrusions in primary fabricating mills, by progressive flooding or high-velocity spraying with cold water. However, parts of complex shape, often with both thin and thick sections (such as die forgings, most castings, impact extrusions, and components formed from sheet) are commonly quenched in a medium that provides somewhat slower cooling. This medium may be water at 65 to 80 °C (150 to 180 °F), boiling water, an aqueous solution of polyalkylene glycol, or some other fluid medium such as forced air or mist. If appreciable precipitation during cooling is to be avoided, two requirements must be satisfied. First, the time required for transfer of the load from the furnace to the quenching medium must be short enough to preclude slow precooling into the temperature range where very rapid precipitation takes place. For alloy 7075, this range was determined to be 400 to 290 °C (750 to 550 °F), and some sources quote this range (or a slightly different range) as the most critical range for quenching of any aluminum alloy. Later work has shown that the most critical range is alloy-dependent, and as will be discussed in detail under "Quench-Factor Analysis," significant errors can result from the assumption that precipitation is negligible outside of a so-called "critical range.

The second requirement for avoidance of appreciable precipitation during quenching is that the volume, heat-absorption capacity, and rate of flow of the quenching medium be such that little or no precipitation occurs during cooling. Any interruption of the quench that might allow reheating into a temperature range where rapid precipitation can occur must be prohibited. For maximum dimensional stability, some forgings and castings are fan cooled or still-air cooled. In such instances, precipitation-hardening response is limited, but satisfactory values of strength and hardness are obtained. Extrusions produced without separate solution heat treatment can be air or mist quenched, but thicker sections may require water quenching by immersion or spraying. Alloys that are relatively dilute, such as 6063 and 7005, are particularly well suited to air quenching, and their mechanical properties are not greatly affected by its low cooling rate. Lower quenching rates are also employed for forgings, castings, and complex shapes to minimize warpage or other distortion and the magnitude of residual stresses developed as a consequence of temperature nonuniformity from surface to interior.

Age Hardening

After solution treatment and quenching, hardening is achieved either at room temperature (natural aging) or with a precipitation heat treatment (artificial aging). In some alloys, sufficient precipitation occurs in a few days at room temperature to yield stable products with properties that are adequate for many applications. These alloys sometimes are precipitation heat treated to provide increased strength and hardness in wrought or cast products. Other alloys with slow precipitation reactions at room temperature are always precipitation heat treated before being used. In some alloys, notably those of the 2xxx series, cold working of freshly quenched material greatly increases its response to later precipitation heat treatment. Mills take advantage of this phenomenon by applying a controlled amount of rolling (sheet and plate) or stretching (extrusion, bar, and plate) to produce higher mechanical properties. However, if the higher properties are used in design, reheat treatment must be avoided.

Natural Aging. The more highly alloyed members of the 6xxx wrought series, the copper-containing alloys of the 7xxx group, and all of the 2xxx alloys are almost always solution heat treated and quenched. For some of these alloys--particularly the 2xxx alloys--the precipitation hardening that resuits from natural aging alone produces useful tempers (T3 and

T4 types) that are characterized by high ratios of tensile to yield strength and high fracture toughness and resistance to fatigue. For the alloys that are used in these tempers, the relatively high supersaturation of atoms and vacancies retained by rapid quenching causes rapid formation of GP zones, and strength increases rapidly, attaining nearly maximum stable values in four or five days. Tensile-property specifications for products in T3- and T4-type tempers are based on a nominal natural aging time of four days. In alloys for which T3- or T4-type tempers are standard, the changes that occur on further natural aging are of relatively minor magnitude, and products of these combinations of alloy and temper are regarded as essentially stable after about one week.

In contrast to the relatively stable condition reached in a few days by 2xxx alloys that are used in T3- or T4-type tempers, the 6xxx alloys and to an even greater degree the 7xxx alloys are considerably less stable at room temperature and continue to exhibit significant changes in mechanical properties for many years. Because of the relative instability of the 7xxx alloys, the naturally aged temper (after solution heat treatment and quenching) is designated by the suffix letter W. For a specific description of this condition, the time of natural aging should be included (example: 7075-W, 1 month).

Aging characteristics vary from alloy to alloy with respect to both time to initial change in mechanical properties and rate of change, but aging effects always are lessened by reductions in aging temperature (see Fig 21). With some alloys, aging can be suppressed or delayed for several days by holding at a temperature of -18 °C (0 °F) or lower. It is usual practice to complete forming and straightening before aging changes mechanical properties appreciably. When scheduling makes this impractical, aging may be avoided in some alloys by refrigerating prior to forming. It is conventional practice to refrigerate alloy 2024-T4 rivets to maintain good driving characteristics. Full-size wing plates for current-generation jet aircraft have been solution heat treated and quenched at the primary fabricating mill, packed in dry ice in specially designed insulated shipping containers and transported by rail about 2000 miles to the aircraft manufacturer's plant for forming.

Unanticipated difficulties may arise as a result of failure to control refrigerator or part temperature closely enough. If opening of the cold box to insert or remove parts is done too frequently, the cooling capacity of the refrigerator may be exceeded. At times, the rate at which heavy-gage parts can be cooled in a still-air cold box has been found to be insufficient. This problem has been solved in one plant by immersing parts in a solvent at -40 °C (-40 °F) before placing them in the refrigerator.

Heat Treating of Magnesium Alloys

Magnesium alloys usually are heat treated either to improve mechanical properties or as means of conditioning for specific fabricating operations. The type of heat treatment selected depends on alloy composition and form (cast or wrought), and on anticipated service conditions.

Solution heat treatment improves strength and results in maximum toughness and shock resistance. Precipitation heat treatment subsequent to solution treatment gives maximum hardness and yield strength, but with some sacrifice of toughness. As applied to castings, artificial aging without prior solution treatment or annealing is a stress-relieving treatment that also somewhat increases tensile properties. Annealing of wrought products lowers tensile properties considerably and increases ductility, thereby facilitating some types of fabrication. Modifications of these basic treatments have been developed for specific alloys, to obtain the most desirable combinations of properties.

The basic temper designations for magnesium alloys, the same as those applied to aluminum alloys, are used throughout to indicate the various types of heat treatment.

The mechanical properties of most magnesium casting alloys can be improved by heat treatment. Casting alloys can be grouped into six general classes of commercial importance on the basis of composition, as follows:

- Magnesium-aluminum-manganese
- Magnesium-aluminum-zinc
- Magnesium-zinc-zirconium
- Magnesium-rare earth metal-zinc-zirconium
- Magnesium-rare earth metal-silver-zirconium, with or without thorium
- Magnesium-thorium-zirconium, with or without zinc.

In most wrought alloys, maximum mechanical properties are developed through strain hardening, and these alloys generally are either used without subsequent heat treatment or merely aged to a T5 temper. Occasionally, however, solution treatment, or a combination of solution treatment with strain hardening and artificial aging, will substantially improve mechanical properties.

Wrought alloys that can be strengthened by heat treatment are grouped into four general classes according to composition:

- Magnesium-aluminum-zinc
- Magnesium-thorium-zirconium
- Magnesium-thorium-manganese
- Magnesium-zinc-zirconium.

Types of Heat Treatment

Annealing. Wrought magnesium alloys in various conditions of strain hardening or temper can be annealed by being heated at 290 to 455°C (550 to 850°F), depending on alloy, for one or more hours. This procedure usually will provide a product with the maximum anneal that is practical.

Because most forming operations on magnesium are done at elevated temperature, the need for fully annealed wrought material is less than with many other metals.

Stress Relieving of Wrought Alloys. Stress relieving is used to remove or reduce residual stresses induced in wrought magnesium products by cold and hot working, shaping and forming, straightening, and welding.

When extrusions are welded to hard rolled sheet, the lower stress-relieving temperature and the longer time should be used to minimize distortion-for example, 150°C (300°F) for 60 min, rather than 260°C (500°F) for 15 min.

Stress Relieving of Castings. The precision machining of castings to close dimensional limits, the necessity of avoiding warp age and distortion, and the desirability of preventing stress-corrosion cracking in welded magnesium-aluminum casting alloys make it mandatory that cast components be substantially free from residual stresses. Although magnesium castings do not normally contain high residual stresses, the low modulus of elasticity of magnesium alloys means that comparatively low stresses can produce appreciable elastic strains.

Residual stresses may arise from contraction due to mold restraint during solidification, from no uniform cooling after heat treatment, or from quenching. Machining operations also can result in residual stress and require intermediate stress relieving prior to final machining.

Solution Treating and Aging. In solution treating of magnesium-aluminum-zinc alloys, parts should be loaded into the furnace at approximately 260°C (500°F) and then raised to the appropriate solution-treating temperature slowly, to avoid fusion of eutectic compounds and resultant formation of voids. The time required to bring the load from 260°C to the solution-treating temperature is determined by the size of the load and by the composition, size, weight and section thickness of the parts, but 2 h is a typical time.

During aging, magnesium alloy parts should be loaded into the furnace at the treatment temperature, held for the appropriate period of time, and then cooled in still air. There is a choice of artificial aging treatments for some alloys; results are closely similar for the alternative treatments given.

Reheat Treating. Under normal circumstances, when mechanical properties are within expected ranges and the prescribed best treatment has been carried out, reheat treating is seldom necessary. However, if the microstructures of heat treated castings indicate too high a compound rating or if the castings have been aged excessively by slow cooling after solution treating, reheat treating is called for. Most magnesium alloys can be reheating treated with little danger of germination.

Effects of Major Variables

Casting size and section thickness, relation of casting size to volume capacity of the furnace, and arrangement of castings in the furnace are mechanical considerations that can affect heat treating schedules for all metals.

Section Size and Heating Time. There is no general rule for estimating time of heating per unit of thickness for magnesium alloys. However, because of the high thermal conductivity of these alloys, combined with their low specific heat per unit volume, parts reach soaking temperature quite rapidly. The usual procedure is to load the furnace and then begin the soaking period when the loaded furnace reaches the desired temperature.

In the heat treating of magnesium alloy castings with thick sections a good rule is to double the time at the solution treating temperature. For example, the usual solution treatment for AZ63A castings is 12 h at about 385°C (725°F), whereas 25 h at about 385°C is suggested for castings with section thickness greater than 50 mm.

Similarly, the suggested solution-treating schedule for preventing excessive grain growth in AZ92A castings is 6 h at about 405°C, 2 h at about 350°C and 10 h at about 405°C; but for castings with sections more than 50 mm thick, it is recommended that the last soak at 405°C be extended from 10 h to 19 h. The best way to determine whether or not additional solution treating time is required is to cut a section through the thickest portion of a scrap casting and examine the center of the section microscopically: if heat treatment is complete, this examination will reveal a low compound rating.

Protective Atmospheres. Although magnesium alloys can be best treated in air, protective atmospheres are almost always used for solution treating. Government specification for heat treating of magnesium castings requires a protective atmosphere for solution treating above 400°C (750°F). Protective atmospheres serve the dual purpose of preventing surface oxidation (which, if severe, can decrease strength) and of preventing active burning should the furnace exceed proper temperature.

The two gases normally used are sulfur dioxide and carbon dioxide. Inert gases also may be used; however, in most instances, these gases are not practical because of higher cost. Sulfur dioxide is available bottled, while carbon dioxide may be obtained either bottled or as the product of recirculated combustion gases from a gas-fired furnace. A concentration of 0.7% (0.5% min) sulfur dioxide will prevent active burning to a temperature of 565°C (1050°F), provided that melting of the alloy has not occurred. Carbon dioxide in a concentration of 3% will prevent active burning to 510°C (950°F), and a carbon dioxide concentration of 5% will provide protection to about 540°C (1000°F).

Equipment and Processing

In solution treating and artificial aging of magnesium alloys, it is standard practice to use an electrically heated or gas-fired furnace equipped with a high-velocity fan or comparable means for circulating the atmosphere and promoting uniformity of temperature. However, because the atmosphere for solution treating sometimes contains sulfur dioxide, only furnaces that are gastight and that provide an inlet for introducing protective atmosphere are suitable.

Quenching Media. Magnesium alloy products normally are quenched in air following solution treatment. Still air usually is sufficient; forced-air cooling is recommended for dense loads or for parts that have very thick sections.

Dimensional Stability

In normal service up to approximately 95°C (200°F), all magnesium casting alloys exhibit good dimensional stability and can be considered free from additional dimensional changes.

Some cast magnesium-aluminum-manganese and magnesium-aluminum-zinc alloys in certain tempers exhibit slight permanent growth after relatively long exposure to temperatures exceeding 95°C. This growth, although slight, can give rise to problems.

In contrast to the growth characteristics of the magnesium-aluminum-zinc alloys are those of the magnesium alloys containing thorium, rare earth metals and zirconium as major alloying elements. These alloys normally are used in the T5 or T6 temper, and they shrink, rather than grow, on exposure to elevated temperatures.

Heat Treating of Titanium and Titanium Alloys

Titanium and titanium alloys are heat treated in order to:

- Reduce residual stresses developed during fabrication (stress relieving)
- Produce an optimum combination of ductility, machinability, and dimensional and structural stability (annealing)
- Increase strength (solution treating and aging)
- Optimize special properties such as fracture toughness, fatigue strength, and high-temperature creep strength.

Various types of annealing treatments (single, duplex, (beta), and recrystallization annealing, for example), and solution treating and aging treatments, are imposed to achieve selected mechanical properties. Stress relieving and annealing may be employed to prevent preferential chemical attack in some corrosive environments, to prevent distortion (a stabilization treatment) and to condition the metal for subsequent forming and fabricating operations.

Alloy Types and Response to Heat Treatment

The response of titanium and titanium alloys to heat treatment depends on the composition of the metal and the effects of alloying elements on the α-β crystal transformation of titanium. In addition, not all heat treating cycles are applicable to all titanium alloys, because the various alloys are designed for different purposes.

- Alloys Ti-5Al-2Sn-2Zr-4Mo-4Cr and Ti-6Al-2Sn-4Zr-6Mo are designed for strength in heavy sections.
- Alloys Ti- 6Al-2Sn-4Zr-2Mo and Ti-6Al-5Zr-0.5Mo-0.2Si for creep resistance.
- Alloys Ti-6Al-2Nb-1 Ta-1Mo and Ti-6Al-4V, for resistance to stress corrosion in aqueous salt solutions and for high fracture toughness.
- Alloys Ti-5Al-2.5Sn and Ti-2.5Cu for weldability; and
- Ti-6Al-6V-2Sn, Ti-6Al-4V and Ti-10V-2Fe-3Al for high strength at low-to-moderate temperatures.

Effects of Alloying Elements on α-β Transformation. Unalloyed titanium is allotropic. Its close-packed hexagonal structure (α phase) changes to a body-centered cubic, structure (β-phase) at 885°C (1625°F), and this structure persists at temperatures up to the melting point.

With respect to their effects on the allotropic transformation, alloying elements in titanium are classified as α stabilizers or β stabilizers. Alpha stabilizers, such as oxygen and aluminum, raise the α-to-β transformation temperature. Nitrogen and carbon are also stabilizers, but these elements usually are not added intentionally in alloy formulation. Beta stabilizers, such as manganese, chromium, iron, molybdenum, vanadium, and niobium, lower the α-to-β transformation temperature and, depending on the amount added, may result in the retention of some β phase at room temperature.

Alloy Types. Based on the types and amounts of alloying elements they contain, titanium alloys are classified as α, near-α, α-β, or β alloys. The response of these alloy types to heat treatment is briefly described below.

Alpha and near-alpha titanium alloys can be stress relieved and annealed, but high strength cannot be developed in these alloys by any type of heat treatment (such as aging after a solution beta treatment and quenching).

The commercial β alloys are, in reality, metastable β alloys. When these alloys are exposed to selected elevated temperatures, the retained β phase decomposes and strengthening occurs. For β alloys, stress-relieving and aging treatments can be combined, and annealing and solution treating may be identical operations.

Alpha-beta alloys are two-phase alloys and, as the name suggests, comprise both α and β phases at room temperature. These are the most common and the most versatile of the three types of titanium alloys.

Oxygen and iron levels have significant effects on mechanical properties after heat treatment. It should be realized that:

- Oxygen and iron must be near specified maximums to meet strength levels in certain commercially pure grades
- Oxygen must be near a specified maximum to meet strength levels in solution treated and aged Ti-6Al-4 V
- Oxygen levels must be kept as low as possible to optimize fracture toughness. However, the oxygen level must be high enough to meet tensile strength requirements

- Iron content must be kept as low as possible to optimize creep and stress-rupture properties. Most creep-resistant alloys require iron levels at or below 0.05wt%.

Stress Relieving

Titanium and titanium alloys can be stress relieved without adversely affecting strength or ductility.

Stress-relieving treatments decrease the undesirable residual stresses that result from first, nonuniform hot forging or deformation from cold forming and straightening, second, asymmetric machining of plate or forgings, and, third, welding and cooling of castings. The removal of such stresses helps maintain shape stability and eliminates unfavorable conditions, such as the loss of compressive yield strength commonly known as the Bauschinger effect.

When symmetrical shapes are machined in the annealed condition using moderate cuts and uniform stock removal, stress relieving may not be required. Compressor disks made of Ti-6Al-4V has been machined satisfactorily in this manner, conforming with dimensional requirements. In contrast, thin rings made of the same alloy could be machined at a higher production rate to more stringent dimensions by stress relieving 2 h at 540°C (1000°F) between, rough and final machining. Separate stress relieving may be omitted when the manufacturing sequence can be adjusted to use annealing or hardening as the stress-relieving process. For example, forging stresses may be relieved by annealing prior to machining.

Annealing

The annealing of titanium and titanium alloys serves primarily to increase fracture toughness, ductility at room temperature, dimensional and thermal stability, and creep resistance. Many titanium alloys are placed in service in the annealed state. Because improvement in one or more properties is generally obtained at the expense of some other property, the annealing cycle should be selected according to the objective of the treatment.

Common annealing treatments are:

- Mill annealing
- Duplex annealing
- Recrystallization annealing
- Beta annealing

Mill annealing is a general-purpose treatment given to all mill products. It is not a full anneal and may leave traces of cold or warm working in the microstructures of heavily worked products, particularly sheet.

Duplex annealing alters the shapes, sizes, and distributions of phases to those required for improved creep resistance or fracture toughness. In the duplex anneal of the Corona 5 alloy, for example, the first anneal is near the β transus to globularize the deformed α and to minimize its volume fraction. This is followed by a second, lower-temperature anneal to precipitate new lenticular (acicular) α between the globular α particles. This formation of acicular α is associated with improvements in creep strength and fracture toughness.

Recrystallization annealing and β annealing are used to improve fracture toughness. In recrystallization annealing, the alloy is heated into the upper end of the α-β range, held for a time, and then cooled very slowly. In recent years, recrystallization annealing has replaced β annealing for fracture critical airframe components.

β (Beta) Annealing. Like recrystallization annealing, β annealing improves fracture toughness. Beta annealing is done at temperatures above the β transus of the alloy being annealed. To prevent excessive grain growth, the temperature for β annealing should be only slightly higher than the β transus. Annealing times are dependent on section thickness and should be sufficient for complete transformation. Time at temperature after transformation should be held to a minimum to control β grain growth. Larger sections should be fan cooled or water quenched to prevent the formation of a phase at the β grain boundaries.

Straightening, sizing, and flattening of titanium alloys are often necessary in order to meet dimensional requirements. The straightening of bar to close tolerances and the flattening of sheet present major problems for titanium producers and fabricators.

Unlike aluminum alloys, titanium alloys are not easily straightened when cold because the high yield strength and modulus of elasticity of these alloys result in significant springback. Therefore, titanium alloys are straightened primarily by creep straightening and/or hot straightening (hand or die), with the former being considerably more prevalent than the latter.

Straightening, sizing, and flattening may be combined with annealing by the use of appropriate fixtures. The parts, in bulk or in fixtures, may be charged directly into a furnace operating at the annealing temperature. At annealing temperatures many titanium alloys have a creep resistance low enough to permit straightening during annealing.

Creep straightening may be readily accomplished during the annealing and/or aging processes of most titanium alloys. However, if the annealing/aging temperature is below about 540 to 650°C (1000 to 1200°F), depending on the alloy, the times required to accomplish the desired creep straightening can be extended. Creep straightening is accomplished with rudimentary or sophisticated fixtures and loading systems, depending on part complexity and the degree of straightening required.

Creep flattening consists of heating titanium sheet between two clean, flat sheets of steel in a furnace containing an oxidizing or inert atmosphere. Vacuum creep flattening is used to produce stress-free flat plate for subsequent machining. The plate is placed on a large, flat ceramic bed that has integral electric heating elements. Insulation is placed on top of the plate, and a plastic sheet is sealed to the frame.

Stability. In α-β titanium alloys, thermal stability is a function of β-phase transformations. During cooling from the annealing temperature, β may transform and, under certain conditions and in β alloys, may form a brittle intermediate phase known as ω.

A stabilization annealing treatment is designed to produce a stable β phase capable of resisting further transformation when exposed to elevated temperatures in service. Alpha-beta alloys that are lean in β, such as Ti-6Al-4V, can be air cooled from the annealing temperature without impairing their stability. To obtain maximum creep resistance and stability in the near-α alloys Ti-8Al-1 Mo-1 V and Ti-6Al-2Sn-4Zr-2Mo, a duplex annealing treatment is employed. This treatment begins with solution annealing at a temperature high in the α-β range, usually 25 to 55°C (50 to 100°F) below the β transus for Ti-8Al-1Mo-1Vand 15 to 25°C (25 to 50°F) below the α-β transus for Ti-6Al-2Sn-4Zr-2Mo.

Solution Treating and Aging

A wide range of strength levels can be obtained in α-β or β alloys by solution treating and aging. With the exception of the unique Ti-2.5Cu alloy (which relies on strengthening from the classic age-hardening reaction of Ti2Cu precipitation similar to the formation of Guinier-Preston zones in aluminum alloys), the origin of heat-treating responses of titanium alloys lies in the instability of the high-temperature β phase at lower temperatures.

Heating an α-β alloy to the solution-treating temperature produces a higher ratio of β phase. This partitioning of phases is maintained by quenching; on subsequent aging, decomposition of the unstable β phase occurs, providing high strength. Commercial β alloys generally supplied in the solution-treated condition, and need only to be aged.

After being cleaned, titanium components should be loaded into fixtures or racks that will permit free access to the heating and quenching media. Thick and thin components of the same alloy may be solution treated together, but the time at temperature is determined by the thickest section. Although preheating is not essential, it may be used to minimize the distortion of complex parts.

Solution treating of titanium alloys generally involves heating to temperatures either slightly above or slightly below the β transus temperature. The solution-treating temperature selected depends on the alloy type and practical considerations briefly described below.

β (Beta) alloys are normally obtained from producers in the solution-treated condition. If reheating is required, soak times should be only as long as necessary to obtain complete solutioning. Solution-treating temperatures for β alloys are above the β transus; because no second phase is present, grain growth can proceed rapidly.

α-β (Alpha-beta) alloys. Selection of a solution-treatment temperature for α-β alloys is based on the combination of mechanical properties desired after aging. A change in the solution-treating temperature of α-β alloys alters the amounts of β phase and consequently changes the response to aging.

To obtain high strength with adequate ductility, it is necessary to solution treat at a temperature high in the α-β field, normally 25 to 85°C (50 to 150°F) below the β transus of the alloy. If high fracture toughness or improved resistance to stress corrosion is required, β annealing or β solution treating may be desirable. However, heat treating α-β alloys in the β range causes a significant loss in ductility. These alloys are usually solution heat treated below the β transus to obtain an optimum balance of ductility, fracture toughness, creep, and stress rupture properties.

Corrosion: Introduction – Definitions and Types

Corrosion can be viewed as a universal phenomenon, omnipresent and omnipotent. It is there everywhere, air, water, soil and in every environment, we encounter. There is no single figure for loss to the nation due to corrosion. It can be a minimum of 3.5% of the nation"s GDP. Losses due to corrosion could be around Rs. 2.0 lakh crores per annum in India. Corrosion costs manifest in the form of premature deterioration or failure necessitating maintenance, repairs and replacement of damaged parts.

In the US, total direct cost of corrosion is estimated at about 300 billion dollars per year; which is about 3.2% of domestic product. Corrosion has a huge economic and environmental impact on all facets of national infrastructure; from highways, bridges, buildings, oil and gas, chemical processing, water and waste water treatment and virtually on all metallic objects in use. Other than material loss, corrosion interferes with human safety, disrupts industrial operations and poses danger to environment. Awareness to corrosion and adaptation of timely and appropriate control measures hold the key in the abatement of corrosion failures.

Corrosion is the deterioration or destruction of metals and alloys in the presence of an environment by chemical or electrochemical means. In simple terminology, corrosion processes involve reaction of metals with environmental species.

As per IUPAC, “Corrosion is an irreversible interfacial reaction of a material (metal, ceramic, polymer) with its environment which results in its consumption or dissolution into the material of a component of the environment. Often, but not necessarily, corrosion results in effects detrimental to the usage of the material considered. Exclusively physical or mechanical processes such as melting and evaporation, abrasion or mechanical fracture are not included in the term corrosion” With the knowledge of the role of various microorganisms present in soil and water bodies, the definition for corrosion need be further widened to include microbially influenced factors.

Corrosion can be classified in different ways, such as Chemical and electrochemical High temperature and low temperature Wet corrosion and dry corrosion.

Dry corrosion occurs in the absence of aqueous environment, usually in the presence of gases and vapours, mainly at high temperatures. Electrochemical nature of corrosion can be understood by examining zinc dissolution in dilute hydrochloric acid.

$$Zn + 2HCl = ZnCl_2 + H_2$$

Anodic reaction is

$$Zn = Zn^{++} + 2e$$

at cathodic areas on the surface of zinc metal. There are two half reactions constituting the net cell reaction.

Environmental effects such as those of presence of oxygen and other oxidizers, changes in flow rates (velocity), temperature, reactant concentrations and pH would influence rates of anodic and cathodic reactions.

Even though the fundamental mechanism of corrosion involves creation or existence of corrosion cells, there are several types or forms of corrosion that can occur. It should however be borne in mind that for corrosion to occur, there is no need for discrete (physically independent) anodes and cathodes. Innumerable micro level anodic and cathodic areas can be generated at the same (single) surface on which anodic (corrosion) and cathodic (reduction) reactions occur. Each form of corrosion has a specific arrangement of anodes and cathodes and specific patterns and locations depending on the type can exist.

The most important types are Uniform corrosion. Galvanic corrosion, concentration cells, water line attack Pitting. Dezincification, Dealloying (selective leaching) Atmospheric corrosion. Erosion corrosion Fretting Crevice corrosion; cavitation Stress corrosion, intergranular and transgranular corrosion, hydrogen cracking and embrittlement Corrosion fatigue.

Crevice corrosion is a localized attack on a metal adjacent to the crevice between two joining surfaces (two metals or metal-nonmetal crevices). The corrosion is generally confined to one localized area to one metal. This type of corrosion can be initiated by concentration gradients (due to ions or oxygen). Accumulation of chlorides inside crevice will aggravate damage. Various factors influence crevice corrosion, such as

- Materials: alloy composition,
- metallographic structure. Environmental conditions such as pH, oxygen concentration, halide concentrations, temperature.
- Geometrical features of crevices, surface roughness.
- Metal to metal or metal to nonmetal type

Filiform corrosion is a special type of crevice corrosion.

Pitting corrosion is a localized phenomenon confined to smaller areas. Formation of micro-pits can be very damaging. Pitting factor (ratio of deepest pit to average penetration) can be used to evaluate severity of pitting corrosion which is usually observed in passive metals and alloys. Concentration cells involving oxygen gradients or ion gradients can initiate pitting through generation of anodic and cathodic areas. Chloride ions are damaging to the passive films and can make pit formation auto-catalytic. Pitting tendency can be predicted through measurement of pitting potentials. Similarly critical pitting temperature is also a useful parameter.

Uniform corrosion is a very common form found in ferrous metals and alloys that are not protected by surface coating or inhibitors. A uniform layer of „rust" on the surface is formed when exposed to corrosive environments Atmospheric corrosion is a typical example of this type.

Galvanic corrosion often referred to as dissimilar metal corrosion occurs in galvanic couples where the active one corrodes. EMF series (thermodynamic) and galvanic series (kinetic) could be used for prediction of this type of corrosion. Galvanic corrosion can occur in multiphase alloys.

Example:

- Copper containing precipitates in aluminium alloys.
- Impurities such as iron and copper in metallic zinc.

Differential aeration (oxygen concentration cell) and ion concentration (salt concentration) cells create dissimilar polarities (anodic and cathodic areas) Eg:-Pitting of metals. Rusting of iron

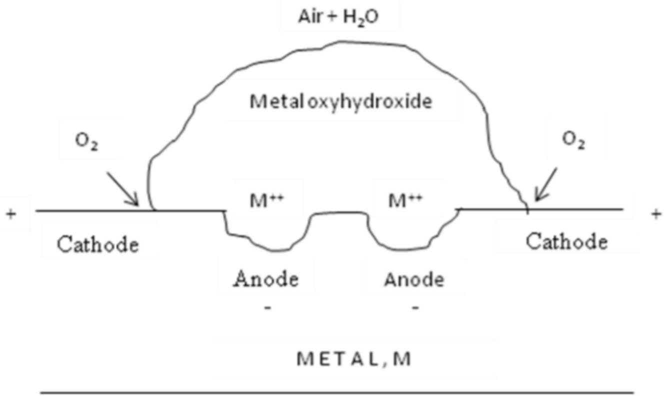


Fig. 9 Differential oxygen cells in rusting of iron

Selective leaching (Dealloying) refers to selective dissolution of active metal phase from an alloy in a corrosive environment.

Examples:

a) Brass containing copper and zinc. Since zinc is anodic to copper, selective dezincification occurs in a corrosive medium, enriching the cathodic copper in the matrix (colour of brass turns red from yellow).

b) Graphitization of grey cast iron-graphite being cathodic enhances dissolution of iron in the matrix, leaving behind a graphite network.

There are several other examples of dealloying besides the above.

- Tin Bronzes in hot brine or steam-Destannification.
- Precious metal alloys such as gold containing copper or silver – strong acids, sulfide environment - preferential dissolution of copper or silver.
- Cupro-nickel alloys in condenser tubes-denickelisation.

Localised attack at or nearer to grain boundaries in a metal or alloy can be termed as intergranular corrosion. Generally the following factors contribute to intergranular corrosion.

- Impurities and precipitation at grain boundaries.
- Depletion of an alloying element (added to resist corrosion) in the grainboundary area

A typical example is sensitized 18-8 stainless steels when chromium carbide is precipitated along grain boundaries. Lowered chromium content in the area adjacent to grain boundaries, leads to formation of anodic and cathodic areas. Such intergranular corrosion is common in stainless steel welded structures and is referred to as weld decay. Intergranular attack can occur in other alloys as well. For example, Duralumin-type alloys (Al – Cu) due to precipitation of $CuAl_2$.

Erosion corrosion is the deterioration of metals and alloys due to relative movement between surfaces and corrosive fluids. Depending on the rate of this movement, abrasion takes place. This type of corrosion is characterized by grooves and surface patterns having directionality. Typical examples are

- Stainless alloy pump impeller,
- Condenser tube walls.

All equipment types exposed to moving fluids are prone to erosion corrosion. Many failures can be attributed to impingement (impingement attack). Erosion corrosion due to high velocity impingement occurs in steam condenser tubes, slide valves in petroleum refinery at high temperature, inlet pipes, cyclones and steam turbine blades.

Cavitation damage can be classified as a special form of erosion corrosion. This is usually caused by formation and collapse of vapour bubbles in liquids closer to a metal surface. Typical examples include ship‟s propellers, pump impellers and hydraulic turbines. Surface damage similar to that of pitting can occur and both corrosion and mechanical factors are involved.

Corrosion occurring at contact regions between materials under load subjected to slip and vibration can be termed Fretting. Such friction oxidation can occur in engine and automotive parts. Fretting is known to occur at bolted tie plates on rails.

Parameters promoting fretting include:

- Relative motion between two surfaces.
- Interface under load.

Both the above produce slip and deformation of surfaces. Wear-oxidation and oxidation-wear theories are proposed to explain fretting corrosion.

Stress corrosion cracking (SCC) refers to failure under simultaneous presence of a corrosive medium and tensile stress. Two classic examples of SCC are caustic embrittlement of steels occurring in riveted boilers of steam-driven locomotives and season cracking of brasses observed in brass cartridge cases due to ammonia in environment. Stress cracking of different alloys does occur depending on the type of corrosive environment. Stainless steels crack in chloride atmosphere. Major variables influencing SCC include solution composition, metal/alloy composition and structure, stress and temperature. Crack morphology for SCC failures consists of brittle fracture and inter - or trans-granular cracking could be observed. Higher stresses decrease time before crack initiation. Tensile stresses of sufficient threshold levels are involved (applied, residual or thermal stresses).

Hydrogen embrittlement although many a time classified under stress corrosion, need be considered separately since the two types respond very differently to environmental factors. Fracture of metals and alloys under repeated cyclic stresses is termed fatigue and corrosion under such circumstances is corrosion fatigue (reduction of fatigue resistance).

Electrochemical factors come into play in many of the above corrosion forms. Both thermodynamic and kinetic aspects of electrochemistry of corrosion are discussed in the following lectures with respect to both corrosion mechanisms and corrosion protection

Corrosion Protection of Al Alloys

Aluminium and its alloys have excellent durability and corrosion resistance, but, like most materials, their behaviour can be influenced by the way in which they are used.(Although pure Aluminium lacks mechanical strength but has higher resistance to corrosion than its alloys). Aluminium is commercially important metal and its alloys are widely used in many industries such as reaction vessels, pipes, machinery and chemical batteries because of their advantages. Aluminium is a well-known sacrificial anode if couple with more passive metal as it is most reliable and cost effective anode. Aluminium sacrificial anode has been used in major project all over the world. It is used in offshore application including structures, platforms, pipelines, jetties and power plants. Aluminium anode is also used for ship-hull and ballast tank protection. Aluminium is an active metal and its resistance to corrosion depends on the formation of the protective oxide film. There are several methods to protect the aluminium and its alloys from corrosion such as coatings (Metallic, Inorganic, conversion and organic coatings), control of environment (operating variable i.e. pH, dissolved oxygen, temperature) and corrosion inhibitors (organic and inorganic additives). Aluminium can be alloyed with different elements like zinc, magnesium, silicon, copper, manganese, as well as lithium. As a result, it can be used for different applications like manufacturing of aluminium foil covering, food packaging industry, food and chemical industry, vehicle panelling, mine cages, air frames, chemical plants, pressure vessels, road tankers, transportation of ammonium nitrate, irrigation pipes and window frames.

Some industrial applications for aluminium alloy are listed below:

1. Aluminium alloys are highly resistant to non-heat treatments. They are good conductors of heat and electricity and that is why they are being used in different chemical industries for preparation of aluminium products.

2. Aluminium alloys get hardened during the process of reactions. That is why they are highly favourable alloys for the factor of weld ability as well as formability. Also, they are superior for the cryogenic uses even in the condition of annealed treatments.

3. Aluminium alloys is its high resistance to corrosion. They deter it from different harsh chemical treatments and help in retaining its lustre and strength. It also has high resistance to sea and ocean water. Therefore, it can be used for different air cages without any hassles.

4. Aluminium alloys can be converted into any form as they are ductile in nature. Be it sheets or wires, they can be drawn into various shapes without any inconvenience.

5. Aluminium alloy also used as sacrificial anode for the cathodic protection system (pipeline cathodic protection, oil tank, ship hull aluminium cp anode,)

Corrosion in nonmetals

Most ceramic materials are almost entirely immune to corrosion. The strong chemical bonds that hold them together leave very little free chemical energy in the structure; they can be thought of as already corroded. When corrosion does occur, it is almost always a simple dissolution of the material or chemical reaction, rather than an electrochemical process.

A common example of corrosion protection in ceramics is the lime added to soda-lime glass to reduce its solubility in water; though it is not nearly as soluble as pure sodium silicate, normal glass does form sub-microscopic flaws when exposed to moisture. Due to its brittleness, such flaws cause a dramatic reduction in the strength of a glass object during its first few hours at room temperature.

Corrosion of glasses: Glass disease is the corrosion of silicate glasses in aqueous solutions. It is governed by two mechanisms: diffusion-controlled leaching (ion exchange) and hydrolytic dissolution of the glass network. Both mechanisms strongly depend on the pH of contacting solution: the rate of ion exchange decreases with pH as 10−0.5pH whereas the rate of hydrolytic dissolution increases with pH as 100.5pH.

Influence of Corrosion on the Structure of Materials

The effect of corrosion on mechanical properties will be shown in uniform corrosion and pitting corrosion, some of the most common mechanisms that we deal with on a daily basis. Uniform corrosion, as the name implies, spreads uniformly over the entire surface of the part in question. Pitting corrosion is a highly localized corrosion where a minute breakage of a protective (passive) layer leads to the creation of a pit where localized galvanic corrosion occurs through a very complex corrosion mechanism, even if the surface layer was resealed.

In the case of uniform corrosion, assessing its influence on the material’s structural strength is straightforward. One should establish the reduction of the thickness and the weight-loss of the plate or beam in question, and calculate the stress for a given load for a new, reduced, cross section of the steel member. The matter is far more complicated for localized corrosion, where the localized stress fields play a much larger role and directly affect the tensile strength of the material itself.

A decrease of nominal tensile strength and total elongation was a function of the thickness (depth of pits) of the corroded layer. With the loss of the member’s thickness due to corrosion, tensile strength decreased gradually, while the maximum elongation decreased in a steeper fashion. Nominal tensile strength is not the same as the ultimate tensile strength (UTS) – one is a property of a structure or its component, and the other is a property of a material. In this case, nominal strength refers to the maximum load over the original cross section area (which is obviously reduced due to corrosion).

This reduction in maximum load carrying capacity was observed to be 2.5 times greater in a member with pitting corrosion compared to a member with uniform corrosion.

When examining small specimens, the maximum allowed tensile load could still be predicted by using the ultimate strength of the material (multiplying the cross section by the ultimate tensile strength).

When wide specimens (plates) are used, stress concentrations caused local plastic deformations and local fracture initiation in the immediate vicinity of the larger pits. This means that the stress threshold for crack initiation for wide specimens is lower than for small specimens.

Besides these effects on tensile strength, pitting corrosion also had a greater negative impact on the buckling strength of a member compared to uniform corrosion.

On the other hand, examination of corroded reinforcing steel bars in a study showed a significant reduction of ductility, and the specimens with more than 12.6% showed considerable brittle behavior compared to the control specimens. In addition, these bars showed reduced elongation. After a certain point (when more than 40% of the material is corroded), these rebars showed preferential corrosion, with a relatively short section showing thinning and notch formation, but these could be attributed to cracked concrete and other environmental factors not directly related to corrosion. This reduction in rebar thickness significantly reduces its load carrying capacity, while the brittle mode of failure makes the situation far more dangerous due to the much faster propagation of cracks in such a material.

Tensile Strength, Ductility and Brittleness

Corrosion damages both physical and mechanical properties of material. The following are effect of corrosion on mechanical properties.

Strength

Thickness reduction due to corrosion directly affects strength of materials. For example, 5” S-135 drill pipes premium class should have tensile strength of 436 klb; however, excessive corrosion damages internal and external surface area of drill pipes. The smaller surface area will result in reduction of tensile strength. Furthermore, it is very difficult to predict the strength of materials when localized corrosion occurs because a surface area of cracking don’t evenly distribute. Moreover, some of corrosive environments such as high temperature, high CO_2&H_2S, high chloride content, etc can dramatically degrade material properties.

Toughness

Corrosion reduces toughness of materials because it can physically and chemically change properties of materials and tough material can be brittle. Additionally, low temperature environment can dramatically decrease toughness. Therefore, equipment used in low temperature conditions must be designed to be able to work in very low temperature environment.

Ductility

Corrosion can change ductile materials into brittle material and this causes failure of structure. Several situations leading to ductility reduction are low temperature, H_2S & CO_2 gas, cyclic load, etc.

Protection against Corrosion

The following methods are used to protect metals against corrosion:

I. Selection of the right material of construction

II. Surface coating.

III. Inhibitors

IV. Proper equipment design

V. Electrical protection

I. Selection of the right material of construction

The right material of construction should have the following properties:

(1) high mechanical strength

(2) high corrosion resistance

(3) low cost

The material selection is carried out through the following steps :

1. Preliminary selection Based on experience, availability and safety aspects

2. Laboratory testing Revaluation of apparently suitable materials under process conditions

3. Interpretation of laboratory results and other data Effect of possible impurities, excess temperature, excess pressure, agitation, and presence of air in equipment

4. Economic comparison of apparently suitable materials Material and maintenance cost, probable life, cost of product degradation, and liability to special hazards

5. Final selection

II. Surface coating

There are two types of surface coating:

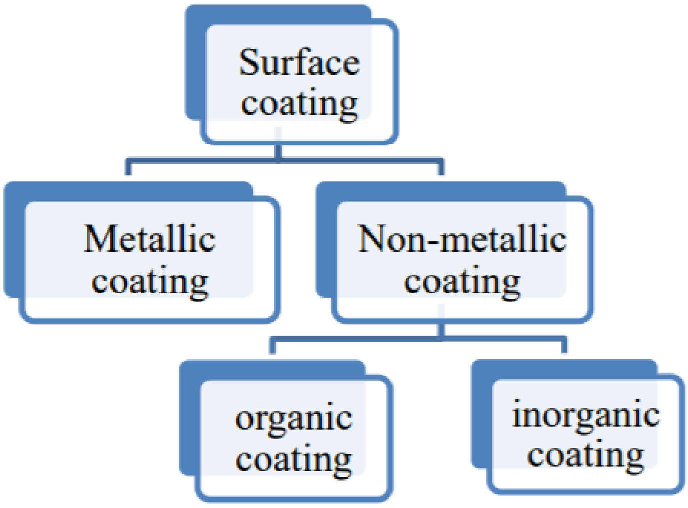


Fig. 10 Types of surface coating

A. Metallic coating

The structure is coated with a layer of other metal which may be more noble than the structure or less noble than it e.g. steel structures can be coated with copper which is more noble than steel or zinc which is less noble. In case of coating the structure with a more noble metal care should be taken that the coat is free from pores or cracks to avoid the formation of dissimilar metal corrosion cells which would lead to corrosion of the structure.

Factors that must be considered in selection of a coating metal:

1. The coating should be able to resist direct attack of the environment.

2. The coating should be nonporous and continuous (no cracks) to a void acceleration of corrosion especially in case of a more noble metal (ex: coating of Fe by Cu).

3. The coating should be hard. In case of coating the structure with a less noble metal the presence of pores and cracks in the coat is not dangerous because in this case the less noble metal will corrode by the formation of dissimilar metal cells while the structure will remain protected.

In practice metallic coating is carried out by different methods such:

(1) Electroplating,

(2) Hot dipping of the work piece in molten metal covered with a flux,

(3) Spraying of the molten metal on the work piece. Surface coating Metallic coating Non-metallic coating organic coating inorganic coating

(1) Electroplating: Electroplating is the method of coating one metal with another. It is most commonly used for¬ decorative purposes, appearance and protection. Electroplated items include chrome bumpers, jewelry, electronics, circuit boards and airplane¬ parts.

Electroplating procedure

1. Preparation of the workpiece. Solutions such as alkaline cleaners, solvent degreasers or acidic pickling mixtures are used to remove dirt, greases, oxidation and contaminants from the piece.

2. The piece to be plated is connected to the negative pole (cathode) of the d. c. power supply while the plating (coating) metal anode is connected to the positive pole (anode). Multi-range ammeter (in series) and voltmeter (in parallel) are connected to the cell to measure the cell current and voltage.

3. The piece is then immersed in the plating solution until coated and rinsed and then buffed or polished, if necessary.

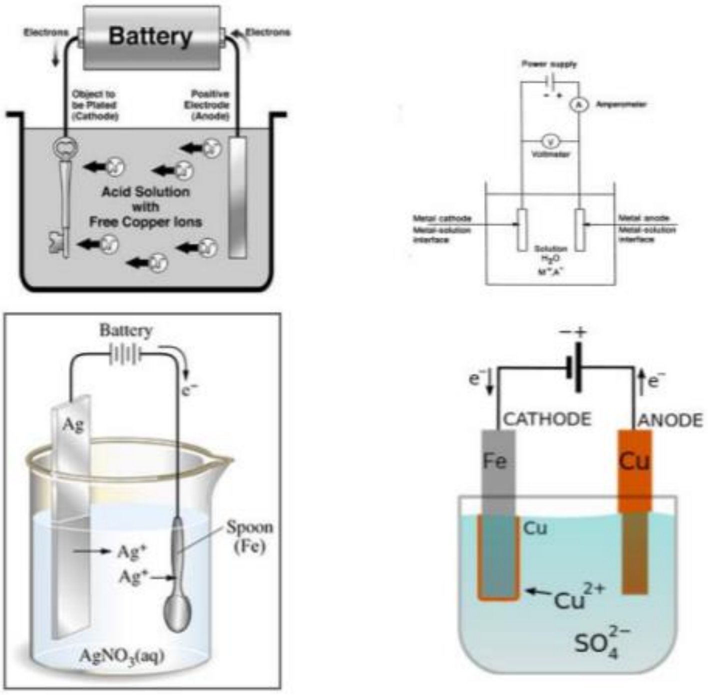


Fig. 11. Examples of Electroplating process

Hot dipping of the work piece in molten metal covered with a flux:

There are two common processes of hot dipping:

1. **Hot dipping galvanizing:** It is a hot coating process whereby the cleaned steel is immersed in molten zinc usually at a¬ temperature of between 445 °C and 450 °C. When the cleaned steel is immersed into the molten zinc, a chemical reaction results, which¬ we refer to as following “metallurgical laws”. As a result of this process the

coating consists of a series of zinc iron alloy layers (intermetallic layer) and usually a top pure zinc layer. The adhesion of the coating to the steel is¬ therefore determined by means of a chemical bond, or a “metallurgical bond”. Such bonding is considered to be far superior to that of a mechanical bond. A hot dipped galvanized coating will provide greater corrosion protection to steel when¬ compared to that of an electroplated product.

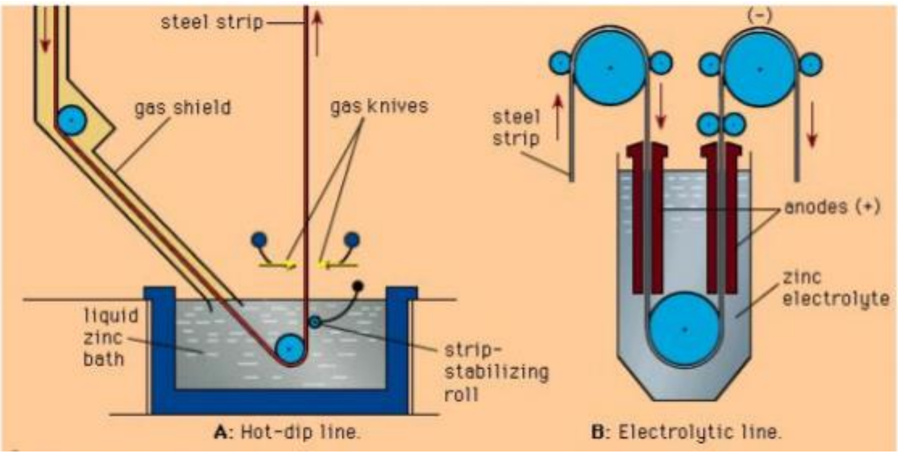


Fig. 12. Galvanizing of steel

2. **Hot-dipped tin plating:**

Tinning is the process of thinly coating sheets of iron or steel with tin, and the resulting product is known as tinplate. It is most often used to prevent rust Tinplate made via hot-dipped tin plating is made by cold rolling steel or iron, pickling or remove any scale, annealing to remove any strain hardening, and then coating it with a thin layer of tin. The attached figure shows the tinning process steps. Fig. 4, Basic concept of hot dipped tin plating process Advantage of hot dipping process No waste from production process.¬ No hazardous substance (such as cyanogens, lead, etc.) is used at all in production process.¬ The coating metal and the base metal are strongly bonded as inter-metallic layer formed.¬ Provide greater corrosion protection to steel when compared to that of an electroplated¬ product.

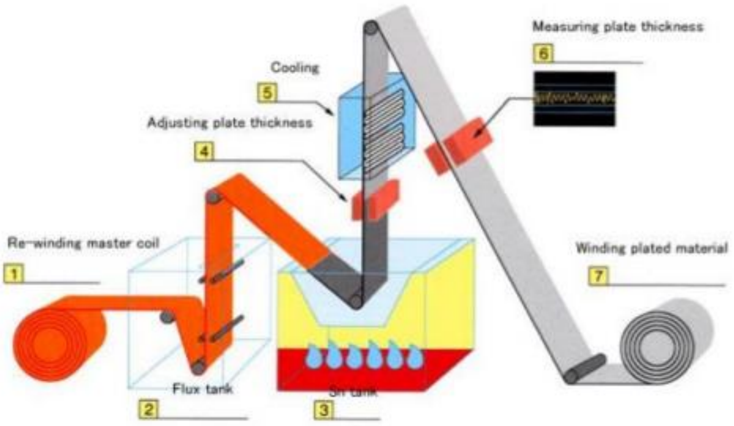


Fig. 13. Hot-dipped tin plating

Advantage of hot dipping process:

¬ No waste from production process.

¬ No hazardous substance (such as cyanogens, lead, etc.) is used at all in production process.

¬ The coating metal and the base metal are strongly bonded as inter-metallic layer formed. ¬ Provide greater corrosion protection to steel when compared to that of an electroplated product.

3. **Spraying of the molten metal on the work piece:**

   - Thermal spraying refers to a process by which a metal wire or powder is melted and¬ sprayed onto a surface to form a coating.
   - A thermal spray gun is used to apply the coatings.
   - The thermal spray gun heats the metallic wire or powder to a molten state and compressed¬ air or other gas propels it onto the surface to form a coating. The compressed gas also aids in division and atomization of the molten coating.
   - The two metals most commonly applied by thermal spray are zinc and aluminum. These metals and their alloys provide excellent protection in a variety of marine and industrial corrosive environments.

Non-metallic coating (the second method of surface coating) :

There are two types of nonmetallic coating:

A. Inorganic coating

   Methods of inorganic coating:

   i. **Oxidation (passivation)**

      Steel can be coated with an oxide film by a). Heating at high temperature, b). Chemical oxidation by treating steel with hot alkaline nitrate, or persulphate or perchlorate c). Anodic oxidation by making the steel structure an anode in electrolytic cell

   ii. **Phosphating**

      Steel is coated with a layer of iron phosphate by dipping in a solution containing phosphoric acid and zinc phosphate. The iron phosphate film is not highly protective because it is porous so it usually covered with paint. The phosphate film improves the bond between the metal and the paint. i.e. Phosphating is a pre-painting step.

   iii. **Enamels**

      Enamels are glassy layer applied to the metal by dipping it in a suspension of powdered glass, and then the metal is heated in a stove (furnace) at high temperature where the glass powder melts and coat the metal.

   iv. **Cement coating**

      It is used to coat the inner side of steel pipelines carrying water or wastewater.

B. Organic coating such as the following:

i. Paints:

Paint consists of;

1. a film forming substance such as linseed oil or a polymer (resin),

2. an organic solvent and

3. a pigment (usually an inorganic oxide or metal powder). Before applying paint to a steel surface, the metal surface should be cleaned of oxides by sand blasting or acid pickling. After cleaning, the metal surface is coated with a thin layer of primer.

Primer

A primer is a paint containing a pigment such as lead oxide (Pb_3O_4 red lead) or zinc chromate which oxidizes the steel surface and inhibits its corrosion. Besides, the primer film increases the strength of the bond between steel and final paint film. When the primer film dries a thick film of the required paint is applied over the primer. Polymers such as alkyd resins, PVC, polyethylene, polyesters, acrylics, polyurethanes, chlorinated rubber, epoxy resins, etc. are used in paint manufacture. For severe conditions such marine and industrial atmosphere, a paint containing epoxy resin (water resistant) is suitable.

ii. Lacquers:

A lacquer consists of a thermoplastic polymer dissolved in an organic solvent. Lacquers can be used to line steel tanks holding corrosive chemicals such as acids.

iii.Coal tar

Coal tar is a brown or black liquid of extremely high viscosity. Coal tar is among the by-products when coal is carbonized to make coke or gasified to make coal gas. Coal tars are complex and variable mixtures of phenols, polycyclic aromatic hydrocarbons (PAHs), and heterocyclic compounds, about 200 substances in all. 12 It is used to protect underground structures. A 50% coal tar + 50% epoxy is superior to coal tar.

iv.Temporary coating:

It is used to protect metallic structures during shipping and storage by coating the structure with layer of lubricating oil which can be removed by an organic solvent when the structure is put to service.

III. Corrosion Inhibitors Required:

Corrosion inhibitors are substances that are added in small amount (e.g 0.1%) to the corrosive medium stop or slow down electrochemical corrosion reactions on a metal surface.

Mechanism:¬

Corrosion inhibitors work by one or more of the following mechanisms.

- They adsorb on metal surfaces to form protective films.
- They combine with corrosion product films to protect metal surfaces.
- They form precipitates, which visibly coat and protect metal surfaces.

Types of inhibitors:

Inhibitors can be divided into two main categories—inorganic and organic.

Inorganic inhibitors are used mainly in boilers, cooling towers, and fractionation units. Organic inhibitors are used mainly in oil field systems.¬

As shown in the figure, inorganic inhibitors are further divided into anodic and cathodic classifications. These classifications describe the part of the electrochemical process that is interrupted by the inhibitor. Organic film-forming inhibitors interrupt both the anodic and cathodic processes.

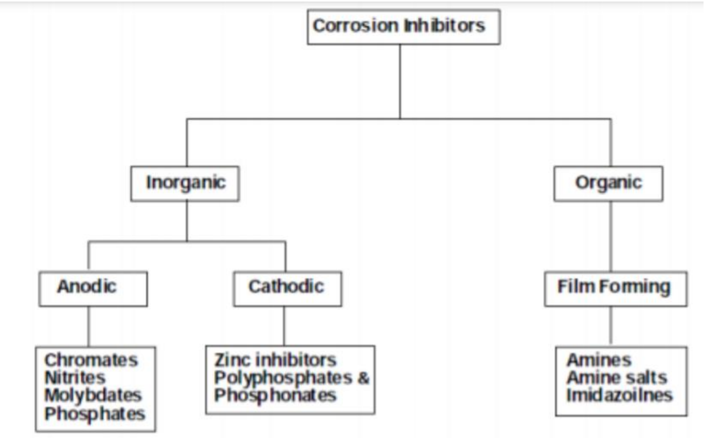


Fig. 14. Classification of Corrosion Inhibitors

Cathodic inhibitors: Cathodic inhibitors are chemical compounds which inhibit the cathodic reaction of the corrosion cell.

Anodic inhibitors: Are chemical compounds which inhibit the anodic reaction of the corrosive cells.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF MECHANICAL ENGINEERING

DEPARTMENT OF AERONAUTICAL ENGINEERING

UNIT – IV - MECHANICAL BEHAVIOUR OF AIRCRAFT MATERIALS

– SAE1616

UNIT 4 Fatigue of Materials

Fatigue of Materials

The concept of fatigue is very simple, when a motion is repeated, the object that is doing the work becomes weak. For example, when you run, your leg and other muscles of your body become weak, not always to the point where you can't move them anymore, but there is a noticeable decrease in quality output. This same principle is seen in materials. Fatigue occurs when a material is subject to alternating stresses, over a long period of time. Examples of where Fatigue may occur are: springs, turbine blades, airplane wings, bridges and bones.

Cyclic Stresses

There are three common ways in which stresses may be applied: axial, torsional, and flexural. Examples of these are seen in Fig. 1. There are also three stress cycles with which loads may be applied to the sample.

1.Reversed stress cycle: The is the simplest form of fatigue and is merely a sine wave where the maximum stress and minimum stress differ by a negative sign. An example of this type of stress cycle would be in an axle, where every half turn or half period as in the case of the sine wave, the stress on a point would be reversed.

2. Repeated stress cycle: This is the most common type of cycle found in engineering applications, where the maximum stress (smax) and minimum stress (smin) are asymmetric (the curve is a sine wave) not equal and opposite.

3. A final type of cycle mode is where stress and frequency vary randomly. An example of this would be automobile shocks, where the frequency magnitude of imperfections in the road will produce varying minimum and maximum stresses.

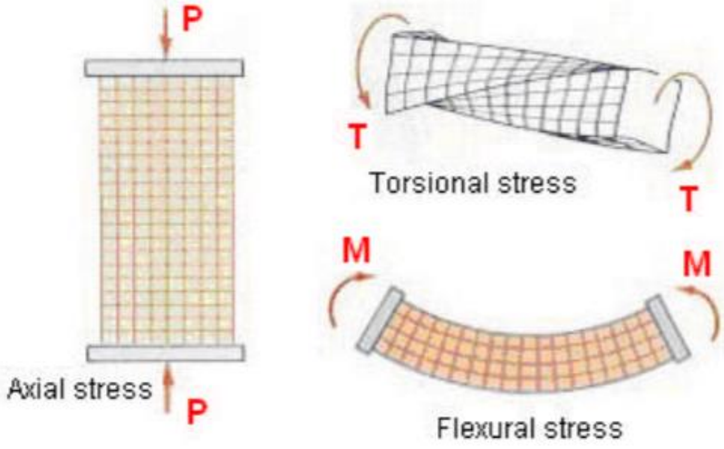


Fig 1. Visual examples of axial stress, torsional stress, and flexural stress.

SN curve

The S-N Curve A very useful way to visualize time to failure for a specific material is with the S-N curve. The "S-N" means stress verse cycles to failure, which when plotted uses the stress amplitude, sa plotted on the vertical axis and the logarithm of the number of cycles to failure. An important characteristic to this plot as seen in Fig. 2 is the fatigue limit

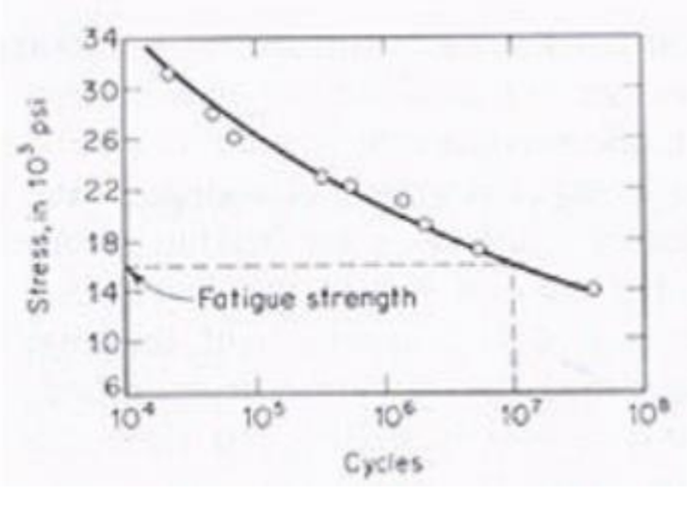


Fig.2 A S-N Plot for an aluminium alloy

The significance of the fatigue limit is that if the material is loaded below this stress, then it will not fail, regardless of the number of times it is loaded. Material such as aluminum, copper and magnesium do not show a fatigue limit, therefore they will fail at any stress and number of cycles. Other important terms are fatigue strength and fatigue life. The stress at which failure occurs for a given number of cycles is the fatigue strength. The number of cycles required for a material to fail at a certain stress in fatigue life.

Crack Initiation and Propagation

Failure of a material due to fatigue may be viewed on a microscopic level in three steps:

1. Crack Initiation: The initial crack occurs in this stage. The crack may be caused by surface scratches caused by handling, or tooling of the material; threads (as in a screw or bolt); slip bands or dislocations intersecting the surface as a result of previous cyclic loading or work hardening.

2. Crack Propagation: The crack continues to grow during this stage as a result of continuously applied stresses

3. Failure: Failure occurs when the material that has not been affected by the crack cannot withstand the applied stress. This stage happens very quickly.

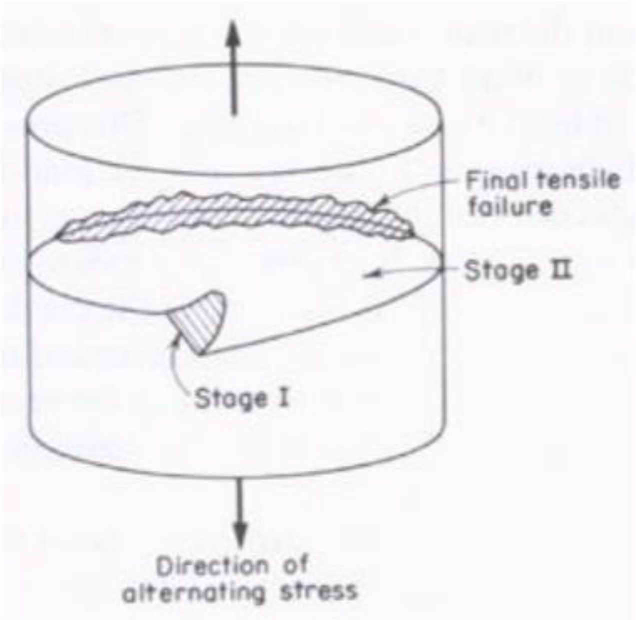


Figure 3 A diagram showing location of the three steps in a fatigue fracture

One can determine that a material failed by fatigue by examining the fracture sight. A fatigue fracture will have two distinct regions; One being smooth or burnished as a result of the rubbing of the bottom and top of the crack(steps 1 & 2); The second is granular, due to the rapid failure of the material. These visual clues may be seen in Fig. 4:

Other features of a fatigue fracture are Beachmarks and Striations. Beachmarks, or clamshell marks, may be seen in fatigue failures of materials that are used for a period of time, allowed to rest for an equivalent time period and the loaded again as in factory usage. Striations are thought to be steps in crack propagation, were the distance depends on the stress range. Beachmarks may contain thousands of striations. Visual Examples of Beachmarks and Striations are seen below in Fig. 5 and 6:

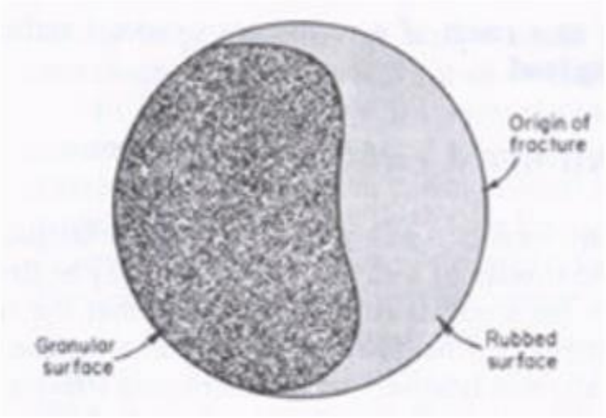


Fig. 4 A diagram showing the surface of a fatigue fracture

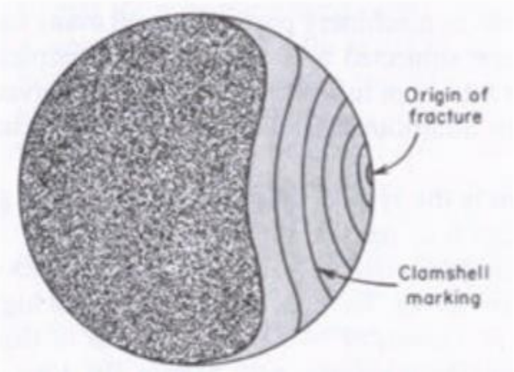


Fig. 5 An example of beachmarks or "clamshell pattern" associated with stress cycles that vary in magnitude and time as in factory machinery

Fig 6 An example of the striations found in fatigue fracture.

Demonstration of Crack Propagation Due to Fatigue

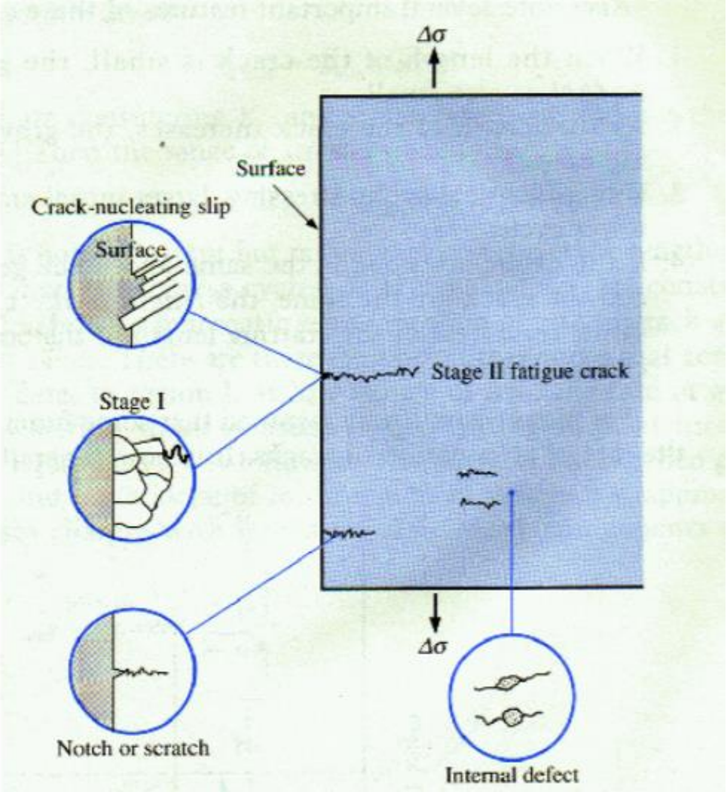


Fig.7 The various ways in which cracks are initiated and the stages that occur after they start

The figure above illustrates the various ways in which cracks are initiated and the stages that occur after they start. This is extremely important since these cracks will ultimately lead to failure of the material if not detected and recognized. The material shown is pulled in tension with a cyclic stress in the y ,or horizontal, direction. Cracks can be initiated by several different causes, the three that will be discussed here are nucleating slip planes, notches. and internal flaws. This figure is an image map so all the crack types and stages are clickable.

Propagation Rate

The rate at which a crack grows has considerable importance in determining the life of a material. The propagation of a crack occurs during the second step of fatigue failure. As a crack begins to propagate, the size of the crack also begins to grow. The rate at which the crack continues to grow depends on the stress level applied. The rate at which a crack grows can be seen mathematically in equation:

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

The variables A and m are properties of the material, da is the change in crack length, and dN is the change in the number of cycles. K is the change in the stress intensity factor or by equation

$$\Delta K = K_{\max} - K_{\min} = Y\Delta\sigma\sqrt{\pi a}$$

Factors That Affect Fatigue Life and Solutions

The Mean Stress is defined as:

$$\sigma_{m} = \frac{\sigma_{\max} - \sigma_{\min}}{2}$$

The Mean stress has the affect that as the mean stress is increased, fatigue life decreases. This occurs because the stress applies is greater. I mentioned previously that scratches and other imperfections on the surface will cause a decrease in the life of a material. Therefore making an effort to reduce these imperfections by reducing sharp corners, eliminating unnecessary drilling and stamping, shot peening, and most of all careful fabrication and handling of the material. Another Surface treatment is called case hardening, which increases surface hardness and fatigue life. This is achieved by exposing the component to a carbon-rich atmosphere at high temperatures. Carbon diffuses into the material filling interstisties and other vacancies in the material, up to 1 mm in depth

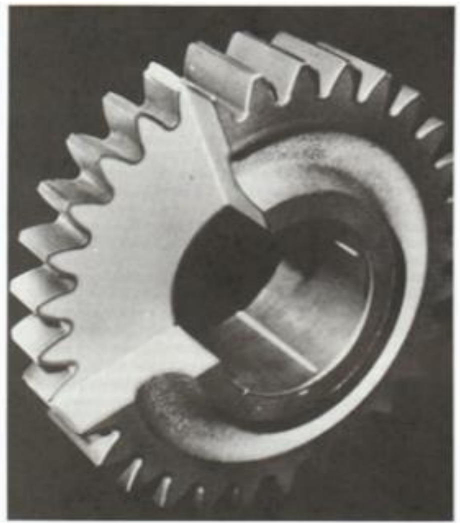


Fig. 8 A case hardened steel gear. Notice the effect of diffusion of Carbon into the material produces a "case" around the gear.

Exposing a material to high temperatures is another cause of fatigue in materials. Thermal expansion, and contraction will weaken bonds in a material as well as bonds between two different materials. For example, in space shuttle heat shield tiles, the outer covering of silicon tetraboride (SiB4) has a different coefficient of thermal expansion than the CarbonCarbon Composite. Upon re-entry into the earth's atmosphere, this thermal mismatch will cause the protective covering to weaken, and eventually fail with repeated cycles. Another environmental affect on a material is chemical attack, or corrosion. Small pits may form on the surface of the material, similar to the effect etching has when trying to find dislocations.

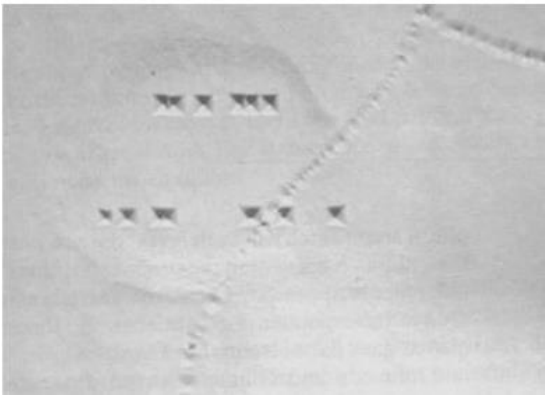


Fig. 9 Example of pits formed by corrosion on the surface of LiF. The "chemical" attacks weak spots on the surface of the material, especially where dislocations intersect the surface.

This chemical attack on a material can be seen in unprotected surface of an automobile, whether it be by road salt in the winter time or exhaust fumes. This problem can be solved by adding protective coatings to the material to resist chemical attack.

High cycle fatigue and Low cycle fatigue

High-cycle fatigue—where stresses and strains are mainly restricted to the elastic region. High-cycle fatigue is related with low loads and long life. The Stress-Life (S-N) or Total Life procedure is extensively used for high-cycle fatigue approach —here the applied stress is within the elastic range of the material and the number of cycles to failure is large.

Low-cycle fatigue—where significant plastic straining occurs. Low-cycle fatigue involves large cycles with significant amounts of plastic deformation and relatively short life. The analytical procedure used to address strain-controlled fatigue is commonly referred to as the Strain-Life, Crack-Initiation, or Critical Location approach.

While low-cycle fatigue is typically related with fatigue life between 10 to 10^5cycles, high-cycle fatigue is related with life greater than 10^5cycle.to analyze both high-cycle fatigue and low-cycle fatigue, a number of methods have been developed by combining the Manson-Coffin mode and Basquin model.

High cycle fatigue

High cycle fatigue could be sub-divided into-

- High cycle fatigue (HCF) – between 10^5 and 10^7
- Very low life cycles VHCF) – over 10^7 cycles to failure

The VHCF of materials is a phenomenon that first became acknowledged and evoked scientific interest only a few decades ago. It was observed that some materials when subjected to a sufficiently high number of load cycles (10^8 – 10^{10}) fail at stress levels that traditionally were considered as safe. Prior to this, it was believed that if a material survives 10^6 – 10^7 load cycles than it would never fail with increasing number of cycles at the same stress level.

High cycle fatigue must consider during design of automobiles, aircraft, compressors, pumps, turbines etc. where vibration occur. High-cycle fatigue data are generally introduced by a plot of stress, S, versus the number of cycles to failure, N. A log scale is associated with the number of cycles. The value of stress, S, can be the maximum stress, σmax, the minimum stress, σmin, or the stress amplitude, σa. The S-N relationship is generally determined for a describe for the value of the mean stress, σm, or one of the two ratios, stress ratio or amplitude ratio.

For determinations of the S-N curve, the usual procedure is to test the first specimen at a high stress where failure is expected in a short number of cycles, e.g., at about two-thirds the static tensile strength of the material. The test stress is decreased for each succeeding specimen until one or two specimens do not fail in the specified numbers of cycles, which is usually at least 10^7 cycles. This method used for presenting fatigue failure in high cycles ($N > 10^5$).

The number of cycle to failure at a identified stress level is referred as fatigue life , while the fatigue strength (also mentioned as the endurance limit) is the stress below which failure does not occur. As the applied stress level decreases, the number of cycles to failure increases.

Normally, the fatigue strength increases as the static tensile strength increases. To the contrary, non-ferrous systems, particularly aluminum and aluminum-based alloys do not possess any well-defined level of stress below which these materials do not fail and therefore, it is critical to define endurance limit for non-ferrous systems. This curve is used to predict the number of cycles sustained under certain stress before failure. The curve gives designers a quick reference of the allowable stress level for an intended service life given bellow in figure

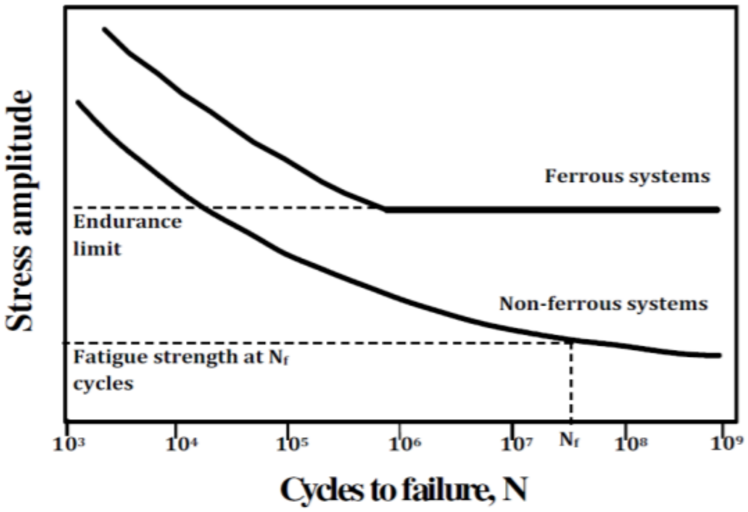


Fig. 10 Schematic representation of the S-N curve.

Low cycle Fatigue

Low cycle fatigue could be also sub-divided into-

- Low cycle fatigue (LCF)– for less than 10^4 cycles to failure.
- Very low life cycles - for less than 100 cycles to failure.

The very low cycle regime is well-known now and is titled as ‘‘extremely low cycle fatigue” (ELCF) to differentiate from the linear part of the life cycle curve that fits the Manson–Coffin law, which occasionally introduced as the ‘‘low cycle fatigue” in a more exact sense. The failure approaches are also dissimilar in these two regimes. For example, in push–pull fatigue tests of cylindrical bars, the fracture in the ELCF range often occurs at the center of the cylindrical bar; while in the strict LCF regime, the crack often starts from the surface. Efforts have been made to control the disadvantage of Manson– Coffin law in the ELCF regime. Low cycle fatigue is the repeated cyclic loadings that cause remarkable plastic distortion in a material.

Low cycle fatigue generally occurs because of repeated localized yielding near stress raisers, such as holes, fillets, and notches, despite the elastic deformation occurring over the bulk of the component. Uniaxial testing is performed on several smooth (unnotched) specimens under different cyclic deformation levels in a typical low cycle fatigue test. The mode of testing is strain control instead of stress control. Stress response monitored during cyclic loading, and the number of cycles to failure is recorded for these tests. The results from several tests are necessary to determine the cyclic stress strain curve and the strain life curve figure 2

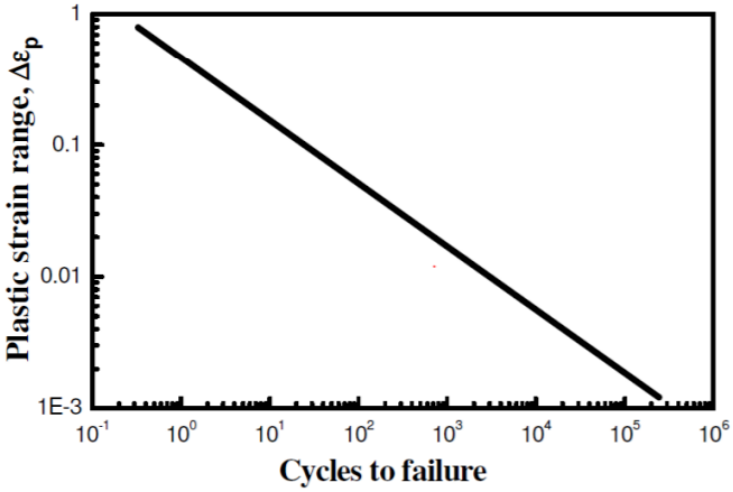


Fig. 11 Low cycle fatigue curve ($\Delta\varepsilon_p$vs. N).

Fatigue Strain-Life Relationship

Under the cyclic loading within elastic region, stress is directly propositional to strain through modulus of elasticity. However, plastic strains which is produced by cyclic loading, and responses are more complicated and create a hysteresis loop in figure 12 from the point O to the point A, the section is in tension. On removing of the load the strain, reaction of the specimen comes after in order to the curve from A to the D. At D, the section is under no stress. As the specimen is subjected to compressive stress, the strain reaction come after in order the curve from D to the point B. delivering the compressive Stress from B and reapplying the tensile stress, the specimen stress-strain condition come back to the point A along the curve explain by B, C, and A. Points A and B represent the cyclic stress and strain limits. The total strain σe, consists of both elastic and plastic components

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p$$

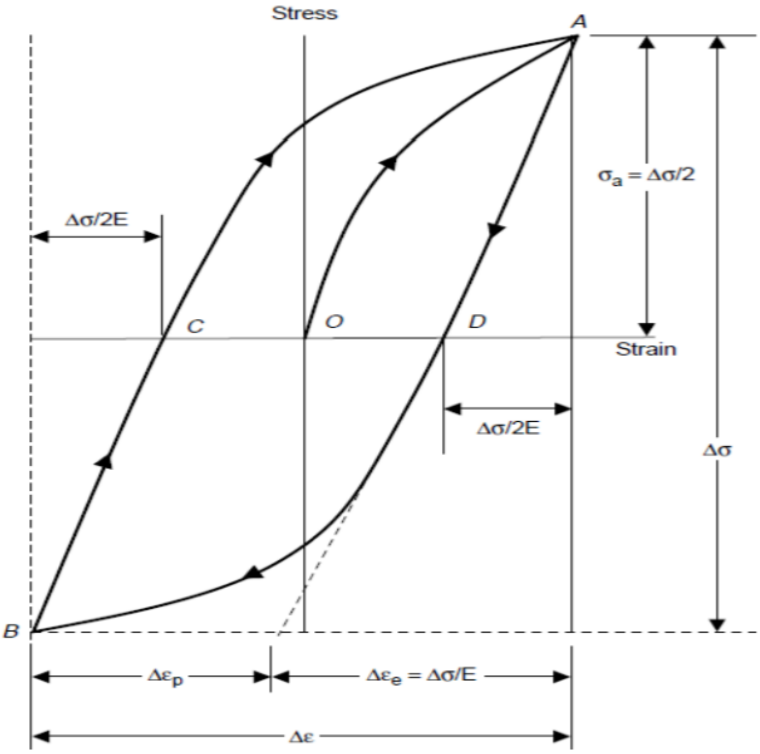


Fig. 12, stable cyclic stress versus strain hysteresis loop

where Δεe is the elastic strain and Δεe=σs/E, and Δεpis the plastic strain and Δεp is the width of the loop at its center, i.e., the distance CD in figure 2.6The area of the hysteresis loop is equal to the energy loss or the work done per cycle.

Low-cycle fatigue test data are commonly shows as a plot of the plastic strain range, Δεp, versus cycles to failure, N. When plotted on log-log coordinates, a straight line is obtained that is described by the Coffin-Manson relation:

$$\frac{\Delta\varepsilon_{\mathrm{p}}}{2}=\varepsilon_{\mathrm{f}}'(2N)^{c}$$

whereΔεp=2 is the plastic strain amplitude, and έf is the fatigue ductility coefficient defined by the strain intercept at 2N=1. For many metals, έf roughly equal to the true fracture strain, εf. 2N is the number of strain turnaround to failure, and c is the fatigue ductility exponent, which is usually varies between -0.5 and - 0.7. A smaller value of c results in longer fatigue lives.

The Basquin equation, which describes the high-cycle fatigue , low-strain regime where the nominal strains are elastic, is:

$$\sigma_{\mathrm{a}}=\frac{\Delta\varepsilon_{\mathrm{e}}}{2}E=\sigma_{\mathrm{f}}'(2N)^{b}$$

Where: σa is the alternating stress amplitude, Δεe=2 is the elastic strain amplitude, E is the young’s modulus, and σ’f is the fatigue strength coefficient, expressed as the stress intercept at 2N=1. σ’f is roughly requal to the true fracture stress, σf. 2N is the number of load turnaround to failure, and b is the fatigue strength exponent, which varies for most metals between -0.05 and -0.12. A smaller b results in a longer fatigue life. Combining Basquin’s equation and the Coffin-Manson equation gives an equation that can be used to estimate the entire range of fatigue lives:

$$\frac{\Delta\varepsilon}{2}=\frac{\Delta\varepsilon_{\mathrm{e}}}{2}+\frac{\Delta\varepsilon_{\mathrm{p}}}{2}$$

For high-cycle strain conditions, ductile metals have the longest lives, while at low-cycle strain conditions; strong metals have the longest lives.

Monotonic Stress Strain behaviour:

A stress strain curve for monotonic uniaxial tensile loading followed by an elastic unloading is shown schematically in Fig. 3. The total strain $\mathscr{E}$ is the sum of the elastic strain, $\mathscr{E}^e$, plus plastic strain, $\mathscr{E}^p$:

$$\epsilon = \epsilon^e + \epsilon^p.$$

…………………….. (1)

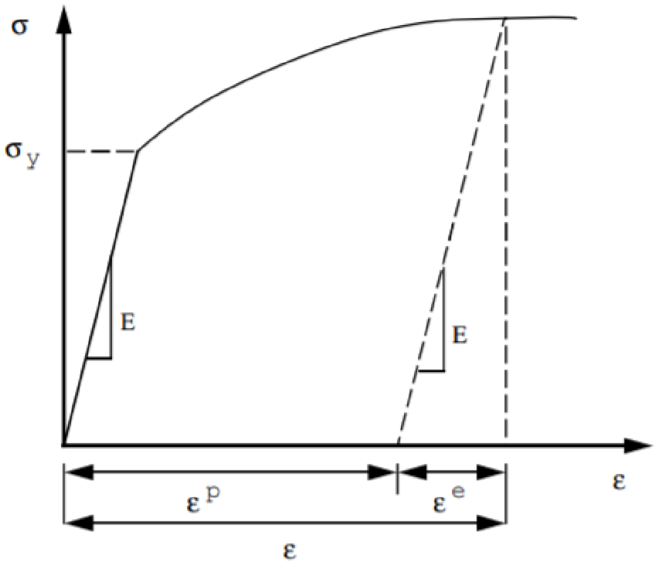


Fig. 3: Schematic of stress strain behavior for monotonic loading

The constitutive assumption which relates the elastic strain to stress is

$$\epsilon^e = \frac{\sigma}{E},$$

………………………….(2)

where E is Young's modulus, and σ is the stress. The plastic strain $\mathscr{E}^p$ is negligibly small for stresses less than a certain value, σy, termed the initial yield strength. For higher values of σ, it is often found that a power law relation of the form,

$$\epsilon^p = \left(\frac{\sigma}{K}\right)^{1/n} \quad \text{or} \quad \sigma = K\left(\epsilon^p\right)^n,$$

…………(3)

connects $\mathscr{E}^p$ and σ very well for $\mathscr{E}^p \geq .002$ or so. The dimensionless material parameter n is called the strain hardening exponent, with typical values $.01 < n \leq .4$, and the stress dimensioned material parameter K is called the strength coefficient. Upon inserting equations (2, 3) into equation (1), we obtain,

$$\boxed{\epsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K}\right)^{1/n}.}$$ …………..(4)

This equation is often referred to as a Ramberg-Osgood stress-strain relation, in recognition of the early proponents of this phenomenological stress-strain relation.

It can be seen from equation (3) that the parameter (σ/K) is raised to a large power, 1/n, and hence for (σ/K) << 1, the plastic strain is exceedingly small. Thus, there is often little error associated with the use of equation (4) even for stresses σ which are less than the initial yield strength, σy.

At the point of fracture in a tension test, the true stress at fracture, σf , is defined by,

$$\sigma_f = \frac{P_f}{A_f},$$ …………………………..(5)

where Pf is the fracture load and A_f is the fracture surface area. The true strain at fracture, $\mathscr{E}^f$, (using the assumption of incompressibility) can be expressed as

$$\epsilon_f = \ln\left(\frac{A_0}{A_f}\right)$$ ……………………..(6)

where A_0 is the initial cross-sectional area.

Since in ductile materials the plastic strain at fracture greatly exceeds the elastic strain, we may say that $\mathscr{E}^f$ is essentially equal to the plastic strain at fracture. We can then substitute $\mathscr{E}^f$ and σf into equation (4) to solve for K as

$$K = \frac{\sigma_f}{(\epsilon_f)^n}.$$

……………………….(7)

On substituting equation (7) into equation (4) we obtain

$$\epsilon = \frac{\sigma}{E} + \epsilon_f \left(\frac{\sigma}{\sigma_f}\right)^{1/n}.$$

The Ramberg-Osgood equation contains four material constants: E, σf , $\mathcal{E}^{f}$, and n.

Mean Stress Effects on Fatigue

We take the mean stress associated with a cyclic loading between stress limits σmax and σmin to be σm = (σmax + σmin)/2, with cyclic amplitude σa = (σmax −σmin)/2. Mean stress effects on fatigue life are most important at long lives, where cyclic plastic straining is small. At low lives, with significant plastic straining, mean stresses quickly relax out under strain-controlled limits, or lead to cyclic ratcheting and “runaway” if it is attempted to enforce unequal stress limits. Although many design approaches have been proposed to account for longlife mean stress effects on fatigue, we will adopt the assumption that a tensile mean stress, σm > 0, will correspondingly reduce the effective fatigue strength coefficient of equation. The modified form of the Basquin relation then becomes,

$$\sigma_a = \left(\sigma'_f - \sigma_m\right) \cdot (2N_f)^b; \quad (\text{for } \sigma_m > 0).$$

The governing equation for the strain life approach is as

$$\epsilon_a = (\frac{\sigma'_f - \sigma_m}{E}) \cdot (2N_f)^b + \epsilon'_f \cdot (2N_f)^c,$$

The effect of a positive mean stress on the strain life curve is to reduce the long life asymptote by $\sigma m/E$. If the mean stress was compressive, $\sigma m < 0$, then the use of above eq. would predict that the long-life fatigue behavior would considerably exceed that obtained under conditions of zero mean stress.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF MECHANICAL ENGINEERING

DEPARTMENT OF AERONAUTICAL ENGINEERING

UNIT – V - MECHANICAL BEHAVIOUR OF AIRCRAFT MATERIALS

– SAE1616

UNIT 5 Fracture Mechanics

Fracture failure

Fracture mechanics is based on the implicit assumption that there exists a crack in a work component. The crack may be man-made such as a hole, a notch, a slot, a re-entrant corner, etc. The crack may exist within a component due to manufacturing defects like slag inclusion, cracks in a weldment or heat affected zones due to uneven cooling and presence of foreign particles. A dangerous crack may be nucleated and grown during the service of the component

Modes of fracture failure

A crack front in a structural component is a line usually of varying curvature. Thus, the state of stress in the vicinity of the crack front varies from one point to another. A segment of the crack front can be divided into three basic modes as shown in Fig. 1. Mode I is the opening mode and the displacement is normal to the crack surface. Mode II is a sliding mode and the displacement is inthe plane of the plate-the separation is antisymmetric and the relative displacement is normal to the crack front. Mode III also causes sliding motion but the displacement is parallel to the crack front, thereby causing tearing.

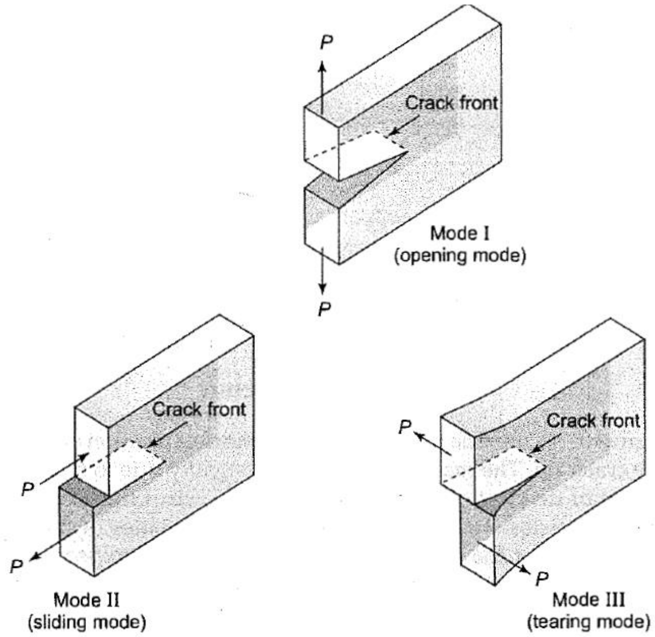


Fig. 1. The three modes of fracture

An inclined crack front in a component can be modeled as a superposition of the three basic modes and then, the effect of loading by each mode can be analyzed separately. Mode I usually play a dominant role in many engineering applications and is considered to be the most dangerous. However, in certain applications, components fail through the dominant roles played by Mode II or Mode III. Mode I has been analyzed most so far. Also, elaborate experimental methods have been developed to determine toughness in Mode I; in fact, detailed codes have been prepared for these experimental methods and they are internationally accepted.

Energy Release Rate

Whether a crack in a component is likely to grow under given load conditions is of vital importance to fracture mechanics. The problem has been analyzed through several approaches-stress, displacement or energy methods. Each approach defines a suitable parameter. A limit on the parameter defines the toughness of the material. For a prescribed load condition, if the value of the parameter exceeds the limit, the crack may grow.

Griffith's dilemma

Griffith, in the early 1920s , developed some basic concepts. He was aware of the analytical solution, developed by Inglis, which determines stress field around an elliptical hole in a large plate, loaded under the tensile stress <r0 as shown in Fig. 2. He noted that the maximum stress develops at point *A* of the ellipse and is given by

$$\sigma_{22}^{\max} = \sigma_0\left(1+\frac{2a}{b}\right) \qquad \text{.................. (1)}$$

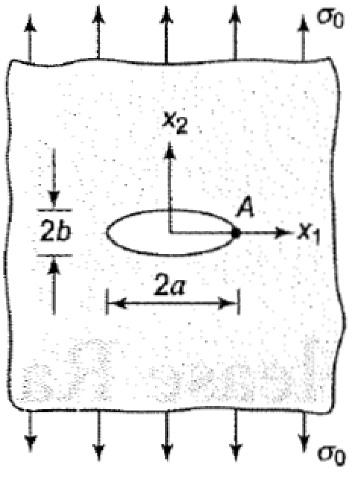


Fig. 2 An elliptical hole in a stretched plate

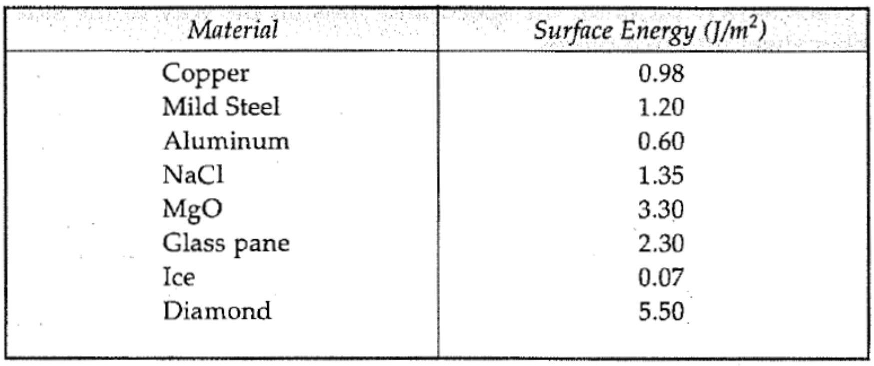
For a circular hole, σ_{22}^{max} is three times greater than σ_0. But for an elliptical hole where the major axis *a* is much longer than the minor axis *b*, σ_{22}^{max} becomes quite large; so much so that for a sharp crack with minor axis tending to be very small (of the order of interatomic distances), no real material can sustain the stress. Thus, even for a small applied stress σ_0, Eq. (1) suggests that $\sigma_{22}{}^{max}$ would be very large and would exceed the ultimate strength of the material. Equation (1) further suggests that even a sharp crack of small length may grow and break the component into two pieces. However, this is contrary to our observations. Griffith thus concluded that some other mechanisms must be existing which helped materials to sustain solid forms.

Griffith's realization

Griffith realized that a crack in a body would not grow unless energy was released to overcome the energy needs of forming two new surfaces, one below and one above the crack plane. The surface energy of a material depends on the material properties. However, its magnitude is rathe1 small, of the order of 1 J /m^2 • Table 1 lists the surface energy of some of commonly encountered solids.

Surface energy of the order of 1 J/m^2 is considered insignificant (1 J energy will raise the temperature of one teaspoon of water by 0.05° Conly). In brittle materials such as silica, glass and diamond, advancing cracks require small energies of the order of surface energies, and, therefore once a crack starts advancing, it runs through the body easily, causing catastrophic failure. But some additional mechanisms operate on most materials which do not allow cracks to grow at low energies.

Table 1 lists the surface energy of some of commonly encountered solids.



Material	*Surface Energy (J/m^2)*
Copper	0.98
Mild Steel	1.20
Aluminum	0.60
NaCl	1.35
MgO	3.30
Glass pane	2.30
Ice	0.07
Diamond	5.50

Griffith's Criterion

Let us consider a plate with no prior crack [Fig. 3a]. It is pulled and then maintained in tension between two rigid supports [Fig. 3b)]. Now, with a knife, a crack is cut at the centre of the plate with the crack plane normal to the tensile stress. The crack length is increased gradually with the help of the knife. A critical stage reaches when the crack starts growing on its own; i.e., without any further need of the knife.

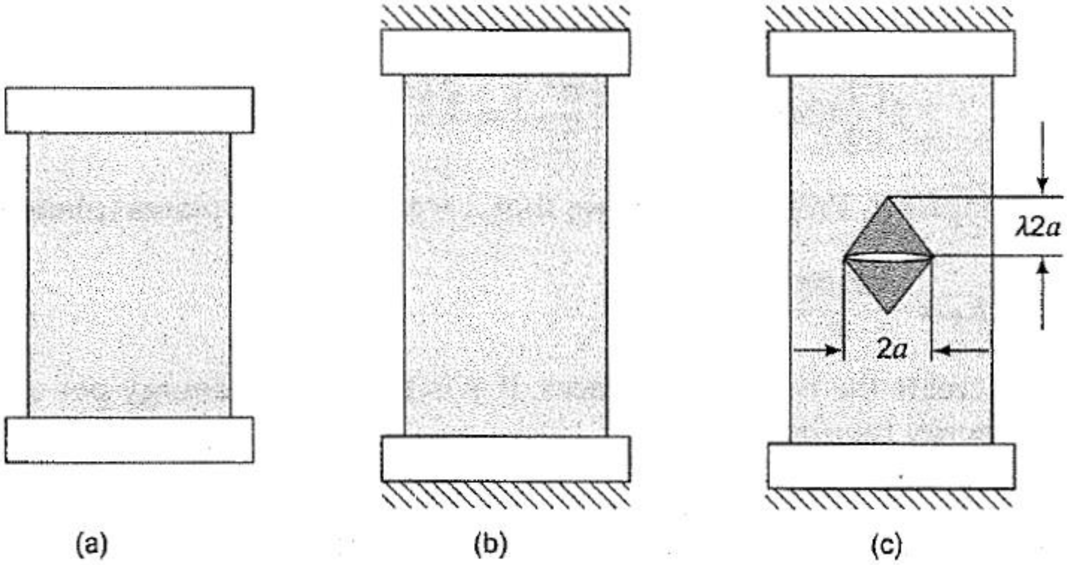


Fig. 3(a) A unstretched plate, (b) the stretched plate, and (c) introducing a crack at the center

The plate becomes less stiffness, as the crack advances. Consequently, for this case of the plate with ends held rigidly, the stress within the plate decreases and the strain energy stored in the plate is reduced. The energy thus released is available for the crack to grow.

To convince ourselves, a simple experiment may be conducted by taking a discarded tube of a bicycle wheel and cutting it along the length to obtain a plate sheet. The sheet is stretched and mounted on a frame. We can even use the arm-rests of a chair for the purpose. A small crack is introduced with a blade or knife normal to the stretching direction and the crack is enlarged gradually. At the critical length, the crack starts advancing slowly on its own and the knife is withdrawn. The crack slowly picks up the speed and runs all the way to the side edges, thus, snapping the rubber sheet into two pieces.

To understand Giriffith's analysis, we will carry an approximate analysis in this section. The plate is chosen to have its dimensions much larger than the longest crack to be considered. Then, the stress at points far away from the crack is assumed to remain constant. Most of the energy release, as the crack advances, comes from those parts of the plate which are adjacent to the cracked surfaces, because they are traction free. For the sake of convenience, the major area of the plate where its strain energy is released may be taken as a triangle on each side of the crack plane as shown in Fig. 3 (c). In fact, other shapes such as a parabola will serve the purpose; we have chosen it to be triangular to keep the algebra simple. With the increase in crack length the base and the height of both triangles increase and, therefore, the area from which the strain energy is released is proportional to the square of the crack length.

The height of a triangle is .$\lambda(2a)$ where A is a constant. Then, the total release of energy E_R is determined by multiplying the area of both triangles with plate thickness B and the strain energy density σ^2/E where σ is the tensile stress and E is the Young's modulus. Thus, released energy is given by

$$E_R = (\text{Volume of triangles}) \times \left(\frac{\sigma^2}{2E}\right)$$

$$= 2\left(\frac{1}{2}(2a)(2\lambda\, a)B\right) \times \left(\frac{\sigma^2}{2E}\right) = \frac{2\lambda a^2 B\sigma^2}{E}$$

Rigorous analysis shows that $\lambda = \pi/2$ for thin plates (plane stress) giving

$$E_R = \frac{\pi\, a^2 B\sigma^2}{E} \qquad \text{………………………(1)}$$

Energy is required to create the two new surfaces. If γ is the surface energy per unit area of one surface, the surface energy required E_s is

$$E_S = 2(2a)B\gamma = 4aB\gamma \qquad \text{………………………(2)}$$

The relations of Equations (1) and (2) are shown graphically in Figure 4. E_R increases parabolically whereas E_s increases linearly with increasing crack length a. Consider a small crack length, $2a_0$ whose length is incremented by Δa. The slope of E_R is smaller than the corresponding slope E_s and, therefore, the energy release ΔE_R in advancing the crack-tip by distance Δa is not sufficient to meet the energy needs of new surfaces (ΔEs). The crack would not grow and would remain subcritical. In fact, the crack would be dormant unless energy is supplied by an external agency, the operator of the knife in this case

As the operator gradually continues cutting the crack further, the slope of E_R increases, while the slope of E_s remains the same. A stage is reached when the slope of E_R becomes equal to the slope of E_s. The crack now becomes critical, because for an incremental advance of crack length, energy release equals the energy required. Therefore, for the crack to become critical

$$\frac{dE_R}{da} \geq \frac{dE_S}{da} \quad \text{…………………(4)}$$

Note that the difference in these two energies has already been supplied by the operator of the knife. In fact, this difference is an extremely useful quantity because of which many small cracks in a body do not grow and they remain dormant. If the plate had been pulled to a higher tension prior to the introduction of the crack, the E_R would increase with the crack length at a faster rate (4) and then even the cracks of smaller lengths would become critical. In order to determine the critical crack length ac we substitute E_R [Eq. (2)] and Es [Eq. (3)] in Inequality[Eq. (4)] to obtain

$$\frac{2\pi\, a_c B\sigma^2}{E} \geq 4B\gamma$$

For a safe crack,

$$a_c \leq \frac{2E\gamma}{\pi\sigma^2} \quad \text{……………..(5)}$$

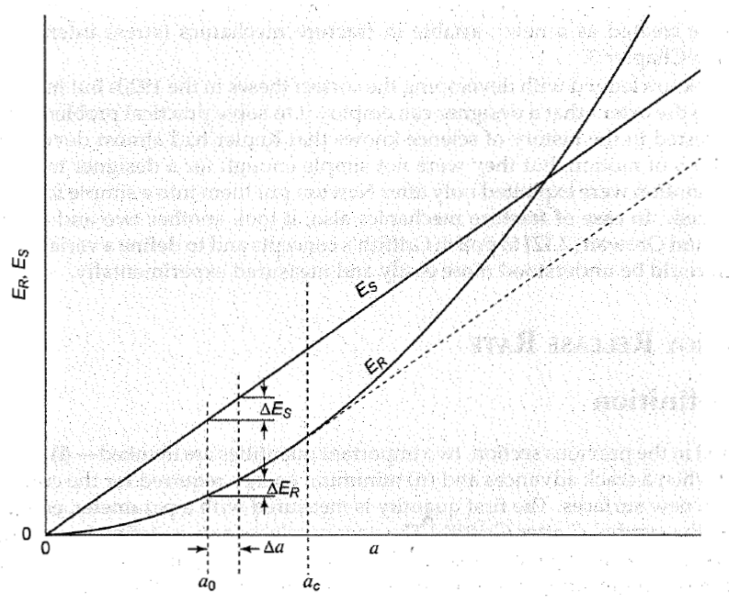


Fig. 4 Variation of energy release *E_R* and required surface energy *E_S* with crack length

If we want to know how much stress is required to advance a given crack for plane stress cases, the inequality is rearranged to,

$$\sigma_c \geq \left[\frac{2E\gamma}{\pi a}\right]^{1/2} \quad \text{……….(6)}$$

For plane strain (thick plate), Eis replaced by E/(l-v^2) and the relation becomes

$$\sigma_c \geq \left[\frac{2E\gamma}{\pi(1-\nu^2)a}\right]^{1/2} \qquad \text{……………(7)}$$

where v is the Poisson's Ratio.

We deduce, from Inequalities 6 and 7, that the critical stress depends on modulus E, surface energy rand crack length a. As expected, the higher value of the surface energy of a material increases the critical stress whereas a longer crack reduces it. Larger modulus means that the plate is capable of storing less energy, thereby resulting into smaller energy release which, in tum requires higher stress for making the crack critical. One may also note from Inequality (6) that the product **σ_c sqrt a** depends only on material properties (elastic constants E and v and surface energy y). Therefore, **σ sqrt** a may be treated as a new variable in fracture mechanics (stress intensity factor).

Griffith acknowledged with developing the correct thesis in the 1920s but his analysis was not developed to the extent that a designer can employ it to solve practical problems. For example, a reader interested in the history of science knows that Kepler had almost developed everything about the laws of motion, but they were not simple enough for a designer to make a machine. The laws of motion were exploited only after Newton put them into a simple form with variables clearly defined. In case of fracture mechanics also, it took another two and a half decades for Irwin and Orowan to polish Griffith's concepts and to define a variable, energy release rate, which could be understood more easily and measured experimentally.

The Irwin Plastic Zone Correction

Irwin suggested a vastly improved expression for the plastic zone size along the x_1-axis through a model which accounts for the absence of high stresses within the yield zone. Consider a case of plane stress with k = 1. As shown in Fig. 5, the plastic zone size beyond the tip T of the actual crack is TL along x_1-axis. The tip of the effective crack is located somewhere inside the plastic zone. There should be an appropriate criterion to find the length of the effective crack.

Note that if the tip of the effective crack is at point M, then the material beyond the tip will not be able to sustain $\sigma_{22} > \sigma_{ys}$· Moreover, the crack-faces of the effective crack are not actually separated between the points T and M and therefore tensile stress equal to σ_{ys} acts on the length δ for an elastic-perfectly plastic material. Irwin proposed that δ is so chosen such that the load not taken beyond the point M, given by the area Pc on the length λ is equal to the load sustained on length δ, given by the area P_A Then, the load sustained on length λ becomes,

$$P_c = B\left[\int_0^{\lambda}\sigma_{22}dx_1 - \sigma_{ys}\lambda\right]$$

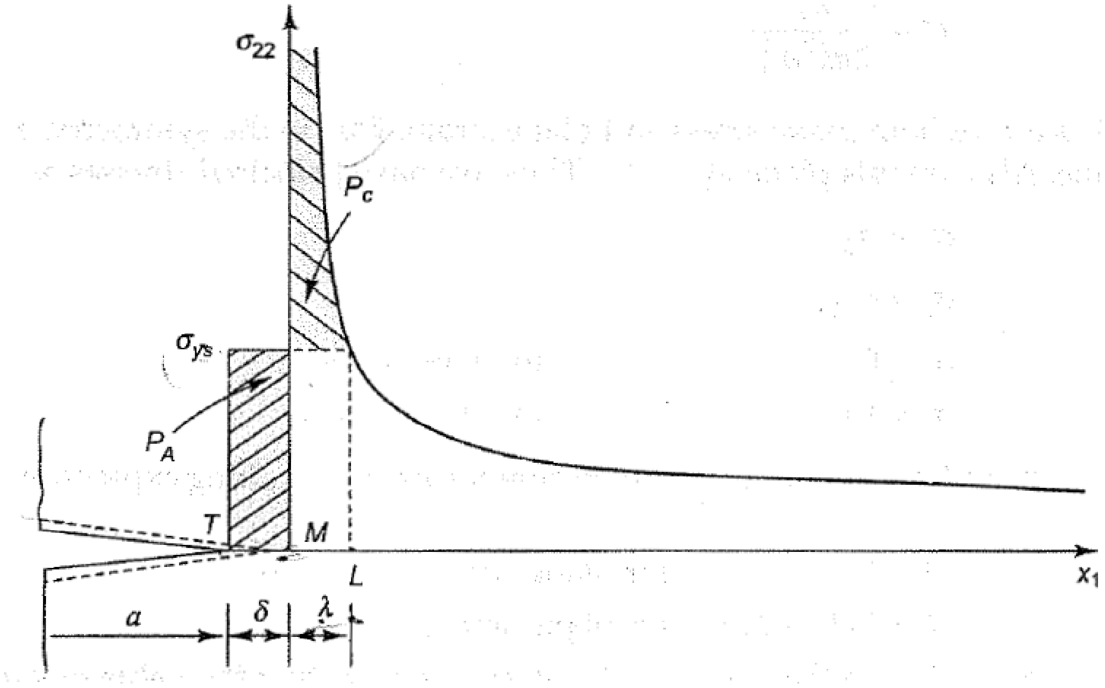


Fig. 5 Irwin plastic zone correction

where, B is the plate thickness. The load sustained (P_A) on length δ is

$$P_A = B\sigma_{ys}\,\delta$$

Irwin's correction ($P_A = P_c$) leads to

$$\sigma_{ys}\,\delta = \int_0^{\lambda}\sigma_{22}dx_1 - \sigma_{ys}\lambda = \int_0^{\lambda}\frac{K_{\mathrm{I}}}{\sqrt{2\pi x_1}}dx_1 - \sigma_{ys}\lambda \qquad \text{…(8)}$$

where, K_{I} is based on the effective crack length ($a + \delta$). λ is determined by noting that at $x_1 = \lambda$, σ_{22} is equal to σ_{ys} for plane stress ($k = 1$), i.e.,

$$\sigma_{ys} = \frac{K_{\mathrm{I}}}{\sqrt{2\pi\lambda}}$$

Rearranging, we have

$$K_I = (2\pi\lambda)^{1/2}\sigma_{ys}$$

Substituting in Eq. (5.7), we have

$$\sigma_{ys}\delta = \int_0^{\lambda} \frac{\sigma_{ys}(2\pi\lambda)^{1/2}}{\sqrt{2\pi x_1}}\,dx_1 - \sigma_{ys}\lambda$$

On solving, we obtain

$$\delta = \lambda$$

The overall plastic zone size becomes,

$$r_p = 2\lambda = \frac{K_I^2}{\pi\sigma_{ys}^2}$$

The effective crack length a_{eff} is given by

$$a_{eff} = a + \delta = a + \lambda = a + \frac{K_I^2}{2\pi\sigma_{ys}^2}$$

Since K_i is based on the effective crack length, a_{eff} is still unknown. If the plastic zone is small in comparison to the crack length, K_i may be assumed to be equal to the SIF of crack length a. However, Ki can also be obtained in a closed form for the case of an infinite plate; the corrected SIF is expressed as,

$$K_{\mathrm{I}} = \sigma\left[\pi(a+\lambda)\right]^{1/2}$$

$$= \sigma\pi^{1/2}\left[a + \frac{K_{\mathrm{I}}^2}{2\pi\,\sigma_{ys}^2}\right]^{1/2}$$

On solving, we obtain

$$K_{\mathrm{I}} = \frac{\sigma\sqrt{\pi\,a}}{\left[1-\frac{1}{2}\left(\sigma/\sigma_{ys}\right)^2\right]^{1/2}}$$

For a finite size specimen,

$$K_{\mathrm{I}} = \sigma\left[\pi(a+\lambda)\right]^{1/2} f\left(\frac{a+\lambda}{W}\right)$$

or

$$K_{\mathrm{I}} = \sigma\pi^{1/2}\left[a + \frac{K_{\mathrm{I}}^2}{2\pi\,\sigma_{ys}^2}\right]^{1/2} f\left(\left(a + \frac{K_{\mathrm{I}}^2}{2\pi\,\sigma_{ys}^2}\right)\Big/ W\right)$$

For certain geometries, function *f* is simple (e.g.,f = 1 for a centre-cracked infinite plate and *f* = 1.12 for an edge-cracked infinite plate) and Ki can easily be obtained from the above equation. Otherwise, it can also be determined by an iterative procedure. For example, Ki on right hand side is taken based on the actual crack length *a* in the first round of iteration. The evaluated value Kr is then fed on the right hand side in the second round. The iterative procedure is repeated until two successive values of Ki are within a small percentage difference.

Irwin's correction to the plane strain case is useful to determine the plastic zone size. Due to the plastic deformation the crack tip becomes rounded. Since the rounded tip acts as a free surface,σ_{11} is released to zero. The effect of the release of σ_{11} is felt for some distance on the X1-axis beyond the crack tip. Irwin found that *k* is no longer 3 but is closer to √2 √2 which may be taken as √3. Then, the plastic zone size for the plane strain becomes $(1/3\pi)Ki^2/\ \sigma_{ys}^2$· For an experimental determination of K_{tc} of material, plane strain conditions are assured by taking plate thickness to be much thicker than the plastic zone size. In fact, it is chosen to be more than 25 times of the plastic zone size and therefore, the ASME codes require

$$B \geq 2.5 \frac{K_I^2}{\sigma_{ys}^2}$$

J-Integral

The J-Integral is also a parameter to characterize a crack. In fact, G is a special case of the J-Integral, i.e., G is usually applied only to linear elastic materials, whereas the J-Integral is not only applicable to linear and non-linear elastic materials, but is considered to be very useful to characterize materials, exhibiting elastic-plastic behavior near the crack tip.

At first sight, the J-Integral looks like a strange term, but as we go along and develop it gradually, it would not look odd. For plane problems, consider a path r around a crack tip (Fig. 6.2) which starts from any point of a crack face and ends on any point on the other crack face. The path can be chosen arbitrarily within the material of the component. It may be smooth or may have corners, but should be continuous. J-Integral was first applied to fracture mechanics by Rice in 1968 for plane problems. It is defined as:

$$J = \int_{r} \left(W dx_2 - T_i \frac{\partial u_i}{\partial x_1} ds \right) \qquad \text{......(9)}$$

where,

$$W = \int \sigma_{ij} \, d\varepsilon_{ij}$$

The Einstein's summation convention is followed in defining the above expressions. For two-dimensional plane cases W may be expressed in full form as:

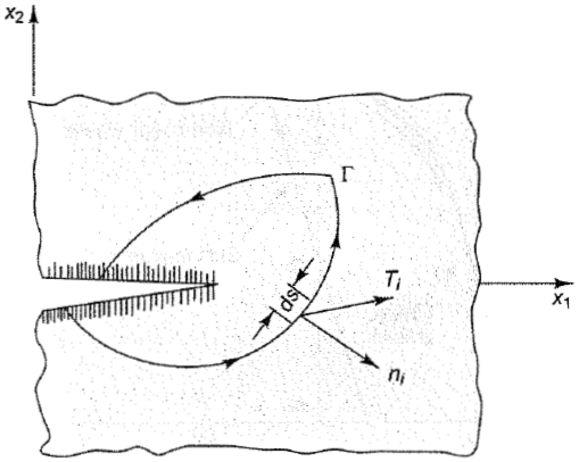


Fig. 6 Path r around the crack tip with outward normal n_i and traction T_i

$$W = \int_0^{\varepsilon_{11}} \sigma_{11}\, d\varepsilon_{11} + \int_0^{\varepsilon_{12}} \sigma_{12}\, d\varepsilon_{12} + \int_0^{\varepsilon_{21}} \sigma_{21}\, d\varepsilon_{21} + \int_0^{\varepsilon_{22}} \sigma_{22}\, d\varepsilon_{22}$$

Summing the middle terms, we have:

$$W = \int_0^{\varepsilon_{11}} \sigma_{11}\, d\varepsilon_{11} + 2\int_0^{\varepsilon_{12}} \sigma_{12}\, d\varepsilon_{12} + \int_0^{\varepsilon_{22}} \sigma_{22}\, d\varepsilon_{22}$$

Also, W which is strain energy per unit volume is a point function, i.e., it varies from point to point

within the body of the component. Other parameters of Eq. (9) are:

T_i = traction vector at a point on the path r,

u_i= displacement vector at a point on the path.

In the expanded form, the second term of Eq. (9) for plane cases is:

$$\int_\Gamma \left(T_1 \frac{\partial u_1}{\partial x_1} + T_2 \frac{\partial u_2}{\partial x_1} \right) ds$$

Crack tip opening displacement

Crack tip opening displacement (CTOD) is another parameter suitable to characterize a crack. Unlike parameters G and K, it can be used for both linear elastic fracture mechanics (LEFM) and elastic plastic fracture mechanics (EPFM). The material cannot withstand very high stresses within the plastic zone, and the usual stress field of the square root singularity no longer exists. However, rigorous analysis is complex and we would like to explore a simple model. The yielding of the material and resulting rearrangement of the stresses around the crack tip can be accounted by an effective crack, which is longer than the actual crack. The tip of the effective crack is located inside the plastic zone. Now, the linear equations of elasticity can be applied to the effective crack.

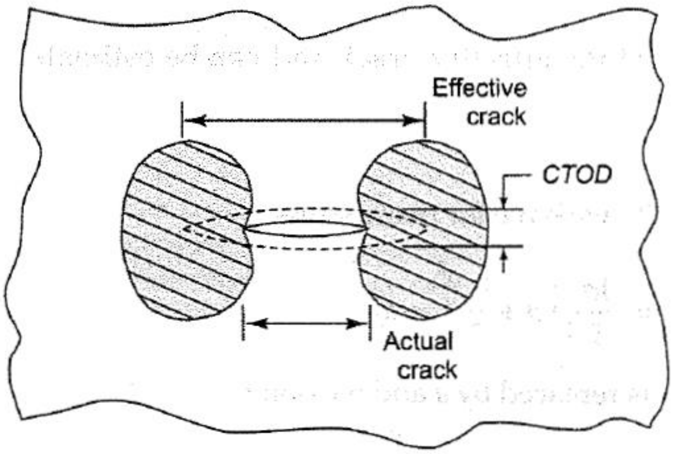


Fig. 7 The effective crack and *CTOD*

One can then visualize that the effective crack has some finite opening at the location of the actual crack tip. The opening is defined as the crack tip opening displacement. In reality, there is hardly any opening of the crack tip; only the tip may become more rounded as the plastic deformation increases. Thus, the definition of the CTOD is based on a model that a beginner may find a bit difficult to accept. Further, the symbol for a crack tip opening displacement is still not standardized. Some people use the symbol δ, while others call it COD (crack opening displacement). However, the term 'COD' is found ambiguous, because at any location of the crack, the opening is called a crack opening displacement. Therefore, we will be using CTOD as the parameter and CTODc, as its critical value. The CTOD parameter has been found more useful for cracks having large plastic zones. Like the J-Integral, the CTOD is another approach to deal with elastic-plastic fracture mechanics (EPFM). However, in the next section, we will explore equivalent relations between K, G and CTOD for cracks with small plastic zone (LEFM).

Relationship between CTOD, K_I and G_I for small scale yielding

We will prove that for small scale yielding in the vicinity of the crack tip

$$CTOD = \frac{K_I^2}{\lambda E \sigma_{ys}}$$

where λ is a constant and its value is close to unity. In fact, its value depends on the plastic zone and, therefore, λ is influenced by the model chosen for determining the plastic zone size. Since $G_I = K_I^2/E$ for plane stress, we can express $CTOD$ in terms of the relation

$$CTOD = \frac{G}{\lambda \sigma_{ys}}.$$

Thus, at distance of x_i, COD becomes:

$$COD = \frac{4\sigma}{E}\sqrt{a_{\text{eff}}^2 - x_1^2} \quad \text{for plane stress}$$

$$COD = \frac{4\sigma(1-\nu^2)}{E}\sqrt{a_{\text{eff}}^2 - x_1^2} \quad \text{for plane strain}$$

where, a_{eff} is the length of the effective crack and can be estimated using Irwin's correction

$$a_{\text{eff}} = a + \frac{r_p}{2}.$$

Substituting a_{eff} in Eq. (7.3), we have for plane stress

$$COD = \frac{4\sigma}{E}\left[\left(a + \frac{r_p}{2}\right)^2 - x_1^2\right]^{1/2}$$

For evaluating $CTOD$, x_1 is replaced by a and we obtain

$$CTOD = \frac{4\sigma}{E}\left[\frac{r_p}{4}\left(4a + r_p\right)\right]^{1/2}$$

For LEFM, *rp* may be neglected in comparison to *fa* and the expression simplifies to

$$CTOD = \frac{4\sigma}{E}\sqrt{ar_p}$$

where r_p is given by

$$r_p = \frac{K_I^2}{\pi\sigma_{ys}^2}.$$

K_i in this expression is based on a_{eff}. In the case of small-scale yielding considered in this section, the approximate relation is used by determining K_i on the actual crack length *a.* Then, the *CTOD* is expressed as

$$CTOD = \frac{4\sigma}{E}\sqrt{a\left(\frac{\sigma^2\pi\, a}{\pi\,\sigma_{ys}^2}\right)} = \frac{4\sigma^2 a}{E\sigma_{ys}}.$$

This expression can be restated by using

$$K_I = \sigma\sqrt{\pi a}$$

as

$$CTOD = \frac{4K_I^2}{\pi E\sigma_{ys}}.$$

If the Dugdale approach is used, it can be shown that

$$CTOD = \frac{K_I^2}{E\sigma_{ys}} = \frac{G_I}{\sigma_{ys}}.$$

Thus, the equivalence relations depend upon the model adopted for finding the effective crack length. Also, the relations depend upon the hardening parameter of work hardening materials. The value of λ can vary between and π/4 and 2.2. However, direct experimental methods find that λ is closer to unity.

Plastic Zone Shape for Plane Stress

Two widely used yield criteria, Mises and Tresca, are applied to Mode I, so as to determine the plastic zone size for plane stress cases. All the three principal stresses and associated Mohr circles are shown in Fig. To ensure the yielding of the material, the Mises criterion states that,

$$(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 \geq 2\sigma_{ys}^2$$

where σ_{ys} is the yield stress. Substituting σ_1, σ_2, and σ_2, in the equation and using the symbol r_{pz} in place of r for the equality sign of the Mises criterion, we obtain,

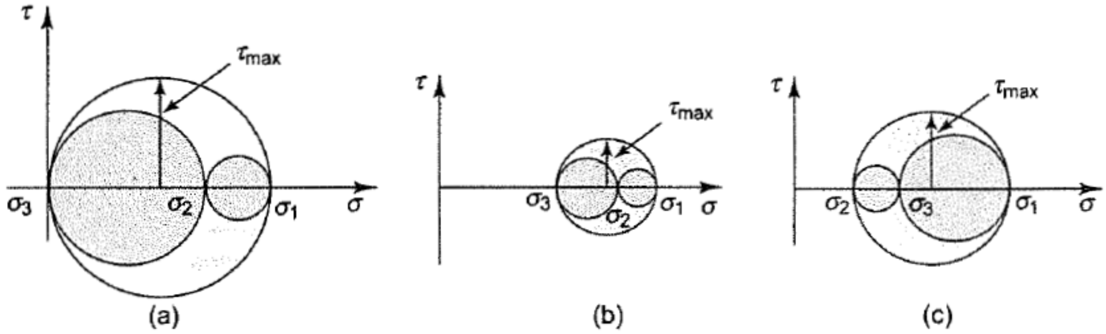


Fig. 8 Mohr circles for (a) plane stress, (b) plane strain for small θ, and (c) plane strain for large θ

$$\frac{K_I^2}{2\pi r_{pz}}\cos^2\frac{\theta}{2}\left[4\sin^2\frac{\theta}{2}+\left(1-\sin\frac{\theta}{2}\right)^2+\left(1+\sin\frac{\theta}{2}\right)^2\right]=2\sigma_{ys}^2$$

which is simplified to

$$r_{pz}=\frac{1}{4\pi}\frac{K_I^2}{\sigma_{ys}^2}\left(1+\frac{3}{2}\sin^2\theta+\cos\theta\right)$$

The resulting shape of the plastic zone size is plotted in Fig. in terms of non-dimensional distance r_{pz} / $(K_I^2/(\pi\sigma^2_{ys}))$. The shape of the plastic zone size is slightly different if the Tresca yield criterion is invoked. In order to ensure yielding, the Tresca yield criterion states that:

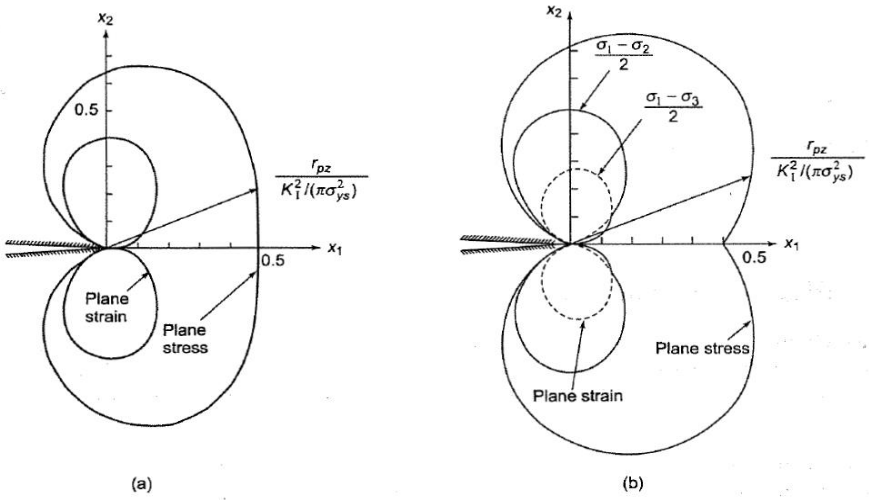


Fig. 9 Plastic zone shape and size for (a) Mises yield criterion, and (b) Tresca yield criterion

$$\tau_{\max} \geq \frac{\sigma_{ys}}{2}$$

For plane stress, rmax is given by the biggest Mohr circle which is between σ_1 and σ_3 • Thus, at $r = r_{pz}$

$$\frac{\sigma_1 - 0}{2} = \frac{\sigma_{ys}}{2}$$

we obtain the approximate shape of the plastic zone as,

$$r_{pz} = \frac{K_I^2}{2\pi \sigma_{ys}^2}\left[\cos\frac{\theta}{2}\left(1+\sin\frac{\theta}{2}\right)\right]^2$$

CRITICAL STRESS INTENSITY. FACTOR

If the SIF of a crack approaches or exceeds an upper limit of the stress intensity factor, the crack may grow. The upper limit is known as the critical stress intensity factor which is a material property and is usually denoted by the symbol K_{Ic} for Mode I cases (K_{IIc} and K_{IIIc} for Mode II and Mode III respectively). In order to provide a feel of stress intensity factor and the critical stress intensity factor, an analogy is made with stress and yield stress of a solid. Stress is a parameter which represents internal loading within the solid and yield stress is the limit on stress, beyond which the material is regarded to have failed by many designers. Similarly, stress intensity factor is a parameter to measure the severity of stress at the crack tip. But, critical stress intensity factor is the limit on the SIF, such that if the SIF exceeds the critical stress intensity factor, the crack may grow.

Thus, in order to predict the growth of a crack in a component, the designer should find two values: (i) the SIF determined through analysis for the geometry of the component, crack configuration and applied loads and (ii) the critical SIF determined through experiments for the material of the component. If the stress intensity factor exceeds the critical stress intensity factor, the designer should do something, such as reducing the loads on the component, modifying the geometry of the component, or choosing a material of higher toughness.

One difficulty, faced during the experimental determination of the critical SIF for the material of the component, is that critical SIF is found to be dependant on the thickness of a plate. In fact, the critical SIF is independent of the thickness only in the case of a thick plate, because the plate is then loaded in the plane strain. One question remains-what is the criterion of assuring that the plane strain conditions prevail? The size of the plastic zone in the vicinity of the crack tip decides it. If the plate thickness is significantly greater than the size of the plastic zone, then the conditions of plane strain exist.

Critical stress intensity factor for thin plates depends on the plate thickness and its value is rarely provided as a function of thickness in literature. However, the critical SIF of a plane stress case is higher than the corresponding value in a plane strain. A designer may find that a component is subjected to plane stress, but critical SIF is available only for plane strain. He may safely use the critical SIF of the plane strain because it would provide a conservative design.

In certain design problems (e.g., components of airplanes, rockets and spaceships), using the critical SIF of a plane strain as material property ma.y be too conservative, because structural plates are mostly used in plane stress. Adopting a too conservative approach is against the philosophy of engineering profession. Engineers should always strive to obtain numbers close to reality. In aerospace applications, where the factor of safety is of the order of 1.1 for many components, using material toughness properties of plane strain is likely to make the machine heavy with poor payload. In such a situation, stress intensity factor is determined by preparing a test-specimen of same thickness as of plates used in the actual application. However, for most down-to-earth problems (such as the designing components of automobiles, roof trusses, locomotive carriages, pipe lines, etc.) the conservative approach of using the critical SIP of the plane strain may be quite practical and useful.

Critical SIF of a material depends on many factors, such as

- Heat treatment which controls the yield stress of the material.
- Speed of the crack.
- Temperature of the specimen.
- Process of manufacturing (e.g., vacuum furnaced or air melted, as cast or rolled).
- Orientation of the crack with respect to the grains at the crack tip.
- Test method.