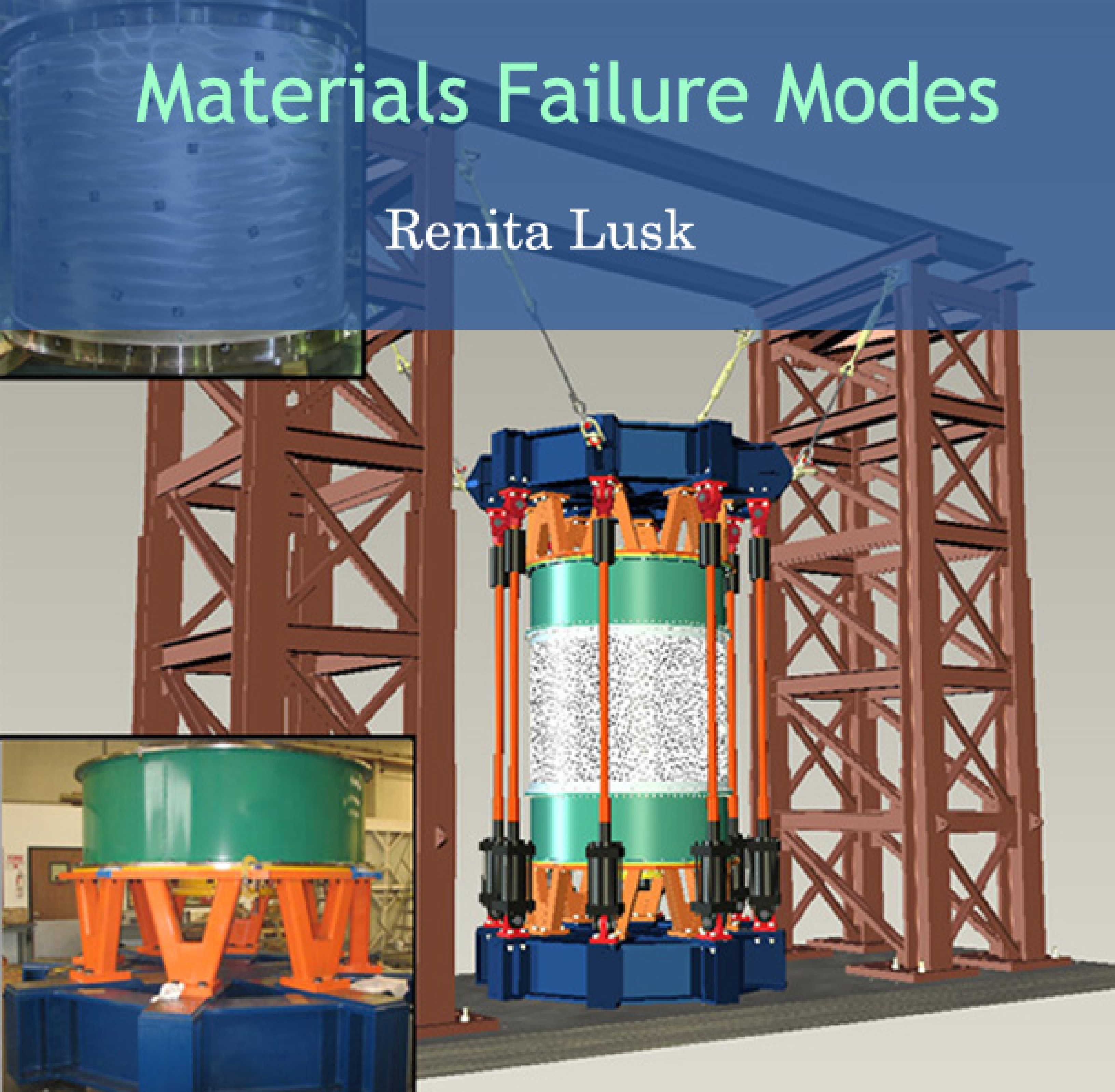
Materials Failure Modes

Renita Lusk



First Edition, 2012

ISBN 978-81-323-4382-0

© All rights reserved.

Published by:
White Word Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

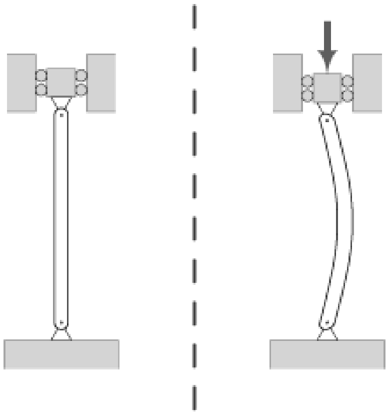
Table of Contents

Chapter 1

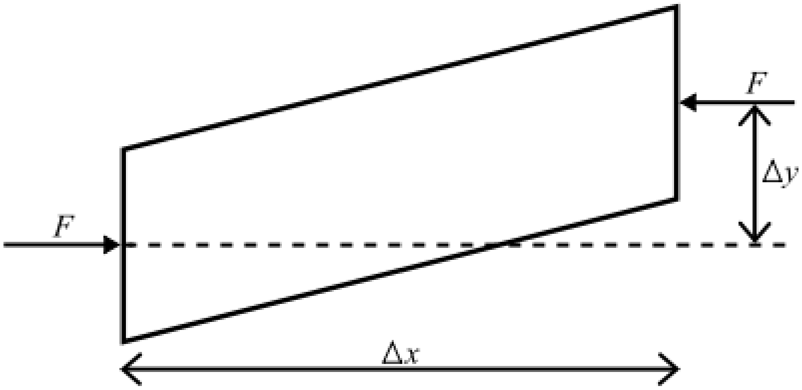
Buckling

In science, **buckling** is a mathematical instability, leading to a failure mode. Theoretically, buckling is caused by a bifurcation in the solution to the equations of static equilibrium. At a certain stage under an increasing load, further load is able to be sustained in one of two states of equilibrium: an undeformed state, or a laterally-deformed state. In practice, buckling is characterized by a sudden failure of a structural member subjected to high compressive stress, where the actual compressive stress at the point of failure is less than the ultimate compressive stresses that the material is capable of withstanding. For example, during earthquakes, reinforced concrete members may experience lateral deformation of the longitudinal reinforcing bars. This mode of failure is also described as failure due to elastic instability. Mathematical analysis of buckling makes use of an axial load eccentricity that introduces a moment, which does not form part of the primary forces to which the member is subjected. When load is constantly being applied on a member, such as column, it will ultimately become large enough to cause the member to become unstable. Further load will cause significant and somewhat unpredictable deformations, possibly leading to complete loss of load-carrying capacity. The member is said to have buckled, to have deformed.

Columns



A column under a concentric axial load exhibiting the characteristic deformation of buckling



The eccentricity of the axial force results in a bending moment acting on the beam element.

The ratio of the effective length of a column to the least radius of gyration of its cross section is called the *slenderness ratio* (sometimes expressed with the Greek letter lambda, λ). This ratio affords a means of classifying columns. Slenderness ratio is important for design considerations. All the following are approximate values used for convenience.

- A short steel column is one whose slenderness ratio does not exceed 50; an intermediate length steel column has a slenderness ratio ranging from about 50 to 200, and are dominated by the strength limit of the material, while a long steel column may be assumed to have a slenderness ratio greater than 200.
- A short concrete column is one having a ratio of unsupported length to least dimension of the cross section not greater than 10. If the ratio is greater than 10, it is a long column (sometimes referred to as a slender column).
- Timber columns may be classified as short columns if the ratio of the length to least dimension of the cross section is equal to or less than 10. The dividing line between intermediate and long timber columns cannot be readily evaluated. One way of defining the lower limit of long timber columns would be to set it as the smallest value of the ratio of length to least cross sectional area that would just exceed a certain constant K of the material. Since K depends on the modulus of elasticity and the allowable compressive stress parallel to the grain, it can be seen that this arbitrary limit would vary with the species of the timber. The value of K is given in most structural handbooks.

If the load on a column is applied through the center of gravity of its cross section, it is called an axial load. A load at any other point in the cross section is known as an eccentric load. A short column under the action of an axial load will fail by direct compression before it buckles, but a long column loaded in the same manner will fail by buckling (bending), the buckling effect being so large that the effect of the direct load may be neglected. The intermediate-length column will fail by a combination of direct compressive stress and bending.

In 1757, mathematician Leonhard Euler derived a formula that gives the maximum axial load that a long, slender, ideal column can carry without buckling. An ideal column is one that is perfectly straight, homogeneous, and free from initial stress. The maximum load, sometimes called the critical load, causes the column to be in a state of unstable equilibrium; that is, any increase in the load, or the introduction of the slightest lateral force, will cause the column to fail by buckling. The formula derived by Euler for columns with no consideration for lateral forces is given below. However, if lateral forces are taken into consideration the value of critical load remains approximately the same.

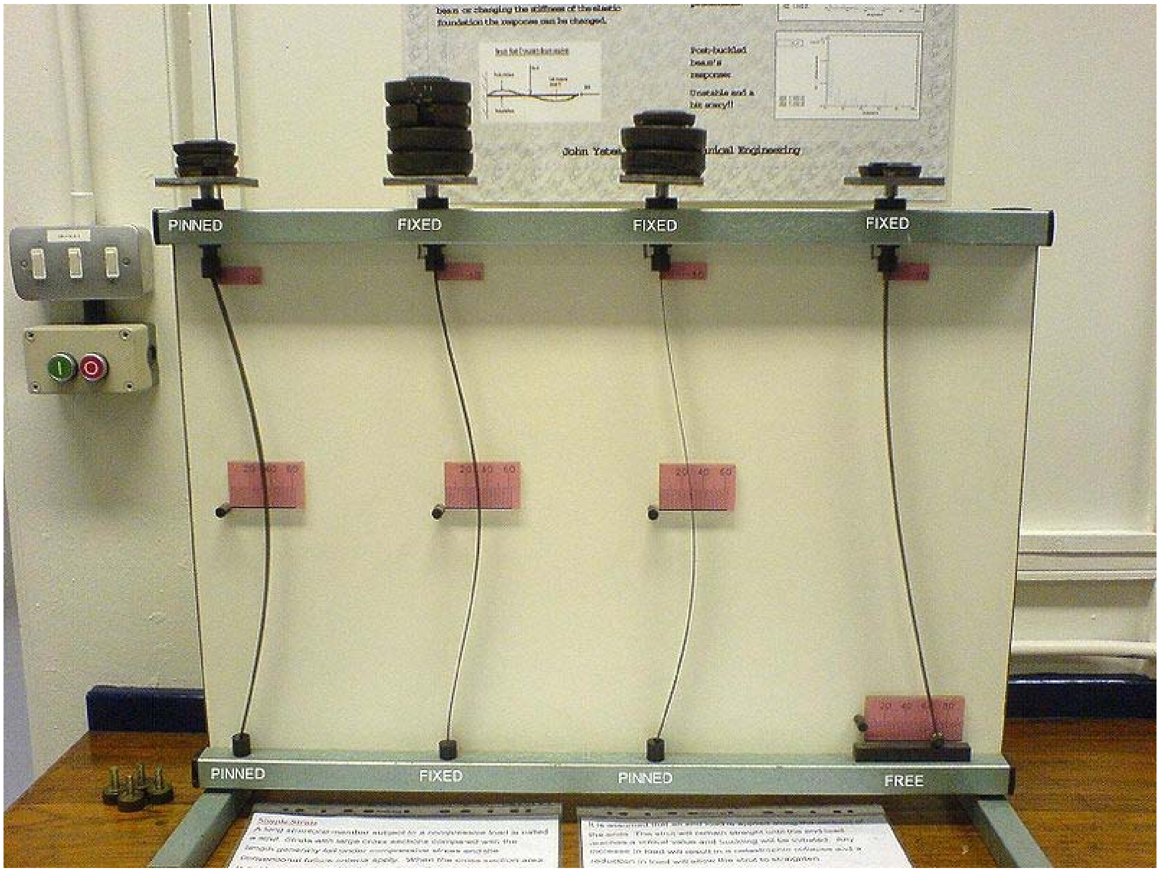
$$F = \frac{\pi^2 EI}{(KL)^2}$$

where

F = maximum or critical force (vertical load on column),
E = modulus of elasticity,
I = area moment of inertia,
L = unsupported length of column,
K = column effective length factor, whose value depends on the conditions of end support of the column, as follows.
For both ends pinned (hinged, free to rotate), $K = 1.0$.
For both ends fixed, $K = 0.50$.
For one end fixed and the other end pinned, $K = 0.699....$
For one end fixed and the other end free to move laterally, $K = 2.0$.
KL is the effective length of the column.

Examination of this formula reveals the following interesting facts with regard to the load-bearing ability of slender columns.

1. Elasticity and not compressive strength of the materials of the column determines the critical load.
2. The critical load is directly proportional to the second moment of area of the cross section.
3. The boundary conditions have a considerable effect on the critical load of slender columns. The boundary conditions determine the mode of bending and the distance between inflection points on the deflected column. The closer together the inflection points are, the higher the resulting capacity of the column.



A demonstration model illustrating the different "Euler" buckling modes. The model shows how the boundary conditions affect the critical load of a slender column. Notice that each of the columns are identical, apart from the boundary conditions.

The strength of a column may therefore be increased by distributing the material so as to increase the moment of inertia. This can be done without increasing the weight of the column by distributing the material as far from the principal axes of the cross section as possible, while keeping the material thick enough to prevent local buckling. This bears out the well-known fact that a tubular section is much more efficient than a solid section for column service.

Another bit of information that may be gleaned from this equation is the effect of length on critical load. For a given size column, doubling the unsupported length quarters the allowable load. The restraint offered by the end connections of a column also affects the critical load. If the connections are perfectly rigid, the critical load will be four times that for a similar column where there is no resistance to rotation (hinged at the ends).

Since the moment of inertia of a surface is its area multiplied by the square of a length called the radius of gyration, the above formula may be rearranged as follows. Using the Euler formula for hinged ends, and substituting $A \cdot r^2$ for I, the following formula results.

$$\sigma = \frac{F}{A} = \frac{\pi^2 E}{(\ell/r)^2}$$

where *F* / *A* is the allowable stress of the column, and *l* / *r* is the slenderness ratio.

Since structural columns are commonly of intermediate length, and it is impossible to obtain an ideal column, the Euler formula on its own has little practical application for ordinary design. Issues that cause deviation from the pure Euler strut behaviour include imperfections in geometry in combination with plasticity/non-linear stress strain behaviour of the column's material. Consequently, a number of empirical column formulae have been developed to agree with test data, all of which embody the slenderness ratio. For design, appropriate safety factors are introduced into these formulae.

Self-buckling

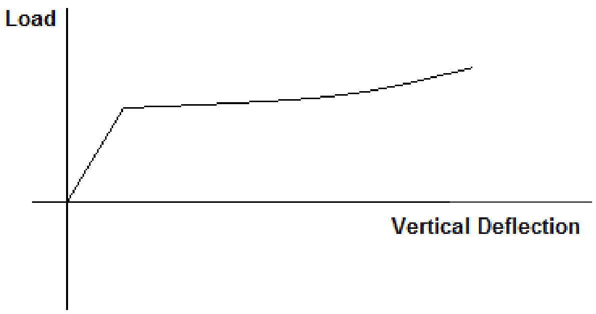
A free-standing, vertical column, with density ρ, Young's modulus *E*, and radius *r*, will buckle under its own weight if its height exceeds a certain critical height:

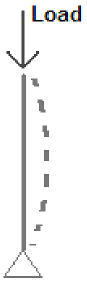
$$h_{crit} = \left(\frac{9B^2}{4}\,\frac{EI}{\rho g \pi r^2}\right)^{1/3}$$

where *g* is the acceleration due to gravity, *I* is the second moment of area of the beam cross section, and *B* is the first zero of the Bessel function of the first kind of order -1/3, which is equal to 1.86635...

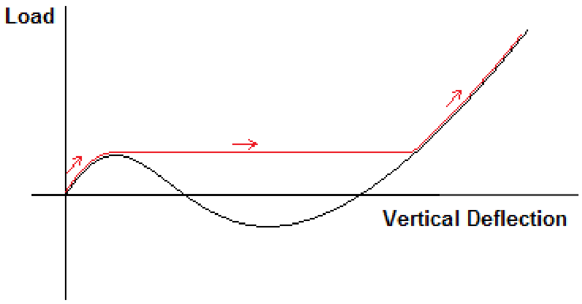
Limit point vs bifurcation buckling

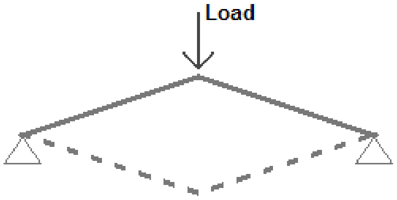
Bifurcation buckling is sometimes called Euler buckling even when applied to structures other than Euler columns. As the applied load is increased by a small amount beyond the critical load, the structure deforms into a buckled configuration which is adjacent to the original configuration. For example, the Euler column pictured will start to bow when loaded slightly above its critical load, but will not suddenly collapse.





In structures experiencing limit point instability, if the load is increased infinitesimally beyond the critical load, the structure undergoes a large deformation into a different stable configuration which is not adjacent to the original configuration. An example of this type of buckling is a toggle frame (pictured) which 'snaps' into its buckled configuration.





Bicycle wheels

A conventional bicycle wheel consists of a thin rim kept under high compressive stress by the (roughly normal) inward pull of a large number of spokes. It can be considered as a loaded column that has been bent into a circle. As such, if spoke tension is increased beyond a safe level, the wheel spontaneously fails into a characteristic saddle shape (sometimes called a "taco" or a "pringle") like a three-dimensional Euler column. This is normally a purely elastic deformation and the rim will resume its proper plane shape if spoke tension is reduced slightly.

Surface materials

Sun kink in rail tracks

Buckling is also a failure mode in pavement materials, primarily with concrete, since asphalt is more flexible. Radiant heat from the sun is absorbed in the road surface, causing it to expand, forcing adjacent pieces to push against each other. If the stress is great enough, the pavement can lift up and crack without warning. Going over a buckled section can be very jarring to automobile drivers, described as running over a speed hump at highway speeds.

Similarly, rail tracks also expand when heated, and can fail by buckling, a phenomenon called sun kink. It is more common for rails to move laterally, often pulling the underlain railroad ties (sleepers) along.

Energy method

Often it is very difficult to determine the exact buckling load in complex structures using the Euler formula, due to the difficulty in deciding the constant K. Therefore, maximum buckling load often is approximated using energy conservation. This way of deciding maximum buckling load is often referred to as the energy method in structural analysis.

The first step in this method is to suggest a displacement function. This function must satisfy the most important boundary conditions, such as displacement and rotation. The more accurate the displacement function, the more accurate the result.

In this method, there are two equations used to approximate the inner energy and outer energy (for small deformations).

$$A_{\text{inner}} = \frac{E}{2} \int I(x)(w_{xx}(x))^2 \, dx$$
$$A_{\text{outer}} = \frac{P_{\text{Crit}}}{2} \int (w_x(x))^2 \, dx$$

where *w*(*x*) is the displacement function and the subscripts *x* and *xx* refer to the first and second derivatives of the displacement. Energy conservation yields:

$$A_{\text{Inner}} = A_{\text{Outer}}$$

Flexural-torsional buckling

Occurs in compression members only and it can be described as a combination of bending and twisting of a member. And it must be consider for design purposes, since the shape and cross sections are very critical. This mostly occurs in channels, structural tees, double-angle shapes, and equal-leg single angles.

Lateral-torsional buckling

Lateral-torsional buckling of an aluminium alloy plate girder

When a simple beam is loaded in flexure, the top side is in compression, and the bottom side is in tension. When a slender member is subjected to an axial force, failure takes place due to bending or torsion rather than direct compression of the material. If the beam is not supported in the lateral direction (i.e., perpendicular to the plane of bending), and the flexural load increases to a critical limit, the beam will fail due to lateral buckling of the compression flange. In wide-flange sections, if the compression flange buckles laterally, the cross section will also twist in torsion, resulting in a failure mode known as *lateral-torsional buckling*.

Plastic buckling

Buckling will generally occur slightly before the theoretical buckling strength of a structure, due to plasticity of the material. When the compressive load is near buckling, the structure will bow significantly and approach yield. The stress-strain behaviour of materials is not strictly linear even below yield, and the modulus of elasticity decreases as stress increases, with more rapid change near yield. This lower rigidity reduces the buckling strength of the structure and causes premature buckling. This is the opposite effect of the plastic bending in beams, which causes late failure relative to the Euler-Bernoulli beam equation.

Dynamic buckling

If the load on the column is applied suddenly and then released, the column can sustain a load much higher than its static (slowly applied) buckling load. This can happen in a long, unsupported column (rod) used as a drop hammer. The duration of compression at the impact end is the time required for a stress wave to travel up the rod to the other (free) end and back down as a relief wave. Maximum buckling occurs near the impact end at a wavelength much shorter than the length of the rod, at a stress many times the buckling stress if the rod were a statically-loaded column. The critical condition for buckling amplitude to remain less than about 25 times the effective rod straightness imperfection at the buckle wavelength is

$$\sigma L = \rho c^2 h$$

where σ is the impact stress, L is the length of the rod, c is the elastic wave speed, and h is the smaller lateral dimension of a rectangular rod. Because the buckle wavelength depends only on σ and h, this same formula holds for thin cylindrical shells of thickness h.

Buckling of thin cylindrical shells subject to axial loads

Solutions of Donnel's eight order differential equation gives the various buckling modes of a thin cylinder under compression. But this analysis, which is in accordance with the small deflection theory gives much higher values than shown from experiments. So it is customary to find the critical buckling load for various structures which are cylindrical in shape from pre-existing design curves where critical buckling load F_{cr} is plotted against the ratio R/t, where R is the radius and t is the thickness of the cylinder for various values of L/R, L the length of the cylinder. If cut-outs are present in the cylinder, critical buckling loads as well as pre-buckling modes will be affected. Presence or absence of reinforcements of cut-outs will also affect the buckling load.

Buckling of pipes and pressure vessels subject to external overpressure

Pipes and pressure vessels subject to external overpressure, caused for example by steam cooling down and condensating into water with subsequent massive pressure drop, risk buckling due to compressive hoop stresses. Design rules for calculation of the required wall thickness or reinforcement rings are given in various piping and pressure vessel codes.

Chapter 2

Corrosion

Corrosion is the disintegration of an engineered material into its constituent atoms due to chemical reactions with its surroundings. In the most common use of the word, this means electrochemical oxidation of metals in reaction with an oxidant such as oxygen. Formation of an oxide of iron due to oxidation of the iron atoms in solid solution is a well-known example of electrochemical corrosion, commonly known as rusting. This type of damage typically produces oxide(s) and/or salt(s) of the original metal. Corrosion can also refer to other materials than metals, such as ceramics or polymers, although in this context, the term degradation is more common.

In other words, corrosion is the wearing away of metals due to a chemical reaction.

Many structural alloys corrode merely from exposure to moisture in the air, but the process can be strongly affected by exposure to certain substances. Corrosion can be concentrated locally to form a pit or crack, or it can extend across a wide area more or less uniformly corroding the surface. Because corrosion is a diffusion controlled process, it occurs on exposed surfaces. As a result, methods to reduce the activity of the exposed surface, such as passivation and chromate-conversion, can increase a material's corrosion resistance. However, some corrosion mechanisms are less visible and less predictable.

Rust, the most familiar example of corrosion.

Volcanic gases have sped the corrosion of this abandoned mining machinery.

Corrosion on exposed metal.

Galvanic corrosion

Galvanic corrosion occurs when two different metals electrically contact each other and are immersed in an electrolyte. In order for galvanic corrosion to occur, an electrically conductive path and an ionically conductive path are necessary. This effects a galvanic couple where the more active metal corrodes at an accelerated rate and the more noble metal corrodes at a retarded rate. When immersed, neither metal would normally corrode as quickly without the electrically conductive connection (usually via a wire or direct contact). Galvanic corrosion is often utilized in sacrificial anodes. What type of metal(s) to use is readily determined by following the galvanic series. For example, zinc is often used as a sacrificial anode for steel structures, such as pipelines or docked naval ships. Galvanic corrosion is of major interest to the marine industry and also anywhere water can contact pipes or metal structures.

Factors such as relative size of anode (smaller is generally less desirable), types of metal, and operating conditions (temperature, humidity, salinity, etc.) will affect galvanic corrosion. The surface area ratio of the anode and cathode will directly affect the corrosion rates of the materials.

Galvanic series

In a given sea environment (one standard medium is aerated, room-temperature seawater), one metal will be either more *noble* or more *active* than the next, based on how strongly its ions are bound to the surface. Two metals in electrical contact share the same electrons, so that the tug-of-war at each surface is translated into a competition for free electrons between the two materials. The noble metal will tend to take electrons from the active one, while the electrolyte hosts a flow of ions in the same direction. The resulting mass flow or electrical current can be measured to establish a hierarchy of materials in the medium of interest. This hierarchy is called a *galvanic series*, and can be a very useful in predicting and understanding corrosion.

Resistance to corrosion

Some metals are more intrinsically resistant to corrosion than others, either due to the fundamental nature of the electrochemical processes involved or due to the details of how reaction products form. If a more susceptible material is used, many techniques can be applied during an item's manufacture and use to protect its materials from damage.

Intrinsic chemistry

Gold nuggets do not naturally corrode, even on a geological time scale.

The materials most resistant to corrosion are those for which corrosion is thermodynamically unfavorable. Any corrosion products of gold or platinum tend to decompose spontaneously into pure metal, which is why these elements can be found in metallic form on Earth, and is a large part of their intrinsic value. More common "base" metals can only be protected by more temporary means.

Some metals have naturally slow reaction kinetics, even though their corrosion is thermodynamically favorable. These include such metals as zinc, magnesium, and cadmium. While corrosion of these metals is continuous and ongoing, it happens at an acceptably slow rate. An extreme example is graphite, which releases large amounts of energy upon oxidation, but has such slow kinetics that it is effectively immune to electrochemical corrosion under normal conditions.

Passivation

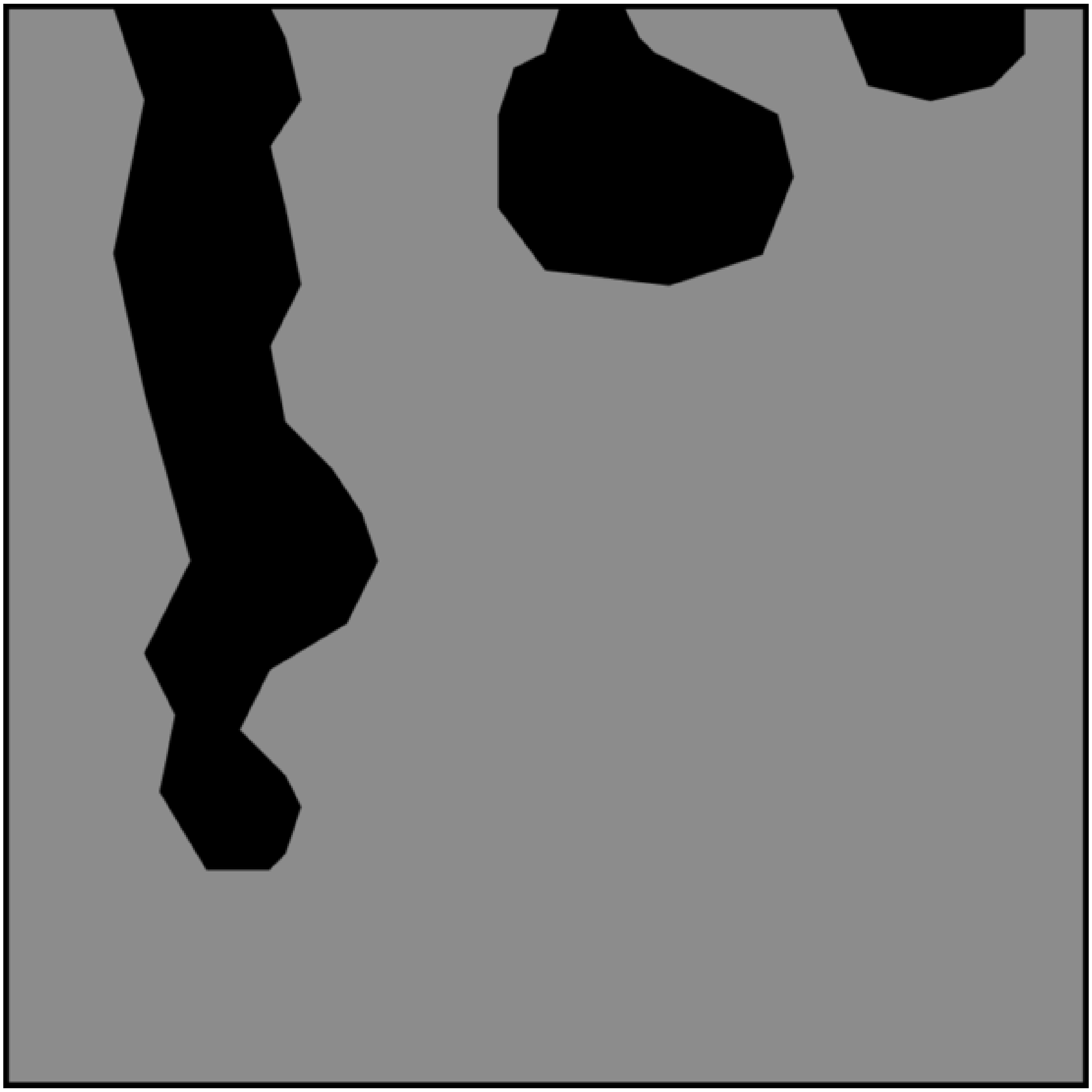
Given the right conditions, a thin film of corrosion products can form on a metal's surface spontaneously, acting as a barrier to further oxidation. When this layer stops growing at less than a micrometre thick under the conditions that a material will be used in, the phenomenon is known as passivation (rust, for example, usually grows to be much thicker, and so is not considered passivation, because this mixed oxidized layer is not protective). While this effect is in some sense a property of the material, it serves as an indirect kinetic barrier: the reaction is often quite rapid unless and until an impermeable layer forms. Passivation in air and water at moderate pH is seen in such materials as aluminium, stainless steel, titanium, and silicon.

These conditions required for passivation are specific to the material. The effect of pH is recorded using Pourbaix diagrams, but many other factors are influential. Some conditions that inhibit passivation include: high pH for aluminium, low pH or the presence of chloride ions for stainless steel, high temperature for titanium (in which case the oxide dissolves into the metal, rather than the electrolyte) and fluoride ions for silicon. On the other hand, sometimes unusual conditions can bring on passivation in materials that are normally unprotected, as the alkaline environment of concrete does for steel rebar. Exposure to a liquid metal such as mercury or hot solder can often circumvent passivation mechanisms.

Corrosion in passivated materials

Passivation is extremely useful in alleviating corrosion damage, but care must be taken not to trust it too thoroughly. Even a high-quality alloy will corrode if its ability to form a passivating film is hindered. Because the resulting modes of corrosion are more exotic and their immediate results are less visible than rust and other bulk corrosion, they often escape notice and cause problems among those who are not familiar with them.

Pitting corrosion



The scheme of pitting corrosion

Certain conditions, such as low concentrations of oxygen or high concentrations of species such as chloride which compete as anions, can interfere with a given alloy's ability to re-form a passivating film. In the worst case, almost all of the surface will remain protected, but tiny local fluctuations will degrade the oxide film in a few critical points. Corrosion at these points will be greatly amplified, and can cause *corrosion pits* of several types, depending upon conditions. While the corrosion pits only nucleate under fairly extreme circumstances, they can continue to grow even when conditions return to normal, since the interior of a pit is naturally deprived of oxygen and locally the pH decreases to very low values and the corrosion rate increases due to an auto-catalytic process. In extreme cases, the sharp tips of extremely long and narrow corrosion pits can cause stress concentration to the point that otherwise tough alloys can shatter; a thin film pierced by an invisibly small hole can hide a thumb sized pit from view. These problems are especially dangerous because they are difficult to detect before a part or structure fails. Pitting remains among the most common and damaging forms of corrosion in passivated alloys, but it can be prevented by control of the alloy's environment.

Weld decay and knifeline attack

Stainless steel can pose special corrosion challenges, since its passivating behavior relies on the presence of a minor alloying component (Chromium, typically only 18%). Due to the elevated temperatures of welding or during improper heat treatment, chromium carbides can form in the grain boundaries of stainless alloys. This chemical reaction robs the material of chromium in the zone near the grain boundary, making those areas much less resistant to corrosion. This creates a galvanic couple with the well-protected alloy nearby, which leads to *weld decay* (corrosion of the grain boundaries near welds) in highly corrosive environments. Special alloys, either with low carbon content or with added carbon "getters" such as titanium and niobium (in types 321 and 347, respectively), can prevent this effect, but the latter require special heat treatment after welding to prevent the similar phenomenon of *knifeline attack*. As its name applies, this is limited to a small zone, often only a few micrometres across, which causes it to proceed more rapidly. This zone is very near the weld, making it even less noticeable[1].

Crevice corrosion

Crevice corrosion is a localized form of corrosion occurring in spaces to which the access of the working fluid from the environment is limited and a concentration cell, areas with different oxygen concentration, will take place with consequent high corrosion rate . These spaces are generally called crevices. Examples of crevices are gaps and contact areas between parts, under gaskets or seals, inside cracks and seams, spaces filled with deposits and under sludge piles.

Microbial corrosion

Microbial corrosion, or bacterial corrosion, is a corrosion caused or promoted by microorganisms, usually chemoautotrophs. It can apply to both metals and non-metallic materials, in both the presence and lack of oxygen. Sulfate-reducing bacteria are common in lack of oxygen; they produce hydrogen sulfide, causing sulfide stress cracking. In presence of oxygen, some bacteria directly oxidize iron to iron oxides and hydroxides, other bacteria oxidize sulfur and produce sulfuric acid causing biogenic sulfide corrosion. Concentration cells can form in the deposits of corrosion products, causing and enhancing galvanic corrosion.

Accelerated Low Water Corrosion (ALWC) is a particularly aggressive form of MIC that affects steel piles in seawater near the low water tide mark. It is characterised by an orange sludge, which smells of Hydrogen Sulphide when treated with acid. Corrosion rates can be very high and design corrosion allowances can soon be exceeded leading to premature failure of the steel pile. Piles that have been coating and have cathodic protection installed at the time of construction are not susceptible to ALWC. For unprotected piles, sacrificial anodes can be installed local to the affected areas to inhibit the corrosion or a complete retrofitted sacrificial anode system can be installed. Affected areas can also be treated electrochemically by using an electrode to first produce chlorine to kill the bacteria, and then to produced a calcareous deposit, which will help shield the metal from further attack.

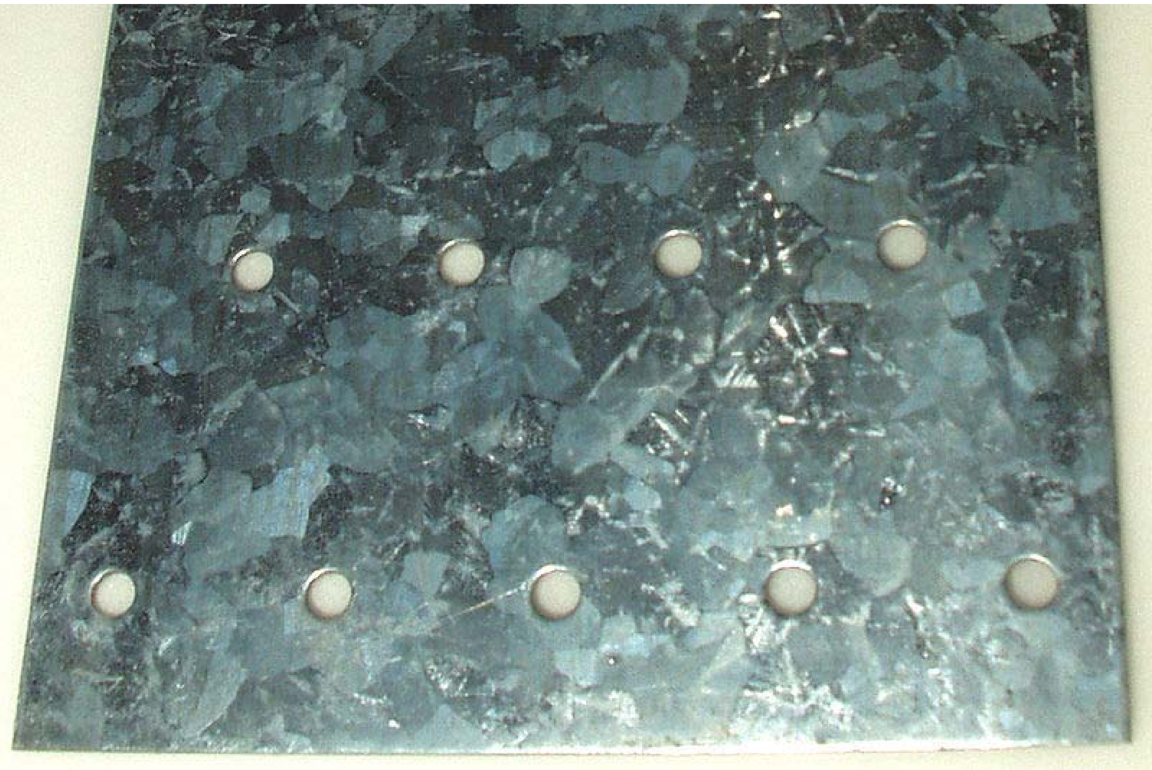
High temperature corrosion

High temperature corrosion is chemical deterioration of a material (typically a metal) under very high temperature conditions. This non-galvanic form of corrosion can occur when a metal is subject to a high temperature atmosphere containing oxygen, sulfur or other compounds capable of oxidising (or assisting the oxidation of) the material concerned. For example, materials used in aerospace, power generation and even in car engines have to resist sustained periods at high temperature in which they may be exposed to an atmosphere containing potentially highly corrosive products of combustion.

The products of high temperature corrosion can potentially be turned to the advantage of the engineer. The formation of oxides on stainless steels, for example, can provide a protective layer preventing further atmospheric attack, allowing for a material to be used for sustained periods at both room and high temperature in hostile conditions. Such high temperature corrosion products in the form of compacted oxide layer glazes have also been shown to prevent or reduce wear during high temperature sliding contact of metallic (or metallic and ceramic) surfaces.

Methods of protection from corrosion

Surface treatments



Galvanized surface

Applied coatings

Plating, painting, and the application of enamel are the most common anti-corrosion treatments. They work by providing a barrier of corrosion-resistant material between the damaging environment and the (often cheaper, tougher, and/or easier-to-process) structural material. Aside from cosmetic and manufacturing issues, there are tradeoffs in mechanical flexibility versus resistance to abrasion and high temperature. Platings usually fail only in small sections, and if the plating is more noble than the substrate (for example, chromium on steel), a galvanic couple will cause any exposed area to corrode much more rapidly than an unplated surface would. For this reason, it is often wise to plate with a more active metal such as zinc or cadmium. Painting either by roller or brush is more desirable for tight spaces; spray would be better for larger coating areas such as steel decks and waterfront applications. Flexible polyurethane coatings, like Durabak-M26 for example, can provide an anti-corrosive seal with a highly durable slip resistant membrane. Painted coatings are relatively easy to apply and have fast drying times although temperature and humidity may cause dry times to vary.

Reactive coatings

If the environment is controlled (especially in recirculating systems), corrosion inhibitors can often be added to it. These form an electrically insulating and/or chemically impermeable coating on exposed metal surfaces, to suppress electrochemical reactions. Such methods obviously make the system less sensitive to scratches or defects in the coating, since extra inhibitors can be made available wherever metal becomes exposed. Chemicals that inhibit corrosion include some of the salts in hard water (Roman water systems are famous for their mineral deposits), chromates, phosphates, polyaniline, other conducting polymers and a wide range of specially-designed chemicals that resemble surfactants (i.e. long-chain organic molecules with ionic end groups).

This figure-8 descender is annodized with a yellow finish. Climbing equipment is available in a wide range of anodized colors.

Anodization

Aluminium alloys often undergo a surface treatment. Electrochemical conditions in the bath are carefully adjusted so that uniform pores several nanometers wide appear in the metal's oxide film. These pores allow the oxide to grow much thicker than passivating conditions would allow. At the end of the treatment, the pores are allowed to seal, forming a harder-than-usual surface layer. If this coating is scratched, normal passivation processes take over to protect the damaged area. Anodizing is very resilient to weathering and corrosion, so it is commonly used for building facades and other areas that the surface will come into regular contact with the elements. Whilst being resilient, it must be cleaned frequently. If left without cleaning Panel Edge Staining will naturally occur.

Controlled permeability formwork

Controlled permeability formwork (CPF) is a method of preventing the corrosion of reinforcement by naturally enhancing the durability of the cover during concrete placement, . CPF has been used in environments to combat the effects of carbonation, chlorides, frost and abrasion.

Cathodic protection

Cathodic protection (CP) is a technique to control the corrosion of a metal surface by making that surface the cathode of an electrochemical cell. Cathodic protection systems are most commonly used to protect steel, water, and fuel pipelines and tanks; steel pier piles, ships, and offshore oil platforms.

Sacrificial anode protection

Sacrificial anode in the hull of a ship.

For effective CP, the potential of the steel surface is polarized (pushed) more negative until the metal surface has a uniform potential. With a uniform potential, the driving force for the corrosion reaction is halted. For galvanic CP systems, the anode material corrodes under the influence of the steel, and eventually it must be replaced. The polarization is caused by the current flow from the anode to the cathode, driven by the difference in electrochemical potential between the anode and the cathode.

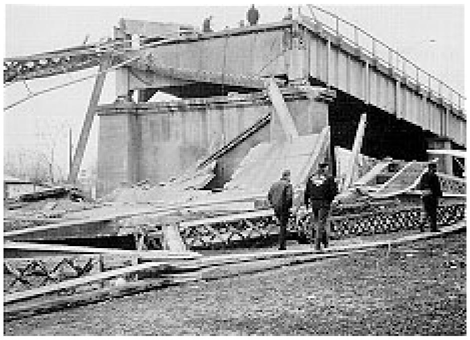
Impressed current cathodic protection

For larger structures, galvanic anodes cannot economically deliver enough current to provide complete protection. Impressed Current Cathodic Protection (ICCP) systems use anodes connected to a DC power source (such as a cathodic protection rectifier). Anodes for ICCP systems are tubular and solid rod shapes of various specialized materials. These include high silicon cast iron, graphite, mixed metal oxide or platinum coated titanium or niobium coated rod and wires.

Anodic protection

Anodic protection impresses anodic current on the structure to be protected (opposite to the cathodic protection). It is appropriate for metals that exhibit passivity (e.g., stainless steel) and suitably small passive current over a wide range of potentials. It is used in aggressive environments, e.g., solutions of sulfuric acid.

Economic impact



The collapsed Silver Bridge, as seen from the Ohio side

The US Federal Highway Administration released a study, entitled *Corrosion Costs and Preventive Strategies in the United States,* in 2002 on the direct costs associated with metallic corrosion in nearly every U.S. industry sector. The study showed that for 1998 the total annual estimated direct cost of corrosion in the U.S. was approximately $276 billion (approximately 3.2% of the US gross domestic product).

Rust is one of the most common causes of bridge accidents. As rust has a much higher volume than the originating mass of iron, its build-up can also cause failure by forcing apart adjacent parts. It was the cause of the collapse of the Mianus river bridge in 1983, when the bearings rusted internally and pushed one corner of the road slab off its support.

Three drivers on the roadway at the time died as the slab fell into the river below. The following NTSB investigation showed that a drain in the road had been blocked for road re-surfacing, and had not been unblocked so that runoff water penetrated the support hangers. It was also difficult for maintenance engineers to see the bearings from the inspection walkway. Rust was also an important factor in the Silver Bridge disaster of 1967 in West Virginia, when a steel suspension bridge collapsed in less than a minute, killing 46 drivers and passengers on the bridge at the time.

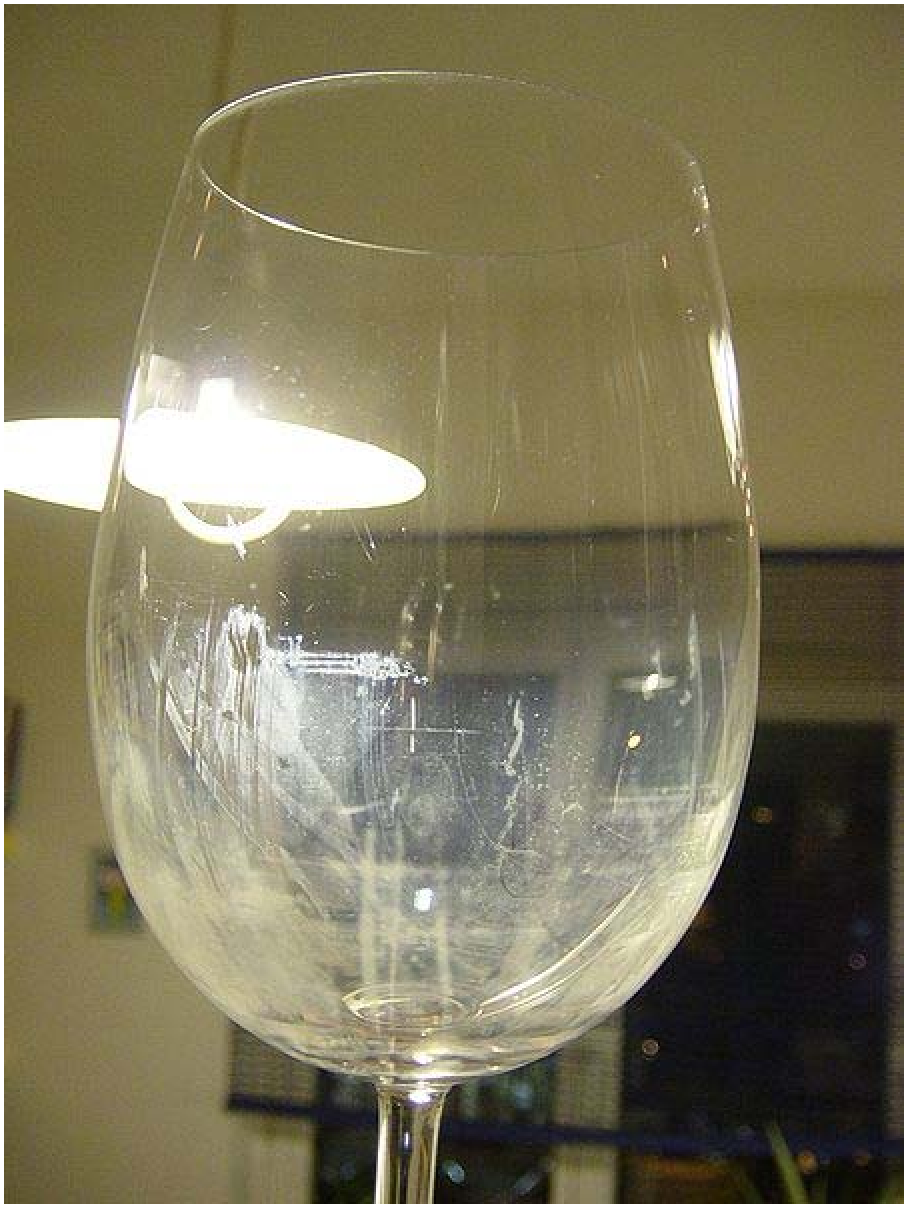
Similarly corrosion of concrete-covered steel and iron can cause the concrete to spall, creating severe structural problems. It is one of the most common failure modes of reinforced concrete bridges. Measuring instruments based on the half-cell potential are able to detect the potential corrosion spots before total failure of the concrete structure is reached.

Corrosion in nonmetals

Most ceramic materials are almost entirely immune to corrosion. The strong ionic and/or covalent 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.

Polymer degradation is due to a wide array of complex and often poorly-understood physiochemical processes. These are strikingly different from the other processes discussed here, and so the term "corrosion" is only applied to them in a loose sense of the word. Because of their large molecular weight, very little entropy can be gained by mixing a given mass of polymer with another substance, making them generally quite difficult to dissolve. While dissolution is a problem in some polymer applications, it is relatively simple to design against. A more common and related problem is *swelling*, where small molecules infiltrate the structure, reducing strength and stiffness and causing a volume change. Conversely, many polymers (notably flexible vinyl) are intentionally swelled with plasticizers, which can be leached out of the structure, causing brittleness or other undesirable changes. The most common form of degradation, however, is a decrease in polymer chain length. Mechanisms which break polymer chains are familiar to biologists because of their effect on DNA: ionizing radiation (most commonly ultraviolet light), free radicals, and oxidizers such as oxygen, ozone, and chlorine. Additives can slow these process very effectively, and can be as simple as a UV-absorbing pigment (i.e., titanium dioxide or carbon black). Plastic shopping bags often do not include these additives so that they break down more easily as litter.

Corrosion of glasses



Glass corrosion

Glass disease is the corrosion of silicate glasses in aqueous solutions. It is governed by two mechanisms: diffusion-controlled leaching (ion exchange) and glass network hydrolytic dissolution. Both corrosion 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 $10^{0.5pH}$

Mathematically, corrosion rates of glasses are characterised by normalised corrosion rates of elements NR_i (g/cm^2 d) which are determined as the ratio of total amount of released species into the water M_i (g) to the water-contacting surface area S (cm^2), time of contact t (days) and weight fraction content of the element in the glass f_i:

$$NR_i = \frac{M_i}{Sf_it}.$$

The overall corrosion rate is a sum of contributions from both mechanisms (leaching + dissolution) NR_i=Nrx_i+NRh. Diffusion-controlled leaching (ion exchange) is characteristic of the initial phase of corrosion and involves replacement of alkali ions in the glass by a hydronium (H_3O^+) ion from the solution. It causes an ion-selective depletion of near surface layers of glasses and gives an inverse square root dependence of corrosion rate with exposure time. The diffusion controlled normalised leaching rate of cations from glasses (g/cm^2 d) is given by:

$$NRx_i = 2\rho\sqrt{\frac{D_i}{\pi t}},$$

where t is time, D_i is the i-th cation effective diffusion coefficient (cm^2/d), which depends on pH of contacting water as $D_i = D_{i0} \cdot 10^{-pH}$, and ρ is the density of the glass (g/cm^3).

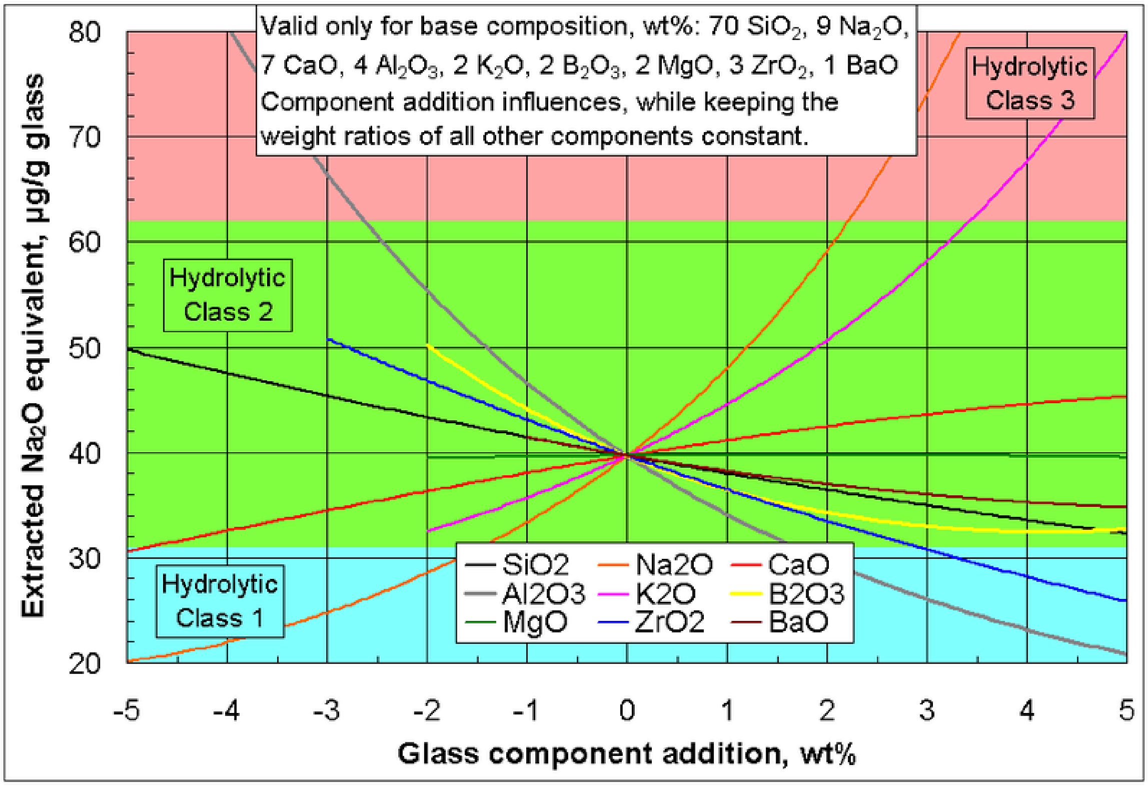
Glass network dissolution is characteristic of the later phases of corrosion and causes a congruent release of ions into the water solution at a time-independent rate in dilute solutions (g/cm^2 d):

$$NRh = \rho r_h,$$

where r_h is the stationary hydrolysis (dissolution) rate of the glass (cm/d). In closed systems the consumption of protons from the aqueous phase increases the pH and causes a fast transition to hydrolysis. However further silica saturation of solution impedes hydrolysis and causes the glass to return to an ion-exchange, e.g. diffusion-controlled regime of corrosion.

In typical natural conditions normalised corrosion rates of silicate glasses are very low and are of the order of 10^{-7} - 10^{-5} g/cm^2 d. The very high durability of silicate glasses in water makes them suitable for hazardous and nuclear waste immobilisation.

Glass corrosion tests



Influences of selected glass component additions on the chemical durability against water corrosion of a specific base glass (corrosion test ISO 719).

There exist numerous standardized procedures for measuring the corrosion (also called **chemical durability**) of glasses in neutral, basic, and acidic environments, under simulated environmental conditions, in simulated body fluid, at high temperature and pressure, and under other conditions.

In the standard procedure ISO 719 a test of the extraction of water soluble basic compounds under neutral conditions is described: 2 g glass, particle size 300-500 µm, is kept for 60 min in 50 ml de-ionized water of grade 2 at 98°C. 25 ml of the obtained solution is titrated against 0.01 mol/l HCl solution. The volume of HCl needed for neutralization is recorded and classified following the values in the table below.

0.01M HCl needed to neutralize extracted basic oxides, ml	Extracted Na_2O equivalent, µg	Hydrolytic class
to 0.1	to 31	1
above 0.1 to 0.2	above 31 to 62	2
above 0.2 to 0.85	above 62 to 264	3
above 0.85 to 2.0	above 264 to 620	4
above 2.0 to 3.5	above 620 to 1085	5

above 3.5	above 1085	>5

Chapter 3

Creep (Deformation)

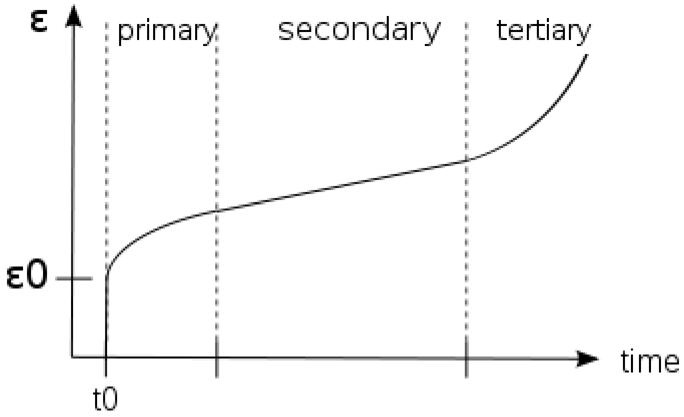
In materials science, **creep** is the tendency of a solid material to slowly move or deform permanently under the influence of stresses. It occurs as a result of long term exposure to high levels of stress that are below the yield strength of the material. Creep is more severe in materials that are subjected to heat for long periods, and near melting point. Creep always increases with temperature.

The rate of this deformation is a function of the material properties, exposure time, exposure temperature and the applied structural load. Depending on the magnitude of the applied stress and its duration, the deformation may become so large that a component can no longer perform its function — for example creep of a turbine blade will cause the blade to contact the casing, resulting in the failure of the blade. Creep is usually of concern to engineers and metallurgists when evaluating components that operate under high stresses or high temperatures. Creep is a deformation mechanism that may or may not constitute a failure mode. Moderate creep in concrete is sometimes welcomed because it relieves tensile stresses that might otherwise lead to cracking.

Unlike brittle fracture, creep deformation does not occur suddenly upon the application of stress. Instead, strain accumulates as a result of long-term stress. Creep is a "time-dependent" deformation.

The temperature range in which creep deformation may occur differs in various materials. For example, tungsten requires a temperature in the thousands of degrees before creep deformation can occur while ice will creep near 0 °C (32 °F). As a rule of thumb, the effects of creep deformation generally become noticeable at approximately 30% of the melting point for metals and 40–50% of melting point for ceramics. Virtually any material will creep upon approaching its melting temperature. Since the minimum temperature is relative to melting point, creep can be seen at relatively low temperatures for some materials. Plastics and low-melting-temperature metals, including many solders, creep at room temperature as can be seen markedly in old lead hot-water pipes. Glacier flow is an example of creep processes in ice.

Stages of creep



Strain as a function of time due to constant stress over an extended period for a viscoelastic material.

In the initial stage, or primary creep, the strain rate is relatively high, but slows with increasing strain. This is due to work hardening. The strain rate eventually reaches a minimum and becomes near constant. This is due to the balance between work hardening and annealing (thermal softening). This stage is known as secondary or steady-state creep. This stage is the most understood. The characterized "creep strain rate" typically refers to the rate in this secondary stage. Stress dependence of this rate depends on the creep mechanism. In tertiary creep, the strain rate exponentially increases with strain because of necking phenomena.

Mechanisms of creep

The mechanism of creep depends on temperature and stress. The various methods are:

- Bulk diffusion (Nabarro-Herring creep)
- Climb — here the strain is actually accomplished by climb
- Climb-assisted glide — here the climb is an *enabling* mechanism, allowing dislocations to get around obstacles
- Grain boundary diffusion (Coble creep)
- Thermally activated glide — e.g., via cross-slip

General creep equation

$$\frac{d\varepsilon}{dt} = \frac{C\sigma^m}{d^b} e^{\frac{-Q}{kT}}$$

where ε is the creep strain, *C* is a constant dependent on the material and the particular creep mechanism, *m* and *b* are exponents dependent on the creep mechanism, *Q* is the activation energy of the creep mechanism, σ is the applied stress, *d* is the grain size of the material, *k* is Boltzmann's constant, and *T* is the absolute temperature.

Dislocation creep

At high stresses (relative to the shear modulus), creep is controlled by the movement of dislocations. For dislocation creep, *Q = Q(self diffusion), m* = 4-6, and *b* = 0. Therefore, dislocation creep has a strong dependence on the applied stress and no grain size dependence.

Some alloys exhibit a very large stress exponent (*n* > 10), and this has typically been explained by introducing a "threshold stress," σ_{th}, below which creep can't be measured. The modified power law equation then becomes:

$$\frac{d\epsilon}{dt} = A\left(\sigma - \sigma_{th}\right)^n e^{\frac{-Q}{RT}}$$

where *A, Q* and *n* can all be explained by conventional mechanisms (so $3 \leq n \leq 10$).

Nabarro-Herring creep

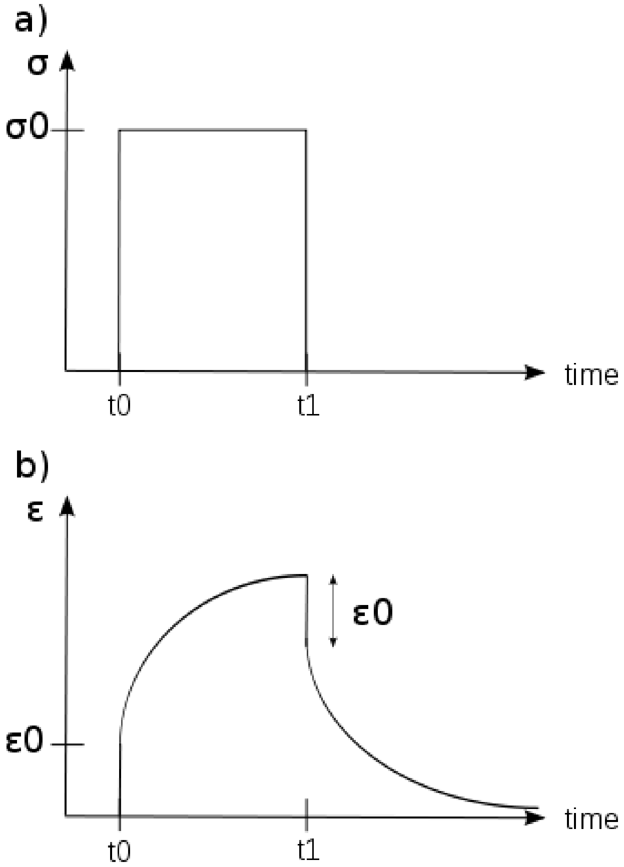
Nabarro-Herring creep is a form of diffusion controlled creep. In Nabarro-Herring creep, atoms diffuse through the lattice causing grains to elongate along the stress axis; *k* is related to the diffusion coefficient of atoms through the lattice, *Q = Q(self diffusion), m* = 1, and *b* = 2. Therefore Nabarro-Herring creep has a weak stress dependence and a moderate grain size dependence, with the creep rate decreasing as grain size is increased.

Nabarro-Herring creep is strongly temperature dependent. For lattice diffusion of atoms to occur in a material, neighboring lattice sites or interstitial sites in the crystal structure must be free. A given atom must also overcome the energy barrier to move from its current site (it lies in an energetically favorable potential well) to the nearby vacant site (another potential well). The general form of the diffusion equation is $D = D_0 exp(E/KT)$ where D_0 has a dependence on both the attempted jump frequency and the number of nearest neighbor sites and the probability of the sites being vacant. Thus there is a double dependence upon temperature. At higher temperatures the diffusivity increases due to the direct temperature dependence of the equation, the increase in vacancies through Schottky defect formation, and an increase in the average energy of atoms in the material. Nabarro-Herring creep dominates at very high temperatures relative to a material's melting temperature.

Coble creep

Coble creep is a second form of diffusion controlled creep. In Coble creep the atoms diffuse along grain boundaries to elongate the grains along the stress axis. This causes Coble creep to have a stronger grain size dependence than Nabarro-Herring creep. For Coble creep *k* is related to the diffusion coefficient of atoms along the grain boundary, *Q = Q(grain boundary diffusion), m = 1*, and *b = 3*. Because *Q(grain boundary diffusion)* < *Q(self diffusion)*, Coble creep occurs at lower temperatures than Nabarro-Herring creep. Coble creep is still temperature dependent, as the temperature increases so does the grain boundary diffusion. However, since the number of nearest neighbors is effectively limited along the interface of the grains, and thermal generation of vacancies along the boundaries is less prevalent, the temperature dependence is not as strong as in Nabarro-Herring creep. It also exhibits the same linear dependence on stress as Nabarro-Herring creep.

Creep of polymers



a) Applied stress and b) induced strain as functions of time over a short period for a viscoelastic material.

Creep can occur in polymers and metals which are considered viscoelastic materials. When a polymeric material is subjected to an abrupt force, the response can be modeled using the Kelvin-Voigt model. In this model, the material is represented by a Hookean spring and a Newtonian dashpot in parallel. The creep strain is given by

$$\varepsilon(t) = \sigma C_0 + \sigma C \int_0^\infty f(\tau)(1 - \exp[-t/\tau])\, d\tau$$

where:

- σ = applied stress
- C_0 = *instantaneous creep compliance*
- C = creep compliance coefficient
- τ = retardation time
- $f(\tau)$ = distribution of retardation times

When subjected to a step constant stress, viscoelastic materials experience a time-dependent increase in strain. This phenomenon is known as viscoelastic creep.

At a time t_0, a viscoelastic material is loaded with a constant stress that is maintained for a sufficiently long time period. The material responds to the stress with a strain that increases until the material ultimately fails. When the stress is maintained for a shorter time period, the material undergoes an initial strain until a time t_1 at which the stress is relieved, at which time the strain immediately decreases (discontinuity) then continues decreasing gradually to a residual strain.

Viscoelastic creep data can be presented in one of two ways. Total strain can be plotted as a function of time for a given temperature or temperatures. Below a critical value of applied stress, a material may exhibit linear viscoelasticity. Above this critical stress, the creep rate grows disproportionately faster. The second way of graphically presenting viscoelastic creep in a material is by plotting the creep modulus (constant applied stress divided by total strain at a particular time) as a function of time. Below its critical stress, the viscoelastic creep modulus is independent of stress applied. A family of curves describing strain versus time response to various applied stress may be represented by a single viscoelastic creep modulus versus time curve if the applied stresses are below the material's critical stress value.

Additionally, the molecular weight of the polymer of interest is known to affect its creep behavior. The effect of increasing molecular weight tends to promote secondary bonding between polymer chains and thus make the polymer more creep resistant. Similarly, aromatic polymers are even more creep resistant due to the added stiffness from the rings. Both molecular weight and aromatic rings add to polymers' thermal stability, increasing the creep resistance of a polymer.

Both polymers and metals can creep. Polymers experience significant creep at temperatures above ca. –200°C; however, there are three main differences between polymeric and metallic creep.

Polymers show creep basically in two different ways. At typical work loads (5 up to 50%) ultra high molecular weight polyethylene (Spectra, Dyneema) will show time-linear creep, whereas polyester or aramids (Twaron, Kevlar) will show a time-logarithmic creep.

Applications

Though mostly due to the reduced yield stress at higher temperatures, the Collapse of the World Trade Center was due in part to creep from increased temperature operation.

The creep rate of hot pressure-loaded components in a nuclear reactor at power can be a significant design-constraint, since the creep rate is enhanced by the flux of energetic particles.

Creep was blamed for the Big Dig tunnel ceiling collapse in Boston, Massachusetts that occurred in July 2006.

An example of an application involving creep deformation is the design of tungsten light bulb filaments. Sagging of the filament coil between its supports increases with time due to creep deformation caused by the weight of the filament itself. If too much deformation occurs, the adjacent turns of the coil touch one another, causing an electrical short and local overheating, which quickly leads to failure of the filament. The coil geometry and supports are therefore designed to limit the stresses caused by the weight of the filament, and a special tungsten alloy with small amounts of oxygen trapped in the crystallite grain boundaries is used to slow the rate of Coble creep.

In steam turbine power plants, pipes carry steam at high temperatures (566 °C or 1050 °F) and pressures (above 24.1 MPa or 3500 psi). In jet engines, temperatures can reach up to 1400 °C (2550 °F) and initiate creep deformation in even advanced-coated turbine blades. Hence, it is crucial for correct functionality to understand the creep deformation behavior of materials.

Creep deformation is important not only in systems where high temperatures are endured such as nuclear power plants, jet engines and heat exchangers, but also in the design of many everyday objects. For example, metal paper clips are stronger than plastic ones because plastics creep at room temperatures. Aging glass windows are often erroneously used as an example of this phenomenon: measurable creep would only occur at temperatures above the glass transition temperature around 500 °C (900 °F). While glass does exhibit creep under the right conditions, apparent sagging in old windows may instead be a consequence of obsolete manufacturing processes, such as that used to create crown glass, which resulted in inconsistent thickness.

Chapter 4

Fatigue (Material)

In materials science, **fatigue** is the progressive and localized structural damage that occurs when a material is subjected to cyclic loading. The nominal maximum stress values are less than the ultimate tensile stress limit, and may be below the yield stress limit of the material.

Fatigue occurs when a material is subjected to repeated loading and unloading. If the loads are above a certain threshold, microscopic cracks will begin to form at the surface. Eventually a crack will reach a critical size, and the structure will suddenly fracture. The shape of the structure will significantly affect the fatigue life; square holes or sharp corners will lead to elevated local stresses where fatigue cracks can initiate. Round holes and smooth transitions or fillets are therefore important to increase the fatigue strength of the structure.

Fatigue life

ASTM defines *fatigue life*, N_f, as the number of stress cycles of a specified character that a specimen sustains before failure of a specified nature occurs.

One method to predict fatigue life of materials is the Uniform Material Law (UML). UML was developed for fatigue life prediction of aluminum and titanium alloys by the end of 20th century and extended to high-strength steels.

Characteristics of fatigue

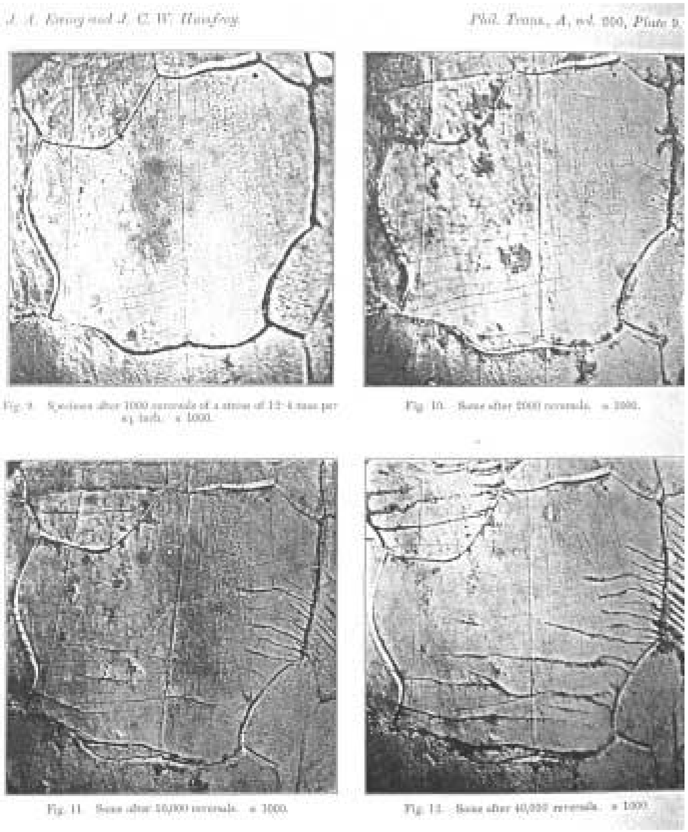
Fracture of an aluminium crank arm. Dark area of striations: slow crack growth. Bright granular area: sudden fracture.

- In metals and alloys, the process starts with dislocation movements, eventually forming persistent slip bands that nucleate short cracks.
- Fatigue is a stochastic process, often showing considerable scatter even in controlled environments.
- The greater the applied stress range, the shorter the life.
- Fatigue life scatter tends to increase for longer fatigue lives.
- Damage is cumulative. Materials do not recover when rested.
- Fatigue life is influenced by a variety of factors, such as temperature, surface finish, microstructure, presence of oxidizing or inert chemicals, residual stresses, contact (fretting), etc.
- Some materials (e.g., some steel and titanium alloys) exhibit a theoretical fatigue limit below which continued loading does not lead to structural failure.

- In recent years, researchers (see, for example, the work of Bathias, Murakami, and Stanzl-Tschegg) have found that failures occur below the theoretical fatigue limit at very high fatigue lives (10^9 to 10^{10} cycles). An ultrasonic resonance technique is used in these experiments with frequencies around 10–20 kHz.
- High cycle fatigue strength (about 10^3 to 10^8 cycles) can be described by stress-based parameters. A load-controlled servo-hydraulic test rig is commonly used in these tests, with frequencies of around 20–50 Hz. Other sorts of machines—like resonant magnetic machines—can also be used, achieving frequencies up to 250 Hz.
- Low cycle fatigue (typically less than 10^3 cycles) is associated with widespread plasticity in metals; thus, a strain-based parameter should be used for fatigue life prediction in metals and alloys. Testing is conducted with constant strain amplitudes typically at 0.01–5 Hz.

Timeline of early fatigue research history

- 1837: Wilhelm Albert publishes the first article on fatigue. He devised a test machine for conveyor chains used in the Clausthal mines.
- 1839: Jean-Victor Poncelet describes metals as being *tired* in his lectures at the military school at Metz.
- 1842: William John Macquorn Rankine recognises the importance of stress concentrations in his investigation of railroad axle failures. The Versailles train crash was caused by axle fatigue.
- 1843: Joseph Glynn reports on fatigue of axle on locomotive tender. He identifies the keyway as the crack origin.
- 1848: Railway Inspectorate report one of the first tyre failures, probably from a rivet hole in tread of railway carriage wheel. It was likely a fatigue failure.
- 1849: Eaton Hodgkinson is granted a *small sum of money* to report to the UK Parliament on his work in *ascertaining by direct experiment, the effects of continued changes of load upon iron structures and to what extent they could be loaded without danger to their ultimate security*.
- 1854: Braithwaite reports on common service fatigue failures and coins the term *fatigue*.
- 1860: Systematic fatigue testing undertaken by Sir William Fairbairn and August Wöhler.
- 1870: Wöhler summarises his work on railroad axles. He concludes that cyclic stress range is more important than peak stress and introduces the concept of *endurance limit*.



Micrographs showing how surface fatigue cracks grow as material is further cycled. From Ewing & Humfrey (1903)

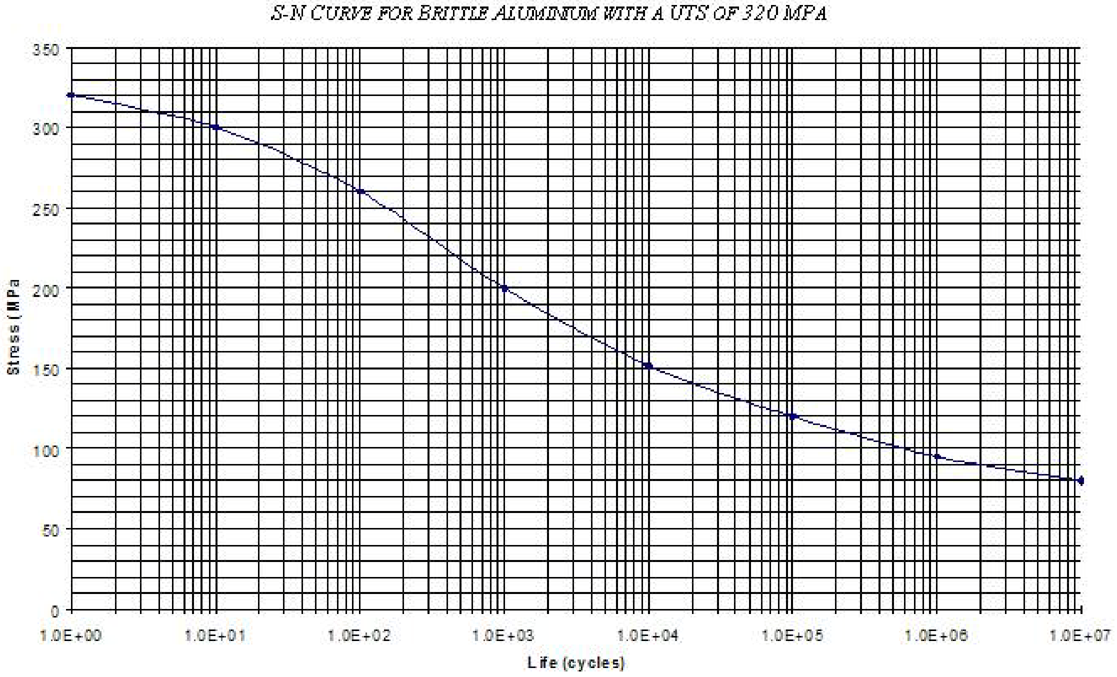
- 1903: Sir James Alfred Ewing demonstrates the origin of fatigue failure in microscopic cracks.
- 1910: O. H. Basquin proposes a log-log relationship for SN curves, using Wöhler's test data.
- 1945: A. M. Miner popularises A. Palmgren's (1924) linear damage hypothesis as a practical design tool.
- 1954: L. F. Coffin and S. S. Manson explain fatigue crack-growth in terms of plastic strain in the tip of cracks.
- 1961: P. C. Paris proposes methods for predicting the rate of growth of individual fatigue cracks in the face of initial scepticism and popular defence of Miner's phenomenological approach.
- 1968: Tatsuo Endo and M. Matsuishi devise the rainflow-counting algorithm and enable the reliable application of Miner's rule to random loadings.
- 1970: W. Elber elucidates the mechanisms and importance of crack closure in slowing the growth of a fatigue crack due to the wedging effect of plastic deformation left behind the tip of the crack.

High-cycle fatigue

Historically, most attention has focused on situations that require more than 10^4 cycles to failure where stress is low and deformation primarily elastic.

The S-N curve

In high-cycle fatigue situations, materials performance is commonly characterised by an *S-N curve*, also known as a *Wöhler curve* . This is a graph of the magnitude of a cyclic stress (*S*) against the logarithmic scale of cycles to failure (*N*).

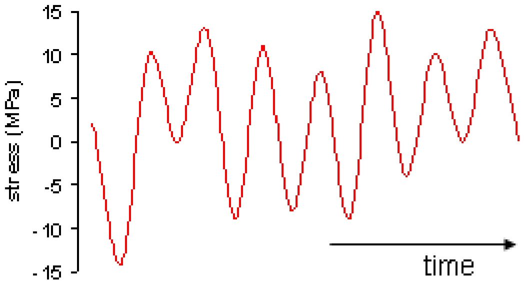


S-N curves are derived from tests on samples of the material to be characterised (often called *coupons*) where a regular sinusoidal stress is applied by a testing machine which also counts the number of cycles to failure. This process is sometimes known as *coupon testing*. Each coupon test generates a point on the plot though in some cases there is a *runout* where the time to failure exceeds that available for the test. Analysis of fatigue data requires techniques from statistics, especially survival analysis and linear regression.

Probabilistic nature of fatigue

As coupons sampled from a homogeneous frame will manifest variation in their number of cycles to failure, the S-N curve should more properly be an *S-N-P curve* capturing the probability of failure after a given number of cycles of a certain stress. Probability distributions that are common in data analysis and in design against fatigue include the lognormal distribution, extreme value distribution, Birnbaum–Saunders distribution, and Weibull distribution.

Complex loadings



Spectrum loading

In practice, a mechanical part is exposed to a complex, often random, sequence of loads, large and small. In order to assess the safe life of such a part:

1. Reduce the complex loading to a series of simple cyclic loadings using a technique such as rainflow analysis;
2. Create a histogram of cyclic stress from the rainflow analysis to form a fatigue damage spectrum;
3. For each stress level, calculate the degree of cumulative damage incurred from the S-N curve; and
4. Combine the individual contributions using an algorithm such as *Miner's rule*.

Miner's rule

In 1945, M. A. Miner popularised a rule that had first been proposed by A. Palmgren in 1924. The rule, variously called *Miner's rule* or the *Palmgren-Miner linear damage hypothesis*, states that where there are k different stress magnitudes in a spectrum, S_i $(1 \leq i \leq k)$, each contributing $n_i(S_i)$ cycles, then if $N_i(S_i)$ is the number of cycles to failure of a constant stress reversal S_i, failure occurs when:

$$\sum_{i=1}^{k} \frac{n_i}{N_i} = C$$

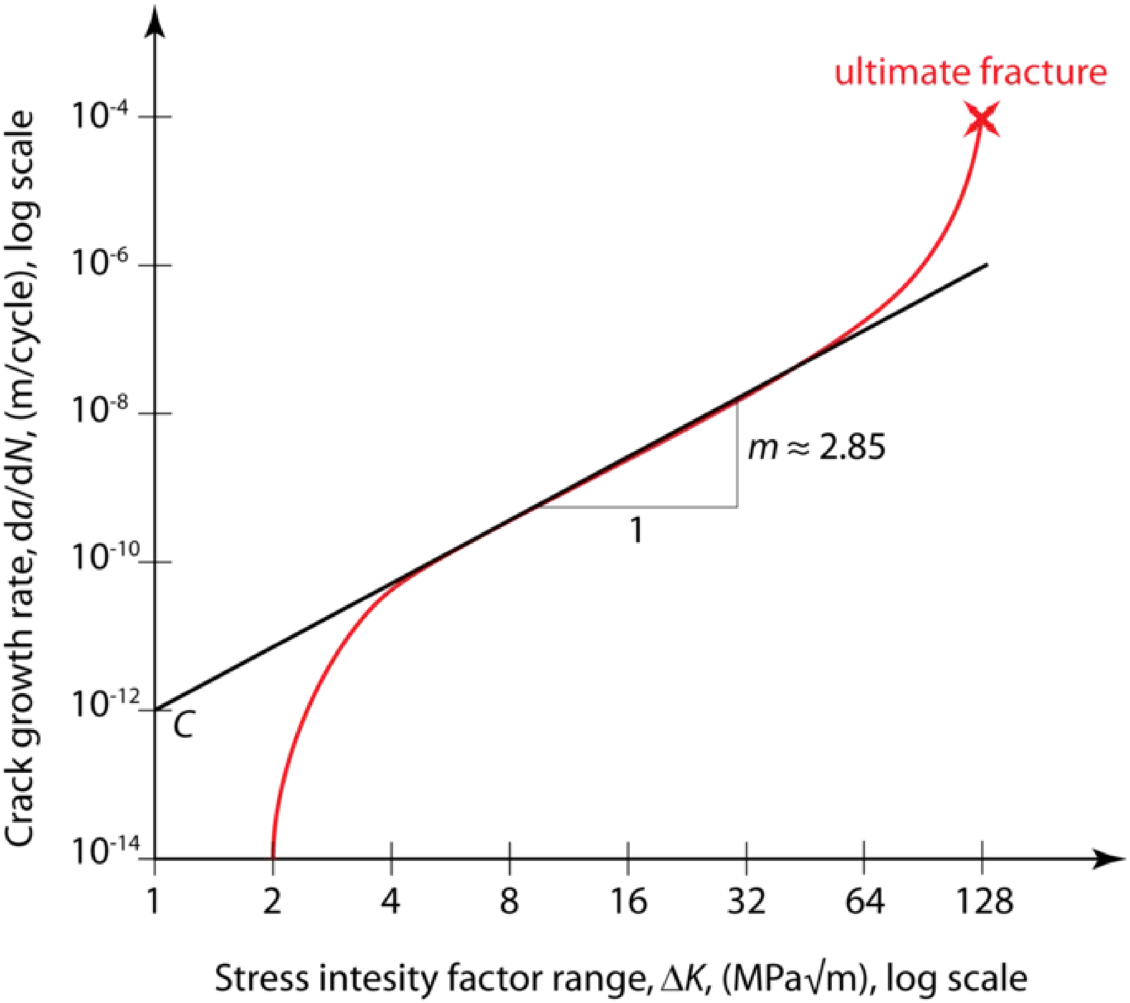
C is experimentally found to be between 0.7 and 2.2. Usually for design purposes, C is assumed to be 1.

This can be thought of as assessing what proportion of life is consumed by stress reversal at each magnitude then forming a linear combination of their aggregate.

Though Miner's rule is a useful approximation in many circumstances, it has several major limitations:

1. It fails to recognise the probabilistic nature of fatigue and there is no simple way to relate life predicted by the rule with the characteristics of a probability distribution. Industry analysts often use design curves, adjusted to account for scatter, to calculate $N_i(S_i)$.
2. There is sometimes an effect in the order in which the reversals occur. In some circumstances, cycles of low stress followed by high stress cause more damage than would be predicted by the rule. It does not consider the effect of overload or high stress which may result in a compressive residual stress. High stress followed by low stress may have less damage due to the presence of compressive residual stress.

Paris' Relationship



Typical fatigue crack growth rate graph.

In Fracture mechanics, Anderson, Gomez and Paris derived relationships for the stage II crack growth with cycles N, in terms of the cyclical component ΔK of the Stress Intensity Factor K

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C(\Delta K)^m$$

where a is the crack length and m is typically in the range 3 to 5 (for metals).

This relationship was later modified (by Forman, 1967) to make better allowance for the mean stress, by introducing a factor depending on (1-R) where R = min stress/max stress, in the denominator.

Low-cycle fatigue

Where the stress is high enough for plastic deformation to occur, the account in terms of stress is less useful and the strain in the material offers a simpler description. Low-cycle fatigue is usually characterised by the *Coffin-Manson relation* (published independently by L. F. Coffin in 1954 and S. S. Manson 1953):

$$\frac{\Delta\epsilon_p}{2} = \epsilon'_f(2N)^c$$

-where:

- $\Delta\varepsilon_p /2$ is the plastic strain amplitude;
- ε_f' is an empirical constant known as the *fatigue ductility coefficient*, the failure strain for a single reversal;
- $2N$ is the number of reversals to failure (N cycles);
- c is an empirical constant known as the *fatigue ductility exponent*, commonly ranging from -0.5 to -0.7 for metals in time independent fatigue. Slopes can be considerably steeper in the presence of creep or environmental interactions.

A similar relationship for materials such as Zirconium, used in the nuclear industry.

Fatigue and fracture mechanics

The account above is purely empirical and, though it allows life prediction and design assurance, life improvement or design optimisation can be enhanced using fracture mechanics. It can be developed in four stages.

1. Crack nucleation;
2. Stage I crack-growth;
3. Stage II crack-growth; and
4. Ultimate ductile failure.

Factors that affect fatigue-life

- **Cyclic stress state:** Depending on the complexity of the geometry and the loading, one or more properties of the stress state need to be considered, such as stress amplitude, mean stress, biaxiality, in-phase or out-of-phase shear stress, and load sequence,
- **Geometry:** Notches and variation in cross section throughout a part lead to stress concentrations where fatigue cracks initiate.
- **Surface quality.** Surface roughness cause microscopic stress concentrations that lower the fatigue strength. Compressive residual stresses can be introduced in the surface by e.g. shot peening to increase fatigue life. Such techniques for producing surface stress are often referred to as *peening*, whatever the mechanism used to produce the stress. Low Plasticity Burnishing, Laser peening, and ultrasonic impact treatment can also produce this surface compressive stress and can increase the fatigue life of the component. This improvement is normally observed only for high-cycle fatigue.

- **Material Type:** Fatigue life, as well as the behavior during cyclic loading, varies widely for different materials, e.g. composites and polymers differ markedly from metals.

- **Residual stresses:** Welding, cutting, casting, and other manufacturing processes involving heat or deformation can produce high levels of tensile residual stress, which decreases the fatigue strength.

- **Size and distribution of internal defects:** Casting defects such as gas porosity, non-metallic inclusions and shrinkage voids can significantly reduce fatigue strength.

- **Direction of loading:** For non-isotropic materials, fatigue strength depends on the **direction of the principal stress**.

- **Grain size:** For most metals, smaller grains yield longer fatigue lives, however, the presence of surface defects or scratches will have a greater influence than in a coarse grained alloy.

- **Environment:** Environmental conditions can cause erosion, corrosion, or gas-phase embrittlement, which all affect fatigue life. Corrosion fatigue is a problem encountered in many aggressive environments.

- **Temperature:** Extreme high or low temperatures can decrease fatigue strength.

Design against fatigue

Dependable design against fatigue-failure requires thorough education and supervised experience in structural engineering, mechanical engineering, or materials science. There are three principal approaches to life assurance for mechanical parts that display increasing degrees of sophistication:

1. Design to keep stress below threshold of fatigue limit (infinite lifetime concept);
2. Design (conservatively) for a fixed life after which the user is instructed to replace the part with a new one (a so-called *lifed* part, finite lifetime concept, or "safe-life" design practice);
3. Instruct the user to inspect the part periodically for cracks and to replace the part once a crack exceeds a critical length. This approach usually uses the technologies of nondestructive testing and requires an accurate prediction of the rate of crack-growth between inspections. This is often referred to as damage tolerant design or "retirement-for-cause".

Stopping fatigue

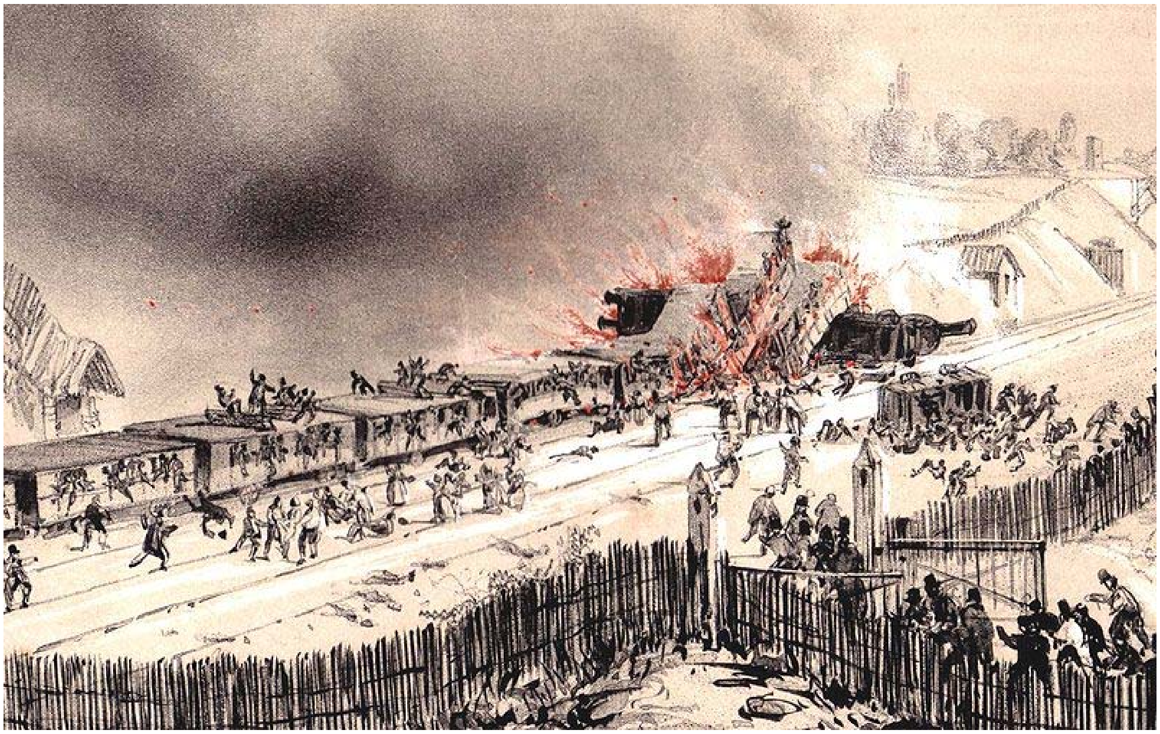
Fatigue cracks that have begun to propagate can sometimes be stopped by drilling holes, called *drill stops*, in the path of the fatigue crack. This is not recommended as a general practice because the hole represents a stress concentration factor which depends on the size of the hole and geometry. There is thus the possibility of a new crack starting in the side of the hole. It is always far better to replace the cracked part entirely.

Material change

Changes in the materials used in parts can also improve fatigue life. For example, parts can be made from better fatigue rated metals. Complete replacement and redesign of parts can also reduce if not eliminate fatigue problems. Thus helicopter rotor blades and propellers in metal are being replaced by composite equivalents. They are not only lighter, but also much more resistant to fatigue. They are more expensive, but the extra cost is amply repaid by their greater integrity, since loss of a rotor blade usually leads to total loss of the aircraft. A similar argument has been made for replacement of metal fuselages, wings and tails of aircraft.

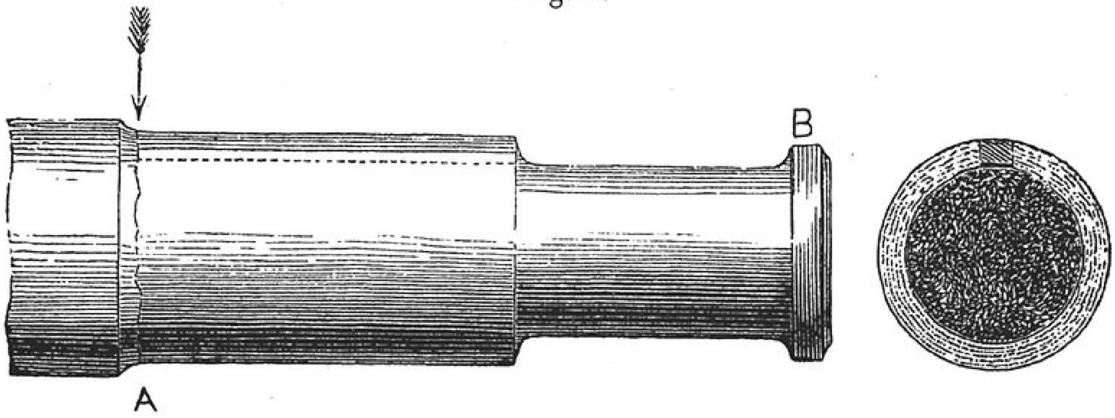
Infamous fatigue failures

Versailles train crash



Versailles train disaster

Fig. 1.

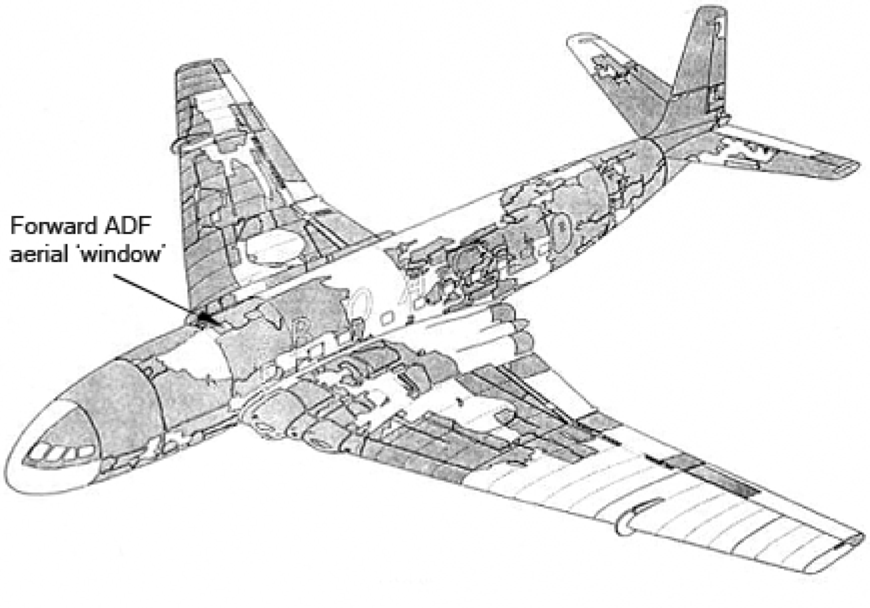


Drawing of a fatigue failure in an axle by Joseph Glynn, 1843.

Following the King's fete celebrations at the Palace of Versailles, a train returning to Paris crashed in May 1842 at Meudon after the leading locomotive broke an axle. The carriages behind piled into the wrecked engines and caught fire. At least 55 passengers were killed trapped in the carriages, including the explorer Jules Dumont d'Urville. This accident is known in France as the "Catastrophe ferroviaire de Meudon". The accident was witnessed by the British locomotive engineer Joseph Locke and widely reported in Britain. It was discussed extensively by engineers, who sought an explanation.

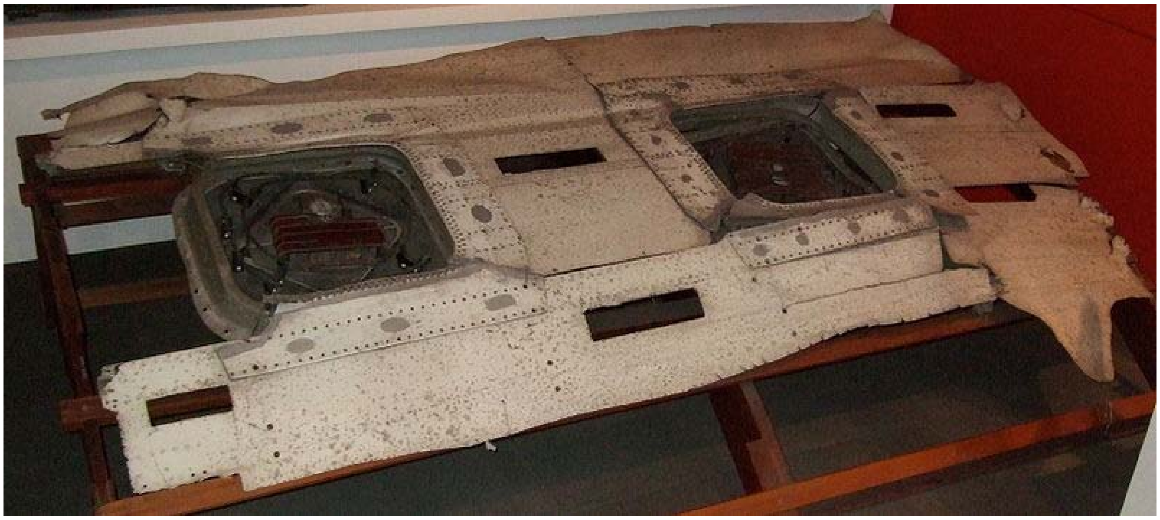
The derailment had been the result of a broken locomotive axle. Rankine's investigation of broken axles in Britain highlighted the importance of stress concentration, and the mechanism of crack growth with repeated loading. His and other papers suggesting a crack growth mechanism through repeated stressing, however, were ignored, and fatigue failures occurred at an ever increasing rate on the expanding railway system. Other spurious theories seemed to be more acceptable, such as the idea that the metal had somehow "crystallized". The notion was based on the crystalline appearance of the fast fracture region of the crack surface, but ignored the fact that the metal was already highly crystalline.

de Havilland Comet



The recovered (shaded) parts of the wreckage of *G-ALYP* and the site (arrowed) of the failure

Two de Havilland Comet passenger jets broke up in mid-air and crashed within a few months of each other in 1954. As a result systematic tests were conducted on a fuselage immersed and pressurised in a water tank. After the equivalent of 3,000 flights investigators at the Royal Aircraft Establishment (RAE) were able to conclude that the crash had been due to failure of the pressure cabin at the forward Automatic Direction Finder window in the roof. This 'window' was in fact one of two apertures for the aerials of an electronic navigation system in which opaque fibreglass panels took the place of the window 'glass'. The failure was a result of metal fatigue caused by the repeated pressurisation and de-pressurisation of the aircraft cabin. Another fact was that the supports around the windows were riveted, not bonded, as the original specifications for the aircraft had called for. The problem was exacerbated by the punch rivet construction technique employed. Unlike drill riveting, the imperfect nature of the hole created by punch riveting caused manufacturing defect cracks which may have caused the start of fatigue cracks around the rivet.

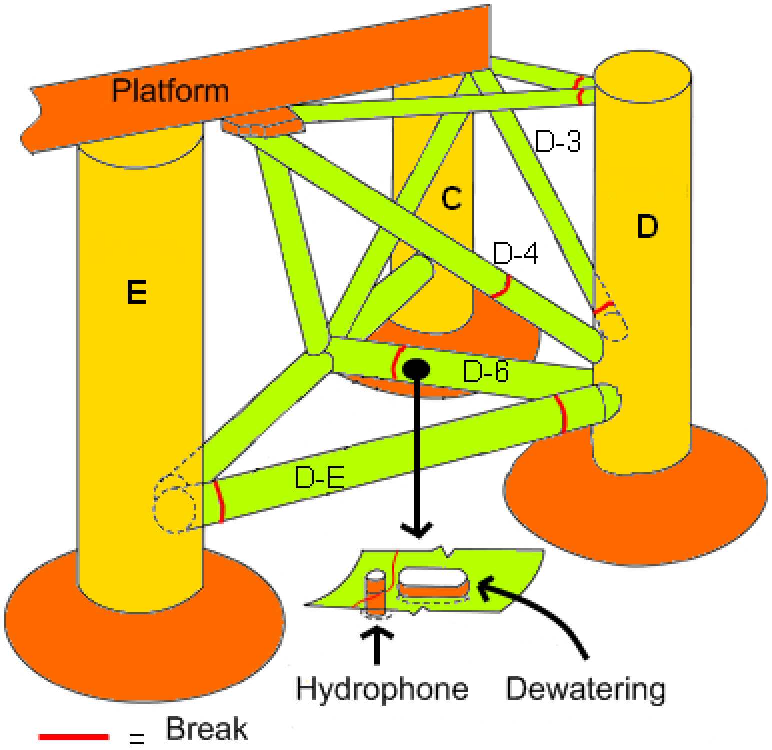


The fuselage fragment of *G-ALYP* on display in the Science Museum in London

The Comet's pressure cabin had been designed to a safety factor comfortably in excess of that required by British Civil Airworthiness Requirements (2.5 times the cabin proof pressure as opposed to the requirement of 1.33 times and an ultimate load of 2.0 times the cabin pressure) and the accident caused a revision in the estimates of the safe loading strength requirements of airliner pressure cabins.

In addition, it was discovered that the stresses around pressure cabin apertures were considerably higher than had been anticipated, especially around sharp-cornered cut-outs, such as windows. As a result, all future jet airliners would feature windows with rounded corners, the curve eliminating a stress concentration. This was a noticeable distinguishing feature of all later models of the Comet. Investigators from the RAE told a public inquiry that the sharp corners near the Comets' window openings acted as initiation sites for cracks. The skin of the aircraft was also too thin, and cracks from manufacturing stresses were present at the corners.

Alexander L. Kielland oil platform capsize



Fractures on the right side of the Alexander L. Kielland rig

The *Alexander L. Kielland* was a Norwegian semi-submersible drilling rig that capsized whilst working in the Ekofisk oil field in March 1980 killing 123 people. The capsize was the worst disaster in Norwegian waters since World War II. The rig, located approximately 320 km east from Dundee, Scotland, was owned by the Stavanger Drilling Company of Norway and was on hire to the U.S. company Phillips Petroleum at the time of the disaster. In driving rain and mist, early in the evening of 27 March 1980 more than 200 men were off duty in the accommodation on the *Alexander L. Kielland*. The wind was gusting to 40 knots with waves up to 12 m high. The rig had just been winched away from the *Edda* production platform. Minutes before 18:30 those on board felt a 'sharp crack' followed by 'some kind of trembling'. Suddenly the rig heeled over 30° and then stabilised. Five of the six anchor cables had broken, the one remaining cable preventing the rig from capsizing. The list continued to increase and at 18.53 the remaining anchor cable snapped and the rig turned upside down.

A year later in March 1981, the investigative report concluded that the rig collapsed owing to a fatigue crack in one of its six bracings (bracing D-6), which connected the collapsed D-leg to the rest of the rig. This was traced to a small 6 mm fillet weld which joined a non-load-bearing flange plate to this D-6 bracing. This flange plate held a sonar device used during drilling operations. The poor profile of the fillet weld contributed to a reduction in its fatigue strength. Further, the investigation found considerable amounts of lamellar tearing in the flange plate and cold cracks in the butt weld. Cold cracks in the welds, increased stress concentrations due to the weakened flange plate, the poor weld profile, and cyclical stresses (which would be common in the North Sea), seemed to collectively play a role in the rig's collapse.

Others

- The 1919 Boston Molasses Disaster has been attributed to a fatigue failure.
- The 1957 "Mt. Pinatubo", presidential plane of Philippine President Ramon Magsaysay, crashed due to engine failure caused by metal fatigue.
- The 1968 Los Angeles Airways Flight 417 lost one of its main rotor blades due to fatigue failure.
- The 1985 Japan Airlines Flight 123 crashed after the aircraft lost its vertical stabilizer due to faulty repairs on the rear bulkhead.
- The 1988 Aloha Airlines Flight 243 suffered an explosive decompression due to fatigue failure.
- The 1989 United Airlines Flight 232 lost its tail engine due to fatigue failure.
- The 1992 El Al Flight 1862 lost both engines on its right-wing due to fatigue failure with the #3 Engine.
- The 1998 Eschede train disaster was caused by fatigue failure of a single composite wheel.
- The 2002 China Airlines Flight 611 had disintegrated in-flight due to fatigue failure.
- The 2005 Chalk's Ocean Airways Flight 101 lost its right wing due to fatigue failure brought about by inadequate maintenance practices.

Chapter 5

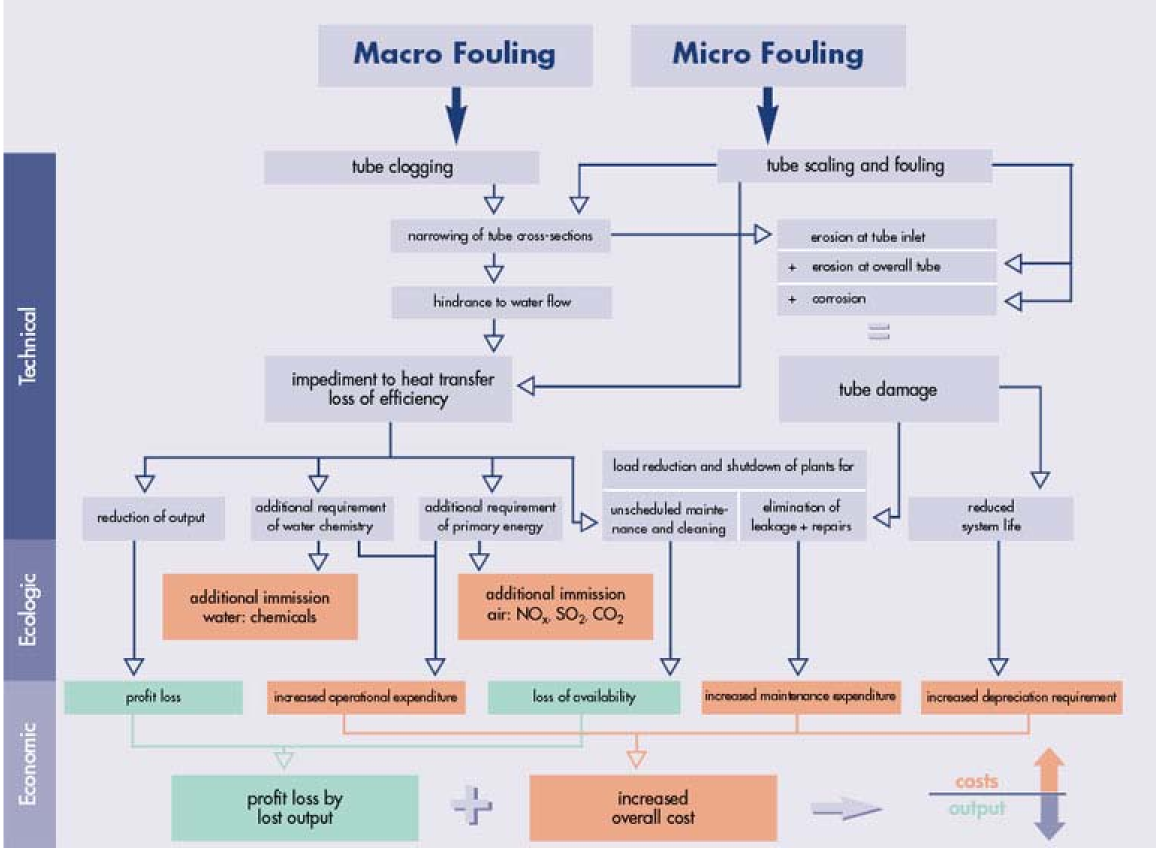
Fouling

Heat exchanger in a steam power plant, fouled by macro fouling

Condenser tube with residues of biofouling (cut open)

Condenser tube with calcium carbonate scaling (cut open)

Brass tube with corrosion traces (cut open)



Cost relations between the individual types of fouling

Fouling refers to the accumulation of unwanted material on solid surfaces, most often in an aquatic environment. The fouling material can consist of either living organisms (biofouling) or a non-living substance (inorganic or organic). Fouling is usually distinguished from other surface-growth phenomena in that it occurs on a surface of a component, system or plant performing a defined and useful function, and that the fouling process impedes or interferes with this function.

Other terms used in the literature to describe fouling include: deposit formation, encrustation, crudding, deposition, scaling, scale formation, slagging, and sludge formation. The last six terms have a more narrow meaning than fouling within the scope of the fouling science and technology, and they also have meanings outside of this scope; therefore, they should be used with caution.

Fouling phenomena are common and diverse, ranging from fouling of ship hulls, natural surfaces in the marine environment (marine fouling), fouling of heat-transfer components through ingredients contained in the cooling water or gases, and even the development of plaque or calculus on teeth, or deposits on solar panels on Mars, among other examples.

Components subject to fouling

The following are examples of components that may be subject to fouling and the corresponding effects of fouling:

- Heat exchanger surfaces - reduces thermal efficiency, decreases heat flux, increases temperature on the hot side, decreases temperature on the cold side, induces under-deposit corrosion, increases use of cooling water;
- Piping, flow channels - reduces flow, increases pressure drop, increases upstream pressure, increases energy expenditure, may cause flow oscillations, slugging in two-phase flow, cavitation; may increase flow velocity elsewhere, may induce vibrations, may cause flow blockage;
- Ship hulls - creates additional drag, increases fuel usage, reduces maximum speed;
- Turbines - reduces efficiency, increases probability of failure;
- Solar panels - decreases the electrical power generated;
- Reverse osmosis membranes - increases pressure drop, increases energy expenditure, reduces flux, membrane failure (in severe cases);
- Electrical heating elements - increases temperature of the element, increases corrosion, reduces lifespan;
- Nuclear fuel in pressurized water reactors - axial offset anomaly, may need to de-rate the power plant;
- Injection/spray nozzles (e.g., a nozzle spraying a fuel into a furnace) - incorrect amount injected, malformed jet, component inefficiency, component failure;
- Venturi tubes, orifice plates - inaccurate or incorrect measurement of flow rate;
- Pitot tubes in airplanes - inaccurate or incorrect indication of airplane speed;
- Spark plug electrodes in cars - engine misfiring;
- Production zone of petroleum reservoirs and oil wells - decreased petroleum production with time; plugging; in some cases complete stoppage of flow in a matter of days;
- Teeth - promotes tooth or gum disease, decreases aesthetics;

- Living organisms - deposition of excess minerals (e.g., calcium, iron, copper) in tissues is (sometimes controversially) linked to aging/senescence.

Macro fouling

Macro fouling is caused by coarse matter of either biological or inorganic origin, for example industrially produced refuse. Such matter enters into the cooling water circuit through the cooling water pumps from sources like the open sea, rivers or lakes. In closed circuits, like cooling towers, the ingress of macro fouling into the cooling tower basin is possible through open canals or by the wind. Sometimes, parts of the cooling tower internals detach themselves and are carried into the cooling water circuit. Such substances can foul the surfaces of heat exchangers and may cause deterioration of the relevant heat transfer coefficient. They may also create flow blockages, redistribute the flow inside the components, or cause fretting damage.

Examples

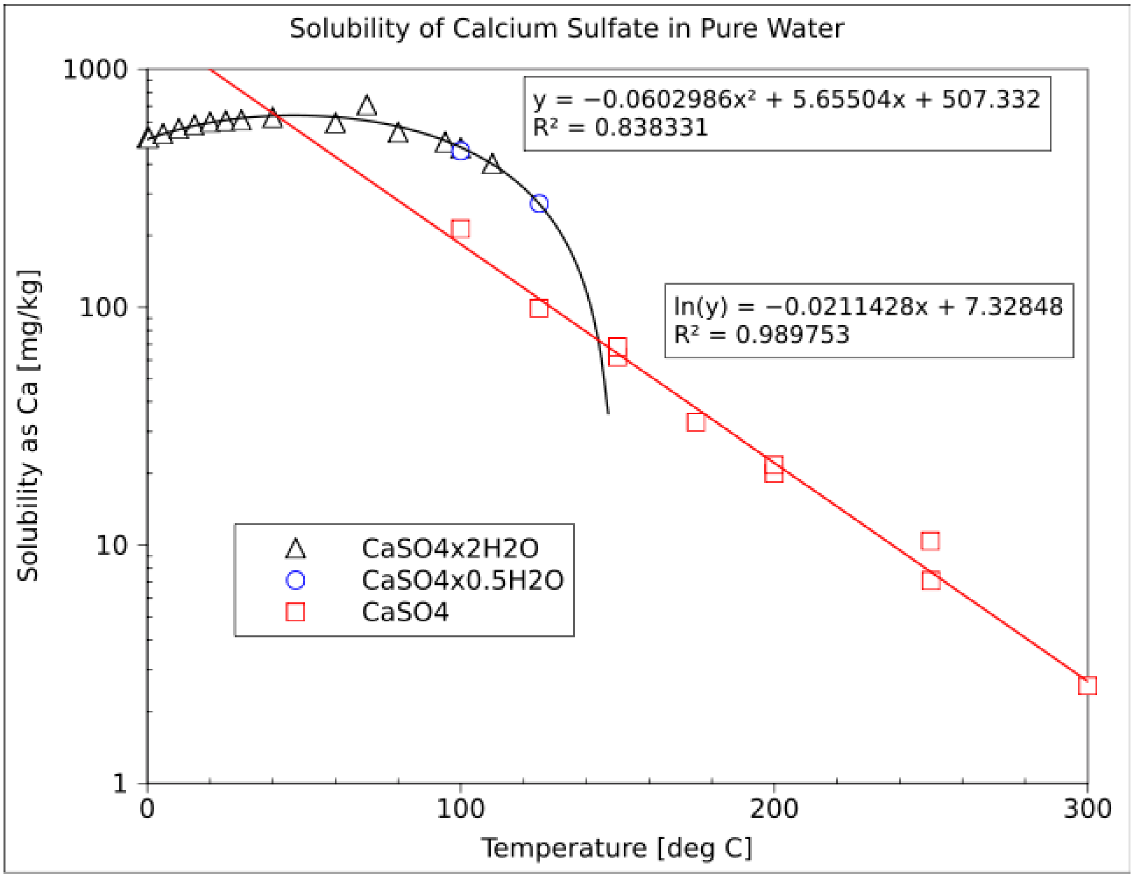
- Manmade refuse;
- Detached internal parts of components;
- Algae;
- Mussels;
- Leaves, parts of plants up to entire trunks.

Micro fouling

As to micro fouling, distinctions are made between:

- Scaling or precipitation fouling, as crystallization of solid salts, oxides and hydroxides from water solutions, for example, calcium carbonate or calcium sulfate;
- Particulate fouling, i.e., accumulation of particles, typically colloidal particles, on a surface;
- Corrosion fouling, i.e., in-situ growth of corrosion deposits, for example, magnetite on carbon steel surfaces;
- Chemical reaction fouling, for example, decomposition or polymerization of organic matter on heating surfaces;
- Solidification fouling - when components of the flowing fluid with a high-melting point freeze onto a subcooled surface;
- Biofouling, like settlements of bacteria and algae;
- Composite fouling, whereby fouling involves more than one foulant or fouling mechanism.

Precipitation fouling



Temperature dependence of the solubility of calcium sulfate (3 phases) in pure water. The water is pressurized so that it can be maintained in the liquid state at the elevated temperatures.

Scaling or precipitation fouling involves crystallization of solid salts, oxides and hydroxides from solutions. These are most often water solutions, but non-aqueous precipitation fouling is also known. Precipitation fouling is a very common problem in boilers and heat exchangers operating with hard water and often results in limescale.

Through changes in temperature, or solvent evaporation or degasification, the concentration of salts may exceed the saturation, leading to a precipitation of solids (usually crystals).

As an example, the equilibrium between the readily soluble calcium bicarbonate - always prevailing in natural water - and the poorly soluble calcium carbonate, the following chemical equation may be written:

$$\mathrm{Ca(HCO_3)_2}(aqueous) \Longrightarrow \mathrm{CaCO_3} \downarrow + \mathrm{CO_2} \uparrow + \mathrm{H_2O}$$

The calcium carbonate that forms through this reaction precipitates. Due to the temperature dependence of the reaction, and increasing volatility of CO_2 with increasing temperature, the scaling is higher at the hotter outlet of the heat exchanger than at the cooler inlet.

In general, the dependence of the salt solubility on temperature or presence of evaporation will often be the driving force for precipitation fouling. The important distinction is between salts with "normal" or "retrograde" dependence of solubility on temperature. The salts with the "normal" solubility increase their solubility with increasing temperature and thus will foul the cooling surfaces. The salts with "inverse" or "retrograde" solubility will foul the heating surfaces. An example of the temperature dependence of solubility is shown in the figure. Calcium sulfate is a common precipitation foulant of heating surfaces due to its retrograde solubility.

Precipitation fouling can also occur in the absence of heating or vaporization. For example, calcium sulfate decreases it solubility with decreasing pressure. This can lead to precipitation fouling of reservoirs and wells in oil fields, decreasing their productivity with time. Fouling of membranes in reverse osmosis systems can occur due to differential solubility of barium sulfate in solutions of different ionic strength. Similarly, precipitation fouling can occur because of solubility changes induced by other factors, e.g., liquid flashing, liquid degassing, redox potential changes, or mixing of incompatible fluid streams.

The following lists some of the industrially common phases of precipitation fouling deposits observed in practice to form from aqueous solutions:

- Calcium carbonate (calcite, aragonite usually at t > ~50 °C, or rarely vaterite);
- Calcium sulfate (anhydrite, hemihydrate, gypsum);
- Calcium oxalate (e.g., beerstone);
- Barium sulfate (barite);
- Magnesium hydroxide (brucite); magnesium oxide (periclase);
- Silicates (serpentine, acmite, gyrolite, gehlenite, amorphous silica, quartz, cristobalite, pectolite, xonotlite);
- Aluminium oxide hydroxides (boehmite, gibbsite, diaspore, corundum);
- Aluminosilicates (analcite, cancrinite, noselite);
- Copper (metallic copper, cuprite, tenorite);
- Phosphates (hydroxyapatite);
- Magnetite or nickel ferrite ($NiFe_2O_4$) from extremely pure, low-iron water.

Particulate fouling

Fouling by particles suspended in water ("crud") or in gas progresses by a mechanism different than precipitation fouling. This process is usually most important for colloidal particles, i.e., particles smaller than about 1 μm in at least one dimension (but which are much larger than atomic dimensions). Particles are transported to the surface by a number of mechanisms and there they can attach themselves, e.g., by flocculation or coagulation.

Note that the attachment of colloidal particles typically involves electrical forces and thus the particle behaviour defies the experience from the macroscopic world. The probability of attachment is sometimes referred to as "sticking probability", which for colloidal particles is a function of both the surface chemistry and the local thermohydraulic conditions. Being essentially a surface chemistry phenomenon, this fouling mechanism can be very sensitive to factors that affect colloidal stability, e.g., zeta potential. A maximum fouling rate is usually observed when the fouling particles and the substrate exhibit opposite electrical charge, or near the point of zero charge of either of them.

Particles larger than those of colloidal dimensions may also foul e.g., by sedimentation ("sedimentation fouling") or straining in small-size openings.

With time, the resulting surface deposit may harden through processes collectively known as "deposit consolidation" or, colloquially, "aging".

The common particulate fouling deposits formed from aqueous suspensions include:

- iron oxides and iron oxyhydroxides (magnetite, hematite, lepidocrocite, maghemite, goethite);
- Sedimentation fouling by silt and other relatively coarse suspended matter.

Fouling by particles from gas aerosols is also of industrial significance. The particles can be either solid or liquid. The common examples can be fouling by flue gases, or fouling of air-cooled components by dust in air. The mechanisms are discussed in article on aerosol deposition.

Corrosion fouling

Corrosion deposits are created in-situ by the corrosion of the substrate. They are distinguished from fouling deposits, which form from material originating ex-situ. Corrosion deposits should not be confused with fouling deposits formed by ex-situ generated corrosion products. Corrosion deposits will normally have composition related to the composition of the substrate. Also, the geometry of the metal-oxide and oxide-fluid interfaces may allow practical distinction between the corrosion and fouling deposits. An example of corrosion fouling can be formation of an iron oxide or oxyhydroxide deposit from corrosion of the carbon steel underneath. Corrosion fouling should not be confused with fouling corrosion, i.e., any of the types of corrosion that may be induced by fouling.

Chemical reaction fouling

Chemical reactions may occur on contact of the chemical species in the process fluid with heat transfer surfaces. In such cases, the metallic surface sometimes acts as a catalyst. For example, corrosion and polymerization occurs in cooling water for the chemical industry which has a minor content of hydrocarbons. Systems in petroleum processing are prone to polymerization of olefins or deposition of heavy fractions (asphaltenes, waxes, etc.).

High tube wall temperatures may lead to carbonizing of organic matter. Food industry, for example milk processing, also experiences fouling problems by chemical reactions.

Fouling through an ionic reaction with an evolution of an inorganic solid is commonly classified as precipitation fouling (not chemical reaction fouling).

Solidification fouling

Solidification fouling occurs when a component of the flowing fluid "freezes" onto a surface forming a solid fouling deposit. Examples may include solidification of wax (with a high melting point) from a hydrocarbon solution, or of molten ash (carried in a furnace exhaust gas) onto a heat exchanger surface. The surface needs to have a temperature below a certain threshold; therefore, it is said to be subcooled in respect to the solidification point of the foulant.

Biofouling

Biofouling or biological fouling is the undesirable accumulation of micro-organisms, algae and diatoms, plants, and animals on surfaces, for example ships' hulls, or piping and reservoirs with untreated water. This can be accompanied by microbiologically influenced corrosion (MIC).

Bacteria can form biofilms or slimes. Thus the organisms can aggregate on surfaces using colloidal hydrogels of water and extracellular polymeric substances (EPS) (polysaccharides, lipids, nucleic acids, etc.). The biofilm structure is usually complex.

Bacterial fouling can occur under either aerobic (with oxygen dissolved in water) or anaerobic (no oxygen) conditions. In practice, aerobic bacteria prefer open systems, when both oxygen and nutrients are constantly delivered, often in warm and sunlit environments. Anaerobic fouling more often occurs in closed systems when sufficient nutrients are present. Examples may include sulfate-reducing bacteria (or sulfur-reducing bacteria), which produce sulfide and often cause corrosion of ferrous metals (and other alloys). Sulfide-oxidizing bacteria (e.g., Acidithiobacillus), on the other hand, can produce sulfuric acid, and can be involved in corrosion of concrete.

Zebra mussel can serve as an example of a larger animal that caused widespread fouling problem in North America.

Composite fouling

Composite fouling is common. This type of fouling involves more than one foulant or more than one fouling mechanism working simultaneously. The multiple foulants or mechanisms may interact with each other resulting in a synergistic fouling which is not a simple arithmetic sum of the individual components.

Fouling on Mars

NASA Mars Exploration Rovers (Spirit and Opportunity) experienced (presumably) abiotic fouling of solar panels by dust particles from the Martian atmosphere. Some of the deposits subsequently spontaneously cleaned off. This illustrates the universal nature of the fouling phenomena.

Quantification of fouling

The most straight-forward way to quantify fairly uniform fouling is by stating the average deposit surface loading, i.e., kg of deposit per m^2 of surface area. The fouling rate will then be expressed in kg/m^2s, and it is obtained by dividing the deposit surface loading by the effective operating time. The normalized fouling rate (also in kg/m^2s) will additionally account for the concentration of the foulant in the process fluid (kg/kg) during preceding operations, and is useful for comparison of fouling rates between different systems. It is obtained by dividing the fouling rate by the foulant concentration. The fouling rate constant (m/s) can be obtained by dividing the normalized fouling rate by the mass density of the process fluid (kg/m^3).

Deposit thickness (μm) and porosity (%) are also often used for description of fouling amount. The relative reduction of diameter of piping or increase of the surface roughness can be of particular interest when the impact of fouling on pressure drop is of interest.

In heat transfer equipment, where the primary concern is often the effect of fouling on heat transfer, fouling can be quantified by the increase of the resistance to the flow of heat (m^2K/W) due to fouling (termed "fouling resistance"), or by development of heat transfer coefficient (W/m^2K) with time.

If under-deposit or crevice corrosion is of primary concern, it is important to note packing of confined regions with deposits or creation of occluded "crevices". The non-uniformity of deposit thickness (e.g., deposit waviness) can also be important if underdeposit corrosion of material (e.g., intergranular attack, pitting, stress corrosion cracking) is of concern. Porosity and permeability of the deposits will likely influence the probability of underdeposit corrosion. Deposit composition can also be important-even minor components of the deposits can sometimes cause severe corrosion of the underlying metal (e.g., vanadium in deposits of fired boilers causing hot corrosion).

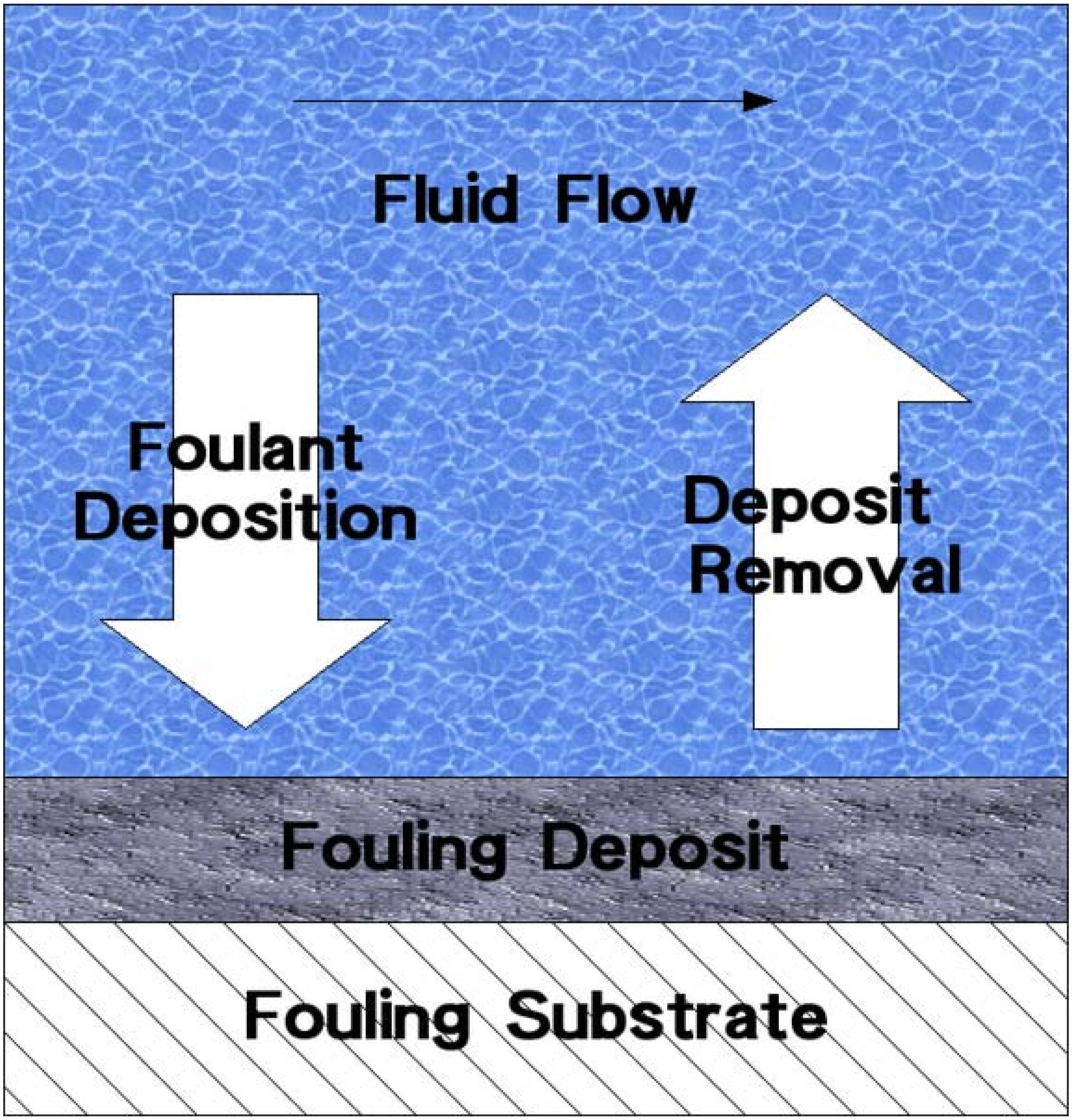
There is no general rule on how much deposit can be tolerated, it depends on the system. In many cases, a deposit even a few micrometers thick can be troublesome. A deposit in a millimeter-range thickness will be of concern in almost any application.

Progress of fouling with time

Deposit on a surface does not always develop steadily with time. The following fouling scenarios can be distinguished, depending on the nature of the system and the local thermohydraulic conditions at the surface:

- **Induction period.** Sometimes, a near-nil fouling rate is observed when the surface is new or very clean. This is often observed in biofouling and precipitation fouling. After the "induction period", the fouling rate increases.
- **"Negative" fouling.** This can occur when fouling rate is quantified by monitoring heat transfer. Relatively small amounts of deposit can improve heat transfer, relative to clean surface, and give an appearance of "negative" fouling rate and negative total fouling amount. Negative fouling is often observed under nucleate-boiling heat-transfer conditions (deposit improves bubble nucleation) or forced-convection (if the deposit increases the surface roughness and the surface is no longer "hydraulically smooth"). After the initial period of "surface roughness control", the fouling rate usually becomes strongly positive.
- **Linear fouling.** The fouling rate can be steady with time. This is a common case.
- **Falling fouling.** Under this scenario, the fouling rate decreases with time, but never drops to zero. The deposit thickness does not achieves a constant value. The progress of fouling can be often described by two numbers: the initial fouling rate (a tangent to the fouling curve at zero deposit loading or zero time) and the fouling rate after a long period of time (an oblique asymptote to the fouling curve).
- **Asymptotic fouling.** Here, the fouling rate decreases with time, until it finally reaches zero. At this point, the deposit thickness remains constant with time (a horizontal asymptote). This is often the case for relatively soft or poorly adherent deposits in areas of fast flow. The asymptote is usually interpreted as the deposit loading at which the deposition rate equals the deposit removal rate.
- **Accelerating fouling.** Under this scenario, the fouling rate increases with time; the rate of deposit buildup accelerates with time (perhaps until it becomes transport limited). Mechanistically, this scenario can develop when fouling increases the surface roughness, or when the deposit surface exhibits higher chemical propensity to fouling than the pure underlying metal.
- **Seesaw fouling**. Here, fouling loading generally increases with time (often assuming a generally linear or falling rate), but, when looked at in more detail, the fouling progress is periodically interrupted and takes the form of sawtooth curve. The periodic sharp variations in the apparent fouling amount often correspond to the moments of system shutdowns, startups or other transients in operation. The periodic variations are often interpreted as periodic removal of some of the deposit (perhaps deposit re-suspension due to pressure pulses, spalling due thermal stresses, or exfoliation due to redox transients). Steam blanketing has been postulated to occur between the partially spalled deposits and the heat transfer surface. However, other reasons are possible, e.g., trapping of air inside the surface deposits during shutdowns, or inaccuracy of temperature measurements during transients ("temperature streaming").

Fouling modelling



Schematics of the fouling process consisting of simultaneous foulant deposition and deposit removal.

Fouling of a system can be modelled as consisting of several steps:

- Generation or ingress of the species that causes fouling ("foulant sourcing");
- Foulant transport with the stream of the process fluid (most often by advection);
- Foulant transport from the bulk of the process fluid to the fouling surface (often by molecular or turbulent-eddy diffusion, but may also occur by inertial coasting (impaction), electrophoresis, thermophoresis, diffusiphoresis, sedimentation, evaporation, vapour diffusion with condensation, and other mechanisms);

- Induction period, i.e., a near-nil fouling rate at the initial period of fouling (observed only for some fouling mechanisms);
- Foulant crystallization on the surface (or attachment of the colloidal particle, or chemical reaction, or bacterial growth);
- Deposit dissolution (or re-entrainment of loosely attached particles);
- Deposit consolidation on the surface (e.g., through Oswald ripening or differential solubility in temperature gradient) or cementation, which account for deposit losing its porosity and becoming more tenacious with time;
- Deposit spalling, erosion wear, or exfoiliation.

Deposition consists of transport to the surface and subsequent attachment. Deposit removal is either through deposit dissolution, particle re-entrainment, or deposit spalling, erosive wear, or exfoliation. Fouling results from foulant generation, foulant deposition, deposit removal, and deposit consolidation.

For the modern model of fouling involving deposition with simultaneous deposit re-entrainment and consolidation, the fouling process can be represented by the following scheme:

$$\begin{bmatrix} \text{rate of} \\ \text{deposit} \\ \text{accumulation} \end{bmatrix} = \begin{bmatrix} \text{rate of} \\ \text{deposition} \end{bmatrix} - \begin{bmatrix} \text{rate of} \\ \text{re-entrainment of} \\ \text{unconsolidated deposit} \end{bmatrix}$$

$$\begin{bmatrix} \text{rate of} \\ \text{accumulation of} \\ \text{unconsolidated deposit} \end{bmatrix} = \begin{bmatrix} \text{rate of} \\ \text{deposition} \end{bmatrix} - \begin{bmatrix} \text{rate of} \\ \text{re-entrainment of} \\ \text{unconsolidated deposit} \end{bmatrix} - \begin{bmatrix} \text{rate of} \\ \text{consolidation of} \\ \text{unconsolidated deposit} \end{bmatrix}$$

Following the above scheme, the basic fouling equations can be written as follows (for steady-state conditions with flow, when concentration remains constant with time):

$$\begin{cases} dm/dt = kC\rho - \lambda_r m_r(t) \\ dm_r/dt = kC\rho - \lambda_r m_r(t) - \lambda_c \cdot m_r(t) \end{cases}$$

where: m is the mass loading of the deposit (consolidated and unconsolidated) on the surface (kg/m^2); t is time (s); k is the deposition rate constant (m/s); ρ is the fluid density (kg/m^3); λ_r is the re-entrainment rate constant (1/s); m_r is the mass loading of the removable (i.e., unconsolidated) fraction of the surface deposit (kg/m^2); and λ_c is the consolidation rate constant (1/s).

This system of equations can be integrated (taking that $m = 0$ and $m_r = 0$ at $t = 0$) to the form:

$$m(t) = \frac{kC\rho}{\lambda}\left(t\lambda_c + \frac{\lambda_r}{\lambda}\left(1 - e^{-\lambda t}\right)\right)$$

where $\lambda = \lambda_r + \lambda_c$.

This model reproduces either linear, falling, or asymptotic fouling, depending on the relative values of k, λ_r, and λ_c. The underlying physical picture for this model is that of a two-layer deposit consisting of consolidated inner layer and loose unconsolidated outer layer. Such a bi-layer deposit is often observed in practice. The above model simplifies readily to the older model of simultaneous deposition and re-entrainment (which neglects consolidation) when $\lambda_c=0$.

The economic and environmental importance of fouling

Fouling is ubiquitous and generates tremendous operational losses, not unlike corrosion. For example, one estimate puts the losses due to fouling of heat exchangers in industrialized nations to be about 0.25% of their GDP. Another analysis estimated (for 2006) the economical loss due to boiler and turbine fouling in China utilities at 4.68 billion dollars, which is about 0.169% the country GDP .

The losses initially result from impaired heat transfer, corrosion damage (in particular under-deposit and crevice corrosion), increased pressure drop, flow blockages, flow redistribution inside components, flow instabilities, induced vibrations (possibly leading to other problems, e.g., fatigue), fretting, premature failure of electrical heating elements, and a large number of other often unanticipated problems. In addition, the ecological costs should be (but typically are not) considered. The ecological costs arise from the use of biocides for the avoidance of biofouling, from the increased fuel input to compensate for the reduced output caused by fouling, and an increased use of cooling water in once-through cooling systems.

For example, "normal" fouling at a conventionally fired 500 MW (net electrical power) power station unit accounts for output losses of the steam turbine of 5 MW and more. In a 1,300 MW nuclear power station, typical losses could be 20 MW and up (up to 100% if the station shuts down due to fouling-induced component degradation). In seawater desalination plants, fouling may reduce the gained output ratio by two-digit percentages (the gained output ratio is an equivalent that puts the mass of generated distillate in relation to the steam used in the process). The extra electrical consumption in compressor-operated coolers is also easily in the two-digit area. In addition to the operational costs, also the capital cost increases because the heat exchangers have to be designed in larger sizes to compensate for the heat-transfer loss due to fouling. To the output losses listed above, one needs to add the cost of down-time required to inspect, clean, and repair the components (millions of dollars per day of shutdown in lost revenue in a typical power plant), and the cost of actually doing this maintenance. Finally, fouling is often a root cause of serious degradation problems that may limit the life of components or entire plants.

Fouling control

The most fundamental and usually preferred method of controlling fouling is to prevent the ingress of the fouling species into the cooling water circuit. In steam power stations and other major industrial installations of water technology, macro fouling is avoided by way of pre-filtration and cooling water debris filters. Some plants employ foreign-object exclusion program (to eliminate the possibility of salient introduction of unwanted materials, e.g., forgetting tools during maintenance). Acoustic monitoring is sometimes employed to monitor for fretting by detached parts. In the case of micro fouling, water purification is achieved with extensive methods of water treatment, microfiltration, membrane technology (reverse osmosis, electrodeionization) or ion-exchange resins. The generation of the corrosion products in the water piping systems is often minimized by controlling the pH of the process fluid (typically alkanization with ammonia, morpholine, ethanolamine or sodium phosphate), control of oxygen dissolved in water (for example, by addition of hydrazine), or addition of corrosion inhibitors.

For water systems at relatively low temperatures, the applied biocides may be classified as follows: inorganic chlorine and bromide compounds, chlorine and bromide cleavers, ozone and oxygen cleavers, unoxidizable biocides. One of the most important unoxidizable biocides is a mixture of chloromethyl-isothiazolinone and methyl-isothiazolinone. Also applied are dibrom nitrilopropionamide and quaternary ammonium compounds. For underwater ship hulls bottom paints are applied.

Chemical fouling inhibitors can reduce fouling in many systems, mainly by interfering with the crystallization, attachment, or consolidation steps of the fouling process. Examples for water systems are: chelating agents (for example, EDTA), long-chain aliphatic amines or polyamines (for example, octadecylamine, helamin, and other "film-forming" amines), organic phosphonic acids (for example, 1-hydroxyethylidene-1,1-diphosphonic acid, known as HEDP), or polyelectrolytes (for example, polyacrylic acid, polymethacrylic acid, usually with a molecular weight lower than 10000). For fired boilers, aluminum or magnesium additives can lower the melting point of ash and promote creation of deposits which are easier to remove.

Magnetic water treatment has been a subject of controversy as to its effectiveness for fouling control since 1950s. The prevailing opinion is that it simply "does not work". Nevertheless, some studies suggest that it may be effective under some conditions to reduce buildup of calcium carbonate deposits.

On the component design level, fouling can often (but not always) be minimized by maintaining a relatively high (for example, 2 m/s) and uniform fluid velocity throughout the component. Stagnant regions need to be eliminated. Components are normally overdesigned to accommodate the fouling anticipated between cleanings. However, a significant overdesign can be a design error because it may lead to increased fouling due to reduced velocities. Periodic on-line pressure pulses or backflow can be effective if the capability is carefully incorporated at the design time. Blowdown capability is always incorporated into steam generators or evaporators to control the accumulation of nonvolatile impurities that cause or aggravate fouling. Low-fouling surfaces (for example, very smooth, implanted with ions, or of low surface energy like Teflon) are an option for some applications. Modern components are typically required to be designed for ease of inspection of internals and periodic cleaning. On-line fouling monitoring systems are designed for some application so that blowing or cleaning can be applied before unpredictable shutdown is necessary or damage occurs.

Chemical or mechanical cleaning processes for the removal of deposits and scales are recommended when fouling reaches the point of impacting the system performance or an onset of significant fouling-induced degradation (e.g., by corrosion). These processes comprise pickling with acids and complexing agents, cleaning with high-velocity water jets ("water lancing"), recirculating ("blasting") with metal, sponge or other balls, or propelling offline mechanical "bullet-type" tube cleaners. Whereas chemical cleaning causes environmental problems through the handling, application, storage and disposal of chemicals, the mechanical cleaning by means of circulating cleaning balls or offline "bullet-type" cleaning can be a more environmentally-friendly alternative. In some heat-transfer applications, mechanical mitigation with dynamic scraped surface heat exchangers is an option. Also ultrasonic or abrasive cleaning methods are available for many specific applications.

Chapter 6

Thermal shock

Thermal shock is the name given to cracking as a result of rapid temperature change. Glass and ceramic objects are particularly vulnerable to this form of failure, due to their low toughness, low thermal conductivity, and high thermal expansion coefficients. However, they are used in many high temperature applications due to their high melting point.

Thermal shock occurs when a thermal gradient causes different parts of an object to expand by different amounts. This differential expansion can be understood in terms of stress or of strain, equivalently. At some point, this stress overcomes the strength of the material, causing a crack to form. If nothing stops this crack from propagating through the material, it will cause the object's structure to fail.

Thermal shock can be prevented by:

1. Reducing the thermal gradient seen by the object, by
    1. changing its temperature more slowly
    2. increasing the material's thermal conductivity
2. Reducing the material's coefficient of thermal expansion
3. Increasing its strength
4. Decreasing its Young's modulus
5. Increasing its toughness, by
    1. crack tip blunting, i.e., plasticity or phase transformation
    2. crack deflection

Effect on materials

Borosilicate glass such as Pyrex is made to withstand thermal shock better than most other glass through a combination of reduced expansion coefficient and greater strength, though fused quartz outperforms it in both these respects. Some glass-ceramic materials (mostly in LAS system) include a controlled proportion of material with a negative expansion coefficient, so that the overall coefficient can be reduced to almost exactly zero over a reasonably wide range of temperatures.

Reinforced carbon-carbon is extremely resistant to thermal shock, due to graphite's extremely high thermal conductivity and low expansion coefficient, the high strength of carbon fiber, and a reasonable ability to deflect cracks within the structure.

To measure thermal shock the impulse excitation technique proved to be a useful tool. It can be used to measure Young's modulus, Shear modulus, Poisson's ratio and damping coefficient in a non destructive way. The same test-piece can be measured after different thermal shock cycles and this way the detoriation in physical properties can be mapped out.

Relative robustness of materials

The robustness of a material to thermal shock is characterized with the **thermal shock parameter**:

$$R_{\mathrm{T}} = \frac{k\sigma_{\mathrm{T}}(1-\nu)}{\alpha E},$$

where

- k is thermal conductivity,
- σ_{T} is maximal tension the material can resist,
- α is the thermal expansion coefficient
- E is the Young's modulus, and
- ν is the Poisson ratio.

Thermal shock parameter in the physics of solid-state lasers

The laser gain medium generates heat. This heat is drained through the heat sink. The transfer of heat occurs at certain temperature gradient. The non-uniform thermal expansion of a bulk material causes the stress and tension, which may break the device even at slow change of the temperature. (for example, continuous-wave operation). This phenomenon is also called **thermal shock**. The robustness of a laser material to the thermal shock is characterized with the **thermal shock parameter**

Roughly, at the efficient operation of laser, the power P_{h} of heat generated in the gain medium is proportional to the output power P_{s} of the laser, and the coefficient q of proportionality can be interpreted as heat generation parameter; then, $P_{\mathrm{h}} = qP_{\mathrm{s}}$. The heat generation parameter is basically determined by the quantum defect of the laser action, and one can estimate $q = 1 - \omega_{\mathrm{s}} / \omega_{\mathrm{p}}$, where ω_{p} and ω_{s} are frequency of the pump and that of the lasing.

Then, for the layer of the gain medium placed at the heat sink, the maximal power can be estimated as

$$P_{\mathrm{s,max}} = 3\frac{R_{\mathrm{T}}}{q}\frac{L^2}{h},$$

where *h* is thickness of the layer and *L* is the transversal size. This estimate assumes the unilateral heat drain, as it takes place in the active mirrors. For the double-side sink, the coefficient 4 should be applied.

Thermal loading

The estimate above is not the only parameter which determines the limit of overheating of a gain medium. The maximal raise ΔT of temperature, at which the medium still can efficiently lase, is also important property of the laser material. This overheating limits the maximal power with estimate

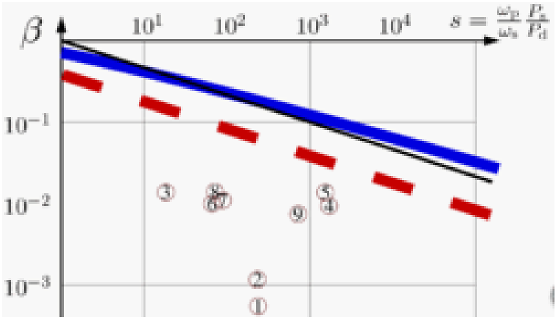
$$P_{\mathrm{s,max}} = 2\frac{k\Delta T}{q}\frac{L^2}{h}$$

Combination of the two estimates above of the maximal power gives the estimate

$$P_{\mathrm{s,max}} = R\frac{L^2}{h}$$

where

$$R = \min \left\{ \begin{array}{c} 3R_{\mathrm{T}}/q \\ 2k\Delta T/q \end{array} \right.$$



Estimates of maximal value of loss β, at which desirable output power *P* is still available in a single disk laser, versus normalized power $s = \frac{\omega_{\mathrm{p}}}{\omega_{\mathrm{s}}}\frac{PQ}{R^2}$, and experimental data (circles)

is **thermal loading**; parameter, which is important property of the laser material. The thermal loading, saturation intensity *Q* and the loss β determine the limit of power scaling of the disk lasers . Roughly, the maximal power at the optimised sizes *L* and *h*, is of order of $P = \frac{R^2}{Q\beta^3}$. This estimate is very sensitive to the loss β. However, the same expression can be interpreted as a robust estimate of the upper bound of the loss β required for the desirable output power *P*:

$$\beta_{\max} = \left(\frac{R^2}{PQ}\right)^{\frac{1}{3}}.$$

All the disk lasers reported work at the round-trip loss below this estimate . The thermal shock parameter and the loading depend of the temperature of the heat sink. Certain hopes are related with a laser, operating at cryogenic temperatures. The corresponding Increase of the **thermal shock parameter** would allow to softer requirements for the round-trip loss of the disk laser at the power scaling.

Examples of thermal shock failure

- Hard rocks containing ore veins such as quartzite were formerly broken down using fire-setting, which involved heating the rock face with a wood fire, then quenching with water to induce crack growth. It is described by Diodorus Siculus in Egyptian gold mines, Pliny the Elder and Georg Agricola.
- Ice cubes placed in a glass of warm water crack by thermal shock as the exterior surface increases in temperature much faster than the interior. As ice has a larger volume than the water that created it, the outer layer shrinks as it warms and begins to melt, whilst the interior remains largely unchanged. This rapid change in volume between different layers creates stresses in the ice that build until the force exceeds the strength of the ice, and a crack forms.
- Incandescent bulbs that have been running for a while have a very hot surface. Splashing cold water on them can cause the glass to shatter due to thermal shock, and the bulb to implode.
- An antique cast iron cookstove is basically a iron box on legs, that has a cast iron top. One builds a wood or coal fire inside the box and cooks on the top outer surface of the box, like a griddle. If one builds too hot a fire, and then tries to cool the stove by pouring water on the top surface, it will crack and perhaps fail by thermal shock.
- The causes of three aircraft incidents in the 1990s (United Airlines Flight 585, USAir Flight 427 and Eastwind Airlines Flight 517). Thermal shock caused their power control unit in the tail to jam and cause rudder hardover, forcing the planes in the direction the rudder turns.
- It is widely hypothesized that following the casting of the Liberty Bell, it was allowed to cool too quickly which weakened the integrity of the bell and resulted in a large crack along the side of it the first time it was rung. Similarly, the strong gradient of temperature (due to the fire) is believed to cause the crash of the Tsar Bell.

Chapter 7

Wear

In materials science, **wear** is erosion or sideways displacement of material from it´s "derivative" and original position on a solid surface performed by the action of another surface.

Wear is related to interactions between surfaces and more specifically the removal and deformation of material on a surface as a result of mechanical action of the opposite surface. The need for relative motion between two surfaces and initial mechanical contact between asperities is an important distinction between mechanical wear compared to other processes with similar outcomes.

The definition of wear may include loss of dimension from plastic deformation if it´s originated at the interface between two sliding surfaces.
However, plastic deformation such as yield stress is excluded from the wear definition if it doesn't incorporates a relative sliding motion and contact against another surface despite the possibility for material removal, because it then lacks the relative sliding action of another surface.
Impact wear is in reality a short sliding motion where two solid bodies interact at an exceptional short time interval. Previously due to the fast execution, the contact found in impact wear was refereed to as an impulse contact by the nomenclature. Impulse can be described as an mathematical model of an synthesised average on the energy transport between two travelling solids in opposite converging contact.
Cavitation wear is a form of wear where the erosive medium or counter-body is a fluid.
Corrosion may be included in wear phenomenons but the damage is amplified and performed by chemical reactions rather than mechanical action.

Wear can also be defined as a process where interaction between two surfaces or bounding faces of solids within the working environment results in dimensional loss of one solid, with or without any actual decoupling and loss of material. Aspects of the working environment which affect wear include loads and features such as unidirectional sliding, reciprocating, rolling, and impact loads, speed, temperature, but also different types of counter-bodies such as solid, liquid or gas and type of contact ranging between single phase or multiphase, in which the last multiphase may combine liquid whit solid particles and gas bubbles.

Measurement

Today there exist a couple of standard test methods for different types of wear to determine the amount of material removal during a specified time period under well-defined conditions.
The ASTM International Committee G-2 attempts to standardise wear testing for specific applications, which are periodically updated. The Society for Tribology and Lubrication Engineers (STLE) has documented a large number of frictional wear and lubrication tests. But all test methods have inbuilt limitations and do not give a true picture in every aspect.
This can be attributed to the complex nature of wear, in particular "industrial wear", and the difficulties associated with accurately simulating wear processes.

A standard result review for wear tests, defined by the ASTM International and respective subcommittees such as Committee G-2, should be expressed as loss of material during wear in terms of volume. The volume loss gives a truer picture than weight loss, particularly when comparing the wear resistance properties of materials with large differences in density.
For example, a weight loss of 14 g in a sample of tungsten carbide + cobalt (density = 14000 kg/m^3) and a weight loss of 2.7 g in a similar sample of aluminium alloy (density = 2700 kg/m^3) both result in the same level of wear (1 cm^3) when expressed as a volume loss. The inverse of volume loss can be used as an comparable index of wear resistance. Standard wear tests are only used for comparative material ranking of a specific test parameter as stipulated in the test method. For more realistic values of material deterioration in industrial applications it's necessary to conduct wear testing under conditions simulating the exact wear process.

The working life of an engineering component is expired when dimensional losses exceed the specified tolerance limits. Wear, along with other ageing processes such as fatigue and creep in association with stress concentration factors such as fracture toughness causes materials to progressively degrade, eventually leading to material failure at an advanced age.
Wear in industrial applications is one of a limited number of fault factors in which an object loses it's usefulness and the economic implication can be of enormous value to the industry.

Stages of wear

Under normal mechanical and practical procedures, the wear-rate normally changes through three different stages:

- Primary stage or early run-in period, where surfaces adapt to each other and the wear-rate might vary between high and low.

- Secondary stage or mid-age process, where a steady rate of ageing is in motion. Most of the components operational life is comprised in this stage.
- Tertiary stage or old-age period, where the components are subjected to rapid failure due to a high rate of ageing.

The secondary stage is shortened with increasing severity of environmental conditions such as higher temperatures, strain rates, stress and sliding velocities etc.
In explicit wear tests simulating industrial conditions between metalic surfaces, there are no clear chronological distinction between different wear-stages due to big overlaps and symbiotic relations between various friction mechanisms. Surface engineering and treatments are used to minimize wear and extend the components working life.

Types

The study of the processes of wear is part of the discipline of tribology. The complex nature of wear has delayed its investigations and resulted in isolated studies towards specific wear mechanisms or processes. Some commonly referred to wear mechanisms (or processes) include:

1. **Adhesive wear**
2. **Abrasive wear**
3. **Surface fatigue**
4. **Fretting wear**
5. **Erosive wear**

A number of different wear phenomena are also commonly encountered and presented in the literature. Impact-, cavitation-, diffusive- and corrosive- wear are all such examples. These wear mechanisms, however, do not necessarily act independently and wear mechanisms are not mutually exclusive. "Industrial Wear" are commonly described as incidence of multiple wear mechanisms occurring in unison. Another way to describe "Industrial Wear" is to define clear distinctions in how different friction mechanisms operate, for example distinguish between mechanisms with high or low energy density. Wear mechanisms and/or sub-mechanisms frequently overlap and occur in a synergistic manner, producing a greater rate of wear than the sum of the individual wear mechanisms.

Adhesive wear

Generally there are two types of adhesive friction.

1. Cohesive adhesive forces, holds two surfaces together even though they are separated by a measurable distance, (i.e., atom/atom or molecules interaction).
2. Adhesive wear, material transfer from one surface to another caused by relative motion, "direct contact" and plastic deformation.

The above description and distinction between "**cohesive adhesive forces**" and its counterpart, such as "**adhesive wear**" are quite common.
Usually "cohesive surface forces" and adhesive potential energy between surfaces are examined as a special field in physic departments.
The "adhesive wear" and material transfer due to direct contact and plastic deformation are examined in engineering science and in industrial research.

However, material transfer is always present when two surfaces are aligned against each other and the cause for material transfer or wear-categorization have been a source for discussion amongst researchers with frequent misinterpretations or misunderstandings due to overlaps and symbiotic relations between "wear" and cohesive attraction, as mentioned. Consequently, the definitions and nomenclature must evolve with the latest science and the objective is to explain the main issue, what causes material transfer and adhesive wear?

Adhesive wear occurs when two bodies slide over or are pressed into, each other which promote material transfer from one surface to another. This can be described as plastic deformation of very small fragments within the surface layers. The asperities or microscopic high points found on each surface, deform each other and develop a plastic zone in the interface between the two surfaces.

The surface roughness gives the initial depth of penetration and break up or remove parts of the oxide layer which eventually cause damage to the underlying bulk material which enhance the possibility for a strong adhesion.. In initial asperity/asperity contact, fragments of one surface are pulled off and adhere to the other, due to the strong adhesive forces between atoms.. Physical-chemical adhesive attraction between the surfaces plays a role in the initial build-up process but the energy absorbed in plastic deformation and relative movement is arguably the main cause for excessive amounts of adhesive material transfer or in other words, the main cause for "adhesive wear".

The outcome can be a growing roughening and creation of protrusions (i.e., lumps) above the original surface which in industrial manufacturing are referred to as galling. If the lump of transferred material grows to a certain height it will penetrate deep down in to the bulk material and create a plastically flowing deformed volume around it.
The geometry and the nominal sliding velocity of the lump defines how the flowing material will be transported and accelerated around the lump and is critical for defining contact pressure or developed temperature during sliding. The mathematical function for acceleration of flowing material is thereby defined by the lumps surface contour.

Given these prerequisites it's clear that contact pressure and developed temperature is highly dependent on the lumps geometry and flowing material exhibits an increase in energy density. Because the initial displacement and acceleration of material demand high pressure. Low pressure in plastic flow is only possible after deceleration of the flowing material and this exposure quickly cools it down. In other words, you can't deform a solid material using direct contact without applying a high pressure and somewhere along the process must acceleration and deceleration take place, i.e., high pressure must be applied on all sides of the deformed material. Flowing material will immediately exhibit energy loss and reduced ability to flow if ejected from high pressure into low pressure, this ability withholds the high pressure in the contact zone and lowers the force needed for further advancement when the sliding continues.

In engineering science, some aspects of adhesive wear is commonly referred to as welding wear due to the exhibited surface characteristics and the tribology process is usually referred to as galling which is a common fault factor in industrial applications such as sheet metal forming (SMF) and is commonly encountered in conjunction with lubricant failures.
The type, mechanism and amplitude of surface attraction varies between different materials but are amplified by an increase in surface energy density. Most solids will adhere on contact to some extent. However, oxidation films, lubricants and contaminants naturally occurring generally suppress adhesion. Spontaneous exothermic chemical reactions between surfaces generally produce a substance with low energy status in the absorbed species.

Abrasive wear

Abrasive wear occurs when a hard rough surface slides across a softer surface. ASTM (American Society for Testing and Materials) defines it as the loss of material due to hard particles or hard protuberances that are forced against and move along a solid surface.

Abrasive wear is commonly classified according to the type of contact and the contact environment. The type of contact determines the mode of abrasive wear. The two modes of abrasive wear are known as two-body and three-body abrasive wear. Two-body wear occurs when the grits, or hard particles, are rigidly mounted or adhere to a surface, when they remove the material from the surface. The common analogy is that of material being removed with sand paper. Three-body wear occurs when the particles are not constrained, and are free to roll and slide down a surface. The contact environment determines whether the wear is classified as open or closed. An open contact environment occurs when the surfaces are sufficiently displaced to be independent of one another

There are a number of factors which influence abrasive wear and hence the manner of material removal. Several different mechanisms have been proposed to describe the manner in which the material is removed. Three commonly identified mechanisms of abrasive wear are:

1. Plowing
2. Cutting
3. Fragmentation

Plowing occurs when material is displaced to the side, away from the wear particles, resulting in the formation of grooves that do not involve direct material removal. The displaced material forms ridges adjacent to grooves, which may be removed by subsequent passage of abrasive particles. Cutting occurs when material is separated from the surface in the form of primary debris, or microchips, with little or no material displaced to the sides of the grooves. This mechanism closely resembles conventional machining. Fragmentation occurs when material is separated from a surface by a cutting process and the indenting abrasive causes localized fracture of the wear material. These cracks then freely propagate locally around the wear groove, resulting in additional material removal by spalling.

Abrasive wear can be measured as loss of mass by the Taber Abrasion Test according to ISO 9352 or ASTM D 1044.

Surface fatigue

Surface fatigue is a process by which the surface of a material is weakened by cyclic loading, which is one type of general material fatigue. Fatigue wear is produced when the wear particles are detached by cyclic crack growth of microcracks on the surface. These microcracks are either superficial cracks or subsurface cracks.

Fretting wear

Fretting wear is the repeated cyclical rubbing between two surfaces, which is known as fretting, over a period of time which will remove material from one or both surfaces in contact. It occurs typically in bearings, although most bearings have their surfaces hardened to resist the problem. Another problem occurs when cracks in either surface are created, known as fretting fatigue. It is the more serious of the two phenomena because it can lead to catastrophic failure of the bearing. An associated problem occurs when the small particles removed by wear are oxidised in air. The oxides are usually harder than the underlying metal, so wear accelerates as the harder particles abrade the metal surfaces further. Fretting corrosion acts in the same way, especially when water is present. Unprotected bearings on large structures like bridges can suffer serious degradation in behaviour, especially when salt is used during winter to deice the highways carried by the bridges. The problem of fretting corrosion was involved in the Silver Bridge tragedy and the Mianus River Bridge accident.

Erosive wear

Erosive wear can be described as an extremely short sliding motion and is executed within a short time interval. Erosive wear is caused by the impact of particles of solid or liquid against the surface of an object. The impacting particles gradually remove material from the surface through repeated deformations and cutting actions. It is a widely encountered mechanism in industry. A common example is the erosive wear associated with the movement of slurries through piping and pumping equipment.

The rate of erosive wear is dependent upon a number of factors. The material characteristics of the particles, such as their shape, hardness, impact velocity and impingement angle are primary factors along with the properties of the surface being eroded. The impingement angle is one of the most important factors and is widely recognized in literature. For ductile materials the maximum wear rate is found when the impingement angle is approximately 30°, whilst for non ductile materials the maximum wear rate occurs when the impingement angle is normal to the surface.

Chapter 8

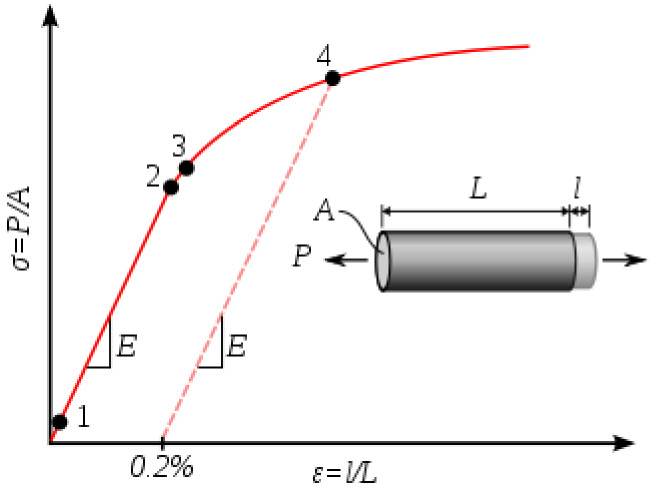
Yield (Engineering)

The **yield strength** or **yield point** of a material is defined in engineering and materials science as the stress at which a material begins to deform plastically. 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.

In the three-dimensional space of the principal stresses (σ_1,σ_2,σ_3), an infinite number of yield points form together a yield surface.

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, this is a soft failure mode which does not normally cause catastrophic failure or ultimate failure unless it accelerates buckling.

Definition



Typical yield behavior for non-ferrous alloys.
1: True elastic limit

2: Proportionality limit
3: Elastic limit
4: Offset yield strength

It is often difficult to precisely define yielding due to the wide variety of stress–strain curves exhibited by real materials. In addition, there are several possible ways to define yielding:

True elastic limit

The lowest stress at which dislocations move. This definition is rarely used, since dislocations move at very low stresses, and detecting such movement is very difficult.

Proportionality limit

Up to this amount of stress, stress is proportional to strain (Hooke's law), so the stress-strain graph is a straight line, and the gradient will be equal to the elastic modulus of the material.

Elastic limit (yield strength)

Beyond the elastic limit, permanent deformation will occur. The lowest stress at which permanent deformation can be measured. This requires a manual load-unload procedure, and the accuracy is critically dependent on equipment and operator skill. For elastomers, such as rubber, the elastic limit is much larger than the proportionality limit. Also, precise strain measurements have shown that plastic strain begins at low stresses.

Yield point

The point in the stress-strain curve at which the curve levels off and plastic deformation begins to occur.

Offset yield point (proof stress)

When a yield point is not easily defined based on the shape of the stress-strain curve an *offset yield point* is arbitrarily defined. The value for this is commonly set at 0.1 or 0.2% of the strain. The offset value is given as a subscript, e.g., $R_{p0.2}$=310 MPa. High strength steel and aluminum alloys do not exhibit a yield point, so this offset yield point is used on these materials.

Upper yield point and lower yield point

Some metals, such as mild steel, reach an upper yield point before dropping rapidly to a lower yield point. The material response is linear up until the upper yield point, but the lower yield point is used in structural engineering as a conservative value. If a metal is only stressed to the upper yield point, and beyond, luders bands can develop.

Yield criterion

A yield criterion, often expressed as yield surface, or yield locus, is a hypothesis concerning the limit of elasticity under any combination of stresses. There are two interpretations of yield criterion: one is purely mathematical in taking a statistical approach while other models attempt to provide a justification based on established physical principles. Since stress and strain are tensor qualities they can be described on the basis of three principal directions, in the case of stress these are denoted by σ_1, σ_2, and σ_3.

The following represent the most common yield criterion as applied to an isotropic material (uniform properties in all directions). Other equations have been proposed or are used in specialist situations.

Isotropic yield criteria

Maximum Principal Stress Theory - Yield occurs when the largest principal stress exceeds the uniaxial tensile yield strength. Although this criterion allows for a quick and easy comparison with experimental data it is rarely suitable for design purposes.

$$\sigma_1 \leq \sigma_y$$

Maximum Principal Strain Theory - Yield occurs when the maximum principal strain reaches the strain corresponding to the yield point during a simple tensile test. In terms of the principal stresses this is determined by the equation:

$$\sigma_1 - \nu(\sigma_2 + \sigma_3) \leq \sigma_y.$$

Maximum Shear Stress Theory - Also known as the Tresca yield criterion, after the French scientist Henri Tresca. This assumes that yield occurs when the shear stress τexceeds the shear yield strength τ_y:

$$\tau = \frac{\sigma_1 - \sigma_3}{2} \leq \tau_{ys}.$$

Total Strain Energy Theory - This theory assumes that the stored energy associated with elastic deformation at the point of yield is independent of the specific stress tensor. Thus yield occurs when the strain energy per unit volume is greater than the strain energy at the elastic limit in simple tension. For a 3-dimensional stress state this is given by:

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 - 2\nu(\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_1\sigma_3) \leq \sigma_y^2.$$

Distortion Energy Theory - This theory proposes that the total strain energy can be separated into two components: the *volumetric* (hydrostatic) strain energy and the *shape* (distortion or shear) strain energy. It is proposed that yield occurs when the distortion component exceeds that at the yield point for a simple tensile test. This is generally referred to as the Von Mises yield criterion and is expressed as:

$$\frac{1}{2}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right] \leq \ \sigma_y^2.$$

Based on a different theoretical underpinning this expression is also referred to as **octahedral shear stress theory**.

Other commonly used isotropic yield criteria are the

- Mohr-Coulomb yield criterion
- Drucker-Prager yield criterion
- Bresler-Pister yield criterion
- Willam-Warnke yield criterion

The yield surfaces corresponding to these criteria have a range of forms. However, most isotropic yield criteria correspond to convex yield surfaces.

Anisotropic yield criteria

When a metal is subjected to large plastic deformations the grain sizes and orientations change in the direction of deformation. As a result the plastic yield behavior of the material shows directional dependency. Under such circumstances, the isotropic yield criteria such as the von Mises yield criterion are unable to predict the yield behavior accurately. Several anisotropic yield criteria have been developed to deal with such situations. Some of the more popular anisotropic yield criteria are:

- Hill's quadratic yield criterion.
- Generalized Hill yield criterion.
- Hosford yield criterion.

Factors influencing yield stress

The stress at which yield occurs is dependent on both the rate of deformation (strain rate) and, more significantly, the temperature at which the deformation occurs. Early work by Alder and Philips in 1954 found that the relationship between yield stress and strain rate (at constant temperature) was best described by a power law relationship of the form

$$\sigma_y = C(\dot{\epsilon})^m$$

where C is a constant and m is the strain rate sensitivity. The latter generally increases with temperature, and materials where m reaches a value greater than ~0.5 tend to exhibit super plastic behaviour.

Later, more complex equations were proposed that simultaneously dealt with both temperature and strain rate:

$$\sigma_y = \frac{1}{\alpha}\sinh^{-1}\left[\frac{Z}{A}\right]^{(1/n)}$$

where α and A are constants and Z is the temperature-compensated strain-rate - often described by the Zener-Hollomon parameter:

$$Z = (\dot{\epsilon}) \exp\left(\frac{Q_{HW}}{RT}\right)$$

where Q_{HW} is the activation energy for hot deformation and T is the absolute temperature.

Strengthening mechanisms

There are several ways in which crystalline and amorphous materials can be engineered to increase their yield strength. By altering dislocation density, impurity levels, grain size (in crystalline materials), the yield strength of the material can be fine tuned. This occurs typically by introducing defects such as impurities dislocations in the material. To move this defect (plastically deforming or yielding the material), a larger stress must be applied. This thus causes a higher yield stress in the material. While many material properties depend only on the composition of the bulk material, yield strength is extremely sensitive to the materials processing as well for this reason.

These mechanisms for crystalline materials include

- Work Hardening
- Solid Solution Strengthening
- Particle/Precipitate Strengthening
- Grain boundary strengthening

Work Hardening

Where deforming the material will introduce dislocations, which increases their density in the material. This increases the yield strength of the material, since now more stress must be applied to move these dislocations through a crystal lattice. Dislocations can also interact with each other, becoming entangled.

The governing formula for this mechanism is:

$$\Delta\sigma_y = Gb\sqrt{\rho}$$

where σ_y is the yield stress, G is the shear elastic modulus, b is the magnitude of the Burgers vector, and ρ is the dislocation density.

Solid Solution Strengthening

By alloying the material, impurity atoms in low concentrations will occupy a lattice position directly below a dislocation, such as directly below an extra half plane defect.

This relieves a tensile strain directly below the dislocation by filling that empty lattice space with the impurity atom.

The relationship of this mechanism goes as:

$$\Delta\tau = Gb\sqrt{C_s}\epsilon^{3/2}$$

where τ is the shear stress, related to the yield stress, G and b are the same as in the above example, C_s is the concentration of solute and ε is the strain induced in the lattice due to adding the impurity.

Particle/Precipitate Strengthening

Where the presence of a secondary phase will increase yield strength by blocking the motion of dislocations within the crystal. A line defect that, while moving through the matrix, will be forced against a small particle or precipitate of the material. Dislocations can move through this particle either by shearing the particle, or by a process known as bowing or ringing, in which a new ring of dislocations is created around the particle.

The shearing formula goes as:

$$\Delta\tau = \frac{r_{\text{particle}}}{l_{\text{interparticle}}}\gamma_{\text{particle}-\text{matrix}}$$

and the bowing/ringing formula:

$$\Delta\tau = \frac{Gb}{l_{\text{interparticle}} - 2r_{\text{particle}}}$$

In these formulas, $r_{\text{particle}}$is the particle radius, $\gamma_{\text{particle}-\text{matrix}}$is the surface tension between the matrix and the particle, $l_{\text{interparticle}}$is the distance between the particles.

Grain boundary strengthening

Where a buildup of dislocations at a grain boundary causes a repulsive force between dislocations. As grain size decreases, the surface area to volume ratio of the grain increases, allowing more buildup of dislocations at the grain edge. Since it requires a lot of energy to move dislocations to another grain, these dislocations build up along the boundary, and increase the yield stress of the material. Also known as Hall-Petch strengthening, this type of strengthening is governed by the formula:

$$\sigma_y = \sigma_0 + kd^{-1/2}$$

where

σ_0 is the stress required to move dislocations,
k is a material constant, and
d is the grain size.

Testing

Yield strength testing involves taking a small sample with a fixed cross-section area, and then pulling it with a controlled, gradually increasing force until the sample changes shape or breaks. Longitudinal and/or transverse strain is recorded using mechanical or optical extensometers.

Indentation hardness correlates linearly with tensile strength for most steels. Hardness testing can therefore be an economical substitute for tensile testing, as well as providing local variations in yield strength due to e.g. welding or forming operations.

Implications for structural engineering

Yielded structures have a lower stiffness, leading to increased deflections and decreased buckling strength. The structure will be permanently deformed when the load is removed, and may have residual stresses. Engineering metals display strain hardening, which implies that the yield stress is increased after unloading from a yield state. Highly optimized structures, such as airplane beams and components, rely on yielding as a fail-safe failure mode. No safety factor is therefore needed when comparing limit loads (the highest loads expected during normal operation) to yield criteria.

Typical yield and ultimate strengths

Note: many of the values depend on manufacturing process and purity/composition.

Material	Yield strength (MPa)	Ultimate strength (MPa)	Density (g/cm^3)	free breaking length (km)
ASTM A36 steel	250	400	7.8	3.2
Steel, API 5L X65	448	531	7.8	5.8
Steel, high strength alloy ASTM A514	690	760	7.8	9.0
Steel, prestressing strands	1650	1860	7.8	21.6
Piano wire		2200–2482	7.8	28.7
Carbon Fiber (CF, CFK)		5650	1.75	
High density polyethylene (HDPE)	26-33	37	0.95	2.8
Polypropylene	12-43	19.7-80	0.91	1.3

Stainless steel AISI 302 - Cold-rolled	520	860		
Cast iron 4.5% C, ASTM A-48	*	172	7.20	2.4
Titanium alloy (6% Al, 4% V)	830	900	4.51	18.8
Aluminium alloy 2014-T6	400	455	2.7	15.1
Copper 99.9% Cu	70	220	8.92	0.8
Cupronickel 10% Ni, 1.6% Fe, 1% Mn, balance Cu	130	350	8.94	1.4
Brass	approx. 200+	550	5.3	3.8
Spider silk	1150 (??)	1200		50
Silkworm silk	500			25
Aramid (Kevlar or Twaron)	3620		1.44	256.3
UHMWPE	21	46	0.97	240
Bone (limb)	104-121	130		3
Nylon, type 6/6	45	75		2
*Grey cast iron does not have a well defined yield strength because the stress-strain relationship is atypical. The yield strength can vary from 65 to 80% of the tensile strength.				

Elements in the annealed state

	Young's modulus (GPa)	**Proof or yield stress (MPa)**	**Ultimate strength (MPa)**
Aluminium	70	15-20	40-50
Copper	130	33	210
Iron	211	80-100	350
Nickel	170	14-35	140-195
Silicon	107	5000-9000	
Tantalum	186	180	200
Tin	47	9-14	15-200
Titanium	120	100-225	240-370
Tungsten	411	550	550-620

Chapter 9

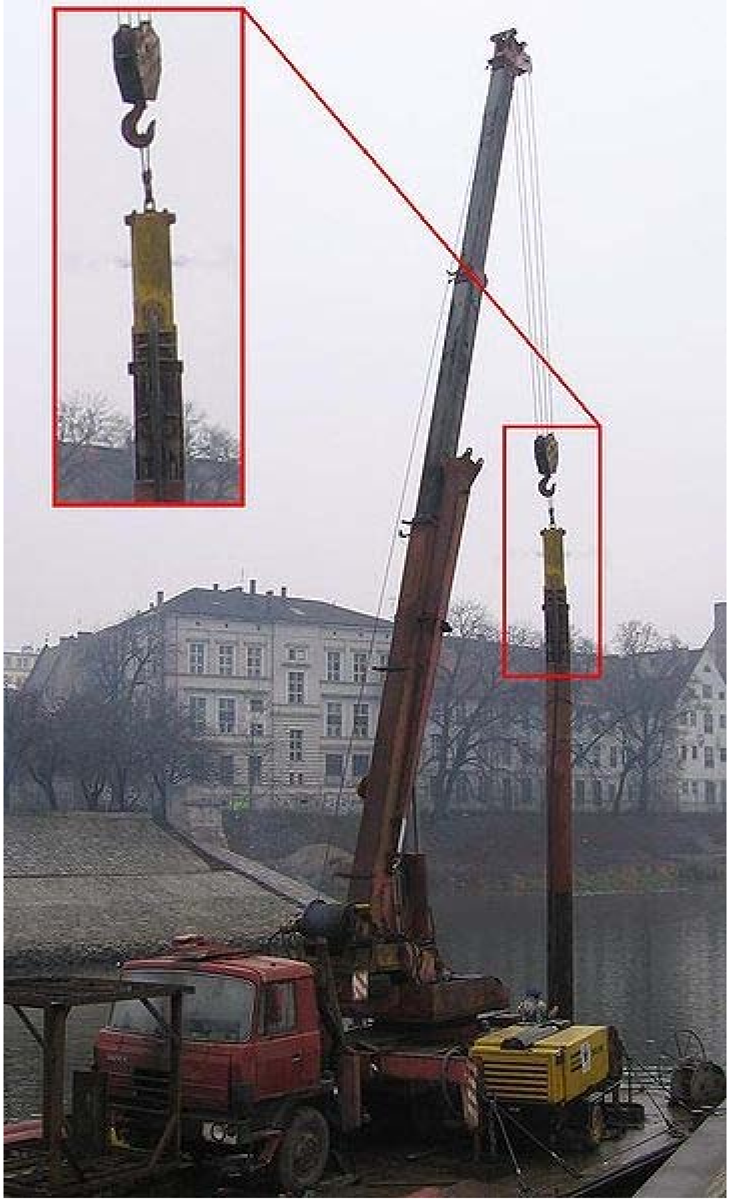
Impact (Mechanics)

In mechanics, an **impact** is a high force or shock applied over a short time period when two or more bodies collide. Such a force or acceleration usually has a greater effect than a lower force applied over a proportionally longer time period of time. The effect depends critically on the relative velocity of the bodies to one another.

At normal speeds, during a perfectly inelastic collision, an object struck by a projectile will deform, and this deformation will absorb most, or even all, of the force of the collision. Viewed from the conservation of energy perspective, the kinetic energy of the projectile is changed into heat and sound energy, as a result of the deformations and vibrations induced in the struck object. However, these deformations and vibrations can not occur instantaneously. A high-velocity collision (an impact) does not provide sufficient time for these deformations and vibrations to occur. Thus, the struck material behaves as if it were more brittle than it is, and the majority of the applied force goes into fracturing the material. Or, another way to look at it is that materials actually are more brittle on short time scales than on long time scales: this is related to time-temperature superposition.

Different materials can behave in quite different ways in impact when compared with static loading conditions. Ductile materials like steel tend to become more brittle at high loading rates, and spalling may occur on the reverse side to the impact if penetration doesn't occur. The way in which the kinetic energy is distributed through the section is also important in determining its response. Projectiles apply a Hertzian contact stress at the point of impact to a solid body, with compression stresses under the point, but with bending loads a short distance away. Since most materials are weaker in tension than compression, this is the zone where cracks tend to form and grow.

Applications



A crane with a pile driver.

A 1/2" drive pistol-grip air impact wrench

A nail is pounded with a series of impacts, each by a single hammer blow. These high velocity impacts overcome the static friction between the nail and the substrate. A pile driver achieves the same end, although on a much larger scale, the method being commonly used during civil construction projects to make building and bridge foundations. An impact wrench is a device designed to impart torque impacts to bolts to tighten or loosen them. At normal speeds, the forces applied to the bolt would be dispersed, via friction, to the mating threads. However, at impact speeds, the forces act on the bolt to move it before they can be dispersed. In ballistics, bullets utilize impact forces to puncture surfaces that could otherwise resist substantial forces. A rubber sheet, for example, behaves more like glass at typical bullet speeds. That is, it fractures, and does not stretch or vibrate.

Accidents involving impact

A Chevrolet Malibu involved in a rollover crash

Road traffic accidents usually involve impact loading, such as when a car hits a traffic bollard, water hydrant or tree, the damage being localised to the impact zone. When vehicles collide, the damage is proportionate to the relative velocity of the vehicles, the damage increasing as the square of the velocity since it is the impact kinetic energy (1/2 mv^2) which is the variable of importance. Much design effort is made to improve the impact resistance of cars so as to minmise user injury. It can be achieved in several ways: by enclosing the driver and passengers in a safety cell for example. The cell is reinforced so will survive in high speed crashes, and so protect the users. Parts of the body shell outside the cell are designed to crumple progressively, absorbing most of the kinetic energy which must be dissipated by the impact.

Various impact test are used to assess the effects of high loading, both on products and standard slabs of material. The Charpy test and Izod test are two examples of standardised methods which are used widely for testing materials. Ball or projectile drop tests are used for assessing product impacts.

The Columbia disaster was caused by impact damage when a chunk of polyurethane foam impacted the carbon fibre composite wing of the space shuttle. Although tests had been conducted before the disaster, the size of the chunks was much smaller than that which fell away from the booster rocket and hit the exposed wing.

A mock-up of a space shuttle leading edge made with an RCC-panel taken from *Discovery* showing impact damage during a test

Chapter 10

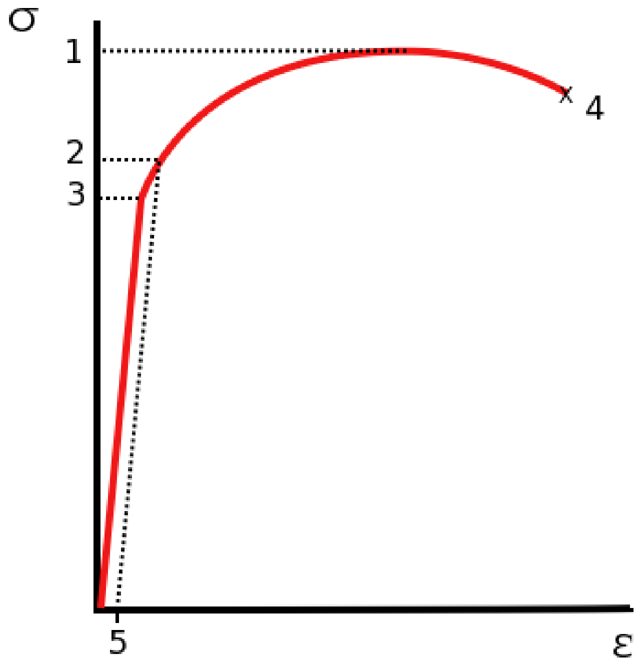
Fracture

A **fracture** is the (local) separation of an object or material into two, or more, pieces under the action of stress.

The word *fracture* is often applied to bones of living creatures (that is, a bone fracture), or to crystals or crystalline materials, such as gemstones or metal. Sometimes, in crystalline materials, individual crystals fracture without the body actually separating into two or more pieces. Depending on the substance which is fractured, a fracture reduces strength (most substances) or inhibits transmission of light (optical crystals).

A detailed understanding of how fracture occurs in materials may be assisted by the study of fracture mechanics.

Fracture strength



Stress vs. strain curve typical of aluminum
1. Ultimate tensile strength

2. Yield strength
3. Proportional limit stress
4. Fracture
5. Offset strain (typically 0.2%)

Fracture strength, also known as **breaking strength**, is the stress at which a specimen fails via fracture. This is usually determined for a given specimen by a tensile test, which charts the stress-strain curve. The final recorded point is the fracture strength.

Ductile materials have a fracture strength lower than the ultimate tensile strength (UTS), whereas in brittle materials the fracture strength is equivalent to the UTS. If a ductile material reaches its ultimate tensile strength in a load-controlled situation, it will continue to deform, with no additional load application, until it ruptures. However, if the loading is displacement-controlled, the deformation of the material may relieve the load, preventing rupture.

If the stress-strain curve is plotted in terms of *true stress* and *true strain* the curve will always slope upwards and never reverse, as true stress is corrected for the decrease in cross-sectional area. The true stress on the material at the time of rupture is known as the breaking strength. This is the maximum stress on the true stress-strain curve, given by point 3 on curve B.

Types

Brittle fracture

Brittle fracture in glass.

Fracture of an Aluminum Crank Arm. Bright: Brittle fracture. Dark: Fatigue fracture.

In *brittle fracture*, no apparent plastic deformation takes place before fracture. In brittle crystalline materials, fracture can occur by *cleavage* as the result of tensile stress acting normal to crystallographic planes with low bonding (cleavage planes). In amorphous solids, by contrast, the lack of a crystalline structure results in a conchoidal fracture, with cracks proceeding normal to the applied tension.

The theoretical strength of a crystalline material is (roughly)

$$\sigma_{\text{theoretical}} = \sqrt{\frac{E\gamma}{r_o}}$$

where: -

E is the Young's modulus of the material,

γ is the surface energy, and
r_o is the equilibrium distance between atomic centers.

On the other hand, a crack introduces a stress concentration modeled by

$$\sigma_{\text{elliptical crack}} = \sigma_{\text{applied}}\left(1 + 2\sqrt{\frac{a}{\rho}}\right) = 2\sigma_{\text{applied}}\sqrt{\frac{a}{\rho}}$$ (For sharp cracks)

where: -

σ_{applied} is the loading stress,
a is half the length of the crack, and
ρ is the radius of curvature at the crack tip.

Putting these two equations together, we get

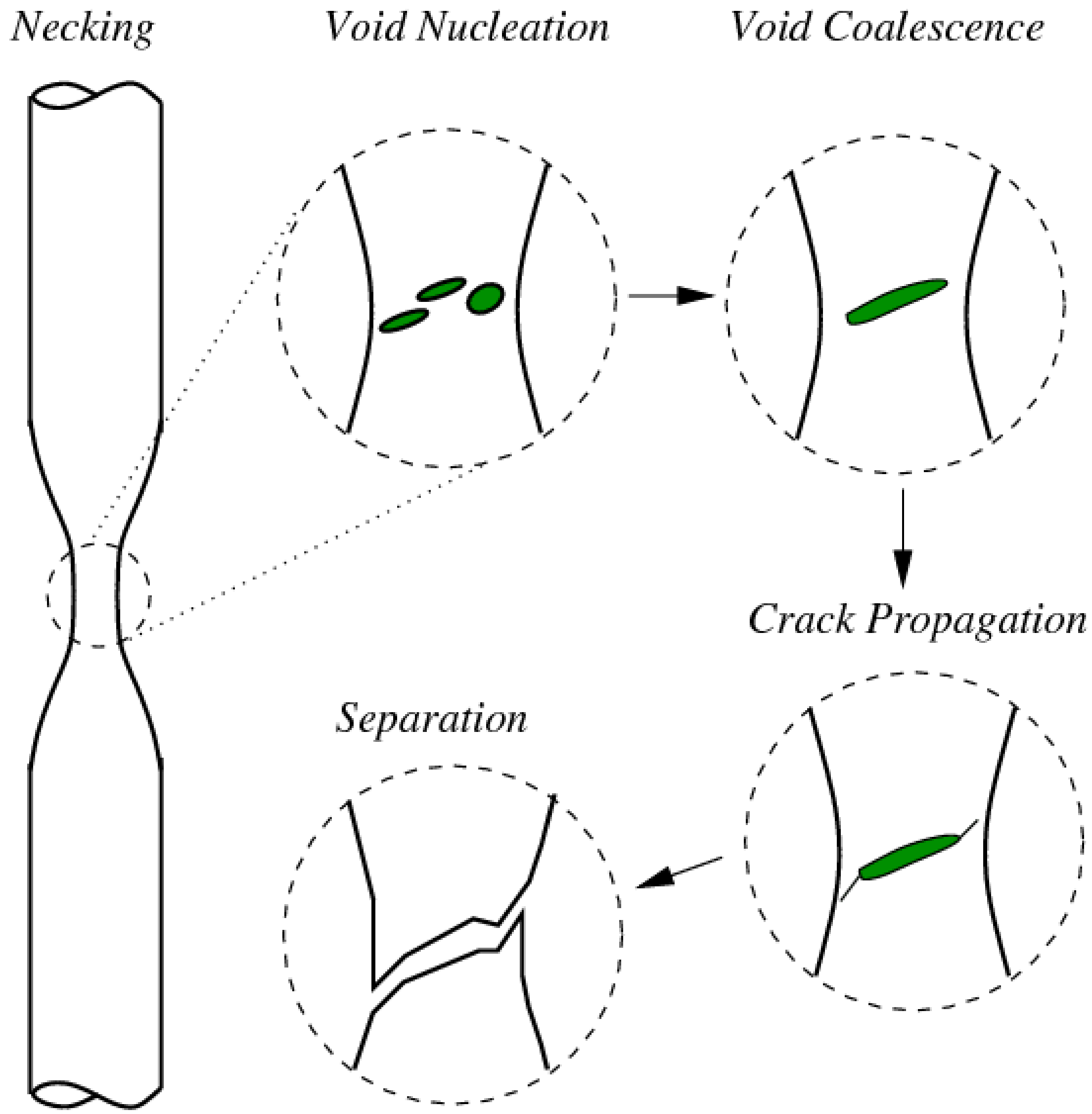
$$\sigma_{\text{fracture}} = \sqrt{\frac{E\gamma\rho}{4ar_o}}.$$

Looking closely, we can see that sharp cracks (small ρ) and large defects (large a) both lower the fracture strength of the material.

Recently, scientists have discovered supersonic fracture, the phenomenon of crack motion faster than the speed of sound in a material. This phenomenon was recently also verified by experiment of fracture in rubber-like materials.

Ductile fracture

Ductile failure of a specimen strained axially.



Schematic representation of the steps in ductile fracture (in pure tension).

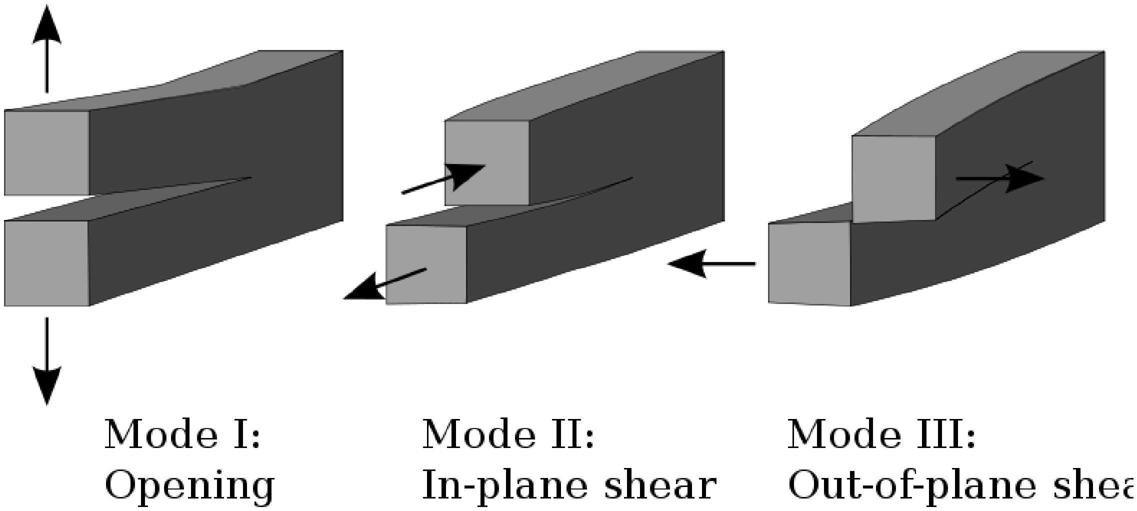
In *ductile fracture*, extensive plastic deformation takes place before fracture. The terms *rupture* or *ductile rupture* describe the ultimate failure of tough ductile materials loaded in tension. Rather than cracking, the material "pulls apart," generally leaving a rough surface. In this case there is slow propagation and an absorption of a large amount energy before fracture.

Many ductile metals, especially materials with high purity, can sustain very large deformation of 50–100% or more strain before fracture under favorable loading condition and environmental condition. The strain at which the fracture happens is controlled by the purity of the materials. At room temperature, pure iron can undergo deformation up to 100% strain before breaking, while cast iron or high-carbon steels can barely sustain 3% of strain.

Because ductile rupture involves a high degree of plastic deformation, the fracture behavior of a propagating crack as modeled above changes fundamentally. Some of the energy from stress concentrations at the crack tips is dissipated by plastic deformation before the crack actually propagates.

The basic steps are: void formation, void coalescence (also known as crack formation), crack propagation, and failure, often resulting in a cup-and-cone shaped failure surface.

Crack separation modes



The three fracture modes.

There are three ways of applying a force to enable a crack to propagate:

- **Mode I crack** – Opening mode (a tensile stress normal to the plane of the crack)
- **Mode II crack** – Sliding mode (a shear stress acting parallel to the plane of the crack and perpendicular to the crack front)
- **Mode III crack** – Tearing mode (a shear stress acting parallel to the plane of the crack and parallel to the crack front)