

Structural Integrity 12

Series Editors: José A. F. O. Correia · Abílio M. P. De Jesus

Jorge Luis González-Velázquez

Mechanical Behavior and Fracture of Engineering Materials

 Springer

Structural Integrity

Volume 12

Series Editors

José A. F. O. Correia, Faculty of Engineering, University of Porto, Porto, Portugal
Abílio M. P. De Jesus, Faculty of Engineering, University of Porto, Porto, Portugal

Advisory Editors

Majid Reza Ayatollahi, School of Mechanical Engineering, Iran University of Science and Technology, Tehran, Iran

Filippo Berto, Department of Mechanical and Industrial Engineering, Faculty of Engineering, Norwegian University of Science and Technology, Trondheim, Norway

Alfonso Fernández-Canteli, Faculty of Engineering, University of Oviedo, Gijón, Spain

Matthew Hebdon, Virginia State University, Virginia Tech, Blacksburg, VA, USA

Andrei Kotousov, School of Mechanical Engineering, University of Adelaide, Adelaide, SA, Australia

Grzegorz Lesiuk, Faculty of Mechanical Engineering, Wrocław University of Science and Technology, Wrocław, Poland

Yukitaka Murakami, Faculty of Engineering, Kyushu University, Higashiku, Fukuoka, Japan

Hermes Carvalho, Department of Structural Engineering, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Shun-Peng Zhu, School of Mechatronics Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

The *Structural Integrity* book series is a high level academic and professional series publishing research on all areas of Structural Integrity. It promotes and expedites the dissemination of new research results and tutorial views in the structural integrity field.

The Series publishes research monographs, professional books, handbooks, edited volumes and textbooks with worldwide distribution to engineers, researchers, educators, professionals and libraries.

Topics of interested include but are not limited to:

- Structural integrity
- Structural durability
- Degradation and conservation of materials and structures
- Dynamic and seismic structural analysis
- Fatigue and fracture of materials and structures
- Risk analysis and safety of materials and structural mechanics
- Fracture Mechanics
- Damage mechanics
- Analytical and numerical simulation of materials and structures
- Computational mechanics
- Structural design methodology
- Experimental methods applied to structural integrity
- Multiaxial fatigue and complex loading effects of materials and structures
- Fatigue corrosion analysis
- Scale effects in the fatigue analysis of materials and structures
- Fatigue structural integrity
- Structural integrity in railway and highway systems
- Sustainable structural design
- Structural loads characterization
- Structural health monitoring
- Adhesives connections integrity
- Rock and soil structural integrity.

Springer and the Series Editors welcome book ideas from authors. Potential authors who wish to submit a book proposal should contact Dr. Mayra Castro, Senior Editor, Springer (Heidelberg), e-mail: mayra.castro@springer.com

More information about this series at <http://www.springer.com/series/15775>

Jorge Luis González-Velázquez

Mechanical Behavior and Fracture of Engineering Materials



Springer

المنارة للاستشارات

Jorge Luis González-Velázquez
Department of Metallurgy and Materials
Instituto Politécnico Nacional
Mexico City, Mexico

ISSN 2522-560X

Structural Integrity

ISBN 978-3-030-29240-9

<https://doi.org/10.1007/978-3-030-29241-6>

ISSN 2522-5618 (electronic)

ISBN 978-3-030-29241-6 (eBook)

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

المنارة للاستشارات

Preface

Mechanical behavior is a discipline that studies the performance of materials under the action of external forces, such as load, impact, bending, or pressure. Basically, it is a part of physics aimed at the study of the stress-strain phenomena of solid bodies. For the engineer, the knowledge of mechanical behavior is fundamental as it provides the basics to analyze and understand the reaction of physical components under the action of forces. The components specifically intended for bear, transmit, or resist forces are called mechanical or structural, for example: beams, shafts, piping, pressure vessels, gears, and etcetera; therefore, they have to be designed, operated, and maintained to fulfill this requirement, and the Mechanical Behavior discipline provides the basis for these tasks in addition to characterize the mechanical properties of the materials used for the fabrication and construction. Even components that are not intended to support loads as their primary function must be designed to resist the action of eventual o secondary forces, such as dead weight, impacts, wind, earthquakes, and etcetera. For example, a utility cable may have the primary function of conducting electricity, but it has to resist the weight of ice and snow, the wind pressure, and the weight of hanging objects during service.

This book describes the fundamentals of the stress and strain theory introduced by Cauchy and others in the nineteenth century, as well as the mechanisms of deformation, strengthening and fracture, applied to materials commonly used in engineering, with emphasis on metals and alloys, but also entails polymers and composite materials. Chapters 1 and 2 have to do with the definition of stress and strain under the assumption that the solid body is continuous, homogeneous, and isotropic; hence, this subject is known as *Continuum Mechanics*. In addition, the method for stress transformation, known as Mohr's Circle, the most yield criteria and the elastic stress-strain relations are described, with emphasis on its practical use in engineering. A brief description of the finite element method to perform the numerical analysis of stress and strain and the experimental methods to measure them are included and at the end of Chap. 2, and finally the hardness test is briefly described.

Chapter 3 describes the plastic strain mechanisms, primarily by dislocation slip, aimed to understand the strengthening mechanisms of metals and alloys discussed in Chap. 4. A section describing the basics of transmission electron microscopy as the most used method for direct observation of dislocations is included in Chap. 3. Since both plastic strain and strengthening mechanisms depend on the microstructure, these chapters introduce basic concepts of metallurgy and heat treatment as the means to control microstructure, and thus the mechanical properties. The mechanical behavior of polymer and composite materials is described in Chap. 5.

The fracture phenomenon is studied in Chap. 6, presenting the basic concepts of fracture along with a description of the mechanisms of brittle and ductile fracture, followed by a comprehensive introduction to fracture mechanics and completed with an introduction to structural integrity, as the in-field application of fracture mechanics. The chapter concludes with a description of the Charpy impact test. Chapter 7 entirely focused on fatigue, due to its importance as fracture mechanism of in-service mechanical components, the interpretation of the S-N curve and the fatigue life prediction methods are explained. The chapter concludes with a brief introduction to fatigue crack growth.

Finally, Chap. 8 deals with high-temperature behavior, specifically the creep phenomenon. The Larson-Miller method to determine creep life is described, as well as the creep mechanisms of dislocation climb and grain boundary sliding. The chapter concludes with the application of the phenomenological creep equations to the design of high-temperature resistant materials.

Mexico City, México
April 2019

Jorge Luis González-Velázquez

Contents

1	Stress	1
1.1	Mechanical Behavior and Engineering Materials	1
1.2	Definition of Stress	4
1.3	Mechanical Behavior in Uniaxial Tension	7
1.4	Mechanical Design by the Stress Definition	11
1.5	The Stress Tensor	13
1.6	Non-uniform Stress Distribution	20
1.7	Stress Transformation and the Mohr's Circle	22
1.8	Yield Criteria	33
1.9	Stress Concentration	40
2	Strain	43
2.1	Definition of Strain	43
2.2	Elastic Stress-Strain Relations	51
2.3	Constitutive Equations for Non-isotropic Elastic Materials (Elastic Anisotropy)	54
2.4	Plasticity	58
2.5	Real Stress and Strain	65
2.6	Plastic Stress-Strain Relations	68
2.7	Numeric Analysis of Stress and Strain	69
2.8	Experimental Measurement of Stress and Strain	73
2.9	Hardness	77
3	Plastic Deformation Mechanisms	81
3.1	Crystalline Defects and Deformation Mechanisms	81
3.2	Definition of Cold and Hot Work	83
3.3	Deformation by Dislocation Slip	85
3.4	Direct Observation of Dislocations	94
3.5	Slip Modes	96
3.6	Deformation by Twinning	100

4	Strengthening Mechanisms	103
4.1	Strengthening of Engineering Materials	103
4.2	Strain Hardening	104
4.3	Grain Boundary Strengthening	108
4.4	Solute Strengthening (Solid Solution)	113
4.5	Second Phase Strengthening	116
4.6	Fine Particle Strengthening	122
4.7	Upper Yield Strength and Strain Aging	127
4.8	Martensite Hardening in Steels	129
5	Mechanical Behavior of Composites and Polymers	135
5.1	Fundamentals of Fiber Reinforced Composite Materials	135
5.2	Mechanical Behavior of Fiber Reinforced Composites	137
5.3	Mechanical Behavior of Polymers	144
5.4	Mechanical Behavior of Elastomers	150
5.5	Failure Mechanisms of Polymers	152
	Reference	154
6	Fracture	155
6.1	Basic Concepts of Fracture	155
6.2	Static Fracture Mechanisms	157
6.3	Fracture Mechanics	159
6.4	Fracture Toughness	169
6.5	Structural Integrity	176
6.6	The Charpy Impact Test	181
	References	186
7	Fatigue	187
7.1	Definition and History of Fatigue	187
7.2	Fatigue Fracture	192
7.3	Mechanical Characterization of Fatigue	196
7.4	Factors Affecting Fatigue	202
7.5	Cyclic Strain Behavior	205
7.6	Fatigue Life	209
7.7	The Universal Slopes Method of Fatigue	215
7.8	Fatigue Crack Growth	218
	References	223
8	High Temperature Mechanical Behavior	225
8.1	High Temperature Deformation	225
8.2	Secondary Creep	229
8.3	Creep Life Prediction	231
8.4	Deformation Mechanisms in Creep	234

8.5 Creep Fracture	236
8.6 High Temperature Materials	240
8.7 Superplastic Behavior	241
References	244

About the Author

Jorge Luis González-Velázquez has been Professor of the Metallurgy and Material Department at the Superior School Chemical Engineering and Extractive Industries of the Instituto Politecnico Nacional (National Polytechnic Institute) of Mexico since 1990. He holds a B.S. degree in metallurgical engineering and a M.S. degree in metallurgy conferred by the same institution. He obtained a Ph.D. from The University of Connecticut, U.S.A. in 1989, working on fatigue crack growth under the supervision of Prof. Arthur J. McEvily. Later on, he joined the IPN and started teaching the subject of Physical Metallurgy, which in 2010 was divided into two courses: Mechanical Behavior and Fracture. He also started the courses of Fracture Mechanics and Fractography at graduate level, for the first time in México.

Alongside his teaching activity, Dr. Gonzalez has worked over the last 30 years on research and technological development in the fields of mechanical behavior, fracture mechanics, physical metallurgy, and structural integrity. He is author of over 340 technical publications and has been director of more than 80 theses at undergraduate and graduate levels. He also has conducted over 180 lectures in international conferences and meetings. He is author of the books titled “Fracture Mechanics” and “Mechanical Metallurgy”, both of them published in spanish by Editorial Limusa, Mexico, and “Fractography and Failure Analysis”, published by Springer Nature.

Chapter 1

Stress



Abstract A brief introduction of the field of Mechanical Behavior and Engineering Materials is given at the beginning of this chapter, where the importance of this field of study is emphasized, followed by the scope of continuum mechanics study and the definition of stress. By using these ideas, the mechanical behavior in uniaxial tension and the design of structural components is described. A comprehensive description of the Cauchy's stress tensor is provided along with a simplified procedure to determine the state of stress, illustrated with practical examples, to continue with a full description of the stress transformation methods, both by matrix algebra and the Mohr's Circle methods. The description of the Tresca and Von Mises yield criteria is given, including solved problems and the use of two-dimension yield maps. The chapter finalizes with a brief introduction to the stress concentration phenomenon.

1.1 Mechanical Behavior and Engineering Materials

Mechanical Behavior refers to the study of the relation among the loads that act upon a solid body and the internal forces and strains produced as a result of such action. It is "mechanical" because it deals with the analysis of forces and their reactions on a solid body, regardless of the origin of the loads and without alteration of the material. According to Newton's third law, which is the foundation of mechanics, to every action corresponds a reaction of equal magnitude and opposite direction; so, whenever a force is applied on a solid static body, an internal reaction force is produced in order to balance the external force and maintain the equilibrium. The magnitude of the internal reaction is the stress and the immediate effect of the presence of stresses in a solid is the strain, therefore, the mechanical behavior discipline analyses the stresses and strains within solids, and determines whether the material have enough strength to withstand such forces without excessively deforming, nor fracturing.

The study of stresses and strains in solid bodies that do not have cracks, voids or discontinuities is known as *continuum mechanics*. Commonly, continuum mechanics is referred as "Strength of Materials", but such term is rather imprecise,

because what is found in Strength of Materials textbooks are methods and formulas meant for the design of structural components, such as beams and columns, among others. Whereas the study of the mechanical behavior and the strengthening mechanisms, is viewed in textbooks with titles such as: Mechanical Metallurgy, Physical Metallurgy, or Materials Science.

Mechanical Behavior is a very important part of engineering, because any solid component of a machine, a structure, a tool or any object that will bear loads, should be designed and fabricated to withstand stresses and strains produced by its use. For instance, a gas tank must be able to resist the internal pressure without distortion or rupture, a railroad track must resist the pass of trains without excessive deformation and wear, the columns of a building must bear the dead weight plus earthquake and wind loads without excessively bending or breaking. The mechanical behavior is important in nature as well, a tree stem must support the weight of its branches, leaves and fruits, the bones of a vertebrate animal must withstand the forces of walking and jumping and so on. Even when the main function of a solid component is not to withstand or transmit forces, they must have enough mechanical strength to resist the forces that eventually may act over them; for example, the main function of a conductor cable is to transport electricity, but if it does not have enough strength it will not resist the bending, hanging and twisting loads that it experiences in service and will fail. A glass window has the main function of separating the internal environment of a building from the external one, while letting the light to pass, but if it does not have mechanical strength, it will not resist the wind load and small hits caused by tiny stones, bird crashes and so on. In addition to analyze the stresses and strains in loaded bodies, the field of mechanical behavior has the task of assessing the strength of materials through the mechanical tests such as the tension, hardness and impact tests.

Another important aspect of the Mechanical Behavior is to understand and control the strengthening mechanisms of the materials, both engineering and natural. This study is generally at a microscopic level, since the mechanisms of deformation and strengthening are related to the crystalline structure, the crystallographic defects and the microstructure. The field of deformation and strengthening mechanisms provides the basis for the design of new materials and manufacturing processes that allow to obtain specific mechanical properties and the optimal combinations of them.

Considering the above, the field comprising the Mechanical Behavior of materials has been divided into three parts which are:

- **Continuum Mechanics:** It is the study of the stress-strain relations, from the macroscopic point of view, assuming the solid is a continuum and homogeneous body.
- **Deformation and Strengthening mechanisms:** This deals with the study on how plastic deformation takes place at an atomic and microstructural level and the origins of the mechanical properties of materials.
- **Fracture:** Which is divided into two branches: Fracture Mechanics, which is the study of the mechanical behavior of cracked bodies to determine the fracture

resistance and the dynamics of crack propagation, from a macroscopic point of view; and Fractography, which is the study of fracture surfaces and fracture mechanisms, at both macro and microscopic levels.

In Continuum Mechanics, the distribution of stresses and strains of a loaded solid body are determined as a function of geometry, points and directions of applied loads and the mechanical properties determine whether or not the body will withstand such stresses and strains. The main mechanical properties are: resistance to stresses (yield and tensile strength), resistance to penetration (hardness), stiffness (elastic modulus), maximum attainable deformation (ductility), resistance to strain (resilience) and strain energy (toughness).

The study of deformation and strengthening mechanisms is aimed to learn about the specific relations between the crystalline structure and microstructure with the mechanical properties, whereas fracture mechanics is aimed to assess the performance of cracked materials, which are the basis for the damage tolerance design and structural integrity or fitness-for-service assessments.

The materials used for the manufacturing of components for structural and mechanical applications in which the mechanical behavior is the fundamental design criteria are known as *engineering materials* and they are classified according to the Table 1.1.

Table 1.1 Classification of engineering materials

Class	Characteristics	Example
Ferrous alloys	High mechanical strength, ductile, good resistance to heat, good electrical conductivity, high formability, high impact and fracture resistance	Low, medium and high alloy steels, cast irons
Nonferrous alloys	High to medium mechanical strength, ductile, mild resistance to heat, high electrical conductivity, high formability, mild impact and fracture resistance	Cooper, Aluminum, Nickel, Titanium, Magnesium, Zinc and their alloys
Polymers	Low mechanical strength, elastic, poor resistance to heat, electrical insulators, high formability, high corrosion resistance	Plastics (thermoset y thermoplastic), elastomers, foams
Composites	High mechanical strength, poor resistance to heat, low density, high formability	Polymer matrix reinforced with fibers (glass, carbon, Kevlar), reinforced concrete
Technical ceramics	High wear strength, refractory, high corrosion resistance, low impact and fracture resistance	Alumina, magnesia, silica, tungsten carbide
Glass	High corrosion resistance, transparent, poor impact resistance	Borosilicate, silica and soda glasses
Rocks	High compression strength, refractory, low impact and fracture resistance	Clay brick, granite, basalt
Bio-materials	Low density, easy to work, low mechanical strength, low cost	Wood, leather, bone, cotton, silk, vegetable fibers

1.2 Definition of Stress

The concept of stress was introduced by the French scientist and mathematician Augustin Louis Cauchy in 1833, based on the Euler's movement laws, Newton's mechanics and the infinitesimal calculus. Cauchy was a remarkable mathematician, catholic and a devoted royalist, that made numerous and important contributions to virtually every branch of mathematics. He was member of the France Academy and professor at Ecolé Polytechnique Parisien and the Paris Sorbone University, where he studied the deformations in loaded solid bodies, introducing the concept of the stress tensor. The next paragraphs describe the Cauchy's stress tensor concept, simplified to meet the objective of this book.

In order to define stress, first it is necessary to identify the types of forces that act on a solid body, such forces are:

- Surface forces. They act throughout an external surface, such as loads, tractions and pressure.
- Body Forces. They are exercised on the entirety of particles of a solid body, such as: gravity, magnetism, inertia, thermal forces, etc.

The study of mechanical behavior only considers surface forces, that is to say, load, tractions and pressure, since body forces do not significantly deform the materials, unless they are extremely high.

To analyze the mechanical behavior of a solid body, it is necessary to simplify the system, because materials are complex arrays of atoms, crystalline defects, grains, second phases and heterogeneities. The simplifications consider that the solid body is:

- Continuum. Matter occupies the entire volume and there are no voids or interruptions.
- Homogeneous. The entire volume is made up of the same type of material.
- Isotropic. The properties are the same in any direction.

Figure 1.1 shows an idealized body under the action of an external force. If the body remains static, the external force F is balanced by an internal reaction force Fr , of the same magnitude, but opposite direction. Physically speaking the internal force is the vector resultant of the many forces that act along the links of the particles that

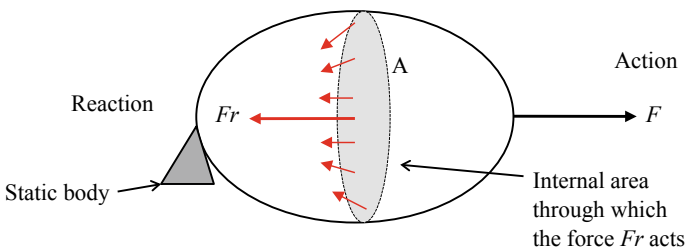


Fig. 1.1 Schematic representation of a static body under an external force

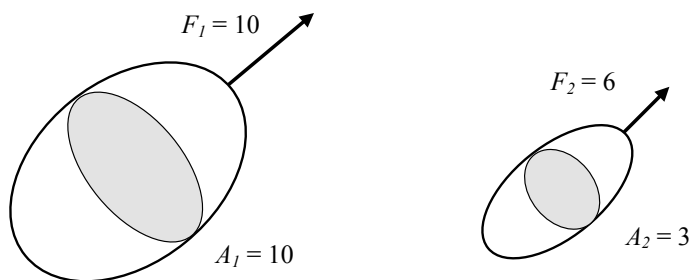


Fig. 1.2 Force/area relation in a static solid under the action of a force

make up the solid (atoms or molecules), thus the effect of the reaction is the stretching, shortening or twisting these links. The magnitude of the internal reaction in the solid, obviously depends on the magnitude of the applied force, but also on the amount of links resisting such force. The number of particle's links resisting these internal forces is proportional to the cross section area A , so the magnitude of the internal effect can be measured by the ratio F/A . If F/A is big, the effect is big and vice versa.

The above concept is illustrated in Fig. 1.2, which shows how the magnitude of the internal effect produced when an external force is applied on a static solid is directly proportional to the applied force and inversely proportional to the cross-section area. The bigger body has a cross area $A_1 = 10$ and an applied force $F_1 = 10$, so it experiences a force unit per area equal to 1.0 force units per unit area. The smaller body has a cross area $A_2 = 3$, with an applied force $F_2 = 6$, so it withstands 2.0 force units per unit area, which is twice of what the bigger body is withstanding, thus it experiences a greater effect.

The magnitude of the internal reaction calculated by the ratio F/A is the *stress* and mathematically is defined by the equation:

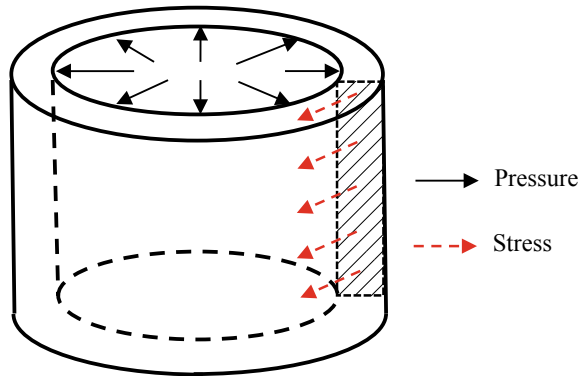
$$\sigma = \frac{F}{A}$$

Therefore, *stress is the magnitude of the internal reaction produced in a static solid under the action of an external load or force*. According to this definition, stress has derived units, and the typical ones are shown in Table 1.2, along with the conversion factors.

At this point, it is important to differentiate between pressure and stress since they are calculated in a similar way and have the same units, so they are often mistaken. Stress, as it has been stated, is the measure of the effect of an external force a body; hence, the effect is internal and actuates over a cross-section area, while pressure is a force evenly distributed on an external surface. Therefore, stress is internal, and pressure is external. In order to visualize this, consider a hollow cylinder containing a pressurized fluid as depicted in Fig. 1.3. Note that the pressure is acting on an external surface, since the definition of "external" refers to a

Table 1.2 Typical units of stress

System	Units	Common multiple
International	Pascal (Pa = Nw/m ²)	MPa = 10 ⁶ Pa
US customary	psi (psi = lbf/plg ²)	ksi = 1000 psi
Metric	Kgf/mm ²	Kgf/cm ² = 100 kg/mm ²
Conversion factors:		1 MPa = 10.5 kgf/cm ² 1 kgf/cm ² = 41.22 psi 1 ksi = 6.895 MPa

Fig. 1.3 Stress produced in a hollow cylinder with internal pressure

location outside of the body's volume, so even though the pressure is in the internal side of the cylinder, it is applied outside the thickness of the cylinder. Now, it is easy to visualize that the internal pressure will expand the cylinder, and this expansion will generate internal forces in the circumferential direction, represented by the dotted arrows in Fig. 1.3. These circumferential forces act across the shaded area of Fig. 1.3, so it can be foreseen that the stress produced by the circumferential force will have a different value from that of the pressure.

As stresses come from forces, and forces are vectors, they can be decomposed into components; in order to avoid dealing with inclined vectors, Cauchy considered one component perpendicular to the cross-section area F_n and other component F_t parallel to or tangential to it, as shown in Fig. 1.4. Since the effects of these forces on the body are different, so are the stresses resulting from them.

Thus, the stress produced by a force normal to the cross-section area is called *normal* and is designated by:

$$\sigma_N = \frac{F_N}{A}$$

Normal stresses, are divided into two types: when internal forces tend to separate particles or stretch the body they are called *tension stresses* and are of positive sign (+), and when the force brings particles close to each other, or shorten the body,

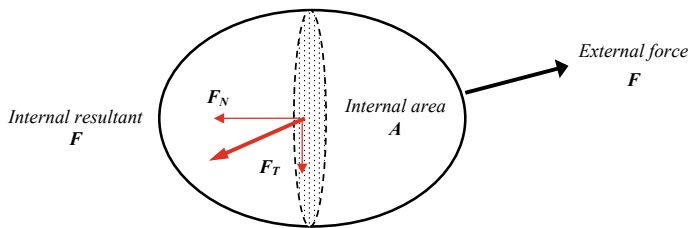


Fig. 1.4 Normal component (F_N) and Tangential component (F_T) of the internal forces

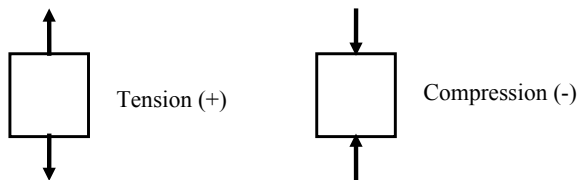


Fig. 1.5 Free body diagrams showing the types of normal stresses

they are *compression stresses* and are of negative sign (-). Such stresses are represented in a free body diagram, as shown in Fig. 1.5.

The stress produced by tangential forces is called *shear stress* and is designated by:

$$\tau = \frac{F_T}{A}$$

The sign of the shear stress is defined in terms of the direction of the bending moment, thus, it is positive if it generates a counter clock moment and vice versa.

1.3 Mechanical Behavior in Uniaxial Tension

The uniaxial tension test is the universally accepted method to characterize the mechanical behavior of engineering materials. Customary, the test is carried out in accordance with the standard ASTM E8, which consists of stretching a test specimen of uniform cross-section (round or rectangular) fixed with a set of grips attached to a load frame. During the test, the specimen's load and elongation are continuously recorded, and the results are plotted in a Load vs. Elongation plot, which for a typical metallic material, looks like the one shown in Fig. 1.6.

The Load-Elongation plot shows that, when a tensile load is applied, the immediate effect is an elongation. Initially, if the load is removed, the test piece recovers its initial shape and dimensions; then, it is said that the deformation is *elastic*. After exceeding a specific load value, a permanent elongation occurs, and then it is said that there is a *plastic strain*. It is important to bear in mind that the

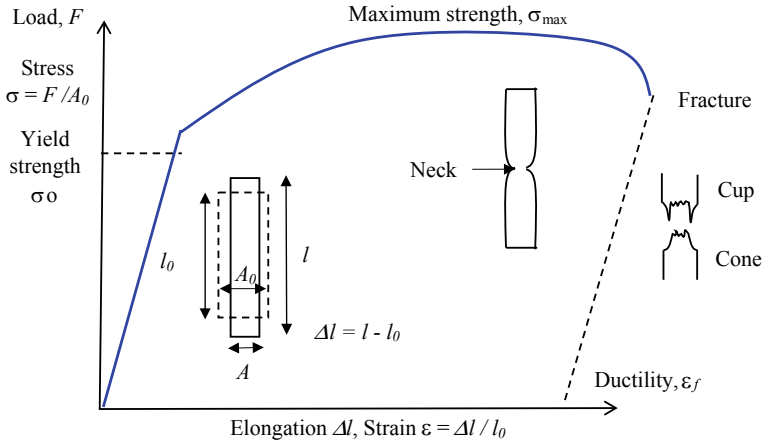


Fig. 1.6 Schematic representation of the Load-Elongation and Stress-strain curves in uniaxial tension of a typical engineering material

elastic deformation does not disappear during plastic deformation, but both are added, therefore the behavior is *elastic-plastic*.

As mentioned before, the result of the uniaxial tension test is a Load vs. Elongation curve, however, the load is divided by the area to give the *engineering stress* (σ), which is defined as:

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

where F is the applied force and A_0 is the initial cross-section area of the test specimen. Similarly, the *engineering deformation* (ϵ), is defined as:

$$\epsilon = \frac{\Delta l}{l_0}$$

where Δl is the elongation and l_0 is the initial length of the test specimen. As A_0 and l_0 are constants, the shape of the Load-Elongation curve does not change when it is transformed into the Engineering Stress-Strain curve, as illustrated in Fig. 1.6. By this way the main mechanical properties of engineering materials are determined from this curve, as described in the next paragraphs.

The stress level at which plastic strain initiates is referred as *yield strength*, represented by the symbol σ_0 . The yield strength is a property of the material, and it is of great importance in many engineering applications, since if a component gets plastically deformed, usually it will not perform properly, constituting a failure; for that reason, most engineering designs are done in such a way that the acting stresses do not exceed the yield strength.

The maximum point in the Stress-Strain curve is the *tensile strength*, identified by the symbols σ_{\max} or σ_u and it is also a property of the material. It may be noticed that after the tensile strength is reached, the Stress-Strain curve falls down; this is due to the formation of a geometrical contraction of the test specimen, called *neck*. The neck rapidly reduces the cross-section area and induces high local tensile stresses, causing a reduction of the load necessary to continue straining the material. The neck also indicates the initiation of the fracture process, which consists on the formation of an internal crack by the nucleation, growth and coalescence of internal voids. The internal crack reduces the cross-section area to a level where the remaining ligament fails by ductile shear fracture, forming the typical *cup and cone* fracture.

The maximum elongation after failure is called *ductility* and is represented by the symbol ε_f , usually it is a material property, but it is strongly influenced by microstructural and environmental factors, so, often it is a non-mandatory or secondary requirement in material's specifications.

For most engineering materials, the elastic part of the Stress-Strain curve linear, as seen in Fig. 1.7. The ratio between stress and elastic strain is the *Young's modulus*, which is defined as:

$$E = \left(\frac{\sigma}{\varepsilon} \right)_{\text{Elastic}}$$

where σ is the stress and ε is the elastic elongation.

Bearing in mind, that $\varepsilon = \Delta l / l_0$ and $\sigma = F/A_0$, it can be stated that:

$$\sigma = E \left(\frac{\Delta l}{l_0} \right) = \frac{F}{A}$$

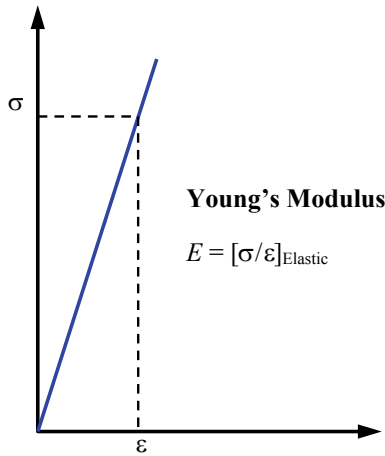


Fig. 1.7 Elastic portion of the stress-strain curve in uniaxial tension

Therefore:

$$F = \left(\frac{AE}{l_0} \right) \Delta l$$

The term (AE/l_0) is known as the *Elastic Coefficient* and it determines how rigid or flexible a structure is. High values of the elastic coefficient will result in rigid structures, less likely to deform; whereas low values of the elastic coefficient will give flexible structures, easily deformed under load.

The application of the Elastic Coefficient is widespread in the design of, both, mechanical and structural components, where it is desirable to set limits or controls over elastic strain. The most frequent cases of elastic design are:

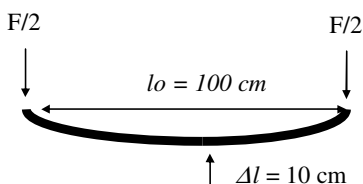
- **Flexible or elastic components:** those are that should feature fairly large elastic strains or flexions, but not as much as to reach yield, as it is the case of helical and leaf springs. In other words, flexible components are designed to have large controlled elastic strains under the applied loads. According to the elastic coefficient, this can be achieved by long lengths, small cross-section areas and materials with a low Young's modulus. The first two conditions make flexible components long and slender.
- **Rigid components:** those are where an excessive elastic strain is adverse for their performance, such as building structures, supports, gears and machine parts. In these components, the magnitude of the elastic strain must be limited to a minimum, so the elastic coefficient must be high. This can be attained by widening the cross-section, shortening the length and by selecting high Young's modulus materials. The first two characteristics make rigid components short and thick.

The following examples illustrate the use of the elastic coefficient.

Example 1 A leaf spring of length $l_0 = 100$ cm long and width W 10 cm must have an elastic vertical displacement δ $l < 10$ cm under a load $F = 1000$ Nw (Approx. 100 kg). Determine the leaf spring thickness B if it is made of steel ($E = 211.4$ GPa).

Solution A free body diagram of the problem is:

According to the force-elongation elastic formula: $F = (AE \Delta l) / l_0$, thus:



$$A = (F l_o) / (E \Delta l)$$

However, $A = W \times B$. Solving for B and substituting values:

$$\begin{aligned} B &= (F l_o) / (E \Delta l W) = (1000 \text{ Nw}) \times (1 \text{ m}) / (211.4 \times 10^9 \text{ Nw/m}^2)(0.1 \text{ m}) \times (0.1 \text{ m}) \\ &= 4.73 \times 10^{-3} \text{ m} = 4.73 \text{ mm} \end{aligned}$$

Example 2 An engineering design came out with a hollow square column of 10 cm external width, 2.5 mm thickness and 4 m long, to support a load $F = 490,500 \text{ Nw}$ (50 000 kg). The typical out-of-plumbness tolerance for building columns is 1 mm per meter of height, neglecting curvature effects, it is a tensile strain of $\Delta l/l_o = 5.73 \times 10^{-3}$. A novel engineer suggests to substitute steel by aluminum in order to save weight. Is this a good idea? Assume that the Young's modulus in compression is identical to that in tension. Steel $E = 211.4 \times 10^9 \text{ Nw/m}^2$; aluminum $E = 70.3 \times 10^9 \text{ Nw/m}^2$.

Solution According to the force-elongation elastic formula: $F = (AE \Delta l) / l_o$, thus:

$$(\Delta l / l_o) = F / (AE)$$

$A = 9.75 \text{ cm}^2 = 0.000975 \text{ m}^2$. Substituting values for steel:

$$(\Delta l / l_o)_{\text{Steel}} = (490\,500 \text{ Nw}) / (0.000975 \text{ m}^2 \times 211.4 \times 10^9 \text{ Nw/m}^2) = 2.38 \times 10^{-3}$$

Substituting values for aluminum:

$$\begin{aligned} (\Delta l / l_o)_{\text{Aluminum}} &= (490\,500 \text{ Nw}) / (0.000975 \text{ m}^2 \times 70.3 \times 10^9 \text{ Nw/m}^2) \\ &= 7.16 \times 10^{-3} \end{aligned}$$

Notice that the aluminum column has three times more elastic deformation than steel and it will not meet the specification for out-of-plumbness. This is why aluminum is seldom used in load bearing structures such as buildings and bridges.

1.4 Mechanical Design by the Stress Definition

The main contribution of the stress concept is that the design of mechanical or structural components can be straight forward, once the stress is known. Based on the definition of stress, the design variables can be set as follows:

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

(Allowable stress) (Applied load) (Cross section size)

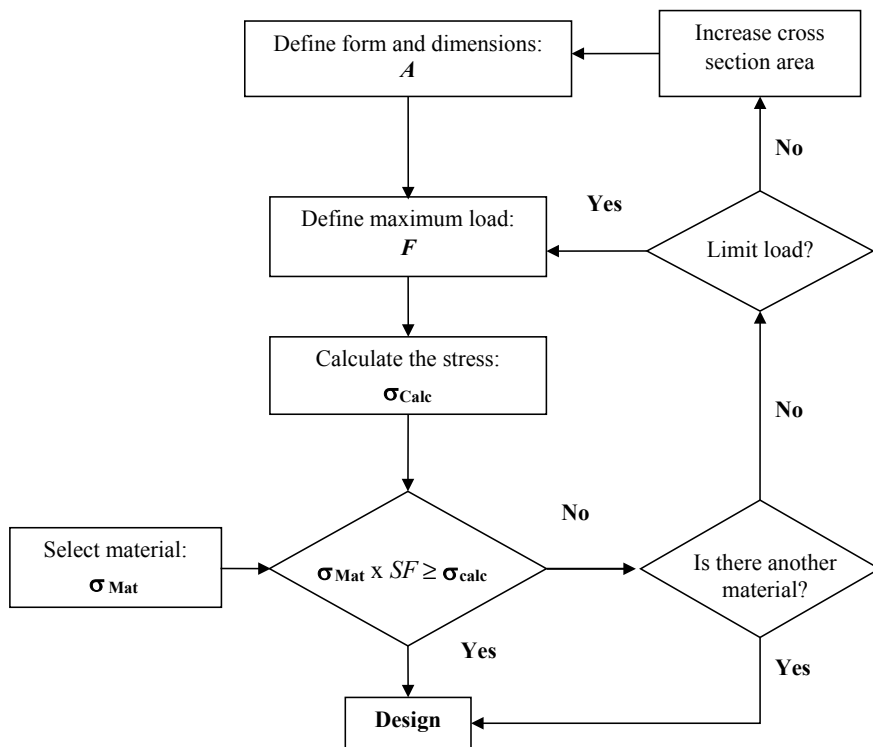


Fig. 1.8 Flow chart of mechanical design based on the definition of stress

According to this equation, any increment or decrement in one variable has to be compensated by adjusting the other two variables. Thus, given a load F , which has to be supported by the component and knowing the material’s mechanical resistance (for example, the yield strength), the cross-section area A that the component needs to have in order to withstand the stress can be calculated. If both A and F are set, then an appropriate material can be selected, being that which strength is greater than the calculated stress, and finally if the material’s strength and A are given, then the maximum applied load can be established. As mentioned before, in many designs, the calculated stress should be below the yield strength, so a safety factor less than 1.0 is applied to the material’s strength. This process is presented in Fig. 1.8.

The fraction of the yield strength is the maximum allowable stress, called *design stress*. The magnitude of the difference between the design stress and the material’s strength is the *safety factor* (SF), which value depends on several factors, being the most important: the load uncertainties, the presence of defects that reduce the strength and the severity of failure consequences.



Example A cylindrical bar has to support a load of 1000 kg. Due to design limitations, the diameter cannot exceed 10 mm. If the fabrication material has a yield strength of 3500 kg/cm², what is the safety factor?

Solution

$$\sigma_{Calc} = F/A = F/(\pi R^2) = (1000 \text{ kg})/(\pi(0.5)^2 \text{ cm}^2) = 1273 \text{ kg/cm}^2$$

$$SF = \sigma_{Calc}/\sigma_{Mat} = 0.363$$

Notice that in the previous example SF is less than 1.0, because it is applied to the materials strength, however, many design codes use safety factors greater than 1.0, so in such cases it is applied to the applied loads or calculated stresses.

1.5 The Stress Tensor

So far, the stress has been calculated by dividing the total internal reaction force by the cross section area, however, internal force is the resultant of the vectorial sum of many force components acting across the unit area elements that compose the total area. If the body is a continuum, this concept leads to the idea that stress may exist in a point. To demonstrate this, consider an area element a_i of very small size, which is under the action of an internal force component f_i , therefore:

$$A = \sum a_i$$

$$F = \sum f_i$$

If the equity $a_i / f_i = F / A$, is true, then it is said that the stress is *uniform*. Furthermore, if the body is a continuum, the following limit exists:

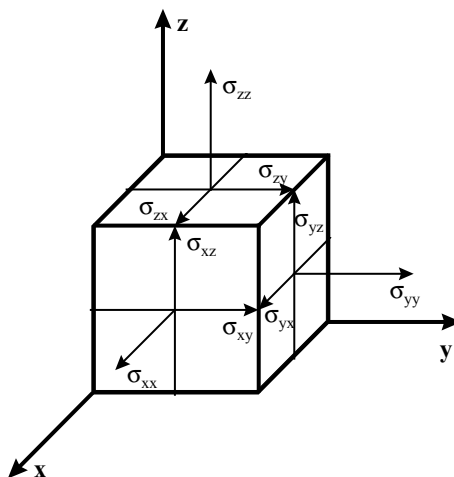
$$\lim_{a_i \rightarrow 0} \frac{f_i}{a_i}$$

And then, the stress can be determined as:

$$\sigma = \frac{dF}{dA}$$

Since dF and dA represent differential quantities (a differential is a value close to zero, but never becomes zero), the previous equation represents the stress in a point, and therefore:

Fig. 1.9 Stress components in a volume element in an orthogonal coordinated system



$$F = \int \sigma dA$$

Cauchy applied this concept to a cube shaped element of differential volume placed in a coordinated Cartesian system (x, y, z) , where the forces are decomposed into three components, each one parallel to the directions x, y and z on each face of the cube, so each face of the cube will have one normal component and two tangential ones. Graphically, the stress components on each cube face would be as shown in Fig. 1.9. The opposite force components that balance the system are not taken into consideration because once the acting components are defined, the opposite are automatically defined, so there is no need to mathematically describe twice as much vectors.

Notice that the following index notation has been introduced, in order to identify the stress components:

σ_{ij} is a stress component,

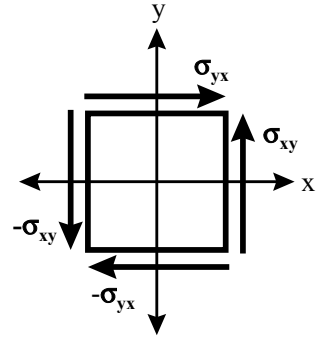
where:

- i is the cube's face
- j is the direction of the force.

As seen in Fig. 1.9 there are nine stress components, three per cube face, and they define the total number of reactions in a static body subject to external loads. Since each component is a vector, by writing the nine components in matrix form, the *stress tensor* is obtained.



Fig. 1.10 Shear stress components acting on the face of a volume element



$$\sigma = \begin{pmatrix} \sigma_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_{yy} & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_{zz} \end{pmatrix}$$

The stress tensor is made up of three normal components (σ_{xx} , σ_{yy} and σ_{zz}) and six shear components. Now, the shear components also create a momentum that has to be balanced as well in order to keep static balance. To determine the magnitude of the momentum, consider a two dimensional body under the action of shear stresses, as shown in Fig. 1.10.

In static conditions, the in-plane momentum balance is:

$$Myx + Mxy = 0$$

where M is the momentum, defined by:

$$Myx = \sigma_{yx} dx dz (dy)$$

In the latter equation, $dx dz$ is the area where force actuates and dy is the leverage length. Similarly:

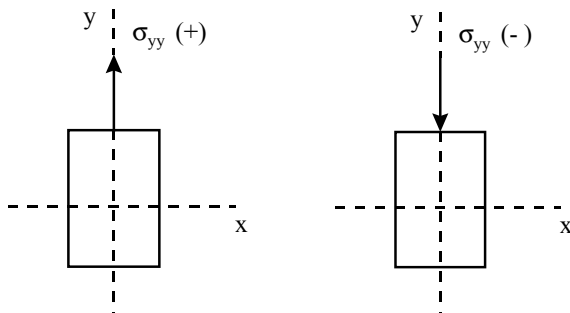
$$Mxy = \sigma_{xy} dy dz (dx)$$

Which leads to:

$$\sigma_{xy} = \sigma_{yx}$$

In general, $\sigma_{ij} = \sigma_{ji}$, therefore, it is said that the stress tensor is *symmetric*. Symmetry reduces the stress tensor to six independent components.

Fig. 1.11 Free body diagrams showing tension and compression stresses



To determine the sign of the stress components the following rule applies:

A force in (+) direction, acting on a (+) face = + stress

A force in (-) direction, acting on a (-) face = + stress

A force in (+) direction, acting on a (-) face = - stress

A force in (-) direction, acting on a (+) face = - stress

The positive faces and directions are those to the right, upwards and downwards and the negative are to the left, backwards and downwards. Figure 1.11 shows examples of this rule.

The combination of non-zero stress components in the stress tensor define the *stress state*. The most common stress states are shown in Table 1.3. Notice that in this table, both zero and non-zero components of stress have been written down, however the common practice is to eliminate the rows and columns that have all zero values.

A simple suggested methodology to qualitatively determine the stress state is as follows:

1. Draw a free body diagram of the component with the applied loads as vectors (arrows with magnitude, direction and point of application), indicating the supports and displacement restrictions.
2. Set a coordinated system of orthogonal axes (x , y , z) at the body's center of gravity, preferably with one of the axes aligned with the body's symmetry axis or parallel to the direction of the main applied load.
3. Draw a cube at the origin of the coordinated system with its edges aligned to the axes.
4. Identify internal reaction forces on the cube faces as vectors of the same magnitude and opposite direction to the external forces.
5. Identify the normal and shear components of the stresses that are generated on the cube's faces, applying the index notation. Bear in mind that the first index is the face where the force acts and the second index is the direction of the force and apply the stress symbol rule.
6. Write down in matrix form the non-zero stress components, the rows are the faces and the columns are the directions; this matrix is the state of stresses.

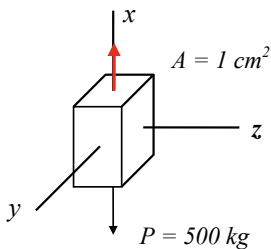
Table 1.3 Common stress states

Name	Stress tensor	Example
Uniaxial tension	$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	A vertical bar with a hanging weight
Simple compression	$\boldsymbol{\sigma} = \begin{pmatrix} -\sigma_{xx} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	Open forging
Biaxial tension	$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & 0 & 0 \\ 0 & \sigma_{yy} & 0 \\ 0 & 0 & 0 \end{pmatrix}$	Hollow cylinder under internal pressure
Pure torsion	$\boldsymbol{\sigma} = \begin{pmatrix} 0 & \tau_{xy} & 0 \\ \tau_{xy} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	Transmission axle
Tension–compression	$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & 0 & 0 \\ 0 & -\sigma_{yy} & 0 \\ 0 & 0 & 0 \end{pmatrix}$	Wire drawing
Plane stress	$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \tau_{xy} & 0 \\ \tau_{xy} & \sigma_{yy} & 0 \\ 0 & 0 & 0 \end{pmatrix}$	Thin plates and free surfaces
Triaxial tension	$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & 0 & 0 \\ 0 & \sigma_{yy} & 0 \\ 0 & 0 & \sigma_{zz} \end{pmatrix}$	Stress concentrators and plane strain
General triaxial stress	$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{xy} & \sigma_{yy} & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \sigma_{zz} \end{pmatrix}$	General case (complex geometry and loads)

The following examples illustrate the above procedure:

Example 1 A square bar supports a tensile load of 500 kg. If the cross section is 1 cm^2 , determine the stress tensor at a point on the center of the bar.

Solution The free body diagram is:



The reactions are:

$$\sigma_{xx} = Px/Ax = 500 \text{ kg}/1 \text{ cm}^2 = 500 \text{ Kg}/\text{cm}^2$$

$$\sigma_{yy} = Py/Ay = 0$$

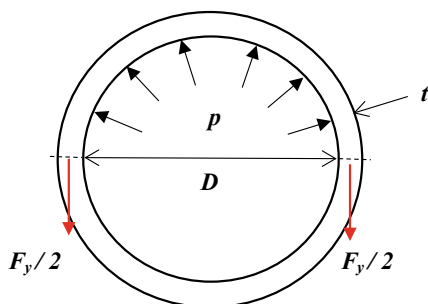
$$\sigma_{zz} = Pz/Az = 0$$

The stress state is:

$$\sigma = \begin{pmatrix} 500 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ kg}/\text{cm}^2$$

Example 2 A thin walled cylinder of length L , wall thickness $t = 1$ cm, and internal diameter $D = 50$ cm, is under internal pressure $p = 50$ kg/cm². Determine the stress state and the acting stresses.

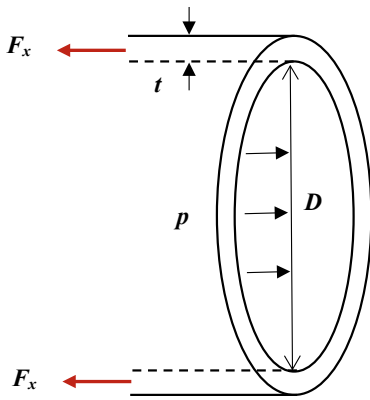
Solution The free body diagram for the circumferential direction y is shown in the figure below. Notice that the internal pressure pushes the half cylinder with a force $F_y = pAc$; by applying the Pascal's principle, Ac can be the mid-section area of the cylinder, which is $Ac = DL$. Therefore, the force induced by the pressure is $F_y = pDL$. Now this force is balanced by two identical forces in each side of the half-cylinder of magnitude $F_y/2$, which acts over a cross-section area equal to tL .



The stress, according to the reactions described above is:

$$\sigma_{yy} = \frac{F_y}{2A_y} = \frac{pDL}{2tL} = \frac{pD}{2t}$$

The free body diagram for the longitudinal direction x is shown in the figure below. Notice that the internal pressure expands the cylinder with a force $F_x = p(\frac{1}{4}\pi D^2)$; the balancing force acts over the annulus cross-section area equal to $\frac{1}{4}\pi [(D + 2t)^2 - D^2]$.



The stress, according to the reactions described above (neglecting the term $4t^2$) is:

$$\sigma_{yy} = \frac{F_x}{A_x} = \frac{\frac{1}{4}p\pi D^2}{\frac{1}{4}\pi [(D+2t)^2 - D^2]} = \frac{pD}{4t}$$

According to the chosen orientation of the coordinate system, the stress tensor is:

$$\sigma = \begin{pmatrix} \sigma_{xx} & 0 \\ 0 & \sigma_{yy} \end{pmatrix}$$

And the stresses are:

$$\sigma_{xx} = (50 \text{ kg/cm}^2)(50 \text{ cm})/2(1 \text{ cm}) = 1250 \text{ kg/cm}^2$$

$$\sigma_{yy} = (50 \text{ kg/cm}^2)(50 \text{ cm})/4(1 \text{ cm}) = 625 \text{ kg/cm}^2$$

$$\sigma = \begin{pmatrix} 625 & 0 \\ 0 & 1250 \end{pmatrix} \text{ kg/cm}^2$$

Deviatoric and hydrostatic component of stress tensor. Many mechanical behavior problems require the subtraction of the hydrostatic components of the stress tensor, because they do not contribute to plastic strain. The hydrostatic component is implicit within the stress tensor, in other words, any stress state can be broken down into a hydrostatic state, and the remaining state is called *deviatoric*. Mathematically the deviatoric component is:

$$\sigma_C = \sigma - \sigma_H$$

where σ_H is the hydrostatic stress and σ_C is the deviatoric component of the stress.

The hydrostatic component is calculated by:

$$\sigma_H = \frac{\sigma_{xx} + \sigma_{yy} + \sigma_{zz}}{3}$$

Therefore, the hydrostatic stress state is:

$$\sigma_H = \begin{pmatrix} \sigma_H & 0 & 0 \\ 0 & \sigma_H & 0 \\ 0 & 0 & \sigma_H \end{pmatrix}$$

Subtracting the hydrostatic stress, the remaining stress tensor is obtained:

$$\begin{aligned} \sigma_C = \sigma - \sigma_H &= \begin{pmatrix} \sigma_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{xy} & \sigma_{yy} & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \sigma_{zz} \end{pmatrix} - \begin{pmatrix} \sigma_H & 0 & 0 \\ 0 & \sigma_H & 0 \\ 0 & 0 & \sigma_H \end{pmatrix} \\ &= \begin{pmatrix} \frac{2\sigma_{xx} - \sigma_{yy} - \sigma_{zz}}{3} & \tau_{xy} & \tau_{yz} \\ \tau_{xy} & \frac{2\sigma_{yy} - \sigma_{xx} - \sigma_{zz}}{3} & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \frac{2\sigma_{zz} - \sigma_{xx} - \sigma_{yy}}{3} \end{pmatrix} \end{aligned}$$

1.6 Non-uniform Stress Distribution

So far, it has been assumed that stress is uniform over the cross-section area, which fortunately, is quite common, but there are also many cases in which the stress is not uniform, this means that it varies along the cross-section of the body. Calculating non-uniform stresses is a complex process that depends on the geometry, the location of the load application points and, obviously the direction of the applied loads. Many times it is not possible to analytically calculate non-uniform stresses, so it is necessary to apply numerical methods or to measure them by experimental techniques. A few cases of the analytical determination of the stress distribution inside a solid can be found in Elasticity and Strength of Materials textbooks; a classic example is the cantilever beam, whose stress analysis is described as follows:

Consider a straight beam with a rectangular cross-section, fixed at one end and with a load F in the other end, as shown in Fig. 1.12. Intuitively, it may be determined that the load will bend the beam, inducing a tension stress on the top of the beam and a compression stress in the lower side. This causes a change from positive to negative in the stress sign, therefore there should be a plane where the stress is zero, this is called *neutral plane*.

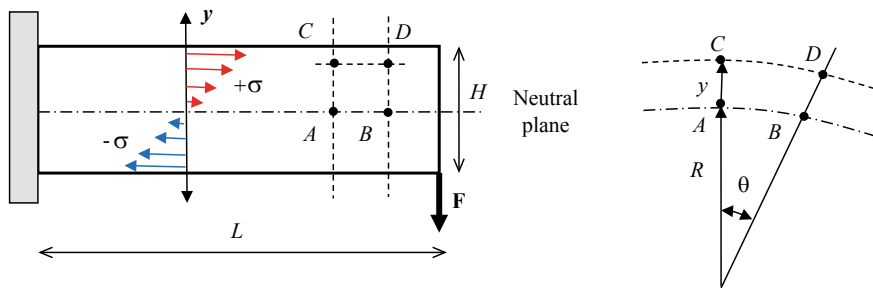


Fig. 1.12 Free body model of a rectangular cantilever beam

The elongation strain on the top fiber of the beam can be calculated from lengths AB , located on the neutral plane and CD located on a fiber a distance y from the neutral plane. The strain ϵ is:

$$\epsilon = \frac{CD - AB}{AB} = \frac{(R + y)\theta - R\theta}{R\theta} = \frac{y}{R}$$

where θ is the angular displacement in radians. If the strain is elastic:

$$\epsilon = \frac{\sigma}{E} = \frac{y}{R}$$

Therefore:

$$\sigma = \frac{E}{R}y$$

Taking a slice of material from the cross-section, with thickness dy and area dA at a distance y from the neutral plane, the force F acting over the slice can be expressed by the following equation:

$$F = \sigma dA = \frac{E}{R}y dA$$

This produces a moment of force in the neutral plane equal to:

$$Fy = \frac{E}{R}y^2 dA$$

Now, the total moment M in the cross-section is:

$$M = \frac{E}{R} \int y^2 dA$$



Notice that E and R come out of the integral because they are constants. The previous integral is termed *momentum of inertia* and is represented by the symbol I , therefore:

$$M = \frac{E}{R} I$$

Bearing in mind that: $\frac{\sigma}{E} = \frac{y}{R}$

It can be written that: $\sigma = \frac{M}{I} y$

Taking $dA = Bdy$, the momentum of inertia of the rectangular beam is:

$$I = \int_{-H/2}^{H/2} y^2 B dy = B \left[\frac{y^3}{3} \right]_{-H/2}^{H/2} = \frac{BH^3}{12}$$

If $M = FL$ and substituting I , the equation for stress at a y distance above the neutral plane is:

$$\sigma = \frac{12FL}{BH^3} y$$

Notice that above the neutral plane, y is positive, so the stress is tensile, and below the neutral plane y is negative and the stress is compressive.

The maximum stress in the rectangular beam can be obtained by substituting $y = H/2$. The result is:

$$\sigma_{max} = \frac{12FL}{BH^3} \left(\frac{H}{2} \right) = \frac{6FL}{BH^2}$$

1.7 Stress Transformation and the Mohr's Circle

The state of stresses not only depends on the loads and body geometry, but also on the orientation of the coordinated axes that are used as a reference to define the stresses. Since coordinate system orientation is arbitrarily selected, the derived stress state shows the components for that particular orientation, but their values in any other direction are unknown. This means that for the same body with the same loads, the stress state is different for each orientation of the coordinate system, in a similar way as the shape of the shadow of an object changes depending on the angle of the light source. For instance, if a cylinder is viewed on the front, its shadow will look like a circle, but if it is viewed on its side, it will look like a rectangle.

The determination of the stress components at different orientations is extremely useful, because in many instances is necessary to determine the values and

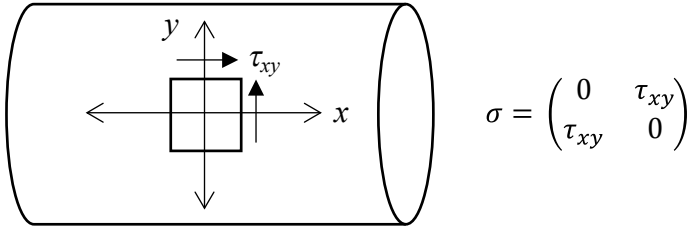


Fig. 1.13 Stress state of a cylindrical bar in pure torsion. The question is: Are there tension and compression stress as well? And what are their values and directions?

orientations of the maximum normal and shear stress, or to determine the stress values at a particular orientation. One example is a cylindrical bar in pure torsion. Obviously, if the x axis of coordinate system is placed along the length of the bar, the stress state is pure shear, and the normal stress components are zero, as shown in Fig. 1.13. The question that arises is: Are there tension and compression stresses as well? And, if yes, what are their values and directions? The following analysis provides the answer.

The schemes of Fig. 1.14 show a body under tensile load, but the coordinate system of the right is inclined with respect to the left figure. In the left figure the x axis is parallel to the applied force P , so the only reaction force produces the stress component σ_{xx} , but if the cross-section plane is inclined, the internal reaction force breaks down into two tension components and two shear components, because the cube faces are now inclined, as shown in Fig. 1.14.

Thus, by changing the orientation of the coordinate system, the stress state changes, because different components can be seen but the resultant force should be the same. In the same way, if the coordinated axes in Fig. 1.13 are inclined 45

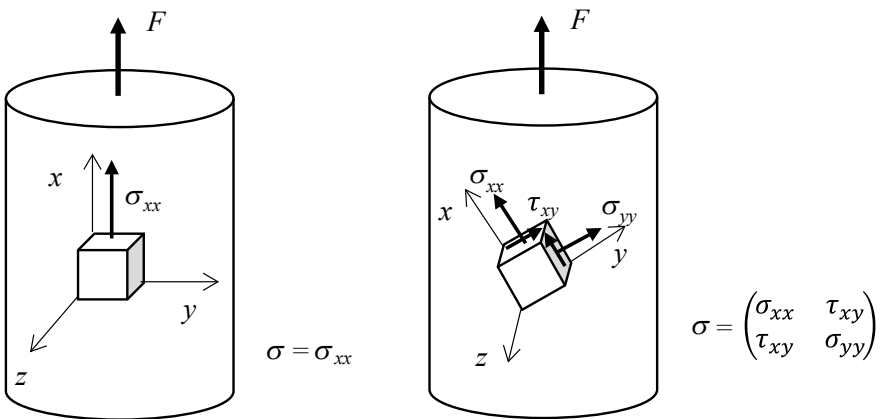


Fig. 1.14 Schematic representation of the stress transformation due to the inclination of the coordinated system. In the left figure, the stress state is uniaxial tension, but when the plane is inclined, the state is plane stress, as seen in the left figure

degrees, it can be seen that a tensile stress and a compressive stress components appear, along with a pair of shear stresses of the same magnitude. The method to calculate a stress state from an orientation to another is called *stress transformation*.

The stress transformation equations in two dimensions, were developed by Cauchy and are:

$$\begin{aligned}\sigma_{x'x'} &= \frac{\sigma_{xx} + \sigma_{yy}}{2} + \left(\frac{\sigma_{xx} - \sigma_{yy}}{2}\right)\cos 2\theta + \tau_{xy}\sin 2\theta \\ \sigma_{y'y'} &= \frac{\sigma_{xx} + \sigma_{yy}}{2} - \left(\frac{\sigma_{xx} - \sigma_{yy}}{2}\right)\cos 2\theta + \tau_{xy}\sin 2\theta \\ \tau_{x'y'} &= \left(\frac{\sigma_{xx} - \sigma_{yy}}{2}\right)\sin 2\theta + \tau_{xy}\cos 2\theta\end{aligned}$$

where the superscript (') indicates the direction of the planes inclined θ degrees. Notice that:

$$\sigma_{xx} + \sigma_{yy} = \sigma_{x'x'} + \sigma_{y'y'}$$

This is known as First Stress Invariant and it applies for three dimensions too.

In 1882, the German civil engineer Christian Otto Mohr published a graphical method to determine the stresses components in directions different from the used to determine an original stress state. Such method is known as “The Mohr’s Circle” and is described next.

By elevating to square power and summing the stress transformation equations for plane stress, the following equation is obtained:

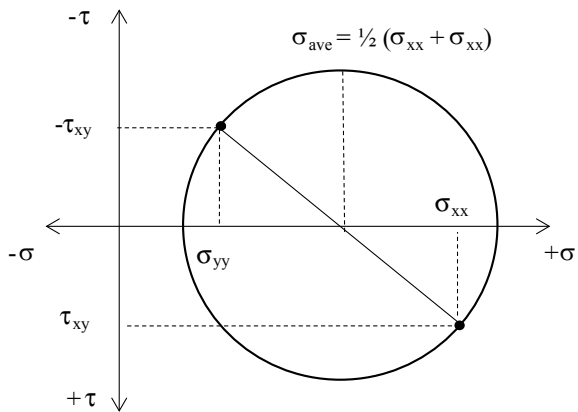
$$\left(\sigma_{x'x'} + \frac{\sigma_{xx} + \sigma_{yy}}{2}\right)^2 + \tau_{x'y'}^2 = \left(\frac{\sigma_{xx} - \sigma_{yy}}{2}\right)^2 + \tau_{xy}^2$$

Notice that this is the equation of a circle: $(x + c)^2 + y^2 = r^2$, when plotted in a coordinated system where the horizontal axis (x) is the normal stress and the vertical axis (y) is the shear stress. The center (c) is the average stress, given by $(\sigma_{xx} + \sigma_{yy})/2$, and the circle’s radius (r) is the right hand side term of the above equation, which correspond to the maximum shear stress. The coordinate points on the circle’s perimeter are the stresses component when the plane is rotated by 2θ degrees.

The guidelines to construct the Mohr’s Circle are:

1. Draw a set of coordinated axes. The abscissas (horizontal axe) are the normal stresses and the ordinates (vertical axe) are the shear stresses. The scale of both axes has to be the same. The positive normal stresses are plotted to the right and the positive shear stresses are plotted downward.
2. Locate the first point of coordinates (σ_{xx}, τ_{xy}) . To locate the second point, the sign of the shear stress is changed $(\sigma_{yy}, -\tau_{xy})$.
3. The two points are joined by a straight line, which the diameter of the circle. The diameter crosscut the normal stress axe at the average normal stress, which is the center of the circle.

Fig. 1.15 Schematic representation of the construction of the Mohr's Circle



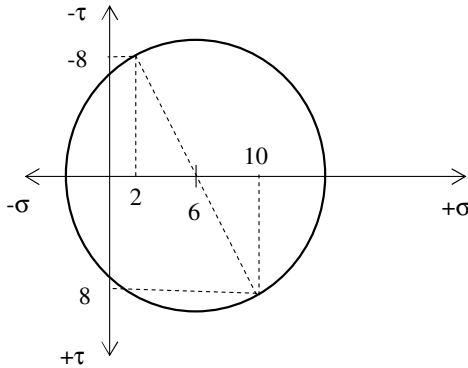
4. Draw a circle, with a ratio equal to half the length of the diameter.
5. The circle's ratio is the maximum shear stress (τ_{\max}).
6. The points at the intersection of the circle with the normal stress axis correspond to the orientations where the shear stress is zero; these stresses are called *principal stresses*.
7. It is important to bear in mind that the rotations in the Mohr's Circle are twice as much their value in the real space (2θ).

Figure 1.15 shows a Mohr's Circle constructed by this method. The stress transformation properties derived from the Mohr's Circle are:

- (a) The maximum shear stress is always at 90° (45° in real space), counter clockwise from the maximum principal stress.
- (b) The minimum principal stress is always at 180° (90° in real space) from the maximum principal stress.
- (c) The maximum shear stress is equal to the maximum principal stress minus the minimum principal stress, divided by two.

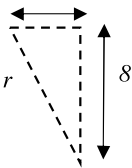
Example Be the stress tensor: $\sigma = \begin{pmatrix} 10 & 8 \\ 8 & 2 \end{pmatrix}$

The initial points of the circle are (10, 8) and (2, -8), the center is at $(10 + 2) / 2 = 6$ us, the circle is.



The ratio is the maximum shear stress and can be calculated using the Pythagoras theorem.

$$10 - 6 = 4$$



$$r = \sqrt{4^2 + 8^2} = 8.944$$

Thus, the maximum shear stress is $\tau_{xy} = 8.944$

The angle θ is given by:

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{\tau_{xy}}{\sigma_{xx} - \sigma_{ave}} \right) = \frac{1}{2} \tan^{-1} \left(\frac{8}{4} \right) = 31.7^\circ$$

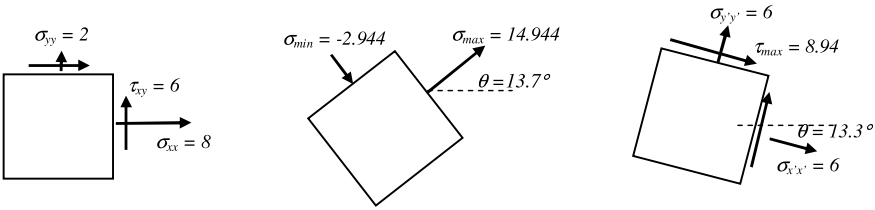
The maximum principal stress is located by adding one ratio to the center, that is:

$$\sigma_{\max} = \sigma_{ave} + r = 6 + 8.944 = 14.944$$

And the minimum principal stress is the center minus one ratio:

$$\sigma_{\min} = \sigma_{ave} - r = 6 - 8.944 = -2.944$$

The next figure shows the physical interpretation of this result:



Another interesting application of the Mohr's Circle is to visualize the effect of the addition of a new stress component. For example, normally a concrete column is under uniaxial compression, but in an earthquake, shear stresses are introduced. The effect is shown in the Mohr's Circle of Fig. 1.16, where it may be seen that the addition of a shear stress makes a tensile stress to appear and the compression stress increases, thus the column behaves as if it were carrying more weight and with a new tension stress. If the earthquake is very strong, the column may buckle due to the high compressive stress and the tension stress may be high enough to cause fractures, since concrete does not resist tension stresses.

Three-Dimension Mohr Circle: The three-dimension Mohr Circle can be easily constructed if the principal stresses are known. The construction consists of drawing a circle between each pair of principal stress values, obtaining three circles, as shown in Fig. 1.17. The shaded region in the larger circle is where the stress components are located when rotating the coordinate axes, but it is not possible to locate them graphically, thus the three-dimension Mohr Circle is used only to observe the behavior of the principal stresses and the maximum shear stress.

As it was explained earlier, the principal stresses correspond to a set of orientations called *principal directions*, additionally, it is known that the principal stresses and the original stress matrix correspond to the same stress content in a

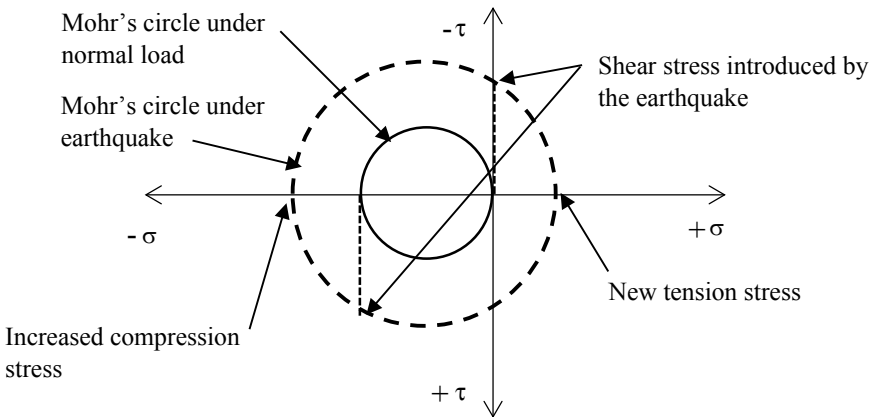
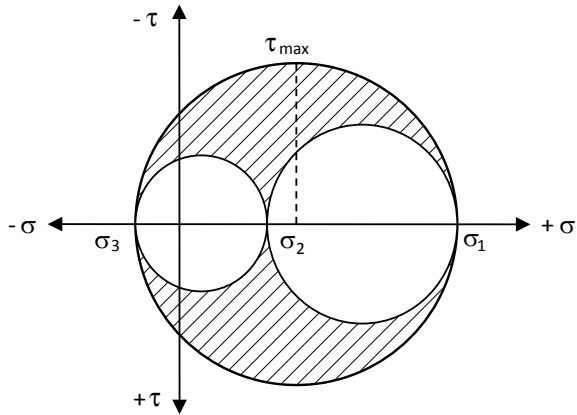


Fig. 1.16 Effect of the introduction of a shear stress in a column working in compression

Fig. 1.17 Mohr Circle in three dimensions. Notice that the maximum shear stress is half-way between the σ_1 and σ_3



body, therefore, the value of the stress tensor has to be the same, regardless the orientation. Mathematically, the determinant represents a tensor value, thus, the following equation can be written:

$$\det|\boldsymbol{\sigma}^P| - \det|\boldsymbol{\sigma}| = \det|\boldsymbol{\sigma}^P - \boldsymbol{\sigma}| = 0$$

where $\boldsymbol{\sigma}^P$ is the principal stress tensor and $\boldsymbol{\sigma}$ is the original stress tensor. The determinant is:

$$\det \begin{vmatrix} \sigma - \sigma_{xx} & -\tau_{xy} & -\tau_{xz} \\ -\tau_{xy} & \sigma - \sigma_{yy} & -\tau_{yz} \\ -\tau_{xz} & -\tau_{yz} & \sigma - \sigma_{zz} \end{vmatrix} = 0$$

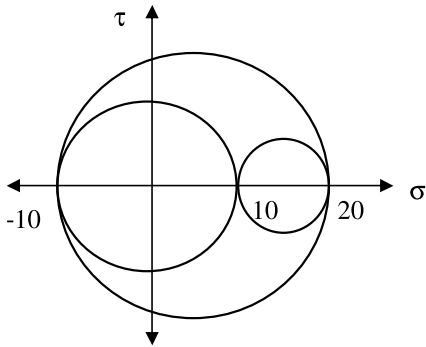
The calculation of this determinant results in the following cubic equation:

$$\begin{aligned} \sigma^3 - (\sigma_{xx} + \sigma_{yy} + \sigma_{zz})\sigma^2 + (\sigma_{xx}\sigma_{yy} + \sigma_{yy}\sigma_{zz} + \sigma_{zz}\sigma_{xx} - \tau_{xy}^2 - \tau_{yz}^2)\sigma \\ - \sigma_{xx}\sigma_{yy}\sigma_{zz} + \tau_{xy}\tau_{yz}\tau_{xz} - \sigma_{xx}\tau_{yz}^2 - \sigma_{yy}\tau_{xz}^2 - \sigma_{zz}\tau_{xy}^2 = 0 \end{aligned}$$

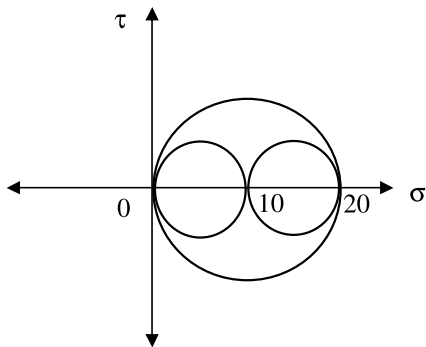
The three roots of this equation are the principal stresses. It is interesting to mention that the coefficients of this equation are the three invariants of the stress tensor, and their value is constant, regardless of the orientation.

Examples Construct a Mohr Circle for each case:

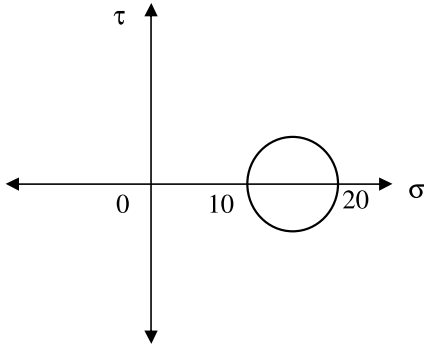
$$\boldsymbol{\sigma} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & -10 \end{pmatrix}$$



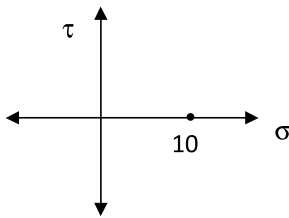
$$\sigma = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$



$$\sigma = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 20 \end{pmatrix}$$



$$\sigma = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$



Example Find the principal stresses of the following state of stress:

$$\sigma = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -6 \end{pmatrix}$$

Applying the determinant equation:

$$\det \begin{vmatrix} \sigma - 10 & -2 & 0 \\ -2 & \sigma - 4 & 0 \\ 0 & 0 & \sigma + 6 \end{vmatrix} = 0$$

Developing:

$$(\sigma + 6)[(\sigma - 10)(\sigma - 4) - (2)(2)] = 0$$

The first root is $(\sigma + 6) = 6$; $\sigma^P = -6$. The other part of the equation is:

$$(\sigma - 10)(\sigma - 4) - (2)^2 = \sigma^2 - 14\sigma + 36 = 0$$

Solving the quadratic equation:

$$\sigma^P = \frac{-(-14) \mp \sqrt{(-14)^2 - 4(1)(36)}}{2}$$

The principal stresses are:

$$\sigma_1 = 10.6$$

$$\sigma_2 = 3.4$$

$$\sigma_3 = -6.0$$

Three-Dimension Stress Transformation. The transformation of the stress tensor in three dimensions is done by the directing cosines method, which consists of the following: Determine the angles between the original coordinated system xyz and the new orientation $x' y' z'$, as shown in Fig. 1.18.

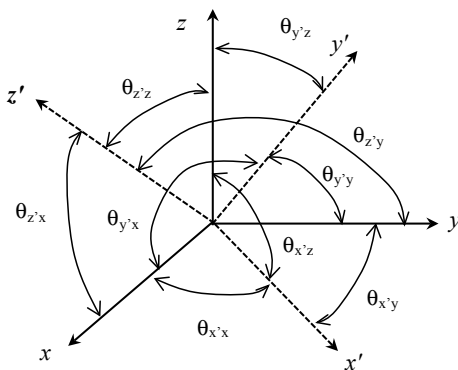
The director cosines matrix is:

$$a_{ij} = \begin{pmatrix} \cos\theta_{x'x} & \cos\theta_{x'y} & \cos\theta_{x'z} \\ \cos\theta_{y'x} & \cos\theta_{y'y} & \cos\theta_{y'z} \\ \cos\theta_{z'x} & \cos\theta_{z'y} & \cos\theta_{z'z} \end{pmatrix}$$

The transformation equation is:

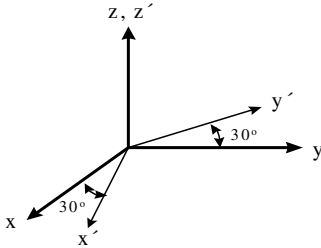
$$\sigma_{i'j'} = a_{i'j} \sigma a_{ji'}$$

Fig. 1.18 Angles to obtain director cosines



The following example illustrate the use of this equation:

Example A coordinate system is rotated 30° with respect to the z axe, as shown in the figure below:



The stress tensor is:

$$\sigma = \begin{pmatrix} 10 & 5 & 0 \\ 5 & -10 & 2 \\ 0 & 2 & 5 \end{pmatrix}$$

The directing cosines matrix is:

$$a = \begin{pmatrix} \cos 30 & \cos 60 & \cos 90 \\ \cos 120 & \cos 30 & \cos 90 \\ \cos 90 & \cos 90 & \cos 0 \end{pmatrix} = \begin{pmatrix} 0.866 & 0.5 & 0 \\ -0.5 & 0.866 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The next step is to multiply the cosine matrixes, but it should be taken into consideration that matrix multiplication is not commutative, so the order of matrices must be the same when doing the multiplication.

The first multiplication is:

$$a \sigma = \begin{pmatrix} 0.866 & 0.5 & 0 \\ -0.5 & 0.866 & 0 \\ 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} 10 & 5 & 0 \\ 5 & -10 & 2 \\ 0 & 2 & 5 \end{pmatrix} = \begin{pmatrix} 11.16 & -0.67 & 1 \\ -0.67 & -11.66 & 1.732 \\ 0 & 2 & 5 \end{pmatrix}$$

And the second multiplication is:

$$\begin{aligned} a \sigma a^T &= \begin{pmatrix} 11.16 & -0.67 & 1 \\ -0.67 & -11.66 & 1.732 \\ 0 & 2 & 5 \end{pmatrix} * \begin{pmatrix} 0.866 & -0.5 & 0 \\ 0.5 & 0.866 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 9.32 & -6.16 & 1 \\ -6.16 & -9.32 & 1.73 \\ 1 & 1.73 & 5 \end{pmatrix} \end{aligned}$$

1.8 Yield Criteria

Up to this point, it is known that when a material reaches its yield strength in uniaxial tension, the plastic strain it will begin, however, the common condition is that the material is under a combined stresses state, different from uniaxial tension, so it is necessary to determine under what combination of stresses the material will yield. The equation to determine if there is yield under a combined state of stress is called *yield criterion*. The yield criterion has to be a stress tensor invariant because it should not depend on the orientation, additionally, it must discount the hydrostatic stress, because they do not produce plastic deformation. The two most widespread yield criteria are Tresca's and Von Mises's, which are described as follows:

Tresca's criterion: Henri Edouard Tresca was a mechanical engineer and professor of the Conservatoire des Arts et Metiers of France. His studies on plasticity led him to develop the first yield criterion in 1860. He was an outstanding engineer, who created the Pattern Metro and his name is among those who contributed to the edification of the Eiffel Tower.

Tresca's criterion, also known as *maximum shear stress*, states that plastic strain initiates when the maximum shear stress exceeds a critical value. For any state of stress, the maximum shear stress is given by the difference between the principal maximum and the minimum principal stresses, as the shown by the Mohr's Circle, thus:

$$\tau_{\text{máx}} = (\sigma_1 - \sigma_3)/2 = k$$

To find the value of k , the previous equation is applied to uniaxial tension, where the maximum principal stress is:

$$\sigma_1 = \sigma_0$$

The other principal stresses are:

$$\sigma_2 = \sigma_3 = 0$$

Then:

$$\tau_{\text{máx}} \sigma_0 / 2 = k$$

Therefore, Tresca's criterion equation is:

$$\sigma_1 - \sigma_3 = \sigma_0$$

Von Mises' criterion: Richard Von Mises was an outstanding Austro-Hungarian scientist, graduated from The University of Vienna. He was director of the Mathematics Institute at the University of Berlin. Due to his Jewish

background, he had to immigrate to Turkey in 1933 and then to The United States, where he became professor at Harvard University. In 1913, he developed the distortion energy theory, which goes by his name.

The von Mises criterion states that yielding initiates when the value of the effective stress, given by the second stress tensor invariant, exceeds a critical value given by k^2 . The effective stress, in terms of the principal stresses, is given by:

$$1/6[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2] = k^2$$

Once again, k is calculated from the uniaxial tension case, where the principal stresses are:

$$\sigma_1 = \sigma_0$$

$$\sigma_2 = \sigma_3 = 0$$

Substituting:

$$\sigma_0 = \sqrt{3} k$$

Which leads to the von Mises's criterion equation:

$$\sigma_0 = \frac{1}{\sqrt{2}} \sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2}$$

The von Mises's criterion can also be expressed in terms of the stress components as:

$$\sigma_0 = \frac{1}{\sqrt{2}} \sqrt{(\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{xx} - \sigma_{zz})^2 + (\sigma_{yy} - \sigma_{zz})^2 + 6(\tau_{xy}^2 + \tau_{xz}^2 + \tau_{yz}^2)}$$

In order to verify the accuracy of the yield criteria, the combined stress tests are used, being the most common the tension-torsion, which consists in stretching and twisting a cylindrical bar.

The main stresses are:

$$\sigma_1 = \sigma_{xx}/2 + (\sigma_{xx}^2/4 + \tau_{xy}^2)^{1/2}$$

$$\sigma_2 = 0$$

$$\sigma_3 = \sigma_{xx}/2 - (\sigma_{xx}^2/4 + \tau_{xy}^2)^{1/2}$$

According to the Tresca criterion:

$$(\sigma_{xx}/\sigma_0)^2 + 4(\tau_{xy}/\sigma_0)^2 = 1$$

According to the von Mises criterion:

$$(\sigma_{xx}/\sigma_0)^2 + 3(\tau_{xy}/\sigma_0)^2 = 1$$

By plotting and comparing the above equations with experimental data, the results, schematically shown in Fig. 1.19, show that von Mises's criterion is more accurate than Tresca's, but the Tresca's criterion gives conservative values, that is, it predicts that yielding will occur before it actually does, which provides a safety margin.

Yield criteria are of great practical relevance, since they allow to know whether a material will plastically deform under a combination of stresses, by using only the yield strength in uniaxial tension as the only experimental required value, which, by the way is broadly available. The yield criteria also allows to determine the yield strength required to prevent plastic deformation. The most important applications of the yield criteria are:

1. **In design:** they are used to determine the maximum allowable stresses to prevent plastic distortion, to select the proper material, or to set a specification for the yield strength.
2. **In forming processes:** such as rolling, forging, extrusion, stamping, etcetera, where yield criteria are applied in order to determine the required work-loads, and to select the most effective combinations of stresses.

The following examples illustrate some of these applications.

Example 1 A cylindrical pressure vessel has to withstand a pressure $p = 50 \text{ kg/cm}^2$. A sheet of $t = 5 \text{ mm}$ thickness is available for fabrication. If the diameter is $D = 100 \text{ cm}$. What should be the minimum yield strength of the fabrication material?

Solution The principal stresses in the vessel are:

$$\sigma_1 = \frac{pD}{2t}$$

$$\sigma_2 = 1/2 \sigma_1$$

$$\sigma_3 = 0$$

Applying the Tresca criterion:

$$\sigma_1 = \sigma_o$$

Therefore:

$$\sigma_o = \frac{50 \text{ kg/cm}^2 \times 100 \text{ cm}}{2 \times 0.5 \text{ cm}} = 5000 \text{ kg/cm}^2$$

By the Von Mises criterion:

$$\sigma_o = \sqrt{\sigma_1^2 + \sigma_2^2 - \sigma_1 \sigma_2}$$

But because $\sigma_2 = \frac{1}{2} \sigma_1$:

$$\sigma_o = \frac{\sqrt{3}}{2} \sigma_1 = \frac{\sqrt{3} p D}{2 \cdot 2t} = 0.866 \times 5000 \frac{\text{kg}}{\text{cm}^2} = 4330 \frac{\text{kg}}{\text{cm}^2}$$

Notice that the von Mises criterion requires a lower yield strength.

Example 2 If in the previous problem a steel plate with $\sigma_o = 3500 \text{ kg/cm}^2$ is used. What should be the minimum thickness to prevent plastic deformation?

Solution The Tresca equation for yield of a cylindrical vessel is:

$$\sigma_1 = \frac{pD}{2t}$$

Solving for t:

$$t = \frac{pD}{2\sigma_o}$$

Substituting data, the required thickness by Tresca is: 7.14 mm.
The Tresca equation for yield of a cylindrical vessel is:

$$\sigma_o = \frac{\sqrt{3} p D}{4 t}$$

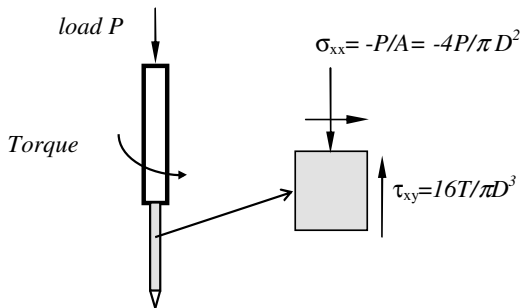
Solving for t:

$$t = \frac{\sqrt{3} p D}{4 \sigma_o}$$

Substituting data, the required thickness by von Mises is: 6.18 mm.

Example 3 A screwdriver should withstand a compressive load of $p = 180$ lbf, a torque $T = 3$ lb-in. One design has a diameter $d = \frac{1}{4}$ in. Determine the minimum required yield strength, considering a safety factor of 2.0.

Solution The free body diagram of the screw driver is shown in the next figure. Notice that the stress state is compression and shear.



The von Mises's criterion is used, to avoid calculating the principal stresses. For compression-torsion, the von Mises equation is reduced to:

$$\sigma_0 = \frac{1}{\sqrt{2}} \sqrt{2\sigma_{xx}^2 + 6\tau_{xy}^2}$$

$$\sigma_{xx} = -\frac{4 \times 180 \text{ lbf}}{\pi(0.25)^2 \text{ in}^2} = 3667 \text{ psi}$$

$$\tau_{xy} = \frac{16 \times 3 \text{ lb-in}}{\pi(0.25)^3 \text{ in}^3} = 6607 \text{ psi}$$

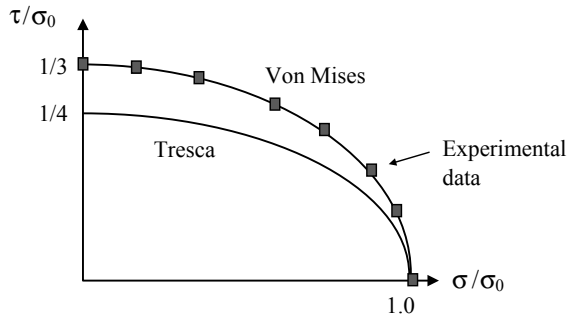
Substituting data, the result is 12,016 psi, but since a safety factor of 2.0 is applied, the minimum required yield strength is: 24,032 psi.

As mentioned before, one use of the yield criteria is to analyze the stress states that give the greater strength or under which a material can be easily deformed. The yield maps are graphic representations of the yield criteria in plane stress, that allow to locate a specific combination of principal stress and visually determine if the deformation is going to be plastic or elastic.

Plane stress yield maps are useful for analyzing stress combinations in cases such as:

- Thin sheets and free surfaces
- Tension-torsion combinations

Fig. 1.19 Experimental data of a yield tension-torsion test as compared to predicted values of the Tresca and Von Mises yield criteria



- Tension-compression combinations
- Thin walled pressure vessels and pipes.

In plane stress, there is always a principal stress equal to zero, so to avoid confusion by the notation convention that states that the maximum principal stress is represented by σ_1 , the intermediate as σ_2 and the minimum or most negative as σ_3 , the principal stresses directions in the yield map are represented as x and y . To construct a Von Mises yield map, it is necessary to write the equation for two dimensions (elevated to the square).

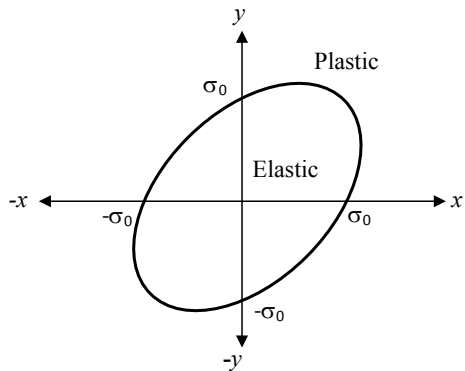
$$\sigma_1^2 + \sigma_2^2 - \sigma_1\sigma_2 = \sigma_0^2$$

This equation is equivalent to:

$$x^2 + y^2 - xy = c^2$$

Which is the equation of an inclined ellipse, that is depicted in Fig. 1.20. Thus, every stress combination falling within the ellipse does not cause yield, while the points out of it mean plastic deformation.

Fig. 1.20 Von Mises yield map for plane stress



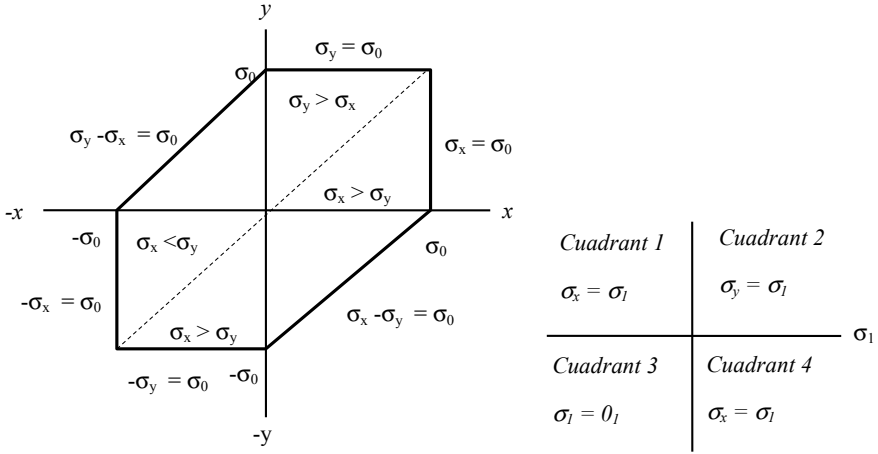


Fig. 1.21 Tresca yield map for plane stress

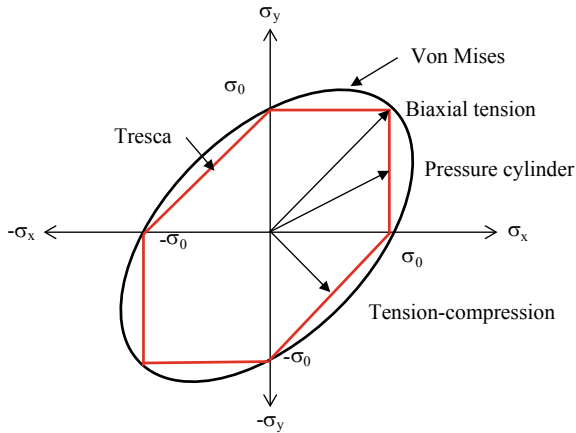


Fig. 1.22 Overlapped Tresca and Von Mises yield maps for plane stress

The plane stress yield map for Tresca’s criterion is shown in Fig. 1.21. It consists of a series of straight lines whose trajectory depends on the quadrant where the stress combination defines which direction has the maximum and minimum principal stresses. By overlapping both maps, the yield criteria can be compared, as shown in Fig. 1.22. In this figure it can be seen that for uniaxial and biaxial tension, both criteria coincide, while in a thin walled pressurized cylinder, the von Mises criterion predicts yield at a value of $\sigma_1 = 1.154 \sigma_0$, whereas Tresca predicts that yield will occur when $\sigma_1 = \sigma_0$. The tension-compression load, located in the second and fourth quadrants, has the shortest load trajectory. According to Tresca, yielding occurs when $\sigma_1 = 0.5\sigma_0$, while Von Mises predicts that $\sigma_1 = 0.577\sigma_0$.



1.9 Stress Concentration

According to the definition of stress, a reduction in the cross-section area means an increment of the stress, in an amount proportional to the area reduction. However, the sharp changes of cross-section, as well as discontinuities, such as holes, grooves, gouges, etcetera, increase the stress beyond the magnitude given by the reduction of area. This phenomenon is called *stress concentration* and it is of great technological importance, because the local stress may be several times greater than the calculated nominal stress, so the component may fail at those locations. The expression of the stress increment due to the stress concentration is the *stress concentration factor*, which is defined by the symbol K_T , its equation is:

$$K_T = \frac{\sigma_{\max}}{\sigma_{app}}$$

where σ_{\max} is the maximum stress value at the root or border of the stress concentrator, and σ_{app} is the applied or nominal stress. One of the first stress concentration factors was obtained by Inglis in 1913, corresponding to a thin sheet with an elliptic hole in the center, as shown in Fig. 1.23.

Inglis found that the maximum stress is given by:

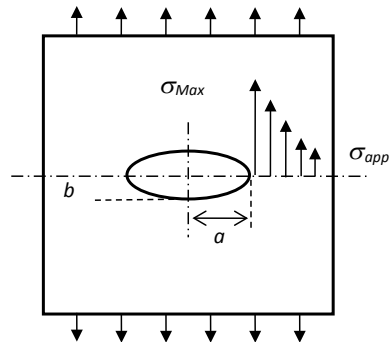
$$\sigma_{\max} = \sigma_{app} \left(1 + 2 \left(\frac{a}{b} \right) \right)$$

Thus:

$$K_T = 1 + 2 \left(\frac{a}{b} \right)$$

Notice that for $a = b$, the hole is circular and $K_T = 3$. Notice as well, that if the ellipse is too sharp ($a \gg b$), the K_T increases to infinite, which actually does not happen because, before that the material will reach its ultimate tensile strength and will break. When discontinuities are very acute, such as cracks, the stress

Fig. 1.23 Inglis's model of a thin sheet with an elliptic hole under uniform stress



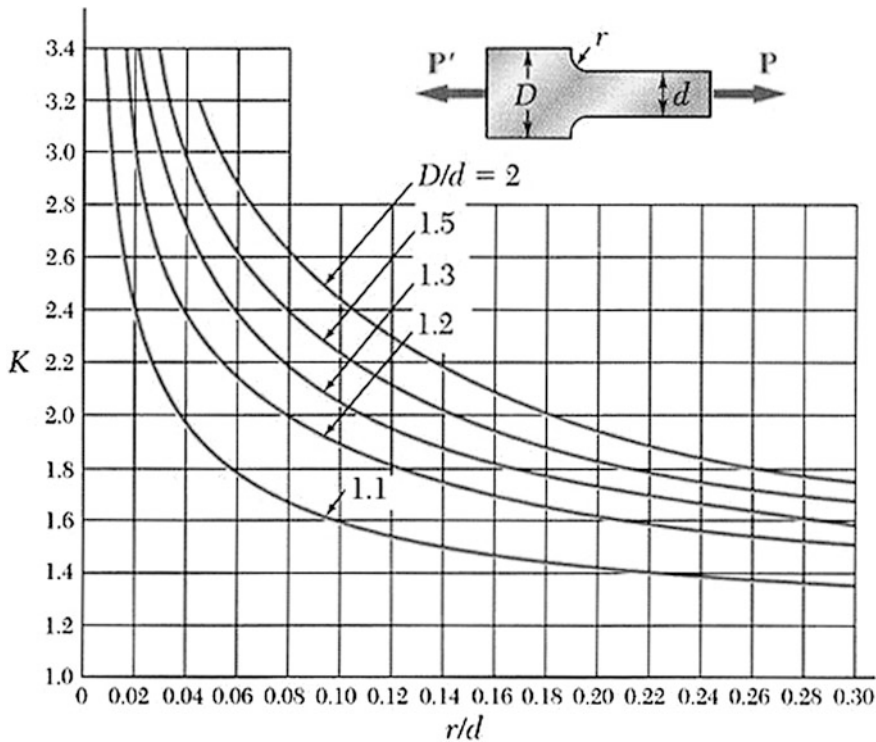


Fig. 1.24 Graph of K_T for a round rod with a diameter reduction. Image taken from Beer, Johnston and DeWolf, Mechanics of Materials 4th Edition, McGraw-Hill, 2006

concentration loses meaning because the mathematical singularity that arises at the tip of the defect (since its ratio is close to zero) undetermines the value of K_T , in this case fracture mechanics is applied to determine the stress values at the crack tip.

In practice, the K_T values are found in charts, like the one shown in Fig. 1.24, which are published in strength of materials textbooks, in mechanical design manuals, as well as the internet. Additionally, there are several low cost applications for mobile devices that provide the stress concentration factors for a wide variety of geometries and load conditions.

In Fig. 1.24 can be noticed that the stress concentration magnitude has a stronger dependency on the sharpness of the stress concentrator, i.e. the radius (r), than on the magnitude of the cross-section area reduction; for instance, a reduction of the diameter (D/d) from 1.1 to 1.5 at $r/d = 0.22$, increases K_T from 1.4, to 1.9, that is about 36%, but if r/d drops to 0.02 at $D/d = 1.1$, K_T is 2.4, that is 71% increment.

Furthermore, the effect of stress concentrators is more severe in brittle materials than in ductile ones, this is because the plastic deformation increases the radius of the stress concentrator, and additionally relaxes the stresses, reducing the value of K_T . This effect is used in glass cutting, which actually is not cutting but breaking, where a sharp blade

makes a deep score on the surface of the glass, then the glass plate is slightly bent and the stress concentration induced by the score overcomes the tensile strength and the glass breaks. Nonetheless, this does not mean that stress concentration may not be a problem in ductile materials, it simply means that it is less dangerous.

Chapter 2

Strain



Abstract This chapter begins with the definition of strain from the continuum mechanics point of view, the description of the strain tensor and the strain transformation concept; then, the elastic stress strain relations, namely the constitutive equations, are described for both isotropic and anisotropic behavior. Next is the description of plasticity and plastic behavior of metallic materials, including tables of typical mechanical properties of engineering materials, the following topics are the Baushingher effect, the concepts of real stress and real strain, and the plastic stress-strain equations. A brief introduction of the numeric analysis of stress and strain is presented, along with a description of the procedure to perform an analysis of mechanical behavior by the Finite Element Method and a brief description of the experimental methods of electric resistance extensometry and polaroscopy to measure stress and strain. At the end of this chapter an introduction to the hardness concept and the standard methods to measure it is presented.

2.1 Definition of Strain

In mechanics of materials, strain is defined as the change of shape in a body due to the action of stresses. Such change in shape means the movement of the particles that constitute the solid body. This movement can be divided into three components: (1) rigid body displacement, (2) rotation, and (3) change of relative position within the body. The vectors that describe the change of shape are: u , v , and w , for the x , y and z directions respectively, as shown in Fig. 2.1. The change of relative position within the body is the only one that produces a change in shape, also called *distortion*, because the other two just mean a change of position in the space but without changing the shape. The mathematic analysis of strain requires to define a vector that describes the distortion only, and therefore the rotation and displacement components should be eliminated.

To calculate the elongation strain, also called *normal strain*, consider a thin bar fixed at one end, and being pulled with a force F_x at the opposite end. As result, the segment AB in the bar will elongate to a final length A'B', as shown in Fig. 2.2.

Fig. 2.1 Vectors describing the change of position of a volume element in a deformed solid body. The dotted line is the initial shape and the solid line is the final deformed geometry

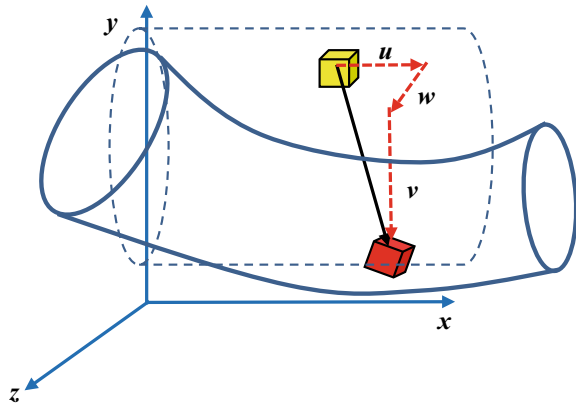
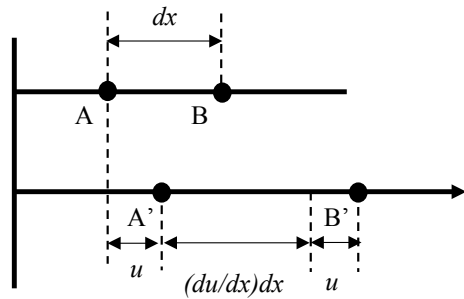


Fig. 2.2 Diagram for the determination of elongation strain of a differential segment of initial length dx that is stretched an amount u along the x direction



The point A' will be displaced a distance u , from its original position, while the point B will be displaced a distance u (because it is being “pushed” by point A), plus the elongation of the segment dx . If the elongation ratio of the segment AB is du/dx , then the total separation between A' and B' will be.

$$dx + \frac{du}{dx} dx$$

If the normal strain is defined as the relative elongation, calculated as the change of length divided by the initial length, the mathematic expression to determine the strain of a segment dx in direction x is:

$$\epsilon_{xx} = \frac{A'B' - AB}{AB} = \frac{dx + \frac{\partial u}{\partial x} dx - dx}{dx}$$

Which results in:

$$\epsilon_{xx} = \frac{du}{dx}$$

Notice that the strain is a vector because it has magnitude and direction. To identify the strain components the following index notation is introduced:

- ϵ_{ij} = strain component
- i = initial direction of the segment
- j = displacement direction

In continuum mechanics it is reasonable to assume that elongation in one direction is independent from elongation in other directions, thus by analogy: if v is the displacement of a segment dy in the y direction and w is the displacement of a segment dz in the z direction, the other components of normal strain are:

$$\epsilon_{yy} = \frac{dv}{dy}; \quad \epsilon_{zz} = \frac{dw}{dz}$$

Shear strain, in the other hand, is defined as the change of an initially straight angle at an edge of a cubic volume element, as shown in Fig. 2.3, where there is a change of shape, but the side dimensions remain constant.

Mathematically, shear strain, identified by the symbol γ_{ij} , is defined by the tangent function of the change of straight angle, thus:

$$\gamma_{xy} = \tan \theta_{xy} = \frac{du}{dy}$$

Notice that shear strain may come along with a rigid body rotation component, as shown in Fig. 2.4, but when $\theta_{xy} = -\theta_{yx}$ there is only rigid body rotation, and there is no shear strain. A simple method to eliminate the rigid body rotation component, is to take the average of the rotation angles.

Fig. 2.3 Definition of shear strain. The dotted figure is the initial shape and the solid line is the shape after shear strain. Notice that the side dimensions do not change

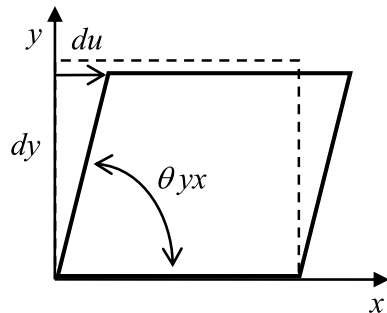
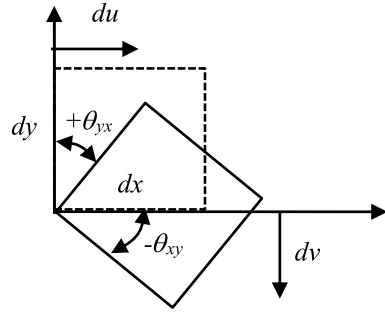


Fig. 2.4 Schematic illustration that a rigid body rotation does not produce shear strain, owing that $\theta_{xy} = -\theta_{yx}$



The shear strains are the tangent of the change of straight angle θ for the x and y edges; thus, they are:

$$\tan \theta_{xy} = \frac{dv}{dx}$$

$$\tan \theta_{yx} = \frac{du}{dy}$$

Taking the average, the shear strain without rigid body rotation is calculated as:

$$\gamma_{xy} = \frac{1}{2} (\tan \theta_{xy} + \tan \theta_{yx}) = \frac{1}{2} \left(\frac{dv}{dx} + \frac{du}{dy} \right)$$

As the order of the addition does not alter the result, it is easy to demonstrate that $\gamma_{xy} = \gamma_{yx}$, which means that the shear strain components are symmetric. Thus, the three independent shear strain components are:

$$\gamma_{xy} = \frac{1}{2} \left(\frac{du}{dy} + \frac{dv}{dx} \right); \gamma_{xz} = \frac{1}{2} \left(\frac{dw}{dx} + \frac{du}{dz} \right); \gamma_{yz} = \frac{1}{2} \left(\frac{dv}{dz} + \frac{dw}{dy} \right)$$

Therefore, the strain is a tensor of nine components, out of which, six are independent, three normal and three of shear. The matrix form of the strain tensor is:

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{xx} & \gamma_{xy} & \gamma_{xz} \\ \gamma_{yx} & \varepsilon_{yy} & \gamma_{yz} \\ \gamma_{zx} & \gamma_{zy} & \varepsilon_{zz} \end{pmatrix}$$

The strain tensor has the same properties as the stress tensor, which are:

(a) The transformation law (the same as the stress transformation law):

$$\varepsilon' = \underline{a}\varepsilon\underline{a}^T$$

(b) The strain tensor has three invariants:

$$I_1 = \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}$$

$$I_2 = (\epsilon_{xy})^2 + (\epsilon_{xz})^2 + (\epsilon_{yz})^2 - \epsilon_{xx}\epsilon_{yy} - \epsilon_{xx}\epsilon_{zz} - \epsilon_{yy}\epsilon_{zz}$$

$$I_3 = \det(\epsilon_{ij})$$

(c) The strain components after rotation of the reference axes are calculated by the Mohr's circle.

The last property means that, just like stresses, strains also produces tensors different from the original one when they are seen from a different angle. This can be physically demonstrated by taking a rectangular piece of latex foil divided into square elements with orientation parallel to the x and y edges (this will be the foil 1) and another identical latex foil divided into square elements oriented at 45° with respect to the x and y edges (this will be the foil 2); then both foils are strained as shown in Fig. 2.5. The Mohr's circle of strains, shows that the principal strains are at 45° from the maximum shear strain, so the squares in foil 1 show the maximum shear strain, but the elements in foil 2 do not show shear strain.

The following example illustrates the determination of the strain tensor and its transformation by the Mohr's circle:

Example The strain tensor of an initially square-shaped sheet, of 1 cm length per side is:

$$\epsilon = \begin{pmatrix} 0.1 & 0.27 \\ 0.27 & -0.2 \end{pmatrix}$$

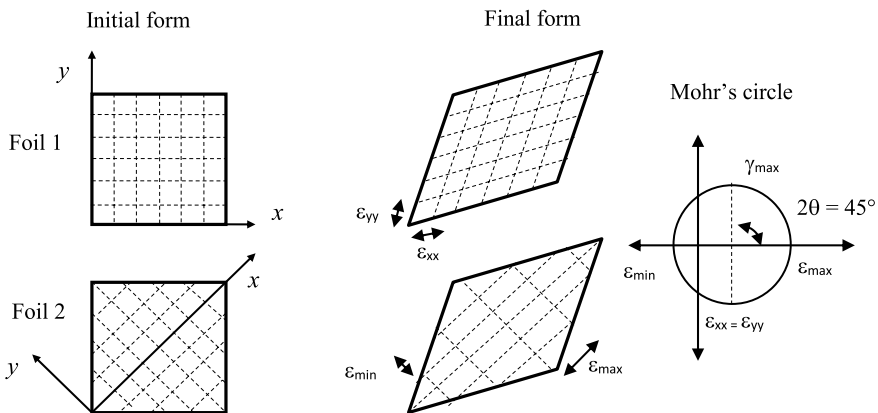


Fig. 2.5 Physical demonstration of the strain transformation



According to the definition of elongation strain, and assuming uniform strain:

$$\varepsilon_{xx} = du/dx = \Delta u/x_o$$

$$\Delta u = x_f - x_o$$

Thus:

$$\varepsilon_{xx} = (x_f - x_o)/x_o = (x_f/x_o) - 1$$

$$x_f = x_o(\varepsilon_{xx} + 1)$$

Substituting:

$$x_f = 1.0(0.1 + 1) = 1.1 \text{ cm}$$

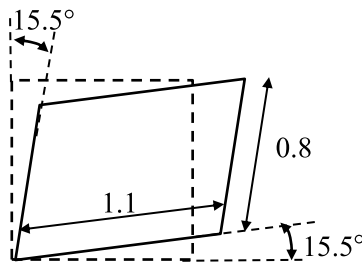
Similarly:

$$y_f = y_o(\varepsilon_{yy} + 1) = 1.0(-0.2 + 1) = 0.8 \text{ cm}$$

The shear strain is the change of straight angle, so:

$$\theta_{xy} = \arctan(0.27) = 15.5^\circ$$

The final shape of a volume element with these strains is:



To construct the Mohr circle, the center is:

$$C = \frac{\varepsilon_{xx} + \varepsilon_{yy}}{2} = \frac{0.1 + (-0.2)}{2} = -0.05$$

The ratio is the maximum shear strain:

$$\gamma_{\max} = \sqrt{(\varepsilon_{xx} - C)^2 + \varepsilon_{xy}^2}$$

$$\gamma_{\max} = \sqrt{0.05^2 + 0.27^2} = 0.275$$

The principal strains are:

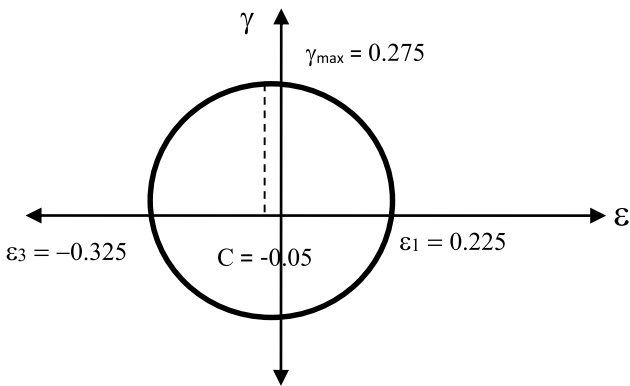
$$\epsilon_1 = C + \gamma_{\max} = -0.05 + 0.275 = 0.225$$

$$\epsilon_3 = C - \gamma_{\max} = -0.05 - 0.275 = -0.325$$

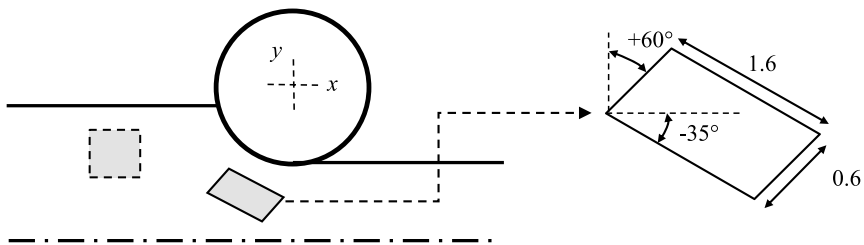
The orientation of ϵ_1 is:

$$2\theta = \tan^{-1}\left(\frac{0.27}{0.15}\right) = 61^\circ$$

The Mohr's circle is:



Example An initially square volume element of unit dimensions is strained at the middle of a rolling pass as shown in the figure below:



Determine the strain tensor of the shaded volume element.

According to the definition of elongation strain:

$$\epsilon_{xx} = (x_f - x_o) / x_o = (1.6 - 1.0) / 1.0 = 0.60$$

$$\epsilon_{yy} = (y_f - y_o) / y_o = (0.6 - 1.0) / 1.0 = -0.40$$

The shear strain is:

$$\gamma_{xy} = 1/2 [\tan \theta_{xy} + \tan \theta_{yx}] = 1/2 [\tan(-35^\circ) + \tan(+65^\circ)] = 0.72$$

The strain tensor is:

$$\boldsymbol{\varepsilon} = \begin{pmatrix} 0.60 & 0.72 \\ 0.72 & -0.40 \end{pmatrix}$$

To construct the Mohr's circle, the center is:

$$C = \frac{\varepsilon_{xx} + \varepsilon_{yy}}{2} = \frac{0.6 + (-0.4)}{2} = +0.10$$

The maximum shear strain is:

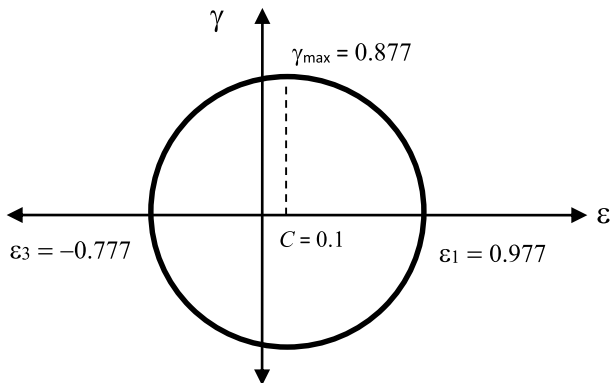
$$\gamma_{\max} = \sqrt{(\varepsilon_{xx} - C)^2 + \varepsilon_{xy}^2} = \sqrt{0.5^2 + 0.72^2} = 0.877$$

The principal strains are:

$$\varepsilon_1 = C + \gamma_{\max} = 0.10 + 0.877 = 0.977$$

$$\varepsilon_3 = C - \gamma_{\max} = 0.10 - 0.877 = -0.777$$

The Mohr's circle is:



2.2 Elastic Stress-Strain Relations

Once stresses and strains have been defined, the next step is to know how much strain is produced by a stress component. If the material is linear-elastic, under uniaxial tension, the stress-strain relation is the Hooke's law:

$$\sigma = E \varepsilon$$

Since the density of most engineering solid materials does not change when they are deformed, as the material elongates, it must become thinner to maintain the volume constant, therefore, a tension stress, say σ_{xx} will produce three strain components: one elongation in the direction of the stress (ε_{xx}) and two of contraction in the other two directions (ε_{yy} , ε_{zz}), as shown in Fig. 2.6.

If the solid is isotropic the lateral contraction is the same in the transverse direction, thus $\varepsilon_{yy} = \varepsilon_{zz}$, and therefore, the ratio $\varepsilon_{yy}/\varepsilon_{xx}$ is constant. This is known as *Poisson's Ratio*, identified by the symbol ν . The Poisson's Ratio is defined as the negative of the lateral transverse strain divided into the elongation strain:

$$\nu = -\frac{\varepsilon_{\text{transverse}}}{\varepsilon_{\text{elongation}}}$$

In the case of shear strains, each shear stress produces only its own corresponding shear strain component, so the Hooke's law for shear strain is:

$$\tau = G \gamma$$

where G is the shear modulus. These three elastic constants for an isotropic material are related by the following equation:

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

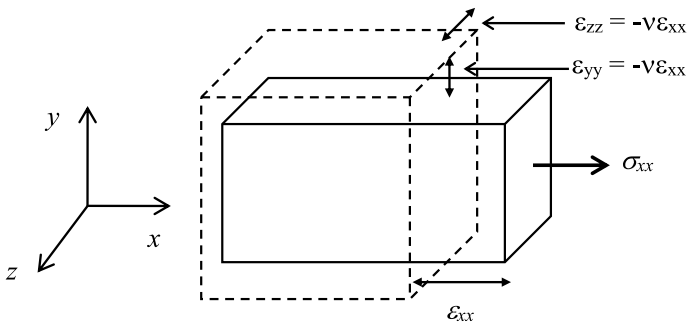


Fig. 2.6 Strain components produced by a uniaxial tension stress

Table 2.1 Elastic constants for isotropic materials

Material	E GPa (10^6 psi)	G GPa (10^6 psi)	ν
Iron	211.4 (30.6)	81.6 (11.8)	0.293
Aluminum	70.3 (10.2)	26.1 (3.8)	0.345
Cooper	129.8 (18.8)	48.3 (7.0)	0.343
Nickel	199.5 (28.9)	76.0 (11.0)	0.3
Magnesium	44.7 (6.5)	17.3 (2.5)	0.291
Gold	78.0 (11.3)	27.0 (3.9)	0.44
Silver	82.7 (12.0)	30.3 (4.4)	0.367
Alumina (Al_2O_3)	415 (60)	–	–
Tungsten carbide (WC)	534.4 (77.5)	219 (31.8)	0.22
Diamond	965 (140)	–	–
Glass	80.1 (11.6)	31.5 (4.6)	0.27
Polycarbonate	2.4 (0.35)	–	–
Polyethylene	0.4 a 1.3	–	–

Therefore, it is said that an isotropic material has two independent elastic constants, while the third one can be calculated by the other two. The typical values of the elastic constants for some engineering materials are shown in Table 2.1.

In Table 2.1, it can be observed that the values of E and G for metals and ceramics are very large (in the order of 10^9 Pa or 10^6 psi), which indicates that these materials are very stiff, whereas the polymers exhibit the lowest values of E and G , so they are able to go through great elastic strains at rather low stress levels, being the most flexible materials. Finally, it is important to mention that the elastic constants vary little for the same material group, regardless the chemical composition and heat treatment, so they are usually reported for material groups rather than specific materials, such as the yield strength. However the elastic constants are highly dependent on temperature, especially for polymers, so care has to be taken on using the corresponding value for the service temperature.

The elastic stress-strain relations for isotropic materials are obtained by applying the *superposition principle*, which states that the strains produced in the same direction can be algebraically summed in order to obtain the total strain.

According to Hooke's law, the magnitude of the strain ε_{xx} produced by the stress σ_{xx} is:

$$\varepsilon_{xx} = \sigma_{xx}/E$$

The transverse strain component produced by the stress σ_{xx} is obtained by the Poisson's ratio:

$$\varepsilon_{yy} = \varepsilon_{zz} = -\nu\varepsilon_{xx} = -\frac{\nu\sigma_{xx}}{E}$$

Table 2.2 Strain components produced by the normal components of stress

Stress/Strain	ϵ_{xx}	ϵ_{yy}	ϵ_{zz}
σ_{xx}	$\frac{\sigma_{xx}}{E}$	$-v \frac{\sigma_{xx}}{E}$	$-v \frac{\sigma_{xx}}{E}$
σ_{yy}	$-v \frac{\sigma_{yy}}{E}$	$\frac{\sigma_{yy}}{E}$	$-v \frac{\sigma_{yy}}{E}$
σ_{zz}	$-v \frac{\sigma_{zz}}{E}$	$\frac{\sigma_{yy}}{E}$	$\frac{\sigma_{zz}}{E}$

Table 2.2 shows the strain components produced by each normal stress component, determined the previous procedure:

By applying the superposition principle, the strain components in each direction (column in the Table 2.2) are summed up to obtain the total strain. The shear strains as obtained for each corresponding Hooke's law for shear $\tau_{ij} = G \gamma_{ij}$. The result is the following set of equations called *Constitutive Equations*.

$\epsilon_{xx} = \frac{1}{E} [\sigma_{xx} - v(\sigma_{yy} + \sigma_{zz})]$	$\epsilon_{xy} = \frac{\sigma_{xy}}{G}$
$\epsilon_{yy} = \frac{1}{E} [\sigma_{yy} - v(\sigma_{xx} + \sigma_{zz})]$	$\epsilon_{xz} = \frac{\sigma_{xz}}{G}$
$\epsilon_{zz} = \frac{1}{E} [\sigma_{zz} - v(\sigma_{xx} + \sigma_{yy})]$	$\epsilon_{yz} = \frac{\sigma_{yz}}{G}$

The following example illustrates the use of the Constitutive Equations.

Example A thin sheet is under the stresses: $\sigma_{xx} = 1400 \text{ kg/cm}^2$, $\sigma_{zz} = 0$, $\tau_{xy} = 500 \text{ kg/cm}^2$ and $\epsilon_{zz} = -0.0003$. Calculate the rest of the stresses and strains if: $E = 20 \times 10^5 \text{ kg/cm}^2$ and $\nu = 0.3$.

Solution The Constitutive Equations for normal strain are:

$$\text{Ec.1} \quad \epsilon_{xx} = E^{-1} [\sigma_{xx} - v(\sigma_{yy} + \sigma_{zz})] = (20 \times 10^5)^{-1} [1400 - 0.3(\sigma_{yy})]$$

$$\text{Ec.2} \quad \epsilon_{yy} = E^{-1} [\sigma_{yy} - v(\sigma_{xx} + \sigma_{zz})] = (20 \times 10^5)^{-1} [\sigma_{yy} - 0.3(1400)]$$

$$\text{Ec.3} \quad \epsilon_{zz} = E^{-1} [\sigma_{zz} - v(\sigma_{xx} + \sigma_{yy})] = (20 \times 10^5)^{-1} [-0.3(1400 + \sigma_{yy})] = -0.0003$$

From Ec. 3:

$$\sigma_{yy} = (-0.0003 \times 20 \times 10^5) / (-0.3) - (1400) = 600 \text{ kg/cm}^2$$

Substituting $\sigma_{yy} = 600 \text{ kg/cm}^2$ into Ecs. 1 and 2

$$\epsilon_{xx} = (20 \times 10^5)^{-1} [1400 - 0.3(600)] = 0.00061$$

$$\epsilon_{yy} = (20 \times 10^5)^{-1} [600 - 0.3(1400)] = 0.00009$$

The shear modulus is:

$$G = \frac{E}{2(1+\nu)} = \frac{20 \times 10^5}{2(1+0.3)} = 7.69 \times 10^5 \text{ kg/cm}^2$$

Therefore, the shear component is:

$$\gamma_{xy} = \tau_{xy}/G = 500 \text{ kg/cm}^2 / 7.69 \times 10^5 = 0.00065$$

2.3 Constitutive Equations for Non-isotropic Elastic Materials (Elastic Anisotropy)

The anisotropic behavior refers to the variation of properties in different directions, which is quite common in engineering materials, and may have different causes, being the main ones:

- *Crystalline orientation*: Single-crystals are naturally anisotropic because the elastic constants depend on the interplanar distances, coordination number, and etcetera; since these characteristics change with direction, so do the elastic constants. In fact, anisotropy levels are classified according to the level of symmetry of the crystalline structures.
- *Preferential orientation (texture)*: Polycrystalline metals are usually isotropic because the orientation of the grains is random, however, some forming processes like rolling, extrusion and wire drawing develop a preferential crystalline orientation of the grains. This preferential orientation is known as *texture* and it may induce anisotropic behavior.
- *Fibrous or banded microstructure*: Many engineering materials feature, naturally or artificially, an alignment of second phase constituents, forming a so-called banded microstructure. Some other materials are constituted of fibers aligned in preferential directions, like wood and fiber reinforced composites. These materials are clearly anisotropic.

In the fully anisotropic case, theoretically each stress component produces its corresponding strain component, so there is an elastic constant for each stress-strain component pair. Mathematically, the constitutive equation for the fully anisotropic case can be written as:

$$\varepsilon_{ij} = S_{ijkl}\sigma_{kl}$$

Since there are six independent components of stress and six independent components of the strain, the total number of anisotropic elastic constants is 36, as demonstrated by the following matrix equation:

$$\begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ \epsilon_{yz} \\ \epsilon_{xz} \\ \epsilon_{xy} \end{pmatrix} = \begin{pmatrix} S_{xxxx} & S_{xxyy} & S_{xxzz} & S_{xxyz} & S_{xxxz} & S_{xxxy} \\ S_{yyxx} & S_{yyyy} & S_{yyzz} & S_{yyyz} & S_{yyxz} & S_{yyxy} \\ S_{zzxx} & S_{zzyy} & S_{zzzz} & S_{zzyz} & S_{zzxz} & S_{zzxy} \\ S_{yzxx} & S_{yzyy} & S_{yzzz} & S_{yzyz} & S_{yzzx} & S_{yzyx} \\ S_{xzxx} & S_{xzyy} & S_{xzzz} & S_{xzyz} & S_{xzxz} & S_{xzxy} \\ S_{xyxx} & S_{xyyy} & S_{xyzz} & S_{xyyz} & S_{xyxz} & S_{xyxy} \end{pmatrix} * \begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \tau_{yz} \\ \tau_{xz} \\ \tau_{xy} \end{pmatrix}$$

Notice that S_{ijkl} is the elastic constant of the ij strain for a given kl stress. S_{ijkl} is a *Stiffness* elastic constant. Whereas if the stresses are calculated from the strains the elastic constant is named *Compliance*. Therefore:

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl}$$

It is important to point out that the stiffness constants are not the inverse of the compliance constants, as it might be thought of, therefore:

$$S_{ijkl} \neq 1/C_{ijkl}$$

As normal stresses do not produce shear strains and shear stresses do not produce normal strains, the number of constants for the fully anisotropic case is reduced to twelve; as the level of anisotropy is reduced, the number of elastic constants is reduced too. Table 2.3 shows the number of constants for different levels of anisotropy.

To simplify the writing of the anisotropic elastic constants, a short notation was introduced, the equivalences are shown in Table 2.4.

Thus, the stiffness tensor in short notation for orthorhombic materials is:

$$S_{ij} = \begin{pmatrix} S_{11} & S_{12} & S_{13} & 0 & 0 & 0 \\ S_{21} & S_{22} & S_{23} & 0 & 0 & 0 \\ S_{31} & S_{32} & S_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{66} \end{pmatrix}$$

Table 2.3 Number of anisotropic elastic constants by symmetry level

Level	No. of independent constants
Triclinic	12
Monoclinic	12
Orthorhombic	9
Tetragonal	6
Hexagonal	5
Cubic	3
Isotropic	2

Table 2.4 Short index notation of the indexes of anisotropic elastic constants

Indexes	Short notation
xx	1
yy	2
zz	3
yz	4
xz	5
xy	6

Table 2.5 Constants of compliance (C) and of stiffness (S) for common metals

Material	C_{11}	C_{12}	C_{44}	S_{11}	S_{12}	S_{44}				
Cubic	10^{10} Pa			10^{-11} Pa ⁻¹						
Aluminum	10.82	6.13	2.85	1.57	-0.57	3.51				
Cooper	16.84	12.14	7.54	1.50	-0.63	1.33				
Iron	23.70	14.10	11.60	0.80	-0.28	0.86				
Nickel	24.65	14.73	12.47	0.73	-0.27	0.80				
Hexagonal	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	S_{11}	S_{12}	S_{13}	S_{33}	S_{44}
Magnesium	5.97	2.62	2.17	6.17	1.64	2.20	-0.79	-0.5	1.97	6.10
Titanium	16.0	9.0	6.6	18.1	4.65	0.97	-0.47	-0.18	0.69	2.61
Zinc	16.1	3.42	5.01	6.10	3.83	0.84	0.05	-0.73	2.84	2.61

The stiffness (S) and compliance (C) constants of some single-crystal materials are given in Table 2.5.

The use of non-isotropic stress-strain relations is quite simple, as the following example shows:

Example Determine the isotropic and anisotropic strains of a material under the following stress state:

$$\sigma = \begin{pmatrix} 50 & 20 & 0 \\ 20 & -50 & 10 \\ 0 & 10 & 100 \end{pmatrix} \text{ ksi}$$

The isotropic elastic constants are:

$E = 20 \times 10^6$ psi, $\nu = 0.35$ subject to the stress tensor:

The anisotropic constants are:

$$S_{ijkl} = \begin{pmatrix} 5 & 1.5 & 1.5 & 0 & 0 & 0 \\ 1.5 & 5 & 1.5 & 0 & 0 & 0 \\ 1.5 & 1.5 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2.5 \end{pmatrix} * 10^{-5} \text{ ksi}^{-1}$$

Solution

(a) Isotropic case:

$$\varepsilon_{xx} = \frac{1}{E} [\sigma_{xx} - \nu(\sigma_{yy} + \sigma_{zz})] = (20\,000)^{-1} [50 - 0.35(50 + 100)] = 0.001625$$

$$\varepsilon_{yy} = \frac{1}{E} [\sigma_{yy} - \nu(\sigma_{xx} + \sigma_{zz})] = (20\,000)^{-1} [-50 - 0.35(50 - 50)] = -0.005125$$

$$\varepsilon_{zz} = \frac{1}{E} [\sigma_{zz} - \nu(\sigma_{xx} + \sigma_{yy})] = (20\,000)^{-1} [100 - 0.35(50 - 50)] = 0.0005$$

$$G = \frac{E}{2(1+\nu)} = \frac{200\,000}{2(1+0.35)} = 7407 \text{ ksi}$$

$$\gamma_{xy} = \frac{\tau_{xy}}{G} = \frac{20}{7407} = 0.0027$$

$$\gamma_{xz} = \frac{\tau_{xz}}{G} = 0$$

$$\gamma_{yz} = \frac{\tau_{yz}}{G} = \frac{10}{7407} = 0.00135$$

(b) Anisotropic case:

$$\begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{yz} \\ \varepsilon_{xz} \\ \varepsilon_{xy} \end{pmatrix} = \begin{pmatrix} 5 & 1.5 & 1.5 & 0 & 0 & 0 \\ 1.5 & 5 & 1.5 & 0 & 0 & 0 \\ 1.5 & 1.5 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2.5 \end{pmatrix} * 10^{-5} \text{ ksi}^{-1} * \begin{pmatrix} 50 \\ -50 \\ 100 \\ 10 \\ 0 \\ 20 \end{pmatrix} \text{ ksi}$$

Performing operations:

$$\epsilon_{xx} = [5(50) + 1.5(-50) + 1.5(100)] * 10^{-5} = 0.00325$$

$$\epsilon_{yy} = [1.5(50) + 5(-50) + 1.5(100)] * 10^{-5} = -0.0025$$

$$\epsilon_{zz} = [1.5(50) + 1.5(-50) + 5(100)] * 10^{-5} = 0.005$$

$$\gamma_{yz} = 2.5(10) * 10^{-5} = 0.0025$$

$$\gamma_{xz} = 0$$

$$\gamma_{xy} = 2.5(20) * 10^{-5} = 0.005$$

2.4 Plasticity

In the field of Mechanical Behavior of Materials, plasticity is the ability of a material to have a permanent and irreversible change of shape and/or dimensions as a result of the application of stresses, without change of volume. In all materials, plastic deformation produces changes in the internal particle arrangement and in the microstructure, and therefore, affects the mechanical properties, and the subsequent plastic behavior. Plastic strain has three important characteristics which can be explained by the process depicted in Fig. 2.7.

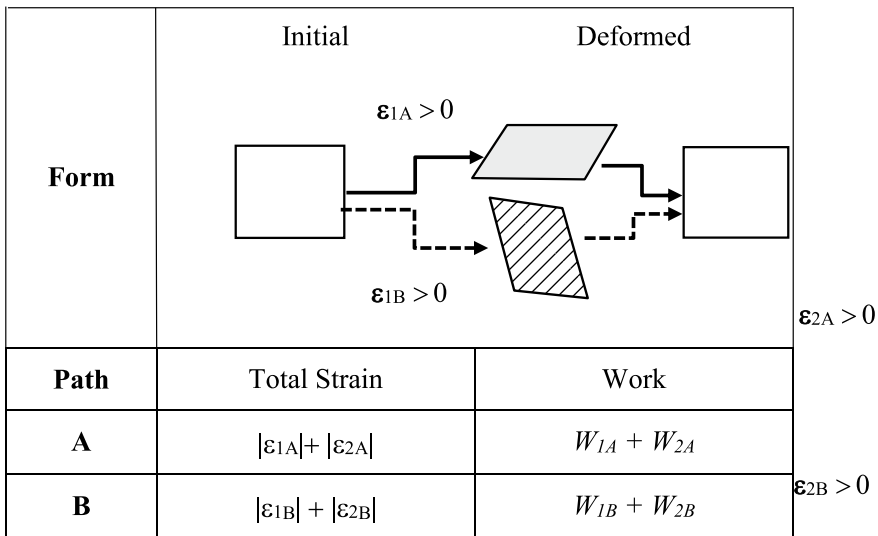


Fig. 2.7 Dependency of plastic strain on the path and the accumulated work

The square on the left is the initial shape, in the deformation path A, indicated by the solid arrow, to go into the intermediate shape, the elongation strain is, say $\epsilon_{1A} = 0.5$, and the shear strain is, say $\gamma_{1A} = 1.0$, and then to go into the final shape, a new elongation strain of, say $\epsilon_{2A} = -0.5$, and a shear strain of, say $\gamma_{1A} = 1.0$ have to be applied; even that the initial and final shapes are identical the total strains are not zero, but are the sum of the absolute values of the strain in every step; in this case $\epsilon_{A\text{total}} = |0.5| + |-0.5| = 1.0$, and $\gamma_{A\text{total}} = |1.0| + |-1.0| = 2.0$. In the other hand, the pass from the initial form to the intermediate one, requires a work W_1 , which produces the strain ϵ_{1A} . Since the system cannot return the introduced work nor elastically recover, an additional work W_2 must be introduced to carry out the strain ϵ_{2A} , so the total spent work will be $W_1 + W_2$. In the deformation path B, indicated by the dotted arrow, obviously the intermediate shape is different to that of path A, therefore the strain components are different to those of path A, let's say: $\epsilon_{B\text{total}} = |-0.5| + |-1.5| = 1.5$, and $\gamma_{B\text{total}} = |-0.5| + |-0.5| = 1.0$; additionally, it is obvious that a different amount of work is required to execute path B, as compared to the work spent in path A.

According to the process above described, the following characteristics of plastic strain can be established:

- (1) *Irreversibility*: A plastically strained body cannot return the work done to deform it, so any change of geometry, even if it returns it to its initial shape, will require additional work. The main causes of irreversibility is the rearrangement of the particles inside the body which, according to the Second Law of Thermodynamics, that states that any system will always go towards a state of maxim entropy, it will not spontaneously return to a more ordered arrangement.
- (2) *Path dependency*: the work required for plastic strain depends on the path followed during the change of shape.
- (3) *Accumulative*: The total strain is the sum of the absolute values of the strains applied in every step to reach a final shape and dimensions.

Since in plastic deformation the volume is constant, the following ratio is true: $\Delta V/V = 0$, now, in a unit size cube element, the volume is $V = dxdydz$, under the strains ϵ_{xx} , ϵ_{yy} , and ϵ_{zz} , the final dimensions of the strained cube are: $(1 + \epsilon_{xx})$, $(1 + \epsilon_{yy})$, and $(1 + \epsilon_{zz})$, thus, the following equation can be written:

$$\Delta V/V = [(1 + \epsilon_{xx})dx(1 + \epsilon_{yy})dy(1 + \epsilon_{zz})dz - dxdydz]/(dxdydz) = 0$$

Simplifying terms:

$$\Delta V/V = (1 + \epsilon_{xx})(1 + \epsilon_{yy})(1 + \epsilon_{zz}) - 1 = 0$$

Neglecting the terms where the strain components multiply to each other, it can be established that in a solid that deforms under constant volume:

$$\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} = 0$$

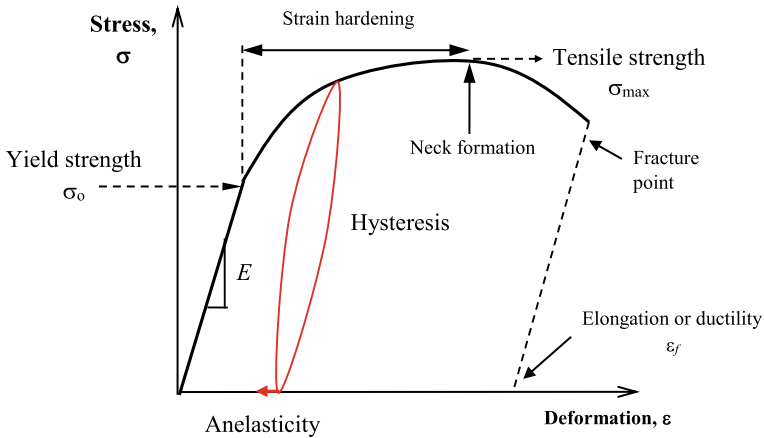


Fig. 2.8 Stress-strain curve in tension showing the main characteristics of plastic strain

This condition is called *strain compatibility* and it implies that normal strains are not independent.

The next important characteristics of plastic behavior of engineering materials are seen in the stress-strain curve in uniaxial tension, which typical form is shown in Fig. 2.8.

Plasticity determine the following characteristics of plastic strain behavior:

1. *Yield strength*. It is the stress value at which the plastic strain begins. The stress-strain curve is no longer linear beyond this point. The yield strength is a material's property.
2. *Elastic-plastic behavior*. It is the condition where plastic strain and elastic strain co-exist. (It must be remembered that the appearance of plasticity does not imply that elasticity is lost).
3. *Strain hardening*. After yield, if it is necessary to increase the stress to continue deforming the material, thus it is say that the material exhibits strain hardening.
4. *Tensile strength*. It is the maximum tension stress that material can withstand. It is a material's property. After reaching the maximum on the stress-strain curve, the stress decreases because the material undergoes to a local contraction called "neck". The neck is responsible for ductile fracture mechanism and also determines the ductility.
5. *Fracture*. It is the culmination of the stress-strain process. Fracture stress is not a material's property, but it depends on other factors such as geometry and the presence of flaws.
6. *Ductility*. It is the maximum normal strain or plastic elongation that the material experiences just before fracture. It is determined by drawing down a straight line parallel to the elasticity line in order to subtract the elastic strain. In practice, it is

determined by joining together the two parts of the fractured specimen and measuring the final length. It is also frequently referred to as the percentage of cross-area reduction.

7. *Plastic hysteresis*. This phenomenon occurs in the plastic regime, if the load is reduced to zero and then the material is reloaded, the stress-strain path in the unload cycle is different to that in the upload stage, forming a hysteresis loop.
8. *Anelasticity*. Some materials show an elastic time-dependent recovery after they have been plastically deformed and then unloaded to zero stress.

Many times the yield point is not well defined on the stress-strain curve, so in order to determine a unambiguous yield strength value it was decided that the design of structures could tolerate a small percentage of plastic strain, so the yield strength may be defined as the stress to reach this small plastic strain. This method, known as “offset”, and it was introduced by the ASTM E8 “Standard Test Method for Tension Testing of Metallic Materials”, where the offset yield strength is defined as the stress where plastic elongation strain is 0.2%. It is determined by the intersection of the stress-strain curve with a straight line, parallel to the elastic line on the stress-strain curve, which starts at 0.2% strain, as shown in Fig. 2.9. When yield strength is obtained by the offset method, its symbol is $\sigma_{0.2\%}$.

The stress-strain curve in uniaxial tension is also used to classify the engineering materials from the mechanical behavior point of view, this classification is shown schematically in Fig. 2.10.

The characteristics of the materials classified according to this criterion are:

- *Hard and brittle*. They feature a high yield strength, low strain hardening and low ductility. The Young’s modulus of these materials is often very high. These materials exhibit high hardness and stiffness as well, but are brittle, so they do not resist impact loads nor high strains. Typical hard and brittle materials are ceramics and glass.
- *Soft and weak*. They feature a low yield strength, low strain hardening and high ductility, so they exhibit high formability. The Young’s modulus may be low, but not necessarily. Typical soft and weak materials are lead and tin alloys.

Fig. 2.9 Determination of the 0.2% offset yield strength

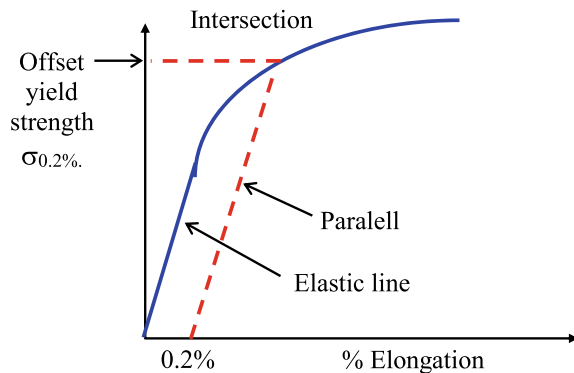
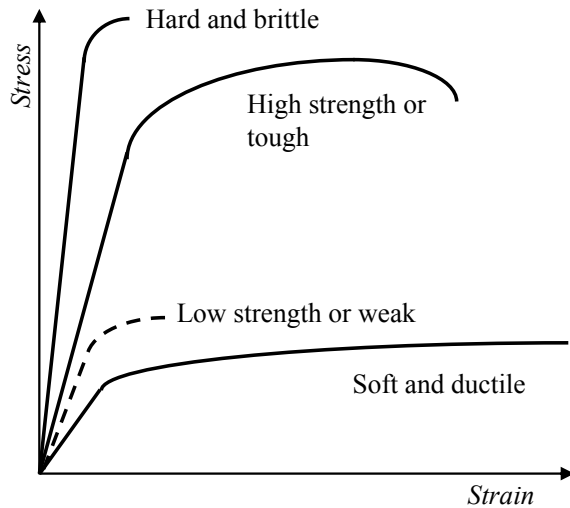


Fig. 2.10 Classification of materials according to their stress-strain curve in uniaxial tension



- *High strength materials or tough.* They feature high yield strength, large strain hardening and moderate to high ductility. The Young's modulus is preferably high. These materials resist heavy loads, strong impacts and absorb a great deal of work before fracturing. Typical tough materials are steel and nickel alloys.
- *Low strength materials or weak.* They feature low mechanical strength and moderated to low ductility. The Young's modulus is usually low. These materials have very low load bearing capabilities and do not resist impacts, they wear and quickly deteriorate in service. Typical weak materials are polymer foams.

Hard and brittle materials are used in cutting tools and wherever high abrasive wear and erosion resistance are required, however, their brittleness makes them easy to get damaged by impacts, and they are not resistant to flexion and thermal shock. Soft and ductile materials are used to fabricate components of intricate forms or that require great plastic deformations, but its low strength limits its application to low load bearing service conditions. To combine the best qualities of the aforementioned materials, scientists and engineers have spent a great deal of effort to combine strength and ductility, so they can fabricate components that may be shaped easily, but once in service can withstand severe work conditions, such as: impacts, extreme loads, shock waves and wear, these are the so called *high strength materials*. Finally, weak materials have applications in one-use components, such as disposable cups, low durability items and as stuffing.

The typical tensile properties of some engineering materials are shown in Table 2.6. It is observed that the strongest materials are the ferrous metals. An interesting material is titanium, which features a mechanical strength similar to low alloy steels, but its high corrosion resistance and low density make it an excellent choice for applications in high-precision machinery and aero spatial equipment. Non-ferrous alloys feature intermediate to low mechanical strengths; nonetheless,

Table 2.6 Tension properties of some common engineering materials

Group	Material	σ_0 (MPa)	σ_{max} (MPa)	% <i>Al</i>
Ferrous metals	Low C steel	250–400	340–580	40–80
	Medium C steel	300–900	400–1200	12–92
	High C steel	400–1200	550–1640	27–92
	Low alloy steel	400–1100	460–1200	14–100
	Stainless steel	170–1000	480–2240	62–180
	Cast iron	215–790	350–1000	10–35
Nonferrous alloys	Aluminum	30–500	60–550	22–35
	Cooper	30–500	100–550	30–90
	Nickel	70–1100	345–1200	80–110
	Titanium	250–1250	300–1600	14–120
	Zinc	80–450	130–520	10–100
Composites	Glass fiber	110–192	130–240	7–23
	Carbon fiber	550–1050	500–1050	6.1–88
Polymers	Natural rubber	20–30	22–32	0.15–0.25
	Neoprene	3–24	3–24	0.1–0.3
	Elastomers (PU)	25–50	25–50	0.2–0.4
	Polycarbonate	60–70	60–70	2.1–4.6
	Polietilene	18–29	21–45	1.44–1.72
	PCV	35–52	40–65	1.46–5.12
	Flexible foams	0.01–3	0.01–3	0.005–0.09
	Rigid foams	0.3–12	0.5–12	0.002–0.91
Ceramics	Alumina	690–5500 ^a	350–665	3.3–4.8
	SiC	1000–5250 ^a	370–680	2.5–5
	WC	3350–6800 ^a	370–550	2–3.8
Glass	Borosilicate	260–380 ^a	22–32	0.5–0.7
	Silica	1100–1600 ^a	45–152	0.6–0.8
	Soda glass	360–420 ^a	30–55	0.55–0.7
Rocks	Brick	50–140 ^a	7–14	1–2
	Concrete	32–60 ^a	2–6	0.35–0.45
	Rock	34–298 ^a	5–17	0.7–1.5
Bio-materials	Wood	30–70	60–100	5–9
	Leather	5–10	20–26	3–5
	Natural fibers		6–200	5–15

^aThe yield strength in this materials is measured in compression, since its tensile strength is very low

they usually have other physical properties such as high electrical conductivity, high thermal conductivity and low density, which make them a good choice for wire conductors, domestic appliances, electrical and illumination devices and light weight vehicles. Polymers have the lowest mechanical strength levels, although their low density, great chemical stability and easy of manufacture makes them ideal candidates

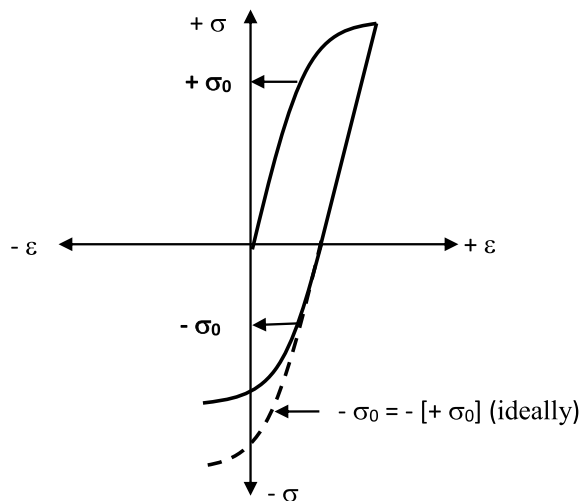
for applications in components, parts and structures that have low load requirements such as household appliances, office equipment, furniture and parts of light machines and body parts of transportation vehicles. Ceramics and glass are highly resistant to compression and exhibit high hardness, but above all, they are chemically inert, which makes them ideal for the fabrication of process equipment, vessels and piping for corrosive substances, biologically sterile utensils and food processing equipment, as well as for high wear resistance applications.

Since the beginning of civilization Rocks have an unmovable place in the construction industry, mainly due to their high availability, high durability, high resistance to compressive stresses, and low cost. Biomaterials used to be neglected by modern engineers, but in recent years their use has gained terrain as they are a good option for manufacturing temporary use and disposable components, because they are inexpensive, easy to work and their environmental impact is minimum. An interesting case is the come back use of vegetal fibers to manufacture airplane seats, temporary shelters, clothes and gear for personal protection due to the already mentioned advantages.

The Bauschinger effect. A phenomenon related to the path dependency of plastic strain, mainly observed in metals, is the *Bauschinger effect*, which is the reduction of the yield strength when a material is loaded in the opposite direction of a previous load that produced certain amount of plastic strain. For example, if the material is plastically deformed in tension, and then is immediately loaded in compression, the yield strength would be smaller than if it had been strained directly in compression. The Bauschinger effect can be observed in the stress-strain curve, as depicted in Fig. 2.11.

Some shaping processes take advantage of the Bauschinger effect to increase its efficiency, one example is bar cold-straightening, where the bars are bent in sequential opposite directions, to save energy and increase the process rate by the reduced yield strength resulting from the Bauschinger effect.

Fig. 2.11 Bauschinger effect in a stress-strain record



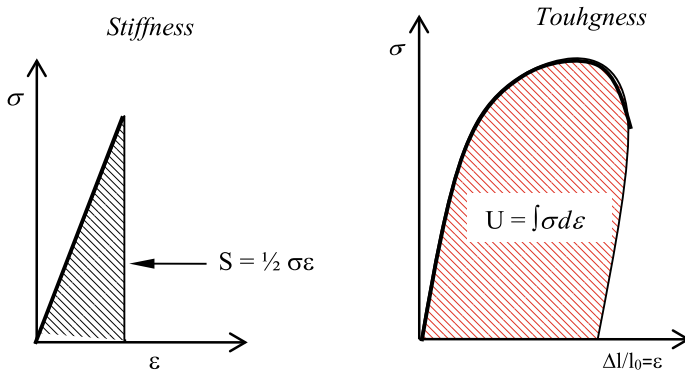


Fig. 2.12 Elastic strain energy (Stiffness) and plastic strain energy (Toughness)

Strain energy: The deformation of a material requires a work supply, and according to the energy conservation principle, the supplied work should transform into stored energy; which is the *strain energy*. The strain energy, represented by U , can be calculated by the following equation:

$$U = \int \sigma d\epsilon$$

If the strain is uniform, the previous equation can be simplified as:

$$U = \left(\frac{F}{A}\right) \left(\frac{\Delta l}{l_0}\right) = \frac{F \Delta l}{A l_0}$$

Notice that the product $F \Delta l$ is force-times-distance, which is the work, while the product $A l_0$, is the volume, therefore the integral $\int \sigma d\epsilon$ represents the work per unit volume supplied to deform a material. The elastic strain energy is called *stiffness*, whereas the plastic strain energy is called *toughness* and both are calculated as the area under the curve, as shown in Fig. 2.12.

2.5 Real Stress and Strain

So far, the calculation of stress has been done by dividing load over the initial cross-section area (F/A_0) and the strain has been calculated as elongation over initial length ($\Delta l/l_0$), they are referred as *engineering stress* and *engineering strain*, respectively. However, in practice both the specimen's cross-section area and length change gradually as load is increased. In the tension test, as the length of the test specimen increases, the cross-section area decreases, so the real values of both

stress and strain in a given instant are different from the engineering ones, since these are calculated by using the initial dimensions. Based on the latter, *real stress* is defined as:

$$\sigma_{real} = \frac{P}{A_i}$$

Where A_i is the instantaneous cross-section area. *Real strain*, represented by the symbol e , can be calculated as the sum of the differential length increments with respect to the immediately previous length. Mathematically, this is an integral between the limits l_o and l , which is:

$$e = \int_{l_o}^l \frac{dl}{l} = \ln\left(\frac{l}{l_o}\right)$$

The following example shows that real strain is a more congruent measure of strain than engineering strain, in terms of the magnitude of strain that has to be applied in order to achieve a determined shape.

Example Calculate engineering and real strains resulting from elongating a body twice its initial length and by compressing it to near-zero thickness.

Solution If the initial length is $l_o = 1.0$ and the final length is $l = 2.0$, the engineering strain is:

$$\varepsilon = \Delta l / l_o = (2.0 - 1.0) / 1.0 = 1.0$$

The real strain is:

$$e = \ln(l/l_o) = \ln(2.0/1.0) = 0.693$$

For the compression case, the engineering strain is:

$$\varepsilon = \Delta l / l_o = (1.0 - 0.0) / 1.0 = -1.0$$

And the real strain is:

$$e = \ln(l/l_o) = \ln(1.0/0.0) = -\infty$$

Notice that engineering strain gives the same value, but of opposite sign for the elongation and compression, whereas the real strain is 70% for the elongation and is infinite to reach a zero thickness. Clearly, real strain is more representative of the amount strain required in each case.

The relations between engineering and real stress and strain, can be obtained as follows:

The engineering strain can be calculated as:

$$\varepsilon = \Delta l / l_0 = l / l_0 - 1$$

Therefore:

$$l / l_0 = \varepsilon + 1$$

Substituting this expression into the definition of real strain:

$$e = \ln(\varepsilon + 1)$$

The real stress can be determined with base on the definitions of both engineering and real stresses:

$$\sigma_{real} = P / A_i$$

$$\sigma_{eng} = P / A_0$$

Multiplying the formula of σ_{real} by A_0 / A_0 :

$$\sigma_{real} = (P / A_i)(A_0 / A_0) = (P / A_0)(A_0 / A_i)$$

If the volume is constant it can be written that:

$$A_0 l_0 = A l$$

Rearranging terms:

$$A_0 / A = l / l_0$$

Taking the relation between real and engineering strain:

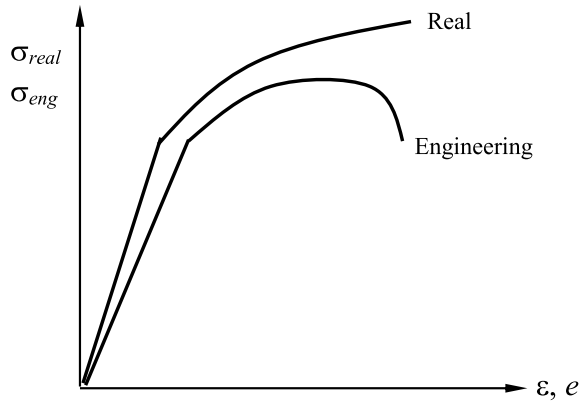
$$e = \ln(1 + \varepsilon) = \ln(l / l_0) = \ln(A_0 / A)$$

Eliminating the logarithms and substituting:

$$\sigma_{real} = \sigma_{eng}(1 + \varepsilon)$$

By plotting real and engineering stress and strain of the same tension test, it can be observed that the real stress-strain curve is located above and to the left of the engineering curve, as shown in Fig. 2.13. Additionally, the real stress-strain curve does not have a drop after the maximum stress.

Fig. 2.13 Real and engineering stress-strain curves in uniaxial tension



2.6 Plastic Stress-Strain Relations

A constitutive equation between real stress and strain was proposed by Ramberg and Osgood in 1943, and has the following form:

$$\frac{e}{e_0} = \frac{\sigma}{\sigma_0} + \alpha \left(\frac{\sigma}{\sigma_0} \right)^n$$

Where σ_0 is the yield strength, e_0 is the real strain at yield, n is the strain hardening exponent and α is an experimental constant. In 1945, Hollomon proposed a simpler constitutive equation, which has been extensively used ever since, and has the form:

$$\sigma_{real} = k e^n$$

Where n is the strain hardening exponent and k is a constant defined as the real stress when $e = 1.0$. Some typical values of n are given in Table 2.7, where it can be observed that steels exhibit the largest strain hardening exponents, being the highest reported value of 0.55, this means that the ultimate tensile strength can double the yield strength; while aluminum alloys and cast irons strain-harden the least. The test method for measuring tensile strain-hardening exponents is described in the ASTM E-646 standard.

Table 2.7 Typical values of the Hollomon's strain hardening exponent

Material	n
Carbon and Low Alloy Steel	0.45–0.55
Pure Cooper	0.3–0.35
Brass and Copper Alloys	0.35–0.4
Aluminum Alloys	0.15–0.25
Cast Iron	0.05–0.15

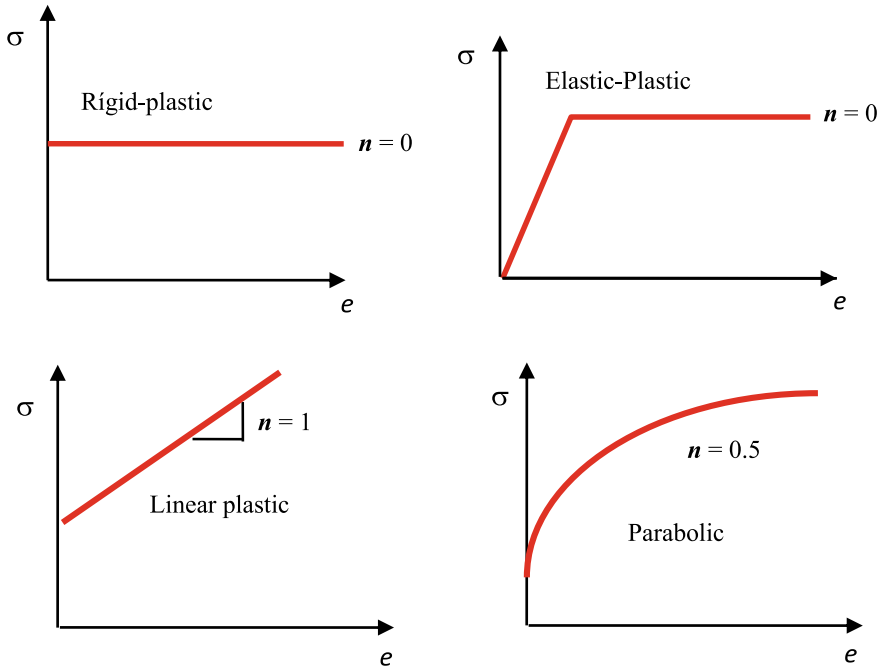


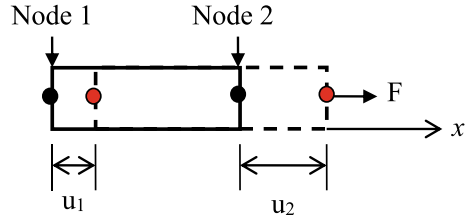
Fig. 2.14 Ideal plastic stress-strain curves

To facilitate the analysis of plastic behavior of specific materials, a few ideal cases of the Hollomon's equation have been introduced, which are schematically depicted in Fig. 2.14. Plaster and raw clay exhibit an ideal rigid-plastic behavior, while high carbon wire can be an ideal elastic-plastic material. Linear-plastic idealizations are used to model the behavior of high strain hardening materials just after yield, where the real stress-strain curve is very steep. Finally, parabolic hardening represents the average plastic behavior of carbon and low alloy steels, so this idealization is particularly used in numerical methods of mechanical behavior modeling.

2.7 Numeric Analysis of Stress and Strain

The finite element method (FEM) is the most frequently used technique to calculate stress and strain of complex geometry bodies, under multiple, non-uniform loads. Currently, this powerful technique is readily available thanks to the great advance and cost reduction of computational equipment and software which allows modelling three-dimensions problems, plastic strain, strain hardening, anisotropy, thermal loads and other complex conditions of mechanical behavior. In some cases FEM modeling is even considered more precise than analytical or experimental stress and strain measurements.

Fig. 2.15 Basic model of the finite element method



The method consists of constructing a virtual model of the body to be analyzed as an assemblage of volume *elements*, which have to be small enough so that the model approaches the real geometry. The smaller the element the more precise is the analysis. Each individual volume element has finite dimensions, hence the name of this method. The elements are connected by *nodes*, so a set of equilibrium equations, one per node, is obtained. The process of discretization of the geometry is called *meshing*, the size of the mesh and the shape of the elements depend on the desired precision and the complexity of the form. The known loads, strains and displacements, are called *boundary conditions*.

In a simple way, FEM can be explained in the model depicted in Fig. 2.15, which consists of an elastic bar formed by one element and two nodes.

In each node, the force F produces a displacement u in the direction x , which is proportional to a constant C_{ij} , the equations that describe these reactions are:

$$F_1 = C_{11}u_1 + C_{12}u_2$$

$$F_2 = C_{21}u_1 + C_{22}u_2$$

The latter equations can be written in matrix form as:

$$\begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix}$$

Notice that for each node, there is an equation, therefore, if a second element is added, there is a new node that experiences a force F_3 , so a new reaction equation has to be added, thus the resulting matrix is:

$$\begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} + C_{22} & C_{23} \\ 0 & C_{32} & C_{33} \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix}$$

If the behavior is elastic, the constants matrix is known as *stiffness matrix* and the constant values depend on the material's elastic properties, whereas the boundary conditions are usually, known values of F and u , so there has to be the same number of equations as unknowns in order to solve the system. Typically, the stiffness matrix is determined by the software so the user only has to focus on constructing a geometric model that closely resembles the actual body and to assign the correct

boundary conditions. Although this looks simple, it may become quite complicated and complex, so a great deal of experience and theoretical knowledge is required, in order to perform correct FEM simulations.

In most problems of mechanical behavior analysis, the loads are known and the stresses, strains and displacements are the sought unknowns. According to the previous description, FEM normally determine the displacements (u , v , w) of each node, with these values the strains are calculated, and finally the stresses are calculated by using the constitutive equations. If the problem includes plastic strain, a plasticity constitutive equation, like the Hollomon's is used.

The procedure to perform an analysis of mechanical behavior by the FEM consists of the following steps:

- (1) *Geometrical model and meshing*. It consists in the making of a computer model of the body geometry, usually by using a drawing software such as AutoCAD, although many commercial FEM software come with its own geometry construction module. The geometrical model must be as close as possible to the real body, although some simplifications can be made, for example, half-models may be used if the symmetry conditions allow it. The meshing process is done automatically by most commercial software, where the mesh size as well as the type of element are selected by the user. The rule is that the mesh has to be finer where a more precise solution is desired, or where the geometry is more complex. Figure 2.16 shows an example of a geometrical model.

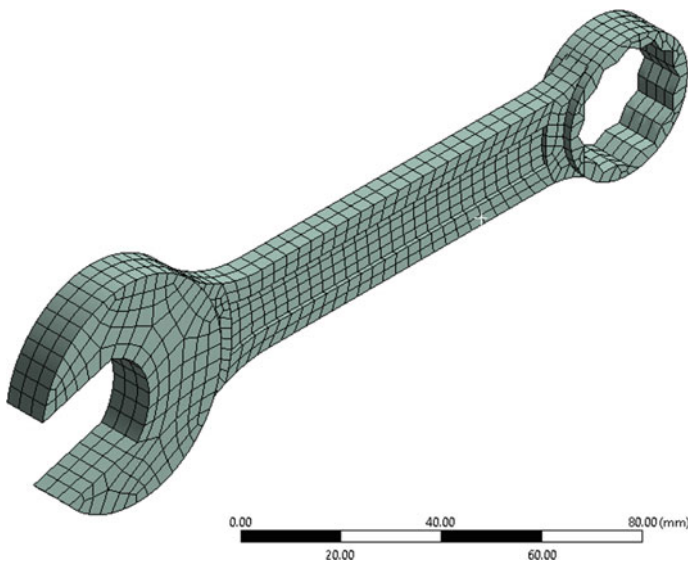


Fig. 2.16 Geometric model and mesh of a Finite Element Method software. Image courtesy of Manuel Alejandro Beltrán Zúñiga

- (2) *Boundary conditions*: Boundary conditions are the known values of displacement, load, stress or strain pertaining to the case to be analyzed. Their selection demands a good knowledge of mechanic behavior, since from these depend the success of the FEM modeling. In most commercial software there are pre-constructed options to assign the boundary conditions, for example, pressure, zero displacements, torque, and etcetera; but there is always the option to assign boundary conditions node by node, which is hard work and must be done by experts. In this stage the mechanical properties are introduced. Most of the commercial software have libraries of common material properties as well as relations to determine the property variations with temperature or models to estimate secondary mechanical properties, such as stiffness or toughness. It just has to be verified that the selected mechanical properties are representative of the actual service conditions of the component to be modeled.
- (3) *Solution and post process*: The output of a FEM model is the solution of the matrix equations by some numeric method, like the Newton-Radson. FEM commercial software automatically construct and solve the matrix equations and verify the validity of the solution through a process called *convergency*, hence, this stage is carried out without the user's intervention. The precision depends basically on the quality of the computer program and the processing capacity of the computer executing the software. The output results of a FEM computer program include:
- Any component of the stress or strain tensor.
 - Any component of displacement.
 - Any of the principal stresses or strains.
 - Maximum shear stress or maximum shear strain.
 - Effective stress (Von Mises) or maximum shear stress (Tresca).

Being 12 independent components of stress and strain, 3 displacement vectors, 3 principal stresses and 3 principal strains and 1 effective stress, there are 22 possible data outputs per node. As it can be foreseen, even a fairly small FEM simulation produces a massive set of data that are practically impossible to read and analyze if they are displayed in tables, this is why the results of a FEM mechanical behavior simulation are usually presented in the form of maps with colored contours painted on the geometric model, as shown in the example of Fig. 2.17, where each color represents a value range of the output variable. By consensus, red is assigned to the maximum values and blue for the smallest or most negative values. The construction of these graphics is called *post-process* and it is usually automatically done by the FEM software for the variable selected by the user. The user can also select the view perspective, like isometric, plant view or any cross section cut, as well as make zooms and rotations of the image to observe in more detail the results.

Another advantage of modern commercial FEM software is that it allows the making of animations in which the development of the stress or strain fields, or displacement movements can be observed, as the increasing loads are applied. Nonetheless, any FEM mechanical behavior simulation should be complemented

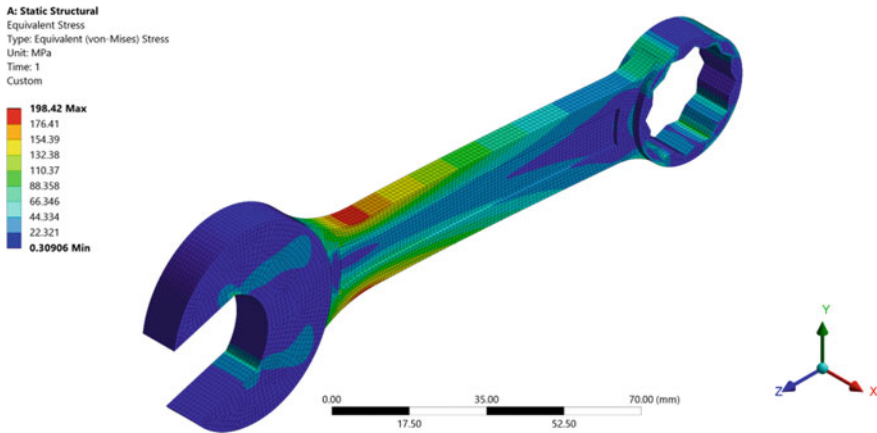


Fig. 2.17 Typical display of FEM results as contour map of stress ranges. Von Mises stress values in MPa. Image courtesy of Manuel Alejandro Beltrán Zúñiga

with an analysis of results done by an expert, where it is verified whether the output data is within the expected range of values, for example, no stress should be higher than the ultimate tensile strength or there cannot be negative values of the von Mises effective stress, additionally it should be verified if the results are congruent from the theoretical point of view or from the reviewer's experience. Finally, it is a good practice to verify the results of a FEM simulation against experimental measurements, either from in-service measurements or from bench-laboratory models, especially if the analysis of the mechanical behavior analysis is part of a large scale project.

2.8 Experimental Measurement of Stress and Strain

Experimental measurement of stresses and strain can be done by several direct or indirect techniques, however, in the present textbook, only the two ones that are mostly used in engineering applications are described, being these *electric resistance extensometry* and *polaroscopy*.

Electric resistance extensometry measures the elongation strain on a surface by means of measuring the change of electric resistance of a very fine wire, which electric resistance varies with the change of length. The wire is placed over a rigid polymer base plate that is glued with a strong adhesive to the surface of the body where the deformation is to be measured. This device is known as *strain gauge* and it has the configuration depicted in Fig. 2.18. The strain gauges may come in one, two, three or more element configurations, and may be especially design for crack analysis, shear strain, pressure sensing or bolt spindle power measurement.

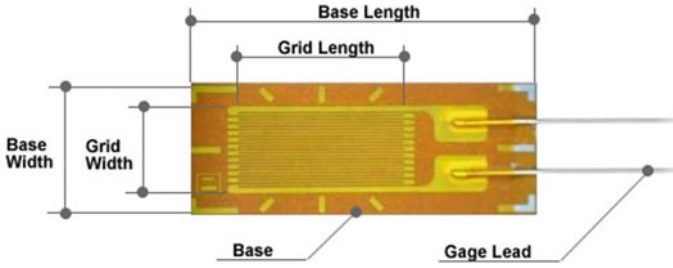


Fig. 2.18 Typical drawing of a single strain gauge. Image taken from: www.showa-sokki.co.jp

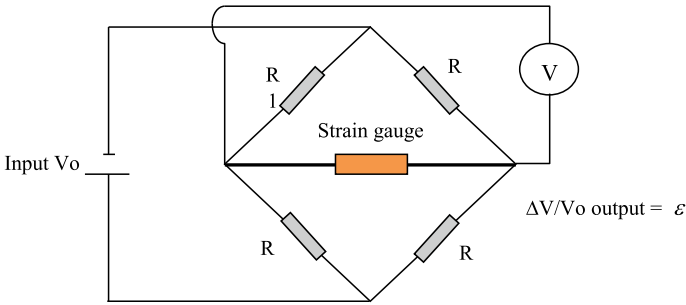


Fig. 2.19 Wheatstone bridge circuit used to measure the strain by electric resistance change

The strain gauge electric resistance is determined by the following equation:

$$R = \left(\frac{\rho}{s}\right)l$$

where ρ is the wire resistivity, usually made of a copper-nickel alloy, s is the cross section area of the wire and l is the grid length of the strain gauge. The change of resistance is done by a Wheatstone bridge circuit, which is depicted in Fig. 2.19, where a power source introduces a constant voltage and a high precision volt-meter measures the output voltage, which is then converted into electrical resistance change. Since the equipment measures the change of resistance (ΔR) with respect to the initial resistance (R), the measurement directly gives the strain, by applying the following equation:

$$\epsilon = \frac{1}{k} \frac{\Delta R}{R}$$

where k is the strain gauge constant, which is provided by the strain gauge manufacturer.

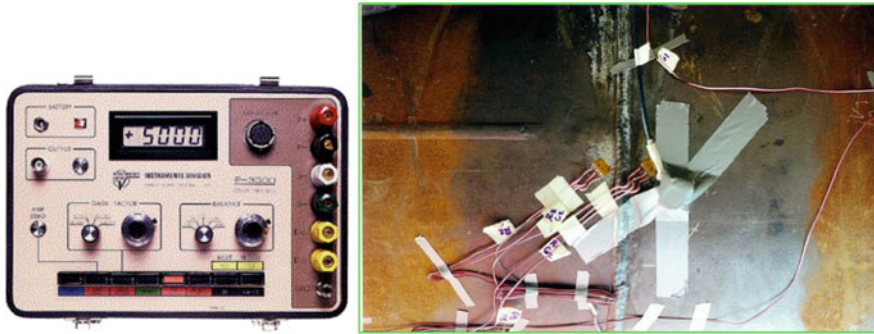


Fig. 2.20 Electric extensometry measuring apparatus and strain gauges installed on the surface of a steel pipe

It is important to mention that electrical resistance extensometry is a differential measurement technique, thus it measures the change of strain with respect to an initial value. So, in order to measure the absolute value of strain, the strain gauges must be installed on the unloaded body, then, the measuring circuit is set to zero and the load is applied to measure the absolute value of strain. An advantage of electrical resistance extensometry is that, even the most economic commercial apparatus have a precision of 10^{-6} , therefore, this is a very precise technique. If the instrument features a high frequency data acquisition device, dynamic strain measurements can be made, for example, in vibrating machinery or under cyclic load conditions. Figure 2.20 shows the typical aspect of an electric extensometry device and a set of strain gauges installed on a steel pipe.

When single strain gauges are used, the normal stress can be calculated directly from the Hooke's law, which is expressed as:

$$\sigma = E \varepsilon$$

where E is Young Modulus. If sets of three strain gauges (called *rosettes*) are used the principal strains and stresses are calculated by the following equations:

$$\varepsilon_{1,2} = \frac{\varepsilon_a + \varepsilon_c}{2} \mp \frac{1}{\sqrt{2}} \sqrt{(\varepsilon_a - \varepsilon_b)^2 + (\varepsilon_b + \varepsilon_c)^2}$$

$$\sigma_{1,2} = \frac{E}{2} \left[\frac{\varepsilon_a + \varepsilon_c}{1-\nu} \mp \frac{\sqrt{2}}{1+\nu} \sqrt{(\varepsilon_a - \varepsilon_b)^2 + (\varepsilon_b - \varepsilon_c)^2} \right]$$

When it is desired to measure plastic strain, the strain gauges should be especially designed for this purpose, since the post-yield behavior is not linear, so it is important to ensure that the strain gauge's range is sufficient to measure the expected elongation plastic strains. Another important concern is the effect of strain gauge bonding and wiring on the stress-strain curve of the material, as well as

considering anisotropy and plastic strain-induced heating and finally, it must be remembered that plastic strain measurement requires the knowledge of the true elastic-plastic strain curve.

The main disadvantage of electric extensometry is that it makes measurements on the free surface, so it only determines plane strain, the second disadvantage is that the measurement is local, only over the area where the strain gauge is placed, so in order to measure a strain field, it is required to install several strain gauges strategically placed to determine the strain field.

Polaroscopy or also known as photoelasticity is an indirect method to measure shear stress. It is based on the birefringency property of certain transparent materials, which is having a double refraction index. The property of birefringency produces a pattern of stripes or fringes when the materials that exhibit this property are stressed and observed with a visor of polarized lenses known as “polaroscope”. The interpretation of this technique is direct because the photoelastic image clearly shows that where the stress concentrate the fringe density is higher. Figure 2.21 shows an example of a photoelastic fringe pattern of a stressed material and a commercial polaroscope.

The procedure of photoelasticity stress measuring consists in making a birefringent material model of the piece or by applying a birefringent coating over the area of the actual piece where the measurement is desired, then the loads are applied and the piece is observed with a polaroscope. The maximum shear stress (τ_{\max}) is determined by the equation:

$$\tau_{\max} = \frac{N\lambda}{2tC}$$

where N is the number of stripes, λ is the wave length of the illumination source, t is the thickness and C is the stress optic coefficient. The value of C is usually provided by the birefringent material supplier.

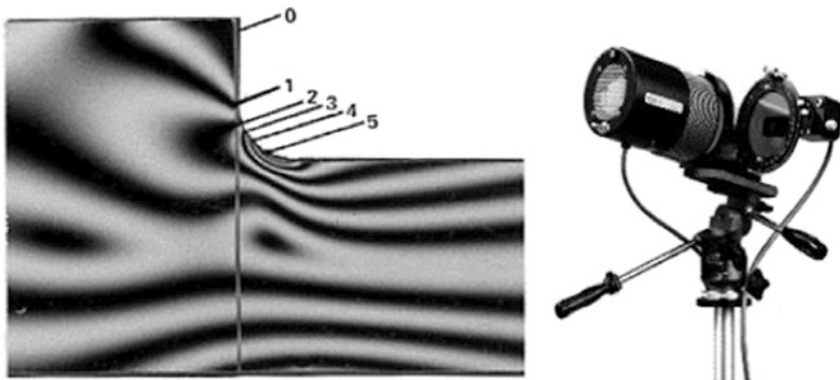


Fig. 2.21 Fringe pattern of a stressed birefringent material and a commercial polaroscope

The main disadvantage of photoelasticity is that it only allows measuring shear stresses and in the case of coating use, the measurement is superficial and it requires a careful surface preparation. Regardless of these limitations, the photoelasticity method is quite useful to improve designs because it immediately detects stress concentration zones.

2.9 Hardness

Hardness is most widely used mechanical test for evaluating the mechanical strength of materials, however hardness is not a fundamental material property, at least in the strict sense, since it depends on a combination of several fundamental mechanical properties, such as the elastic modulus, the tensile and yield strength and the strain hardening exponent, among others. Since the first days of engineering, hardness has been defined as the resistance of a material to be scratched or penetrated. For structural and material engineers, hardness is the resistance to plastic deformation, while for mechanical engineers, is the resistance to machining and wear.

The hardness scale for scratching resistance used for minerals was introduced by Mohs in 1822 and it is a relative scale, being 1 for the softest material (Talc) and 10 for the hardest (Diamond), however, the hardness test for engineering materials, regarded as the resistance to permanent indentation, was introduced by the Swedish engineer Johann A. Brinell. The test, known today as the “Brinell Hardness Test”, and it is described in the ASTM E10 standard, and consists of a ball penetrator, made of very hard material (hardened steel or tungsten carbide), which is pressed on a clean and smooth surface of the test material. A pre-determined load is applied during a standard time, usually 30 s and the size of the indentation mark is measured. The softer the material, the wider the indentation. Soon after the introduction of the Brinell test, the Vickers test using a diamond pyramid indenter was introduced (the Vickers hardness test method is described in the ASTM E92 standard). Brinell hardness (*BHN*) and Vickers hardness (*VHN*) are both based on the measurement of the size of the indentation mark, using the following formulas:

$$BHN = \frac{2L}{(\pi D) \left[D - \sqrt{D^2 - d^2} \right]}, \quad VHN = \frac{1.584L}{d^2}$$

where L is the load in kilos, D is the diameter of the spherical penetrator in millimeters, and d is the diameter of the mark in millimeters.

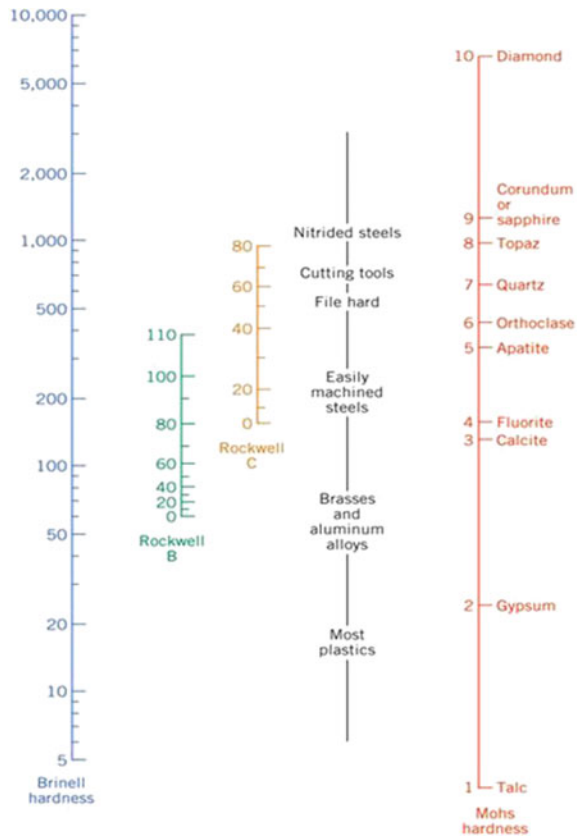
In 1919, a Hartford, Connecticut, heat-treatment engineer named Stanley P. Rockwell introduced a more practical hardness test based on the measurement of the depth of the indentation produced by a constant load penetration, such method carries his name and it is described in the ASTM E18 standard. The typical Rockwell scales are: A, with a 60 kg load and 120° diamond penetrator; B,

with a 100 kg load and 1/16" steel ball penetrator; and C, with 150 kg load and 120° diamond penetrator. To carry out the test, a 10 kg preload is first applied in order to settle the penetrator, then the test load is applied and the device measures depth of the indentation. Each Rockwell scale contains 100 divisions corresponding to a 0.00008" (2.032 micron) deep each one.

While Brinell and Vickers scales are applicable to the whole range of hardness values of metallic materials, Rockwell scales have a limited applicability according to the estimated hardness value. Figure 2.22 shows a comparison of the different hardness scales and groups of materials that exhibit a particular range of hardness values. It can be observed that Rockwell B scale is appropriate for non-ferrous metals and low strength steels, whereas Rockwell C scale is useful for hardened steels and cutting tools.

Despite its limited application range, Rockwell hardness is the most widely used hardness test in industry because it is faster and more precise than Brinell and Vickers scales. The Vickers hardness test has been adapted to use loads no greater than 1 kg, allowing to make very small indentations called "microhardness" testings.

Fig. 2.22 Comparison of hardness scales



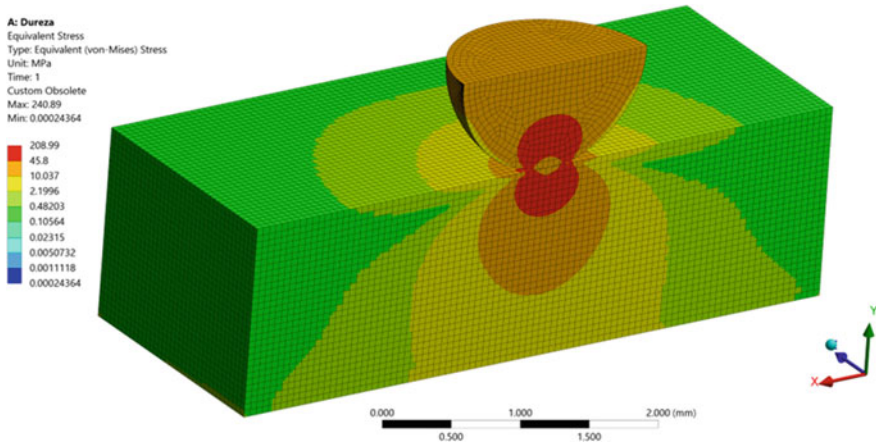


Fig. 2.23 Distribution of Von Mises stress in ksi calculated by the finite element method in a Brinell hardness test. Image courtesy of Manuel Alejandro Beltrán Zúñiga

As mentioned before, hardness is a complex material property that involves a non-uniform distribution of stress and strain, as seen in the MEF model of Fig. 2.23. As seen in this figure, the material deforms plastically beneath the indenter, forming a plastic zone that is surrounded by an elastically strained zone, thus the hardness value depends primarily on the Young's modulus, the tensile properties and the strain hardening behavior. Logically the friction between the indenter and the test material surface plays an important role, so the measured hardness value depends on the surface finish condition, lubrication and other factors. In spite of this, due to its simplicity, quickness and economy, the hardness test is widely used in engineering for materials selection, quality control and to assess a specific resistance such as wear, fatigue and stress corrosion cracking susceptibility, just to mention a few.

For carbon steels, of mid and low alloy and thermally treated steels, the Brinell hardness number can be related directly to the tensile strength, based on the following relation:

$$\sigma_{\max}(\text{ksi}) = 0.5 \text{ HBN}$$

Likewise, semi-quantitative relations can be established with ductility and the yield strength (the greater the hardness, greater σ_0 and less ductility), thus, hardness can be also used as nondestructive test to investigate the condition of an in-service component, which is quite useful in structural integrity assessments and failure analysis.

Because of the necessity to perform in-field hardness measurements, a variety of portable hardness measuring techniques have been introduced since the middle of the XX century, among which outstand the following:

Telebrineller. It is a commercial brand that uses a device consisting of an anvil, a steel pattern bar and a hardened steel ball. The ball is placed onto the body whose hardness is to be measured while the opposite side rests on the pattern bar, over which the anvil is placed. The anvil receives a strong hit with a mallet and then the diameters of the marks on the piece and the pattern bar are measured using a field microscope. Since the pattern bar hardness is known, the hardness of the piece is determined by a rule of three.

Leeb method. Introduced by Leeb and Brandestini in 1975, it is based on the measuring of the backward speed of a sphere driven by an electromagnetically controlled spring after impacting the test surface. Hardness is proportional to the difference between the speed of impact and the bounce back speed of the sphere. Since 1996, this method is ruled by standard ASTM A956 and the commercial equipment feature resulting values in the conventional scales Brinell, Vickers and Rockwell.

TIV Method. It stands for Through-Indenter-Viewing, is a method that measures Vickers hardness and consists of an optic device that allows to visualize directly the penetration of a diamond indenter on the test surface. Such device applies a pre-determined and calibrated load so that the mark size, measured directly on the device's LCD screen, is converted directly into Vickers hardness or into any other scale, by an on-board computer, at user's discretion.

Chapter 3

Plastic Deformation Mechanisms



Abstract This chapter begins with an introduction of the crystalline structure and crystalline defects of solid materials and the definition of hot and cold work. Then, a full description of the dislocation slip mechanism is presented, including an explanation of the direct observation of dislocations by transmission electron microscopy. Based on the preceding ideas, a description of the slip modes and their associated dislocation substructures is given. The chapter concludes with a brief description of the plastic deformation mechanism by twinning.

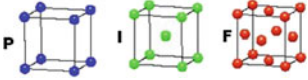
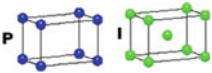
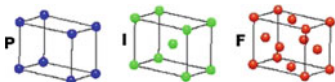
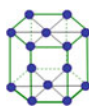

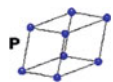
3.1 Crystalline Defects and Deformation Mechanisms

Thanks to the works of the physicists Max Von Laue from Germany and the British W. H. Bragg and W. L. Bragg, in 1913, it is known that most engineering metals and alloys are *crystals*. A crystal is a solid whose atoms are arranged in an order formed by the repetition of a unitary geometry that extends throughout the solid. This idea was proposed by the self-taught English scientist William Barlow in 1883, based on the ideas of the French physicist Auguste Bravais, who introduced in 1848 the field of crystallography, describing the *unitary cell*, a term given to the geometric unit that repeats in a crystalline solid, being also the symmetry-unit that defines the type of crystalline structure. The most important crystalline structures, called *Bravais lattices*, are shown in Table 3.1

The ordered repetition of a group of unitary cells in a solid is called *lattice*. Lattices are tridimensional, so each plane and direction in the cell unit is defined by a vector, described by the *Miller indexes*, (Miller's notation can be found in textbooks of Materials Science or Crystallography). Figure 3.1 illustrates these concepts in two dimensions for further simplicity.

Soon after discovering the crystalline structure, it was found that the crystalline lattices in natural and engineering materials are not perfect; but rather contain flaws called *crystalline defects*. The most common types of crystalline defects are listed in Table 3.2 and a schematic representation of them is shown in Fig. 3.2.

Table 3.1 Common crystalline structures in solid materials

Structure P = Primitive, I = Body centered, F = Face centered	Material
Cubic 	Face centered cubic (<i>fcc</i>): Cu, Ni, Al, Fe- γ , Au, Ag. Body centered cubic (<i>bcc</i>): Fe- α , Mo, W.
Tetragonal 	Fe ₃ C, Martensite (α')
Orthorhombic 	Marcasite, Olivine, Aragonite
Hexagonal 	Compact hexagonal (<i>hc</i>): Cd, Mg, Ti, Zn
Monoclinic 	Gypsum, Orthoclase
Triclinic 	Tantite, Wollastonite

Evidently, the plastic deformation of solid materials is closely related to the crystalline structure and its crystalline defects. The sequence of events that produce plastic deformation in solids is called *deformation mechanism*. At this point it is necessary make clear that the elastic deformation mechanisms is just the result of the stretching or shortening of atomic links, therefore it is quite straight forward,

Fig. 3.1 Schematic two dimensions representation of a single cubic crystal lattice

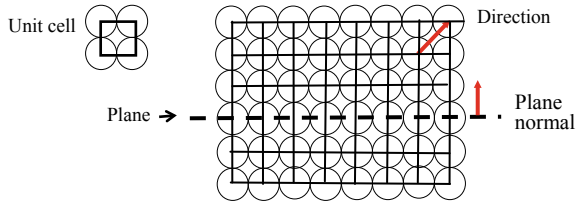


Table 3.2 Classification of crystalline defects

Type	Defect
Punctual	<ul style="list-style-type: none"> • Vacancies • Substitutional impurity atom • Interstitial impurity atom
Lineal	<ul style="list-style-type: none"> • Dislocations—perfect [edge, screw and mix] • Partial dislocations
Planar	<ul style="list-style-type: none"> • Stacking faults: Intrinsic y extrinsic • Twin boundary • Grain boundary
Volume	<ul style="list-style-type: none"> • Precipitates

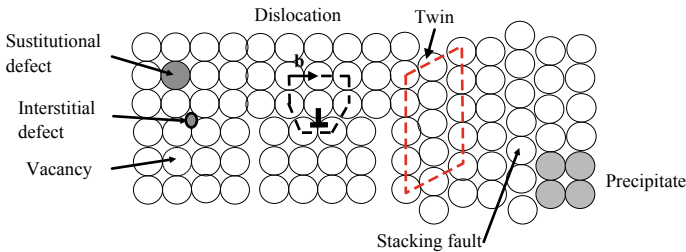


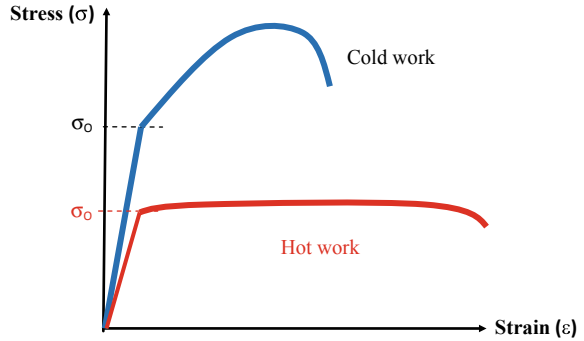
Fig. 3.2 Schematic representation of main crystalline defects

and requires no further study, whereas plastic deformation mechanisms involve the rearrangement of particles within the solid, including defects movement, all of which make them very complex.

3.2 Definition of Cold and Hot Work

Before studying the deformation mechanisms, it is convenient to bear in mind that the mechanical behavior of engineering materials is strongly influenced by the temperature, and metallic materials may feature two types of behavior with respect to it, referred as: low temperature or *cold work*, and high temperature or *hot work*. The main difference is that at high temperature, the deformation mechanisms involve time-dependent thermally activated processes such as diffusion and viscous

Fig. 3.3 Tension Stress–Strain curve of hot and cold work



flow, whereas at low temperature, deformation occurs by processes that are independent of time. The difference between hot and cold work is exhibited in the tension stress-strain curve, as the absence of strain hardening in hot deformation, in addition to reduction of the elastic modulus, yield strength and ultimate tensile strength, as schematically shown in Fig. 3.3.

The parameter to determine whether it is hot or cold work is the *homologous temperature*, which defined as:

$$\Theta = T/T_f$$

where Θ is the homologous temperature, T is the work temperature in absolute degrees and T_f is the material's melting point absolute temperature (in Kelvin). The most widely accepted criterion is:

If $\Theta > 0.4$, it is hot work

If $\Theta < 0.4$, it is cold work

However, the value of Θ that determines the transition from cold work to hot work is not a fixed one, but it is rather a range of values. Furthermore, since Θ depends on the melting point, the condition of hot work does not necessarily indicate that the material is at a temperature much higher than room temperature, as shown in the following example:

Example Determine the Θ value steel and a Pb–Sn alloy at room temperature (25 °C, 298 K)

For steel: $T_f = 1600$ °C = 1873 K, thus: $\Theta = 298/1873 = 0.16$

For the Pb–Sn alloy: $T_f = 200$ °C = 473 K, thus: $\Theta = 298/473 = 0.63$

Notice that at room temperature, steel is in cold work whereas the Pb–Sn alloy is hot work condition.

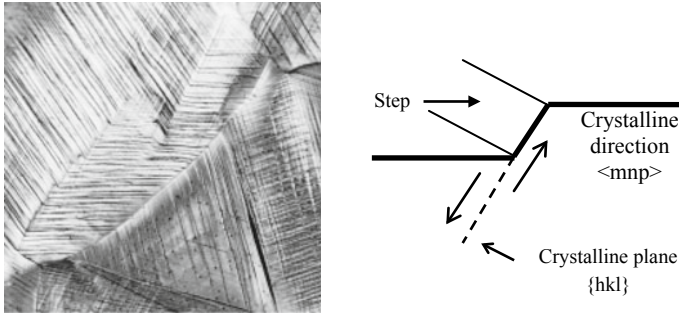


Fig. 3.4 Slip lines on a polished surface as seen in the metallographic microscope and schematic representation of the formation of a surface step by slip

3.3 Deformation by Dislocation Slip

In March, 1899, Ewing and Rosenhain, from St. John’s College, Cambridge, published a short paper, where they reported observations through the metallographic microscope, of the polished surface of a metal strip stretched up to reaching plastic deformation. They observed fine parallel dark lines within the grains, that increased in number and new line systems appeared in different directions, as the specimen was stretched further. After performing a crystallographic analysis, they concluded that the lines were actually tiny surface steps produced by the slip of crystalline planes, as shown in Fig. 3.4. Such lines were called *slip lines* and it was concluded that plastic deformation in metals is caused by the slip of crystalline planes.

Later on, it was demonstrated that slip occurs in specific crystallographic planes and directions for each crystalline structure, where the slip direction is always parallel to the slip plane. The combination of plane and direction of slip is called *slip system*. The slip systems of the main crystalline structures of engineering metals and are shown in Table 3.3.

Notice that the *bcc* structure has a greater number of slip systems than the *fcc*, but metals with *fcc* structure are usually more ductile. This is because the slip planes in *fcc* metals are more compact than those in *bcc* metals, thus facilitating slip. Once experimentally demonstrated that plastic deformation occurs by the slip of

Table 3.3 Main slip systems of common crystalline structures

Structure	Planes	Directions	No. of systems
<i>fcc</i>	{111}	$\langle 110 \rangle$	12
<i>bcc</i>	{111} {112} {123}	$\langle 111 \rangle$	48
<i>hc</i>	{0001}	$\langle 1120 \rangle$	3



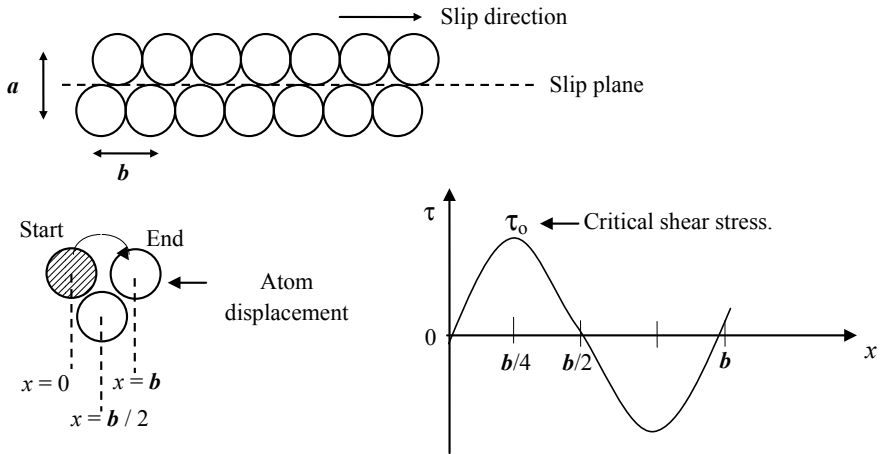


Fig. 3.5 Slip model in a perfect lattice to determine the stress necessary to produce slip

crystalline planes, the next step was to calculate the stress necessary to produce the slip of an atomic plane over another one. Such estimation was done by Frenkel in 1926 by the following procedure:

Consider a model of two parallel rows of atoms as shown in Fig. 3.5, when the upper row of atoms moves over the one below, the shear stress varies according to the following equation:

$$\tau = \tau_0 \sin\left(\frac{2\pi x}{b}\right)$$

For small values of the argument ($2\pi x/b$), the approximation $\sin \theta = \theta$ is valid, therefore the previous equation can be simplified as:

$$\tau = \tau_0 \left(\frac{2\pi x}{b}\right)$$

On the other hand, the shear deformation may be calculated as:

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

Before the upper row of atoms reach the position $x = b/4$, it can be assumed that the deformation is elastic, thus the shear elastic deformation is:

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

Therefore:

$$\frac{Gx}{a} = \tau_0 \left(\frac{2\pi x}{b} \right)$$

Taking $a = b$ (as typical in cubic crystals) and solving for τ_0 , the theoretical shear stress for slip in the perfect lattice is obtained:

$$\tau_0 = \frac{G}{2\pi}$$

The typical values of G for metals range from 6000 to 70,000 MPa (10^6 to 10^7 psi), thus calculating τ_0 with the previous equation, the resulting values are, from 100 up to 10,000 times higher than those obtained experimentally, which are in the range of 7–2000 MPa (10^3 to 3×10^5 psi). The discrepancy between the theoretical and experimental slip stress values does not disregard slip as the plastic deformation mechanism, since it is an experimental observation, but instead it implies there must be “something” that helps slip to occur at a much lower stress.

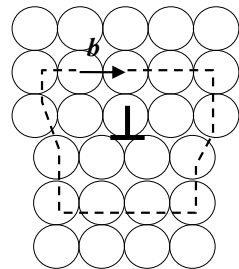
In 1905, the Italian scientist Vito Volterra developed the dislocation theory, from which, later in 1934, Orowan, Taylor and Polanyi, each one separately, proposed that slip could be the result of dislocation movement. Orowan is the most recognized dislocation scientist in the world, but it is Taylor’s model that more clearly illustrates how the dislocation movement can be responsible of the slip. The Taylor’s model of slip is based on an edge dislocation, which is characterized by a Burger’s vector (b) parallel to the dislocation plane, as shown in Fig. 3.6.

To demonstrate that the slip mechanism is caused by dislocation movement, the following conditions must be met:

1. Dislocation movement must produce plastic deformation.
2. Dislocation movement must occur in directions over a plane and produce surface steps.
3. Dislocation movement must take place at a stress lower than the theoretical one.

The following mechanism demonstrate conditions 1 and 2: By applying a shear stress parallel to the Burger’s vector, the atomic bonds unbalance, as shown in Fig. 3.7a, this makes the dislocation move forward one atomic position in the

Fig. 3.6 Edge dislocation with Burgers vector in equilibrium position



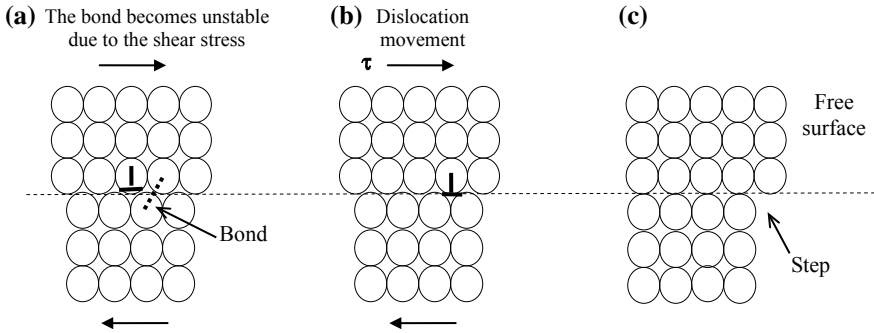


Fig. 3.7 Taylor’s model showing that slip can be produced by the movement of an edge dislocation

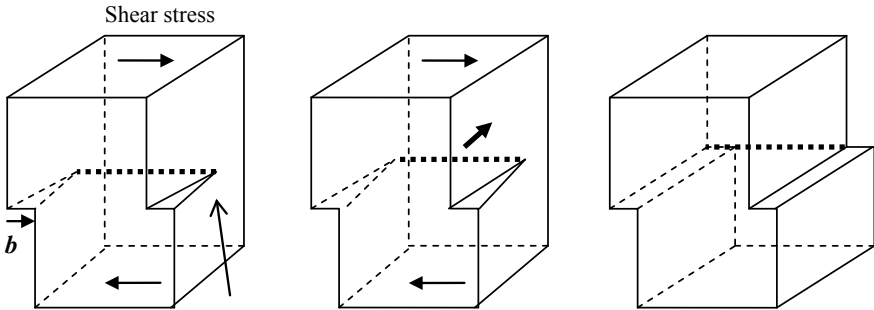


Fig. 3.8 Crystal slip produced by the movement of a screw dislocation

lattice, Fig. 3.7b. By holding the stress, the dislocation keeps moving and shears the crystal until it reaches a free surface, where a step is formed, as seen in Fig. 3.7c. An analog behavior is produced by a shear stressed screw dislocation, as shown in Fig. 3.8. The models depicted in the previous figures demonstrate that indeed the dislocation movement is consistent with the characteristics of plastic deformation by slip observed experimentally.

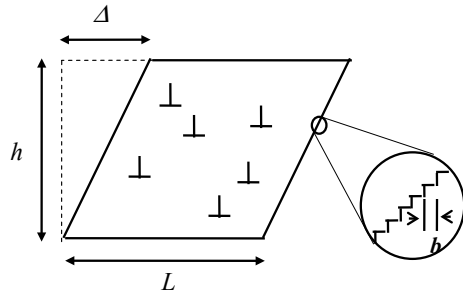
The shear stress to move a dislocation was calculated by Pierls and corrected by Nabarro in 1947, and is given by the equation:

$$\tau_{dis} = \left\{ \frac{2G}{(1 - \nu)} \right\} \exp\left(\frac{-2\pi w}{b} \right)$$

where: ν is Poisson’s ratio, b is Burgers vector, w is the dislocation width, defined as the distance up to which the dislocation distorts the crystal and G is the shear module. This equation combines the elastic stiffness, given by G , while the exponential term is related to an energetic barrier that has to be surpassed in order to



Fig. 3.9 Two dimensional model to estimate the number of dislocations necessary to produce a significant shear strain



reach an equilibrium state of lower energy. It has been estimated that $w = 10b$ in ductile metals and $w = 2b$ in brittle metals, thus if: $G = 11 \times 10^6$ psi, $\nu = 0.35$ and $w = 5b$:

$$\tau_{dis.} = \{2[11 \times 10^6]/(1 - 0.35)\} \exp(-10\pi) = 7.68 \times 10^{-7} \text{ psi}$$

This calculation shows that an extremely low stress is required to move a dislocation, but now it is necessary to explain how the dislocation movement reach the experimentally observed yield strength. What actually happens is that a very large number of dislocations has to be moved in order to achieve a significant amount of plastic strain, in addition to the fact that dislocations have to overcome obstacles, which altogether, increases the stress necessary to produce the slip.

The estimation of the number of dislocations necessary to plastically strain a crystal by slip can be done by the following reasoning: Consider an originally rectangular crystal, of h height and L length, which is deformed by a group of dislocations moving in parallel planes, as depicted in Fig. 3.9.

If the total shear displacement (Δ) produced by the group of dislocations is the sum of the individual displacements produced by each single dislocation, then:

$$\Delta = b \left(\frac{\sum x_i}{L} \right)$$

where x is the slip distance of each individual dislocation, L is the crystal width and b is Burgers vector. On the other hand, the shear strain is equal to:

$$\gamma = \Delta/h$$

Substituting Δ into γ :

$$\gamma = \frac{b}{hL} \sum x_i$$

Introducing the average slip distance (x_{av}) given by:

$$x_{av} = \frac{\sum x_i}{N}$$

where N is the number of dislocations crossing the crystal and combining the previous equations, the total strain produced by the dislocations is:

$$\gamma = \mathbf{b}x_{av} \left(\frac{N}{hL} \right)$$

The term (N/hL) is called *dislocation density* and is represented by the symbol ρ_{dis} , thus the shear strain produced by the dislocation movement is:

$$\gamma = \mathbf{b}\rho_{dis}x_{av}$$

The following example gives an estimation of the dislocation density required to produce an appreciable plastic deformation.

A 1 cm per side cubic crystal, where $\mathbf{b} = 3 \times 10^{-8}$ cm strains at $\gamma = 0.1$. Calculate the associated dislocation density.

Solution If the spatial distribution of the dislocations within the crystal is random, it may be assumed that the average slip length is one half of the crystal width $x_{av} = 0.5$ cm. Taking $\mathbf{b} = 3 \times 10^{-8}$ cm, and solving for ρ_{dis} :

$$\rho_{dis} = \gamma/\mathbf{b}x = (0.1)/(3 \times 10^{-8} \text{ cm})(0.5 \text{ cm}) = 6.7 \times 10^6 \text{ Dislocations/cm}^2$$

This example shows that a great number of dislocations is required in order to deform a crystal. In fact, it has been experimentally determined that heavily cold worked materials have dislocation densities in the range of 10^{12} to 10^{14} Dislocations/cm². Obviously, such number of dislocation does not initially exist, because deformation-free crystals have dislocation densities of 10^3 to 10^4 Dislocations/cm² and this amount is not sufficient to produce an appreciable plastic deformation, therefore, there must be mechanisms to generate high numbers of dislocations in order to plastically deform a crystal.

The best known mechanism of dislocation multiplication is the Frank-Read source, which is schematically described in Fig. 3.10; its mechanism is as follows:

1. A dislocation segment is pinned between two obstacles.
2. Under the action of a shear stress on the slip plane, the dislocation starts to bow out up to a critical curvature, then, the ends of the loop attract each other, because they are of opposite signs, forming a dipole (a pair of opposite sign dislocations close to each other).

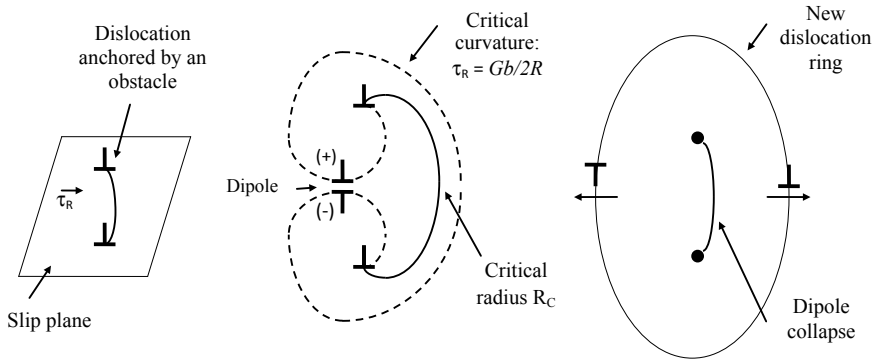


Fig. 3.10 Frank-Read source of dislocation multiplication

3. When the dislocation segments of the dipole are close enough, they collapse and the original dislocation is regenerated while the remaining segment of the loop forms a ring around the original dislocation.
4. The regenerated dislocation segment can continue generating new dislocation loops, while the dislocation loops can slip further into the crystal. This mechanism can repeat continuously until the crystal is saturated with dislocations. In fact, Frank and Read source is considered to be inexhaustible.

Slip strain rate: Although it has been said that cold work is independent of time, it actually occurs at certain rate, because the dislocations take some time to move through a crystal once a shear stress has been applied on the slip plane. The strain rate can be determined in the following way:

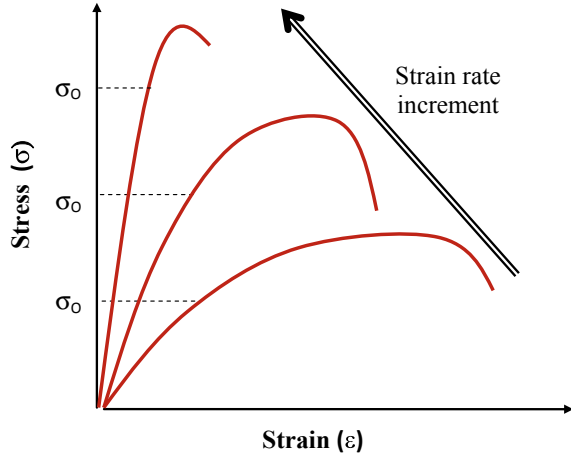
The shear strain rate ($d\gamma/dt$) can be determined by the total derivative:

$$\frac{d\gamma}{dt} = b\rho_{dis} \left(\frac{dx}{dt} \right) + bx \left(\frac{d\rho_{dis}}{dt} \right)$$

In the above equation, the term (dx/dt) is the dislocation slip rate, and the term ($d\rho_{dis}/dt$) is the dislocation multiplication rate, which is the amount of dislocations that have to be generated in order to obtain certain plastic deformation. At high strains, the crystal saturates with dislocations and ($d\rho_{dis}/dt$) comes close to zero, so the strain rate remains only in terms of slip rate:

$$\frac{d\gamma}{dt} = b\rho_{dis} \left(\frac{dx}{dt} \right)$$

Fig. 3.11 Effect of the strain rate on tension behavior



Experimentally, it has been found that:

$$\frac{dx}{dt} = A\tau^{m'}$$

where A and m' are material's constants, thus:

$$\frac{d\gamma}{dt} = b\rho_{dis} \left(A\tau^{m'} \right)$$

The previous equation indicates that the strain rate depends on the applied stress. Experimentally, it has been observed that when a crystalline material is deformed at high strain rates, the flow stress increases proportionally to the strain rate, while ductility is reduced, as shown in Fig. 3.11.

Resolved shear stress. Resolved shear stress (τ_R) is the shear component of a normal stress which acts directly onto a slip plane. It is calculated from the applied tensile stress (P/A), as shown in the diagram of the Fig. 3.12.

$$\tau_R = \frac{P \cos \phi}{\frac{A}{\cos \lambda}} = \frac{P}{A} \cos \phi \cos \lambda$$

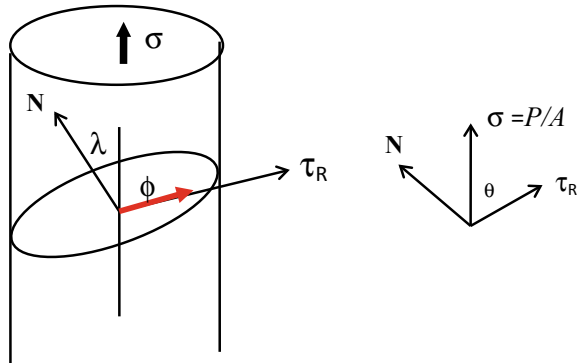
where the term $\cos \phi \cos \lambda$ is called the *Schmid Factor*, therefore:

$$\tau_R = \sigma \cos \phi \cos \lambda$$

When τ_R is sufficiently high as to initiate slip, it is called *critical resolved shear stress* (τ_{RC}). Therefore, in a single crystal:



Fig. 3.12 Definition of resolved shear stress

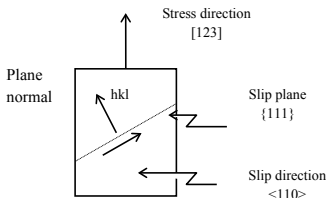


$$\tau_{RC} = \sigma_0 \cos \phi \cos \lambda$$

where σ_o is the single crystal yield strength. When there are several active slip systems, slip will begin in the system with the highest Schmid Factor. If the Schmid Factor value is zero, the dislocation slip is impossible in such system.

Example An fcc single crystal is tension stressed along the [123] direction. Determine which slip system is first one to be active.

Solution A free-body diagram of the single crystal is:



The slip systems are:

- (111)[110]
- (111)[101]
- (111)[110]

In a cubic crystal, the angle between planes or directions is calculated by:

$$\cos \phi = \frac{[hkl] * [mnp]}{\sqrt{h^2 + k^2 + l^2} \sqrt{m^2 + n^2 + p^2}}$$



For the (111) slip plane and the [123] stress direction:

$$\cos \lambda = (1 \times 1 + 2 \times 1 + 3 \times 1) / (\sqrt{14}\sqrt{3}) = 6/\sqrt{42}$$

For the slip direction [110] and the stress direction [123]:

$$\cos \phi = (1 \times 1 - 1 \times 2 + 0 \times 3) / \sqrt{14}\sqrt{2} = -1/\sqrt{28}^2$$

The Schmid Factor is:

$$FS_1 = (6/\sqrt{42}) * (-1/\sqrt{28}) = -0.175$$

$$\text{For system No.2 : } FS_2 = -0.35$$

$$\text{For system No.3 : } FS_3 = 0.52$$

Thus, system (111)[110] is the first to become active.

3.4 Direct Observation of Dislocations

The theory of dislocations succeeded in describing the plastic deformation mechanism of crystalline solids, even when dislocations had not been directly observed. The concept of dislocation was introduced by Volterra in the first years of the XX century, in those days the dislocations were observed in bubble rafts, stacks of pipes, block walls, and so forth. However, for over 40 years after the dislocation theory was published the direct observation of dislocations in engineering materials was not possible, until the work done Hirsch and collaborators in 1956, using the newly developed *transmission electron microscopy* technique.

Transmission electron microscopy (TEM) consists of an electron beam accelerated by hundreds of kilovolts, that is focused by magnetic fields that work as lenses to pass-through a thin foil of the material to be examined (typically of less than 1000 Angstroms thick). The hollow column that contains the electron beam must be at high vacuum to prevent air molecules from deflecting the beam. The electron beam, after passing through the specimen, is amplified and projected on a fluorescent screen placed on the lower part of the column, which can be observed through a transparent window. Figure 3.13 shows a scheme of a TEM.

The Bragg law states that the planes at an angle $\theta = \sin^{-1}(\lambda/2d)$, where λ is the electron beam wave length and d is the interplanar distance, will diffract the electron beam. At high voltages, λ has a very small value (3×10^{-12} m at 100 keV), so only those planes closely parallel to the beam will diffract. Therefore, when an electron beam passes through a crystal having a dislocation, the slight change of inclination of the planes produced by the dislocation will cause diffraction of the electron beam, then by eliminating the diffracted beams with an aperture plate

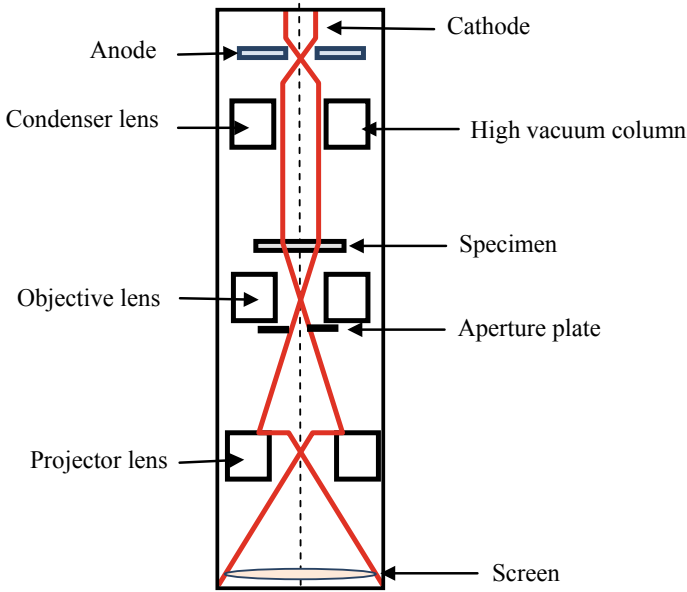
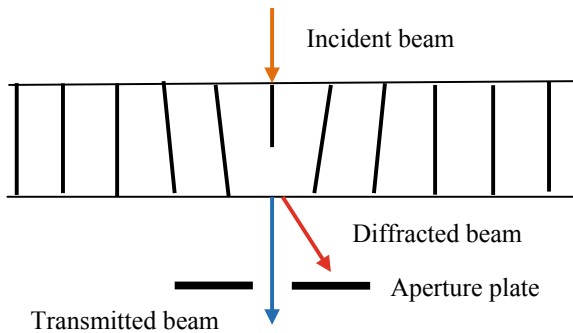


Fig. 3.13 Schematic construction of the transmission electron microscope

Fig. 3.14 Contrast mechanism that makes a dislocation visible in the TEM



placed just below the objective lens, a shadow will be observed in the exact same location where the dislocation line lies. This is schematically illustrated in Fig. 3.14.

Figure 3.15 shows an image of dislocations in an *fcc* metal, obtained with a TEM in bright field, meaning that the image is formed by the transmitted beam and eliminating the diffracted beam. The dislocation lines look wavy due to the ondulatory nature of electron diffraction, therefore if the crystal is tilted, the dislocation line shape may vary, or it may become invisible. The TEM images of dislocation lines are the segments that go from top to the bottom of the thin foil, and since the image is a bi-dimensional projection of a tri-dimensional space they look

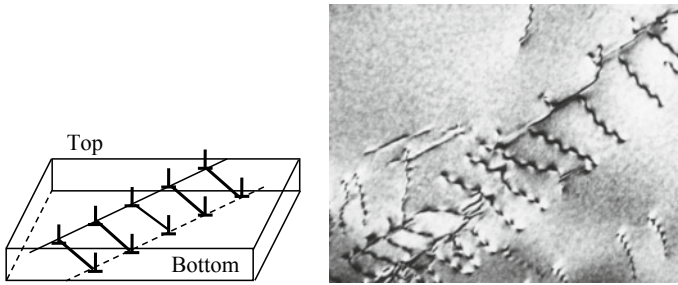


Fig. 3.15 Tri-dimensional lay out of dislocations segments on a thin foil and bi-dimensional projection image as seen in the TEM. Image courtesy of Dr. Héctor Dorantes R

short, as shown in Fig. 3.15. It is important to mention that, at a first glance, it cannot be determined which end is up or down.

3.5 Slip Modes

As it was seen before, plastic deformation is the result of the movement of a large number of dislocations over several crystallographic planes and directions, so there is a large number of interactions among single dislocations, as well as, among dislocation groups, which makes slip a complex phenomenon. The overall result of these interactions within a crystal is a particular configuration of dislocation groups, called *dislocation array*, which greatly influences the mechanical behavior of materials.

There are two basic modes in which a group of dislocations may move in the interior of a crystal, they are called *slip modes* and are described as follows:

Planar slip—In close-packed crystals, the dislocations tend to dissociate into partial dislocations that stay in the plane they were formed so they make up parallel aligned bands of dislocations. When the dislocations bands meet an obstacle, the distance between individual dislocations diminishes as they approach to the obstacle, thus producing arrays called *pile ups*, which look like the example shown in Fig. 3.16. Planar slip produces straight and well defined slip lines, as shown in Fig. 3.16.

Planar slip occurs is typical in *fcc* crystals, since a dislocation splits into two partial dislocations that have an associated stacking fault between them, that makes impossible for the dislocation to move out of the slip plane. The mechanism of partial dislocation formation in a *fcc* crystal is depicted in Fig. 3.17. To move on the (111) plane, a perfect dislocation must follow the [101] direction, thus, the dark atom must go over the gray atom. It is easy to see that the path around the gray atom, given by the directions $a/c[112]$ and $a/c[211]$ is more favorable, so the dislocation “splits” into two partials, forming a stacking fault, as shown in Fig. 3.18.

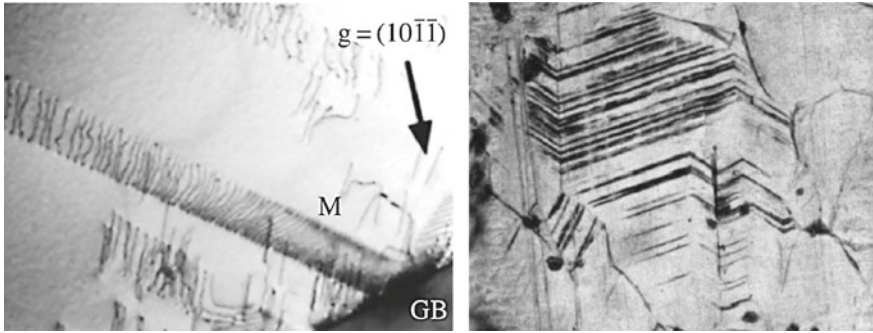


Fig. 3.16 Dislocation bands and straight slip lines produced by planar slip

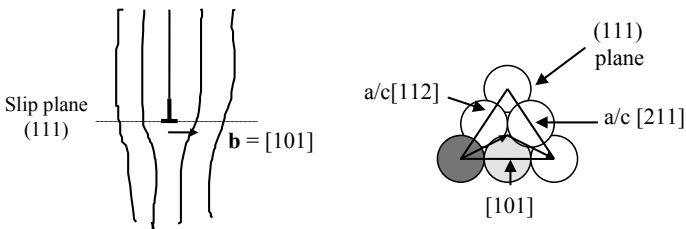


Fig. 3.17 Geometric model of an edge dislocation $b = [101]$ in a (111) plane in a *fcc* crystal, showing the split into two partial dislocations $b = [112]$ and $b = [211]$

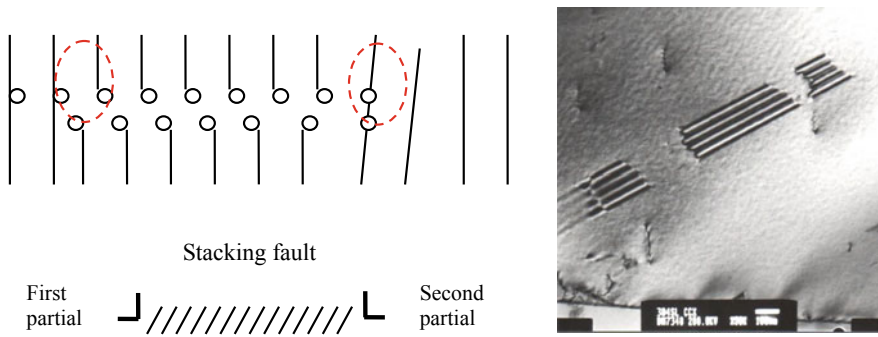


Fig. 3.18 Schematic representation of two partial dislocations with a stacking fault between them. The striped bands are stacking faults between partial dislocations as seen in the TEM

It is clear that stacking faults cannot change from one plane to another, so the partial dislocations are forced to stay on the planes where they were formed, favoring planar slip. The wider the stacking fault is, the less likely that a partial dislocation close. The separation of a stacking fault between two partial dislocations is result of the work balance between the rejection force of the partials (because

they are of the same sign) and the stacking fault energy (SFE). The calculation of the stacking fault width is as follows:

The rejection force F_{re} between two partials is given by the following equation:

$$F_{re} = G \frac{b_1 b_2}{2\pi l}$$

where l is the separation between partial dislocations and b_1 and b_2 are the corresponding Burgers vectors. The SFE is an energy by surface unit, therefore, the work necessary to form it is:

$$SFE = F_{re}(l)$$

Substituting F and solving for r , an equation to calculate l is obtained:

$$l = G \frac{b_1 b_2}{2\pi SFE}$$

This equation indicates that the lower the SFE , the wider the stacking fault between two partial dislocations and so it is more difficult to close them, which favors planar slip. Substitutional solid solutions usually feature lower SFE , thus metal alloys with fcc structure like stainless steels, bronzes, brasses and nickel alloys exhibit planar slip.

Wavy slip—Under normal circumstances, it is quite likely that a dislocation may change of slip plane, this is termed *cross-slip*. This is common in high SFE materials, such as pure metals and interstitial solution alloys, because they do not form partial dislocations and the single dislocations can cross-slip as long as their Burgers vector is compatible between the original and secondary slip planes. The slip lines produced when the dislocations cross-slip are diffuse and wavy, as shown in Fig. 3.19, so that is why this slip mode is called *wavy slip*.

Cross slip allows dislocations to move more freely than planar slip, which leads to the formation of complex dislocation arrays that evolve as the dislocation density

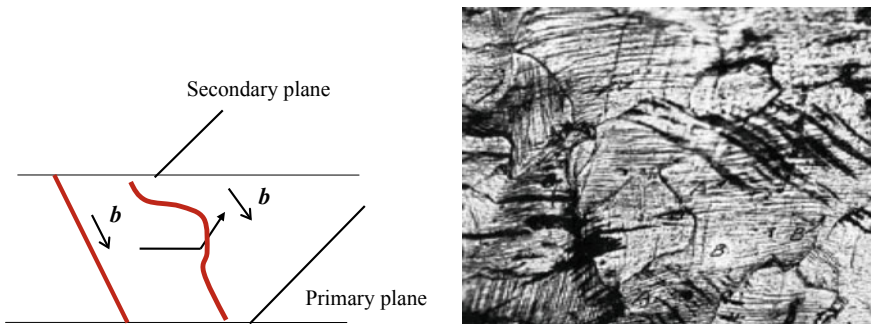


Fig. 3.19 Cross-slip of a dislocation and wavy slip lines as seen in a metallographic microscope

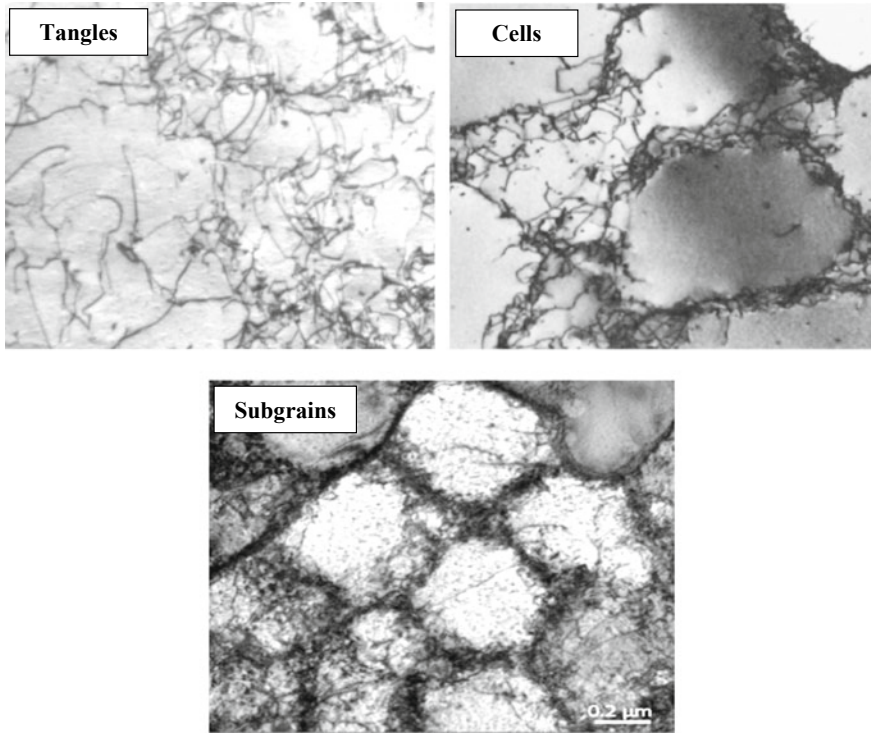


Fig. 3.20 Evolution of wavy-slip dislocation arrays. Initially, tangles are formed, which evolve into cells at higher deformation levels, up to a point where subgrains are formed

increases further. The evolution of dislocation arrays in the wavy slip begins with the formation of *tangles* which are dislocation bundles separated relatively long distances from each other. When the tangles get denser, the dislocation segments get closer to each other, up to a point where they rearrange to form *cells*, that are dense walls of tangled dislocations that surround or enclose low dislocation density zones. Finally, at high deformations, the dislocation cells produce a sufficiently large disorientation between neighboring zones, to lead to the formation of *subgrains*. Figure 3.20 shows the evolution of the wavy slip dislocation arrays as seen in the TEM.

The planar slip mode favors anisotropy of the mechanical behavior, because the plastic deformation is easier in the slip band directions, whereas wavy slip favors more homogeneous and isotropic deformation. Materials with planar slip also feature a more pronounced Bauschinger effect than materials with wavy slip because of the high back stress generated in the dislocation pile ups.

The slip mode also affects recrystallization, which is the formation of new deformation-free grains in cold worked materials after annealing, because the introduced heat causes the rearrangement and annihilation of dislocation arrays.

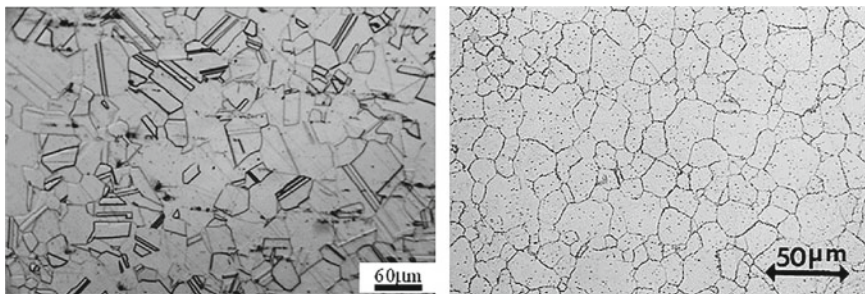


Fig. 3.21 Recrystallized grains of an alloy with planar slip (left) and an alloy with wavy slip (right). Left image taken from Metals 2017, 7(9), 348, <https://doi.org/10.3390/met7090348>

In planar slip materials, the dense dislocation bands favor the formation of annealing twins, and polygonal-shaped recrystallized grains, while in wavy slip materials, the subgrains favor the formation of refined-grain microstructures. Examples of the recrystallized microstructures of planar slip and wavy slip metals are given in Fig. 3.21.

3.6 Deformation by Twinning

Even though dislocation slip is the predominant plastic deformation mechanism at low homologous temperature, there are conditions where dislocation slip is not favored or it is completely inactive. Some of the conditions that inhibit dislocation slip are:

1. Few slip systems.
2. Unfavorable orientation of the slip plane (Schmid factor close to zero).
3. Strong obstacles.
4. High strain rates.
5. Very low temperatures (homologous temperature close to zero).

When any of these conditions is present, the *twinning* deformation mechanism appears, which consists of the shear strain of an entire section of a crystal, produced by the simultaneous movement of atoms in the sheared region. Figure 3.22 shows a schematic illustration of twinning. Deformation by twinning is common in hexagonal metals such as magnesium and zinc, as well as on carbon steels, bronzes and other metals deformed at ballistic speeds, since deformation by twinning occurs at micro second rates, whereas slip rate is slower, in the order of milliseconds.

Twinning occurs on well-defined crystalline planes and crystalline directions, known as *twinning system*. The twinning systems in common crystalline structures are given in Table 3.4. Since twinning produces a region with different crystalline orientation, with plane boundaries, they are clearly visible in a metallographic

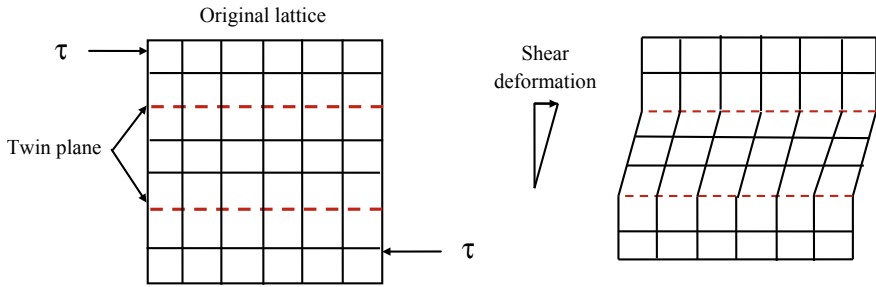
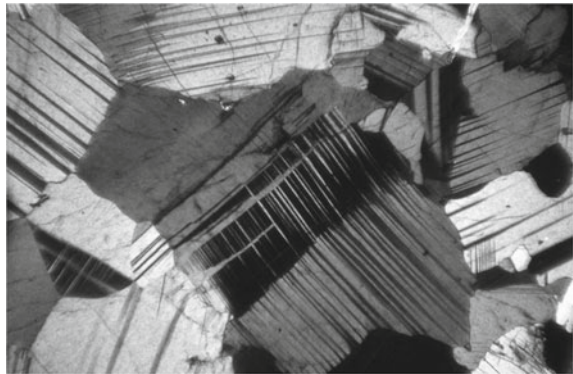


Fig. 3.22 Plastic deformation of a crystal by twinning

Table 3.4 Twinning systems in common crystalline structures

Structure	Twinning plane	Twinning direction	S
<i>fcc</i>	{111}	<112>	0.707
<i>bcc</i>	{112}	<111>	0.707
<i>hc</i>	{1012}, {1121}	<1011>, <1123>, <1126>	0.139

Fig. 3.23 Deformation twins as viewed in the metallographic microscope



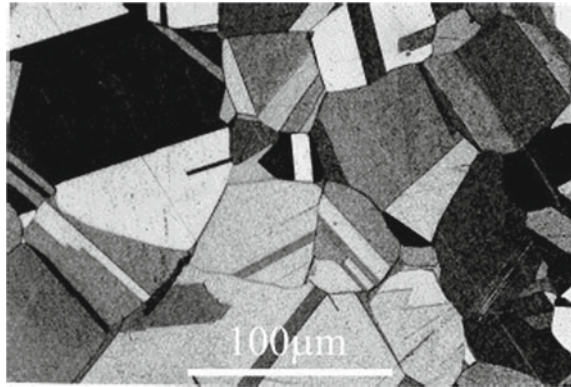
preparation as straight bands across the grains. Deformation twins are fine and numerous, as shown in Fig. 3.23, while annealing twins in planar slip metals are formed by rearrangement of dislocation bands, forming coarser and less numerous bands, as shown in Fig. 3.24.

The magnitude of deformation by twinning can be calculated according to Hall (1954), by the following expression:

$$\Delta l = 1 - (1 + S \tan \chi)^{1/2}$$

$$\tan \chi = 1/2(S \pm \sqrt{S^2 + 4})$$

Fig. 3.24 Annealing twins as viewed in the metallographic microscope



where Δl is the elongation produced by twinning and S is a constant that depends on the structure and the twinning system (see Table 3.3). Based on this expression, the maximum elongation attainable by twinning is about 32%, but the average values obtained experimentally are rarely over 10%.

Twinning is rather unimportant as deformation mechanism because it occurs under quite special conditions and the obtained plastic deformation is not high. However, twinning is very important in the mechanical behavior, because by changing the crystal orientation, it may place some slip systems in favorable slip orientations and others on unfavorable orientations (i.e. low Schmid Factor). Additionally, twins are strong obstacles to dislocation movement, so they contribute to increase the mechanical strength.

Chapter 4

Strengthening Mechanisms



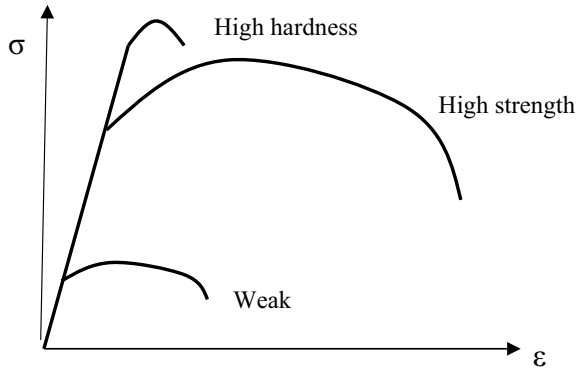
Abstract This chapter starts with an explanation of the concept of strength in engineering materials to introduce to a detailed description of the most important strengthening mechanism of metallic materials, being: strain, grain boundary, solid solution, second phase and fine particle. These mechanisms are explained accordingly to the dislocation theory and the most important rules, along with the relevant equations such as the Hall-Petch law, the rule of mixtures, the interparticle spacing strengthening relation and so on are introduced. The effects of second phase particle shape and distribution on the mechanical strength of materials is discussed. The upper yield strength and strain aging phenomena are described. Furthermore, the martensite hardening behavior in steels is thoroughly described due to its technological importance.

4.1 Strengthening of Engineering Materials

Strengthening of materials can be understood as the increment of mechanical properties in order to increase their capacity to bear loads and resist external forces without failure; specifically, it is achieved by increasing the yield and tensile strength, but without reducing, or better yet, increasing ductility. The property that best describes strengthening is the *toughness*, which is the amount of work necessary to deform a material up to fracture. Occasionally, strengthening is referred to as “hardening” because hardness is directly related to the tensile strength, however a material can be “hardened”, without necessarily make it stronger. For example, a hard material such as quartz is not resistant to mechanical work, since it can be easily broken with a hand hammer; on the other hand, materials termed as “high strength” or “tough”, do feature high hardness values, but combined with high ductility, which make them very strong and resistant to mechanical work. The typical aspect of the tension stress-strain curve of hard, weak and high strength materials is schematically illustrated in Fig. 4.1.

The strengthening of materials can be achieved in several ways, called *strengthening mechanisms*, which all of them are related to the chemical

Fig. 4.1 Schematic stress-strain curves of a high hardness, high strength and weak materials



composition and microstructure to make dislocation movement more difficult, but not to be stopped at all.

The main strengthening mechanisms of engineering materials are:

1. Strain hardening.
2. Alloying.
3. Grain boundaries.
4. Second phases.

Additionally, all materials, whether they are metallic, polymers, ceramics or bio-materials, can be macroscopically combined, in the form of a matrix providing bulk and volume to the piece, and other material, usually of high strength that provides the mechanical strength. These aggregates are called “*composites*”. Their strengthening mechanisms will be treated separately in Chap. 5.

4.2 Strain Hardening

Strain hardening is the increment of the stress necessary to continue plastically straining a material. Since strain hardening always comes with an increment in hardness, but with a ductility reduction, it is adequate to say hardening instead of strengthening for this mechanism. As mentioned in the previous chapter, strain hardening is the cause of the curvature of the stress-strain plot and it is associated with cold work. The strain hardening mechanism involves the interaction among dislocations to generate obstacles that make slip more difficult. As plastic deformation goes further, the dislocation density increases so it does the interactions among dislocations, resulting on an increased difficulty to further deform the material, which is precisely the strain hardening. The absence of strain hardening during hot work is attributed to the annihilation of dislocations by thermally activated processes, taking place at a sufficiently fast rate to prevent the interaction of dislocations.

The types of dislocation interactions that are acknowledged in strain hardening are:

- (1) Interaction of single dislocations.
- (2) Interaction of dislocation groups.
- (3) Interaction of individual and dislocation groups with obstacles.

In most engineering metals and alloys, strain hardening can be so intense that the ductility can be almost totally suppressed, resulting in brittle behavior. As strain hardening increases, the yield limit approach progressively the tensile strength, while ductility is reduced, as shown in the graphs of Fig. 4.2. The term “%Cold work” refers to the maximum plastic deformation attainable before the material has a completely brittle behavior in the next load cycle.

As mentioned above, strain hardening is primarily caused by dislocation interaction, this is due to the fact that when two dislocations moving in different planes intersect each other, a kink is formed in the intersected dislocation line segment, as shown in Fig. 4.3. The dislocation kink has a Burgers vector different from the one of the original dislocation, so, in an edge dislocation, the kink is of the screw type and vice versa, and because the kink's Burgers vector is off plane, it cannot slip along with the dislocation.

Because the dislocation kinks cannot glide along with the dislocations on the plane on which they were formed, they remain rigid (the term used in dislocation theory is “*glissile*”), as the rest of the dislocation keeps on gliding, the segments connected to the glissile segments curve and stretch, forming a pair of dislocation segments of opposite signs separated a short distance, called *dipolar tail*. As the slip continues, the dipolar tail stretches more and more, requiring a higher stress to

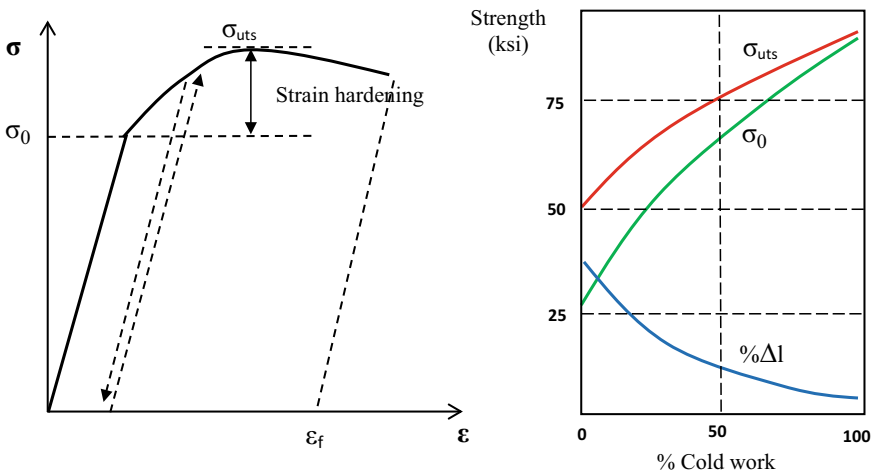


Fig. 4.2 Stress-strain curve of a metallic alloy with strain hardening and variation of the tensile properties of a low carbon steel in terms of % of Cold work

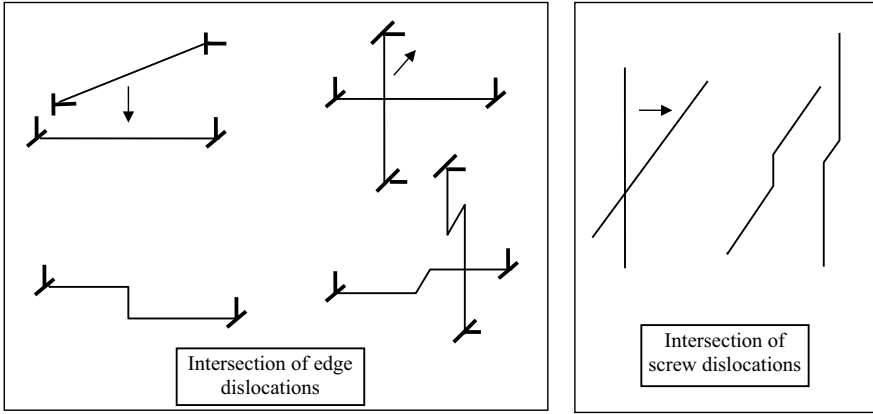


Fig. 4.3 Kink formation by the intersection of dislocation lines

continue moving the dislocation, which appears itself as strain hardening. Finally, when the dipolar tail cannot stretch anymore it collapses, leaving a wake of dislocation rings which contribute to increase the dislocation density and consequently the strain hardening. This process is schematically depicted in Fig. 4.4. Some authors refer to the rings formed by the collapse of dipolar tails as “*dislocation debris*”, which play an important role in other phenomena such as strain aging and fatigue. Figure 4.5 shows a TEM image of a cold worked metal alloy where several dipolar tails (shown with arrows) and dislocation debris can be seen.

The resolved shear stress to lengthen a dipolar tail can be estimated by the following equation:

$$\tau = \alpha Gb/L$$

where L is the dipolar tail length and α is a constant. Notice that the longer the dipolar tail is, the less stress is required to lengthen it, but long dipolar tails will collapse more easily, so a ductile material will harden slowly in the initial stages of deformation, but as the crystal is being filled up with new dislocation rings, strain hardening will increase, simply because more dislocations intersect to each other.

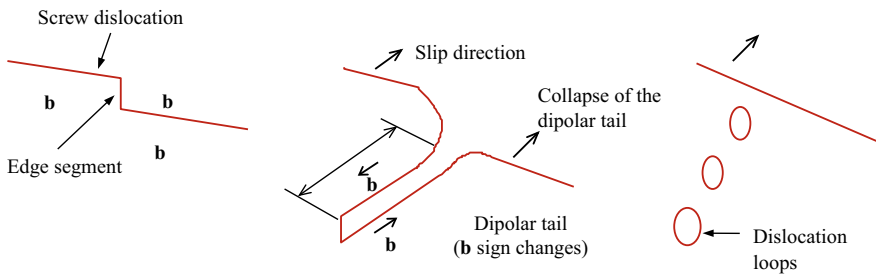


Fig. 4.4 Formation and collapse of a dipolar tail in a screw dislocation with an edge kink

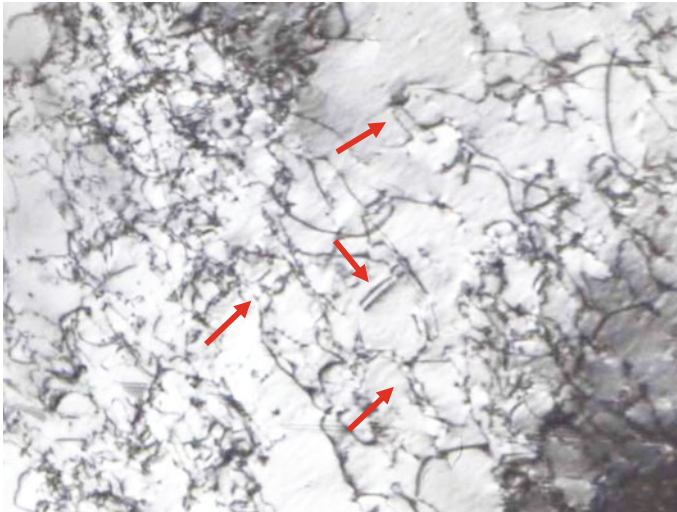


Fig. 4.5 TEM image of a cold worked metal with several dipolar tails (pointed with arrows)

The result of this is that the strain hardening rate is non-constant during the plastic deformation process.

In materials that exhibit planar slip, strain hardening comes from two sources; one is by the rejection forces of the dislocations of the same sign gliding on the same plane and the other is the interaction with a major obstacle where the dislocation band ends, usually a grain boundary; this results in a back stress that has to be overcome by the externally applied stress, resulting in an increment on the applied stress. Upon approaching the obstacle, the spacing among dislocations within the band is gradually reduced, forming a pile-up, as more dislocations incorporate into the pile-up the back stress increases, thus increasing the strain hardening, and at the same time the shear stress in the tip of the pile-up may be so high, that it may induce plastic deformation the neighboring grain. Figure 4.6

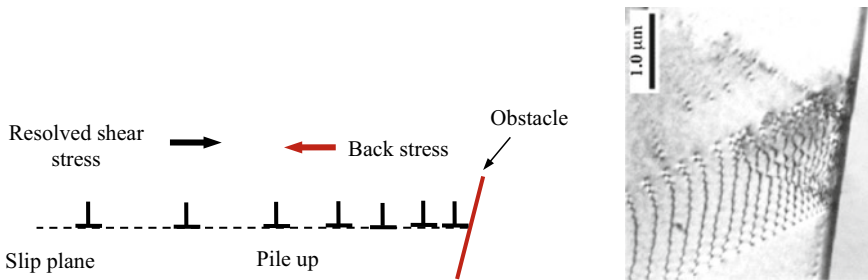


Fig. 4.6 Schematic representation of the formation of a dislocation pile-up and TEM image of a pile-up pushing against a grain boundary

shows a scheme of a dislocation pile-up along with a TEM image of a dislocation pile-up pushing against a grain boundary.

4.3 Grain Boundary Strengthening

Practically since the discovery that engineering metal alloys have a polycrystalline microstructure, where each individual crystal is a “*grain*”, the strengthening effect of the grain size has been studied. The metallurgical engineers quickly discovered that the smaller the size of the grain size, the greater the mechanical strength. This was further explained by the fact grain boundaries are obstacles to the dislocations movement, but this is not the only reason, since grain boundaries have mechanical strength of their own, so there is a mixture strengthening mechanism, similar to that of second phases, which will be discussed further in this Chapter. The plastic deformation of a polycrystalline aggregate can be analyzed as follows. At first, the grain structure can be viewed as a mixture of single crystals, meaning the individual grains, mixed with a material of different mechanical properties located in a narrow strip of material at both sides of the grain boundary and the grain boundary itself, as depicted in Fig. 4.7. Since the material within the grains is naturally anisotropic, and it has a different crystalline orientation with respect to the neighboring grains, it cannot deform freely, but it has to accommodate the deformation of the adjacent grains. Inside the grains, the deformation mechanism is by dislocation slip whereas in the zone close to the grain boundary, the deformation mechanism is more complex, although slip is still active, there is a contribution of grain boundary glide, that has its own critical resolved stress.

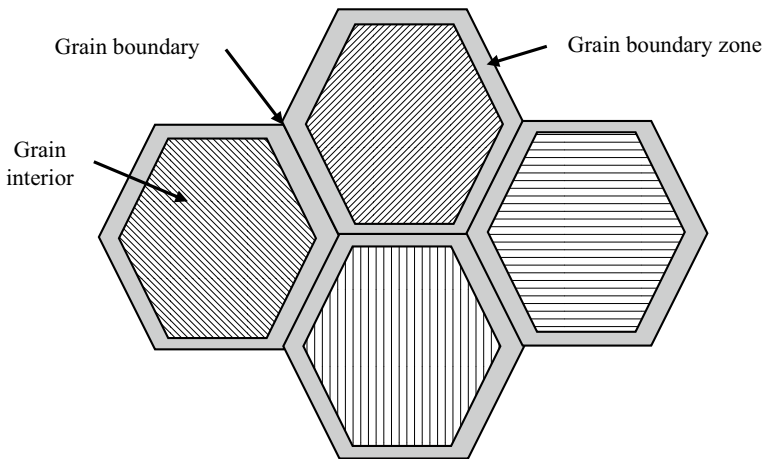


Fig. 4.7 Schematic representation of a polycrystalline microstructure formed by grains (single crystals) with different crystalline orientation, and the grain boundary region

Before going further, it is important to recognize that according to the difference of the crystalline orientation between neighboring grains, there are two types of grain boundaries, which are shown in Fig. 4.8.

- (a) High angle grain boundary: They are boundaries where the misorientation between one grain and another is more than 20° . They have a high surface energy (around 600 erg/cm^2) and it is considered as a region of various atomic layers thick, where the crystalline structure is quite altered, with a high dislocation density and high concentration of vacancies.
- (b) Low angle grain boundary: It is a zone separating two crystal regions with misorientation less than 20° . They have low energy (around 25 erg/cm^2) and they can be viewed as a vertical arrangement of dislocations.

As it was mentioned above, grain boundaries have their own mechanical strength, which is usually measured in tensile test specimens containing only two grains separated by only one boundary called *bicrystals*. The strength of a grain boundary depends on its orientation with respect to the applied stress, by a parameter known as *Taylor's factor*, but more important it depends on the temperature. At low temperatures, grain boundaries are stronger than the material in the grain interior, whereas at high temperatures is the opposite, as shown in Fig. 4.9. The temperature at which both grain boundaries and grain interiors have the same strength is termed as *equicohesive temperature*.

As already mentioned, the effect of grain size in the mechanical strength of engineering metals is well known. As schematically illustrated in Fig. 4.10, a fine grain material is stronger than the same material with coarse grain microstructure.

A rationale of this behavior can be understood from the following analysis. The dislocation slip shear deformation by is calculated by:

$$\gamma = \rho_{\text{dis}} \mathbf{b}x$$

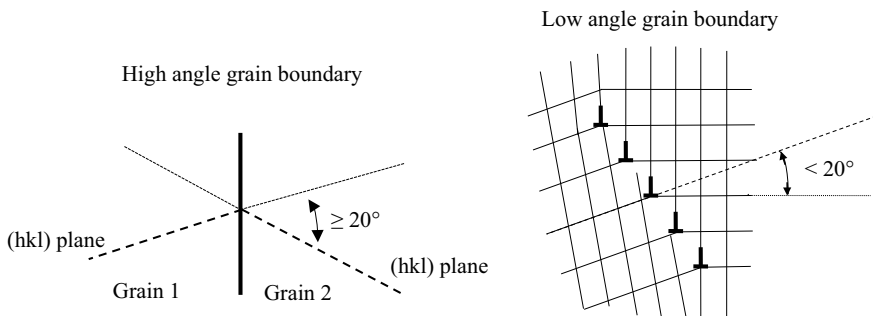


Fig. 4.8 Types of grain boundaries based on the misorientation angle between neighboring grains

Fig. 4.9 Grain (crystal) and grain boundary strength as a function of temperature

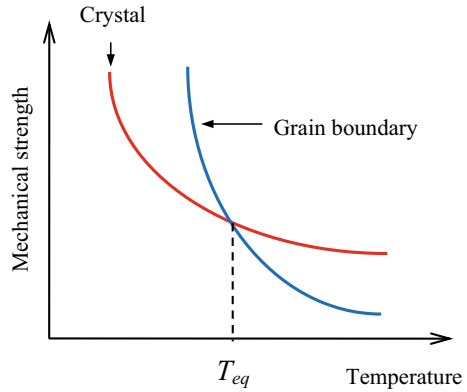
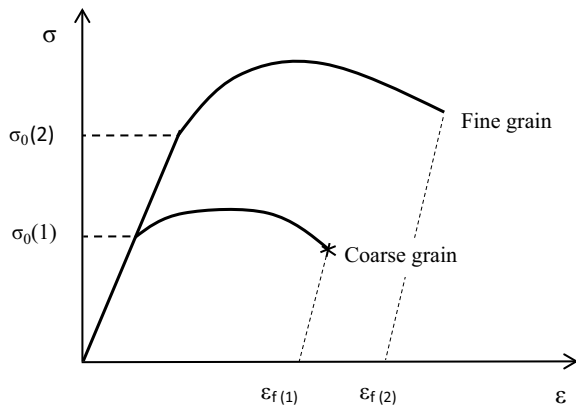


Fig. 4.10 Schematic illustration of the grain size effect on the stress-strain curves on a metallic material. The differences are similar to those observed on experimental curves



Using Hollomon equation for stress-strain $\sigma = k \gamma^n$, and assuming that normal strain e is twice the shear strain γ the following equation can be written:

$$\sigma = 2k(\rho_{dis} b x)^n$$

where x is the average slip distance, d is the grain size and k is a constant. Thus assuming that the maximum value of x is d , substituting $x = d$ and solving for ρ_{dis} , the following equation is obtained:

$$\rho_{dis} = \frac{1}{bd} \left(\frac{\sigma}{k} \right)^n$$

From this equation it is concluded that the smaller the grain size, the greater the dislocation density, consequently it is concluded that a fine grain material will exhibit greater strain hardening, but without losing ductility in comparison with a thick grain material, because it has a greater amount of mobile dislocations.



The Hall-Petch Law. The grain size effect in yield limit has been expressed by the empiric equation known as the Hall-Petch Law, which is expressed as:

$$\sigma_0 = \sigma_i + \frac{k}{\sqrt{d}}$$

where σ_0 is the yield strength and σ_i and k are empiric constants that represent the grain strength and a dislocation release factor respectively. The Hall-Petch law is of great importance in metallurgy and its application has led to the development of ultrafine grain materials that have been a solution to relevant engineering problems such as the construction of steel pipelines in Arctic climates, where high tensile strength, combined with high weldability and low ductile-brittle transition temperature is required, all of this without increasing the alloy content, nor applying heat treatment; this solution was achieved by grain refining techniques. It is worth to mention that this was the case of the Trans Alaskan Pipeline, constructed in the late 1970s, an engineering solution that greatly contributed to end the oil crisis of these years, which otherwise may have led to a global scale conflict.

In addition to increasing the yield strength, the reduction of grain size also improves strain hardening, tensile strength, ductility, fracture toughness, ductile-brittle transition temperature and weldability. All of these makes desirable to always have a fine grained microstructure, however, in grain sizes smaller than 10 nm, the dislocations have little space to move, so the deformation will occur preferably by grain boundary sliding, therefore 10 nm would be the theoretical threshold for fine grain strengthening. On the other hand, fine grain materials have greater internal energy and they will be more reactive in corrosive environments, and at temperatures above T_{eq} , a fine grain microstructure becomes detrimental because the grain boundaries are weaker and suffer cavitation, incipient fusion and preferential oxidation. Thus, for high temperature applications, coarse grain materials or even single crystals are preferred.

Grain size determination: At this point it is important to define grain size. The most common way for determining the grain size is described in the ASTM E112 standard which proposes two methods.

Linear interception (known as the Heyn method). This procedure determines the average grain diameter d , which is defined by the length of one or more straight lines, divided into the number of intersections with the grain boundaries in a bidimensional field, usually in a photograph taken in a metallographic microscope, as shown in Fig. 4.11. In order to obtain a representative measurement, the length of the line must be such that it has at least 50 intersections (if several lines are used, the average is taken). The ends of the lines are counted as $\frac{1}{2}$, an intersection with a triple point is counted as $1\frac{1}{2}$ and the line segments tangent to a grain boundary are counted as one.

ASTM Comparative method. This method is based on the estimation of the number of grains per square inch in a field at $100\times$ magnification. The

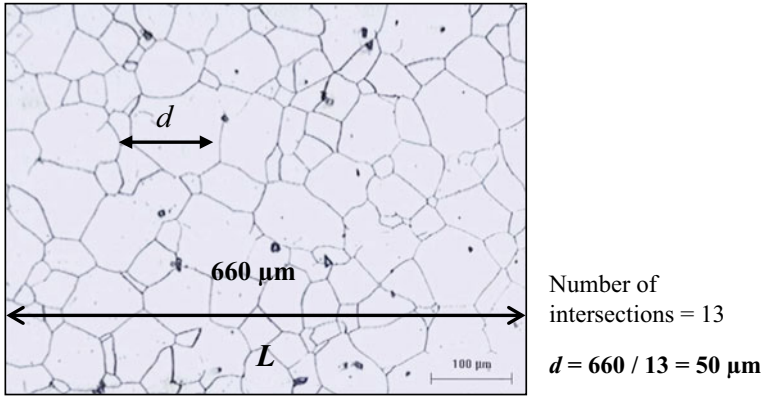


Fig. 4.11 Example of the Heyn method for determining the average grain diameter d

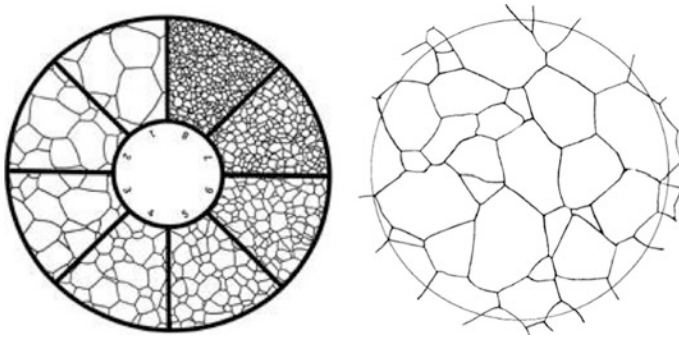


Fig. 4.12 Pattern grid for the determination of the ASTM grain size and example of grain size 3 as viewed in the metallographic microscope at $100\times$

measurement may be done by direct counting or by comparing it with a pattern engraved within the objective lens. The comparison grid is shown in Fig. 4.12 along with an example of a field of one square inch at $100\times$, which has a grain size 3. The ASTM grain size is the exponent G , of the equation that gives the number of grains per square inch N_{AE} , as follows:

$$N_{AE} = 2^{G-1}$$

The approximate equivalence between the ASTM grain size, number of grain per square inch and the average grain diameter is given in Table 4.1.

Table 4.1 Relation between ASTM grain size and average grain diameter

G	Grains/inch ² at 100×	d (μm)
1	1	250
3	4	127
5	16	63.5
8	128	22.5
10	512	11.2
12	2048	5.6
14	8192	2.8

4.4 Solute Strengthening (Solid Solution)

Since the beginning of the use metals in the bronze ages, more than 5000 years ago, it has been known that a combination of metals, meaning to say an alloy, is stronger than the pure metals they are made of. In binary alloys, the maximum strength is obtained at about 50% of alloying element, providing the alloy is a solid solution. The effect of the alloy content in the yield strength and the stress-strain curve is schematically shown in Fig. 4.13.

The Hume-Rothery rules give the conditions that determine the formation of a solid solution of two alloying elements of d_A and d_B atomic diameters:

1. If $d_B < 0.59 d_A$ an interstitial solution will be formed.
2. If d_A is 15% greater than d_B , the solubility of B in A will be less than 1% weight.
3. If the difference in electron affinity is low, a solid solution will be formed.
4. If the difference in electron affinity is high, an intermetallic compound will be formed.
5. If the valence of the solute is greater than the valence of the solvent, the solubility increases.

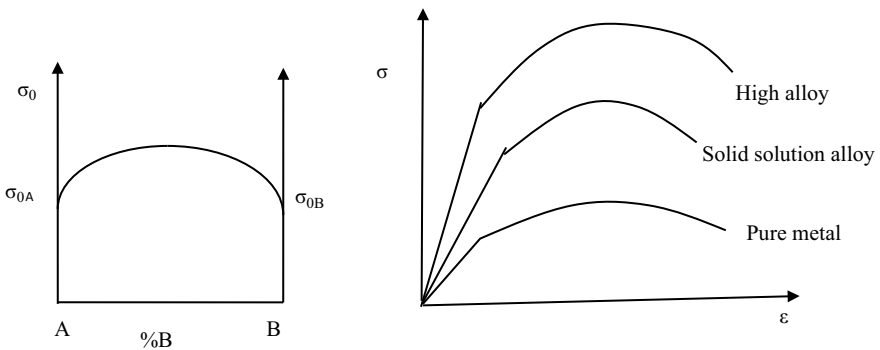


Fig. 4.13 Schematic representation of the effect of alloy content in the tension properties of a solid solution metallic alloy

Solid solution strengthening is due to the interactions among solute atoms and dislocations, it does not matter whether the interaction is attraction or repulsion, because in both cases, strengthening results from the additional stress necessary to continue moving the dislocations over or away from the obstacle. Under attraction forces, the dislocation is initially attached to the solute atom, but when the dislocation moves away, the solute attracts it back, acting as a drag force that increases the resolved shear stress for slip. In case of repulsion, the dislocation is rejected by the solute as it approaches it, thus an additional stress is required to overcome such force.

The interaction among dislocations and solute atoms is controlled by the following factors:

Size factor: it is due to the interaction of the elastic stresses around the dislocation and the stresses caused by the lattice distortion when a foreign atom is introduced, as shown in Fig. 4.14. The magnitude of the distortion within the lattice is proportional to the difference in size between the solute atoms and the solvent, the greater the difference in size, the greater the magnitude of the distortion and therefore more strengthening. Generally, the interstitial atoms produce more strengthening than the substitutional ones (about threefold for the same content).

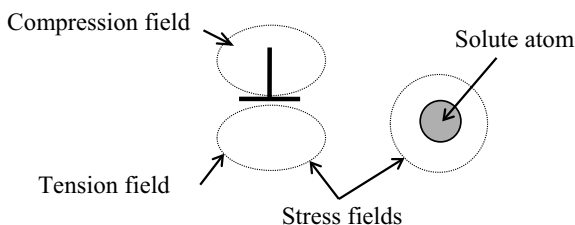
The energy of the elastic field interaction between the dislocation and the solute atom is given by the equation:

$$U_i = 4 G b a^3 \varepsilon' \sin \theta / r$$

where a is the lattice parameter, ε' is the magnitude of the distortion introduced by the foreign atom in the lattice and is calculated by: $\varepsilon' = (a' - a)/a'$, where a is the lattice parameter of the pure matrix, and a' is the lattice parameter of the solid solution; θ and r are the polar coordinates of the distance between the solute atom and the dislocation. According to this equation, the energy of the interaction is directly proportional to the lattice distortion, but notice that the dependence on a is a cubic power, that means that the size factor is much stronger in less-compact lattices, that have greater values of **the lattice parameter**.

Elastic modulus factor: The presence of solute atoms in the lattice changes the elastic modulus, affecting the tension in the dislocation line. As the dislocations move, they rearrange atoms, thus causing variations in the elastic modulus that alter the tension along the dislocation line, as schematically illustrated in Fig. 4.15, which makes the dislocation advance more difficult.

Fig. 4.14 Schematic representation of the interaction between elastic fields of a dislocation and a solute atom



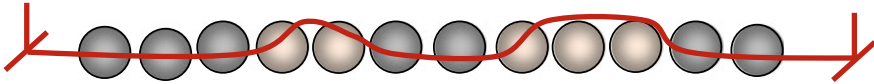


Fig. 4.15 Schematic representation of the alteration in the tension of a dislocation line by the presence of solute atoms

Electrostatic interaction factor: in polyvalent materials, the solute atoms release electrons into the lattice, leaving ions of positive charge. Around the dislocation, electrons tend to migrate to the tension field region, thus forming an electric dipole. Such dipole exerts an electrostatic attraction with positively charged solute atoms, slowing their movement, as depicted in Fig. 4.16.

Chemical interaction: Dislocations, especially partials, have different solubility with respect to the matrix. If the solubility in the dislocation region is higher than in the matrix, a cluster of solute atoms will form around the dislocation core, known as *Cottrell atmosphere*; so, an additional stress will be required to move the dislocation away. In addition, Cottrell atmospheres distort the lattice and increase the energy of the crystal. Such energy increment is obtained from the work done by the external loads, so that the final effect is an increment of the plastic flow stress. This mechanism is schematically shown in Fig. 4.17.

Configurational interaction: A lattice in equilibrium state has a randomly distributed arrangement of solute atoms instead of regularly-ordered groups. The pass of a dislocation alters such arrangements, as depicted in Fig. 4.18. This mechanism increases the lattice free energy, which leads to strengthening, as an additional applied stress is needed to provide the extra work.

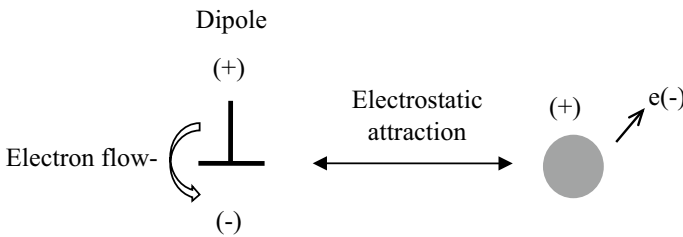


Fig. 4.16 Schematic representation of the electrostatic attraction between a dislocation and a cationic solute atom

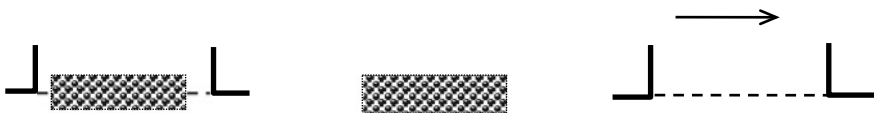


Fig. 4.17 Schematic representation of a partial dislocation with a high concentration of solutes within the stacking fault, which has to be left behind as the dislocation moves away, requiring an additional stress

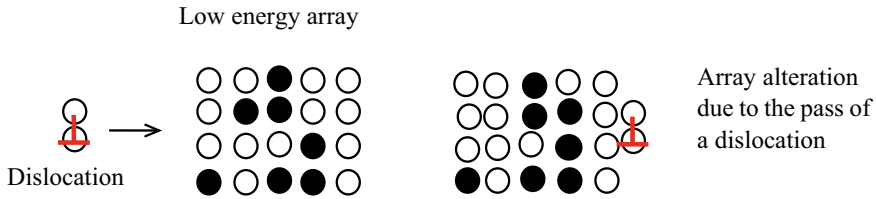


Fig. 4.18 Alteration of the random low energy configuration of a crystal with solute atoms due to the pass of a dislocation

4.5 Second Phase Strengthening

When the content of an alloying element in a solid solution reaches its solubility limit, the solution is divided into two or more constituents of fixed chemical composition and crystalline structure, called *phases*, each phase being a saturated solid solution, with its own mechanical, physical and chemical properties. These mixtures are known as *multiphase* systems. Typically the phase in greater content is the *matrix* and the phases contained within are referred as the *second phases*. From the metallurgical point of view, there are two types of *multiphase* microstructures:

- *Aggregates*, where the size of the second phase is in the same order of the matrix grain size.
- *Dispersions*, where second phases form a dispersion of fine particles contained within the grains and/or over the grain boundaries.

The microscopic aspect of each of these microstructures in metal alloys is shown in Fig. 4.19.

Second phases increase mechanical strength in engineering materials, from moderate increments up to the maximum possible strength, but they may be also detrimental, depending on their physical characteristics, spatial location and the nature of their interaction with the dislocations. This is due to the great number of factors involved in the second phase strengthening, being the most important:

1. Amount of second phase (rule of mixtures).
2. Shape of second-phase particles.

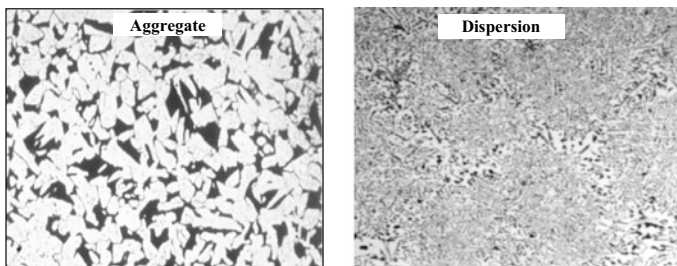


Fig. 4.19 Types of microstructures according to the size of the second phase

3. Size and spatial distribution of second-phase.
4. Mechanical properties of the matrix and second phase.
5. Crystalline structure and orientation of the matrix and second phases.
6. Cohesive strength between the second phase and the matrix.

The amount of second-phase is the most important factor to be considered in second phase strengthening, since the overall strength depends on the contribution of each phase in terms of its volumetric content and mechanical strength. The estimation of the resulting strength of a second phase aggregate can be estimated by *the rule of mixtures*, which is valid under the assumption that the properties of each phase are independent from each other and from the quantity. The rule of mixtures can be expressed in its general form, as:

$$P_{Mix} = \sum f_i P_i$$

where f_i is the volume fraction and P_i is an independent property of the i th component. If two phases (1 and 2) of P_1 and P_2 properties are mixed, the property of the mixture P_{Mix} is:

$$P_{Mix} = f_1 P_1 + f_2 P_2$$

A classic example of the rule of mixtures is the density as seen in the following example.

Example Suppose that two solids of densities $\rho_1 = 10$ and $\rho_2 = 20$ (arbitrary units) are mixed at 50% volume; since it is not expected that the density of one solid affects density of the other, because density is an independent property, the density of the mixture can be calculated as:

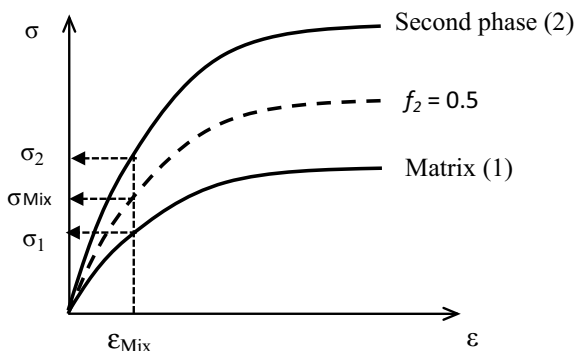
$$\rho_{Mix} = f_1 \rho_1 + f_2 \rho_2 = 0.5(10) + 0.5(20) = 15$$

In the stress-strain behavior of second-phase aggregates, most of the mechanical properties, such as E , σ_o , σ_U , ϵ_f , and etcetera, are independent, therefore the rule of mixtures can be applied to determine the resulting property of the mixture. However, before applying the rule of mixtures for second-phase strengthening it is important to identify the two types of stress-strain behavior between the matrix and the second-phase, these are:

Equal-strain: when the second-phase and the matrix are tightly bonded, it is expected that they will strain equally, therefore the rule of mixtures applies to the strength, as shown in Fig. 4.20. If the hard phase is the second-phase (2) and the soft phase is the matrix (1), and there is a 50% volume mixture, the strain of the mixture is:

$$\epsilon_{Mix} = \epsilon_1 = \epsilon_2$$

Fig. 4.20 Schematic stress-strain curves of a soft matrix (1) reinforced with a hard second phase (2) under equal-strain behavior



And the strength of the mixture is:

$$\sigma_{\text{Mix}} = f_1 \sigma_1 + f_2 \sigma_2$$

As observed in Fig. 4.20, under equal-strain, the stress-strain curve of the 50% volume aggregate is just in the middle of the stress-strain curves of the matrix and the second phase.

Equal-stress: If the matrix and the second phase strain independently from each other, then it is expected that each one will experience the same stress, thus:

$$\sigma_{\text{Mix}} = \sigma_1 = \sigma_2$$

And the strain of the aggregate is:

$$\varepsilon_{\text{Mix}} = f_1 \varepsilon_1 + f_2 \varepsilon_2$$

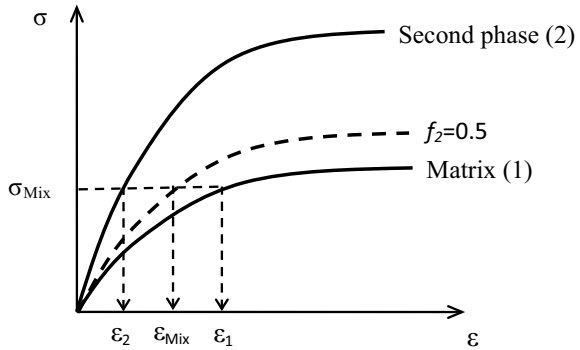
Again, for a 50% volume mixture, the stress-strain curve of the aggregate will be as shown in Fig. 4.21; notice that now, the stress-strain curve of the mixture is closer to the tension curve of the soft phase.

In conclusion, the stress-strain curve of the mixture that behaves in equal-strain behavior is above the stress-strain curve of the equal-stress. In conclusion, when the matrix and the second-phase have equal-strain behavior, the second phase strengthening of the aggregate is more efficient.

It should be mentioned that not all second-phases produce strengthening, even if they are harder than the matrix. For a second phase to produce hardening, at least the following conditions must be met:

- (1) There has to be a strong matrix-second phase bond, so an equal-strain behavior is assured.
- (2) Strain has to initiate in the soft phase, which is usually the matrix, for that, the content of hard phase has to be about 30–50% volume.

Fig. 4.21 Schematic stress-strain curves of a soft matrix reinforced with a hard second phase under equal stress behavior



- (3) If the content of hard phase is over 70% volume, the hard phase will dominate the behavior and the mechanical behavior of the mixture will be similar to that of the hard phase.

The next important controlling factor of second-phase strengthening is the spatial distribution of the second-phase. Table 4.2 summarizes the effect of the second-phase distribution on the strength of a metallic alloy, proving that the second-phase is harder than the matrix. Even when these rules were derived for metals, they apply for most engineering materials.

The shape of the second phase has also an important effect in mechanical strength, as illustrated in Fig. 4.22. In general, spherical particles harden the least, followed by plates and discs, being the needle-shaped (referred as *acicular*) second phases the ones that provide the highest hardness, but they also produce brittleness, therefore, they are not desirable in most microstructures.

Another important aspect in second-phase strengthening is the size and separation among particles. The yield strength is related to a λ parameter, called *interparticle spacing*, according to the following experimental rule:

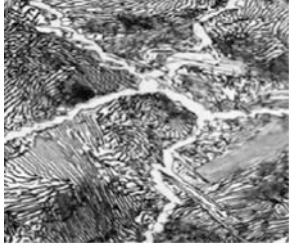
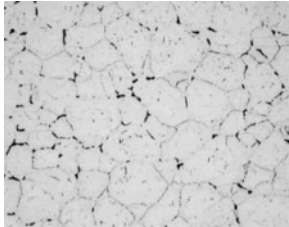
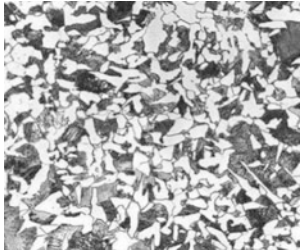
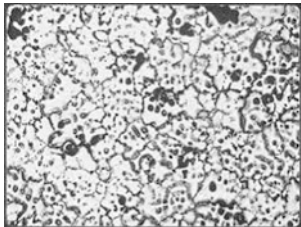
$$\sigma_0 = C \log\left(\frac{1}{\lambda}\right)$$

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

where C is an experimental constant, f is the volume fraction of the second phase particles and r the particle ratio. It is interesting to notice that the term $1/\lambda$ depends both on the number and size of the particles, increasing as f increases, whereas if the particle size is reduced, $1/\lambda$ increases, increasing the yield strength. Therefore, it is preferable to have many particles of small size, instead of few particles and of bigger size, as seen in the next example. Consider the theoretical microstructures of Fig. 4.23.



Table 4.2 Effect of the spatial distribution of the hard phase in the second-phase strengthening

Location of the hard phase	Effect	Example
At grain boundaries as a continuous net	Brittleness	Hypereutectic white irons 
Discontinuous particles along grain boundaries	Moderated brittleness and increased hardness	Sensitized stainless steels 
Aggregate of second phase	High strength and moderated ductility	Medium and high C steels 
Dispersion of fine particles	Optimum combinations of strength and ductility	Aluminum series 7XXX heat treated. 

The microstructure on the left in Fig. 4.23 has two particles with a ratio $r = 0.200$ mm and $f = 0.25$. The value of λ is 0.800 mm, thus $\log(1/\lambda) = 0.0969$. The microstructure on the right of Fig. 4.23 has equal volume fraction of particles, but $r = 0.100$ mm. The value of λ is 0.400 mm, thus $\log(1/\lambda) = 0.3979$. If it is assumed that C is the same for both microstructures, then the reduction of the

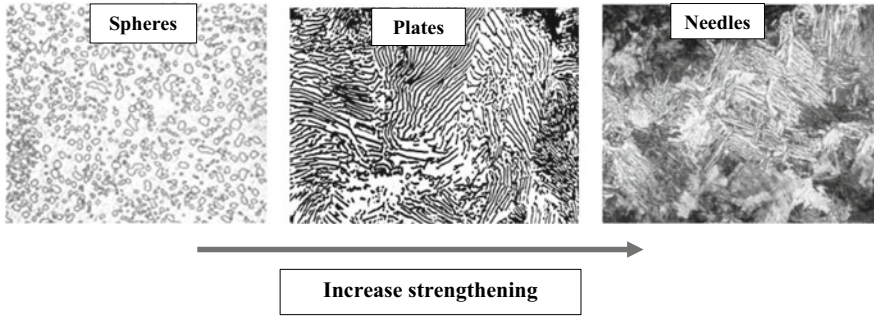


Fig. 4.22 Effect of the second phase shape in the strengthening level

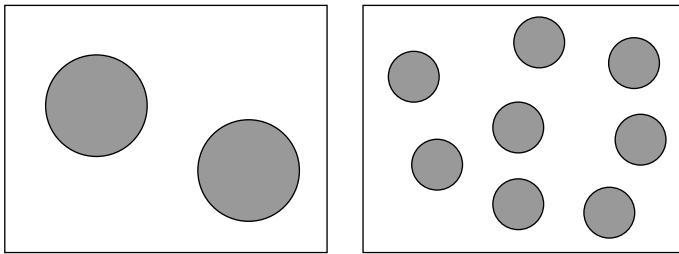


Fig. 4.23 Hypothetical micro structures with equal volume fraction of second phase, but with different size and spacing

particle ratio from 200 to 100 μm , increases the yield limit by $0.3979/0.0965 = 4.1$ fold.

The capability for controlling the mechanical strength through varying the size and shape of second-phase particles by heat treatment, as schematically depicted in Fig. 4.24 has been applied to pearlitic steels to improve productivity, in the manufacture of high strength bolts. Heavy duty bolts are manufactured of medium carbon steel (0.4% C) from grade SAE 5 and up, which have a minimum yield strength of 100 ksi. In addition they have very strict dimensional tolerances and must be made in one-piece, usually by forging. A priori, the fabrication process would be by hot work, but this produces distortion and thick oxide layers that compromise dimensional tolerances. On the other hand, cold shaping in the high hardness condition is very difficult. For that reason, in modern industrial processes, the raw material (coil wire) is ordered in spheroidized condition (the cementite is globular shaped), which is soft and ductile, allowing to fabricate the bolts by cold forging, with excellent dimensional control and neither oxidation nor distortion. Once the bolts are shaped, they are thermally treated by quench and tempering, in controlled atmosphere furnaces, so as to get a microstructure of fine plates, achieving the desired mechanical strength.

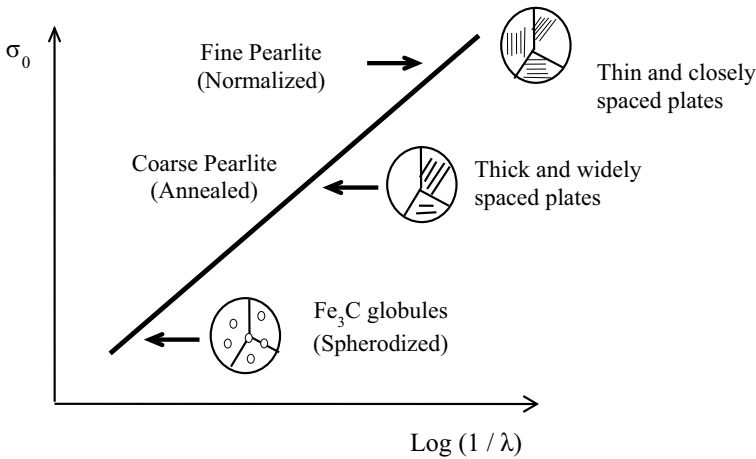


Fig. 4.24 Variation of yield strength in carbon steel as a function of shape and spacing of cementite plates in pearlitic microstructures

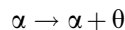
4.6 Fine Particle Strengthening

In the previous section it was stated that the optimum combination of strength and ductility is obtained when the second-phase is in the form of numerous close-spaced small particles dispersed within the matrix. This has led to the development of an entire metallurgical branch meant for the development and manufacture of high strength materials, referred as *fine particle strengthening*, also named: *precipitation hardened* or *dispersion hardened* materials.

There are two ways to obtain a fine particle dispersion in an engineering material:

- (1) Precipitation from a supersaturated solution.
- (2) Mechanical alloying (formerly known as powder metallurgy).

Fine particle precipitation from a supersaturated solution occurs in material systems, metallic and ceramics, that exhibit a phase transformation where the solubility at high temperature decreases during cooling. A common reaction of this type is the Aluminum rich end eutectic transformation in the Aluminum-Copper (Al-Cu) system, which is represented by the equation:



where α is a solid solution poor in Cu and θ is the Cu rich phase (actually θ is a CuAl_2 tetragonal phase with 52 wt% Cu). Figure 4.25 shows the Al-Cu phase diagram of the Al rich zone. The solubility of Cu at $T_1 = 548 \text{ }^\circ\text{C}$ is 5.65 wt% and the solubility of Cu at $T_2 = 21 \text{ }^\circ\text{C}$ is less than 1 wt%. If cooling is rapid enough, the reaction $\alpha \rightarrow \alpha + \theta$ can be suppressed and a supersaturated solution of Cu in Al

(α_{ss}) is obtained, at temperature $T_2 = 21\text{ }^\circ\text{C}$, as shown in the Temperature-Time diagram of Fig. 4.25.

However, the phase α_{ss} is unstable at low temperature, thus after some time, it will start decomposing into $\alpha + \theta$, with the characteristic that the precursor of phase θ will appear in form of fine particles. The precipitation may be accelerated by increasing temperature, but as long as it is lower than the critical temperature, otherwise θ will dissolve into α again.

Another method for obtaining a fine particle dispersion is by a technique called mechanical alloying, which consists in the mixing, pressing and sintering of different metal powders. In this process, a powder of a hard, insoluble and thermally stable material is mixed with a powder of a ductile, low melting point material, which will constitute the matrix. The mixture is compacted and sintered to obtain the final product. By this method, alloy systems that are impossible to obtain by precipitation can be fabricated. For example, an aluminum matrix with tungsten carbide particles. This process is schematically illustrated in the flow diagram of Fig. 4.26.

The precipitates are classified depending on the coincidence of their crystalline planes with those of the matrix, as depicted in Fig. 4.27, as follows:

- (a) *Coherent*. The matrix planes match in all directions with those of the particle.
- (b) *Semi-coherent*. The plane matching is partial.
- (c) *Incoherent*. There is no plane match.

As observed in Fig. 4.27, in coherency, the differences of lattice parameters between the particle and the matrix introduces an elastic deformation field around the particles. Such stress field difficult the dislocation movement by the attraction and repulsion forces between the dislocation and the particle stress fields.

The mechanism of fine particle strengthening depends on whether the particles can be cut by dislocations or not. When particles are cut by dislocations, a shear

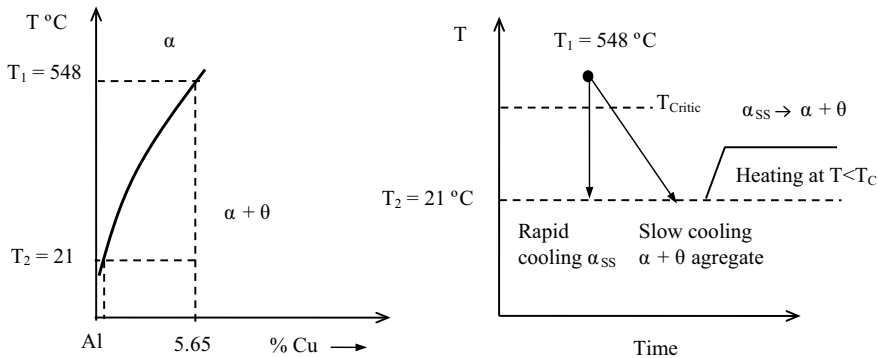


Fig. 4.25 Al rich side of the Cu-Al phase diagram and thermal cycle of a precipitation treatment from a supersaturated solution



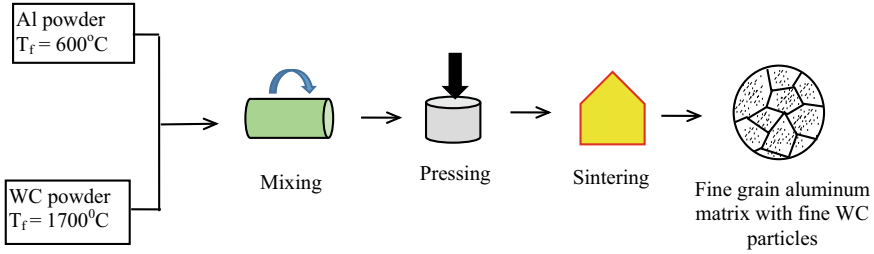


Fig. 4.26 Flow diagram of the mechanical alloying process by mixing, pressing and sintering powders to obtain a fine dispersion of precipitates

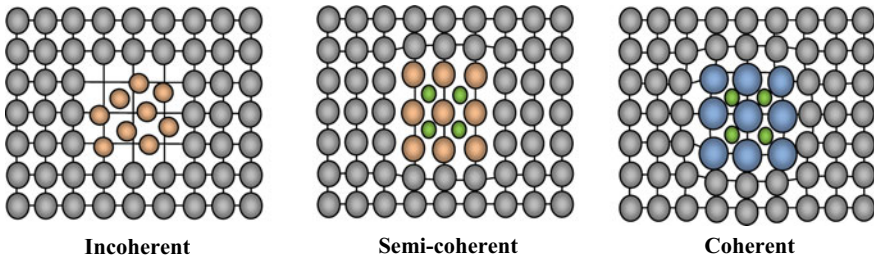


Fig. 4.27 Schematic representation of precipitate coherency

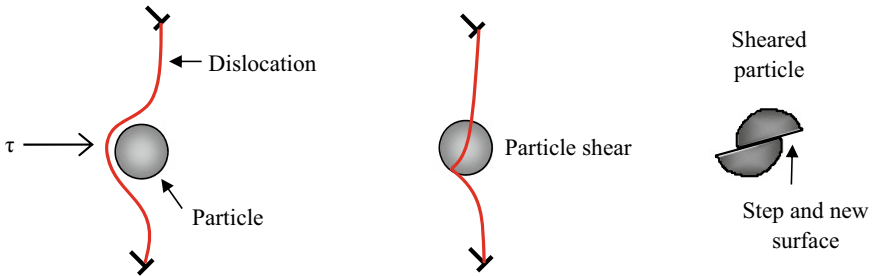


Fig. 4.28 Mechanism of fine particle dislocation cutting

deformation, along with a step in each side of the particle are produced, as illustrated in Fig. 4.28. The deformation strains the surrounding matrix and the new surfaces created by the steps increase the work demand to produce the corresponding surface energy. Since the work is supplied by the load, the result is an increment on the stress required to plastically deform the material.

The strengthening when particles are cut by dislocations comes from the additional stress to deform the particle within the solid, this extra-stress has been calculated by Mott and Nabarro by the following equation:



$$\Delta\sigma = G \varepsilon'^{3/2} \sqrt{rf}$$

$$\varepsilon' = \frac{a'_0 - a_0}{a_0}$$

where r is the particle ratio, f is the volume fraction of particles, G is the shear modulus of the matrix, ε' is a parameter known as *lattice misfit strain*, a'_0 is the lattice parameter of the solid solution around the precipitate and a_0 is the lattice parameter. Additionally, when the particle is sheared, an additional work supplied by the externally applied loads is necessary to create the new surfaces in the steps. This extra-stress has been estimated as:

$$\Delta\sigma = \left(\frac{2G^{1/2}}{\pi} \right) \left(\frac{f\gamma_s}{r} \right)$$

where γ_s is the particle's surface energy. The last contribution comes from the alteration of the dislocation line tension, which increases the internal energy, given by the following equation:

$$\Delta\sigma = \frac{0.8Gb}{\lambda} \left[1 - \frac{E_D}{E_S} \right]^{1/2}$$

where λ is the inter-particle spacing, \mathbf{b} is the Burgers vector, E_D is the Young's modulus of the hard phase and E_S is Young's modulus of the soft phase. By comparing the $\Delta\sigma$ values calculated with the former three equations, it can be seen that the matrix deformation factor dominates, followed by the alteration of the dislocation line tension and the least contribution is by the surface energy factor. It is interesting to notice that in the matrix deformation factor, $\Delta\sigma$ depends on \sqrt{r} , so larger particles at first may produce more strengthening, while for the surface energy contribution, the dependency is on $1/r$, which favors small particles to increase the strengthening. In conclusion, the cutting particle mechanism requires an optimal particle size, not too small, but no too large either, and closely spaced particles in order to maximize strengthening.

When a particle cannot be cut, the dislocation line bows around the particles, forming curved segments that lengthen until they collapse, forming new dislocation rings around the particles. This process is known as the Orowan's mechanism, and is schematically depicted in Fig. 4.29.

The stress necessary to bow a dislocation over a particle to make it collapse can be estimated by the equation:

$$\tau_{RC} = \frac{Gb}{\lambda}$$

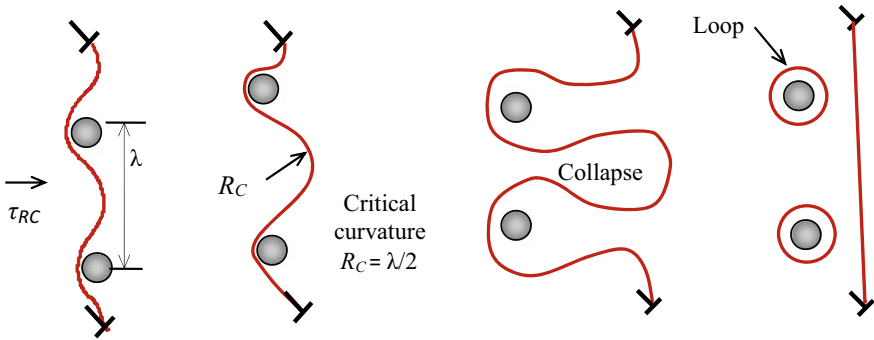


Fig. 4.29 Orowan’s strengthening mechanism by non-cutting particles

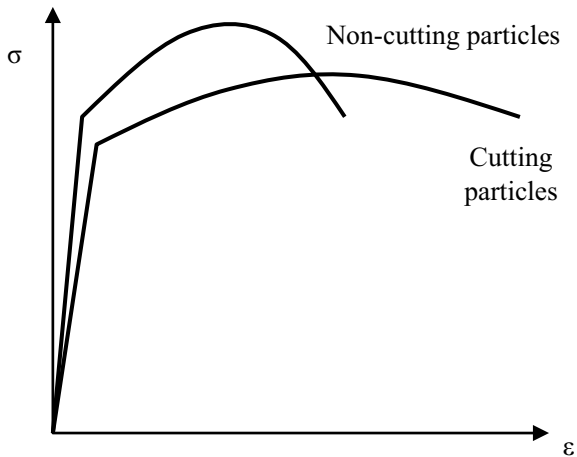
And, the increment of the flow stress $\Delta\sigma$ for the Orowan’s mechanisms is given by the equation:

$$\Delta\sigma = \frac{0.13Gb}{\lambda} \ln\left(\frac{r}{b}\right)$$

According to this equation, non-cutting particle extra-stress linearly increases with the reduction of particle spacing, and logarithmically increases with particle size. At first this may be interpreted that coarse and closely spaced particles will provide the maximum hardening, however the particle size should be small enough to allow the dislocations to bow around them, but not too big to behave as an aggregate, so there is an optimal particle size to achieve maximum strengthening by this mechanism.

As a general rule, materials reinforced with cutting particles will be more ductile and will produce less strain hardening in comparison with materials reinforced with non-cutting particles, as schematically shown in the stress-strain curves of Fig. 4.30.

Fig. 4.30 Schematic stress-strain curves of materials reinforced with fine cutting and non-cutting fine particles



4.7 Upper Yield Strength and Strain Aging

In diluted interstitial alloys at moderate temperatures, like low carbon steel, a peak on the stress-strain curve can be observed, just after the onset of yielding. This peak has been interpreted as two yield strengths: an upper one, which has to be reached to initiate plastic strain, followed by a lower yield strength, at which a localized plastic strain occurs, causing a series of “perturbations” or oscillations in the stress-strain curve, called Lüders deformation, as schematically depicted in Fig. 4.31. The Lüders deformation proceeds until the plastic deformation becomes uniform and the curve recovers its usual parabolic shape. During Lüders deformation a series of surface lines, visible at plain sight, appear in the surface of the tension test specimen, known as Lüders bands, oriented in direction of the maximum shear stress; that is, at 45 degrees from the tensile stress direction.

Lüders bands are regions with the highest Schmid factor, so there the first dislocations are released forming coarse dislocation bands that concentrate slip within a narrow strip of metal. The Lüders bands strain-harden very quickly, so soon after appearing they stop growing and new bands are formed, causing a little drop in the flow stress. Every time this process is repeated, there is a stress perturbation until the material is saturated with Lüders bands, and then uniform deformation initiates.

The upper yield phenomenon may be caused also by the accumulation of interstitial atoms around the dislocations, forming clusters of atoms known as *Cottrell Atmospheres*. The upper yield strength is due to the additional stress necessary to move the dislocations away from their Cottrell Atmospheres, once the dislocations have been released, the plastic strain continues at a lower stress level corresponding to the yield strength of the crystal. When the upper yield is followed by Lüders deformation, the cause is the same, since the first bands to activate are those with the highest Schmid factor.

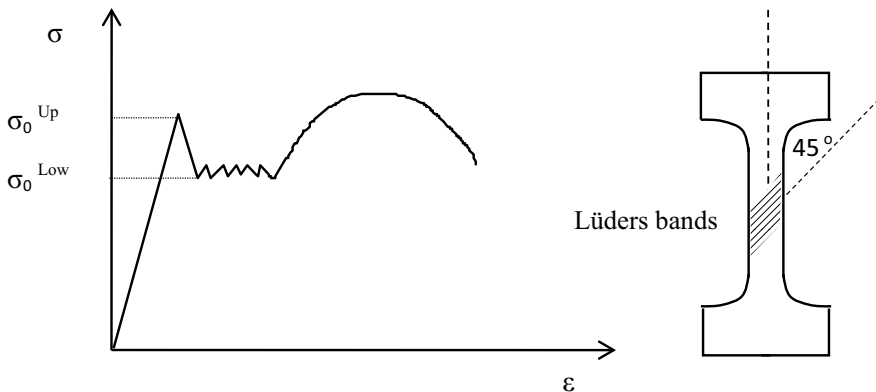


Fig. 4.31 Schematic stress-strain curve of a material with upper yield strength

Lüders bands are considered a flaw in cold formed shapes like bars and sheet plates, so their appearance is undesirable. One way to avoid both upper yield and Lüders bands is to plastically predeform the material by a small amount or apply an annealing heat treatment. In the first case, the dislocations are released from their Cottrell atmospheres so in the next deformation cycle they are free to move at the normal critical resolved shear stress, and in the second, the interstitial atoms are redistributed in the matrix, eliminating the Cottrell atmospheres.

The upper yield phenomenon can also appear in any material if the initial dislocation density is very low and the externally imposed strain rate is very high. This is explained by the following analysis:

The slip strain rate is given by:

$$d\gamma/dt = \mathbf{b} \rho_{\text{dis}} v$$

where v is the dislocation velocity, which depends on stress, as to:

$$v = C \left(\frac{\tau}{\tau_{RC}} \right)^{m'}$$

where C and m' are material constants, τ is the resolved shear stress and, τ_{RC} is the critical resolved shear stress. Then, if the initial dislocation density is low (say 10^3 dislocations/cm²), a greater stress is required to reach the externally imposed strain rate. As the strain goes further, the crystal is quickly filled with dislocations, increasing the dislocation density, so the dislocation velocity is reduced and consequently the required stress drops. In this case, Lüders bands are not formed and the upper yield effect disappears as soon as the strain plastic strain is generalized, as schematically shown in Fig. 4.32.

Another phenomenon that produces a saw tooth appearance of the stress-strain curve, as depicted in Fig. 4.33, is the *Portevin-Lechatelier effect* or also called, *deformation aging*. This phenomenon is caused by the repeated precipitation and dissolution of fine particles during plastic strain, termed *dynamic precipitation*. The Portevin-Lechatelier effect occurs at intermediate temperatures and specific strain rates in metastable alloys.

Fig. 4.32 Schematic strain-stress curve of a material with initial low dislocation density deformed at a high strain rate

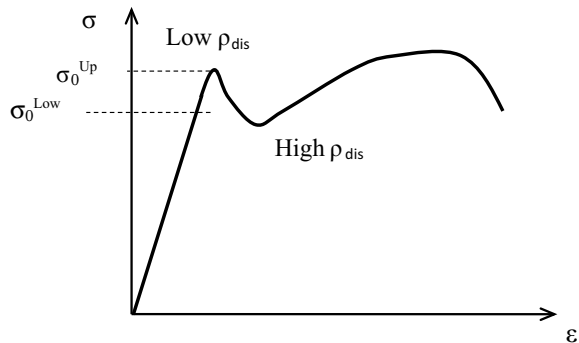
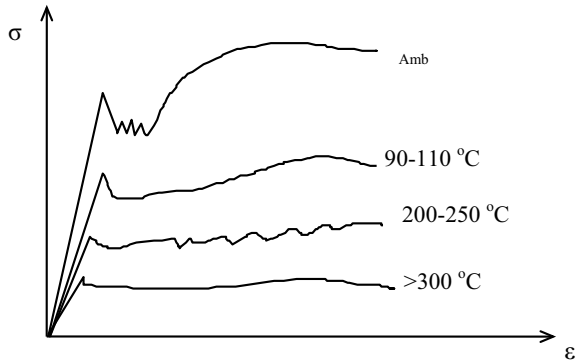


Fig. 4.33 Schematic stress-strain curves of a metastable low alloy steel strained at different temperatures. The curve at 200–250 °C shows the Portevin-Le Chatelier effect



4.8 Martensite Hardening in Steels

Steel is the most extensively used engineering material, mainly because it can display almost every combination of mechanical properties, from soft-and-ductile to hard-and-brittle and, from very soft to ultra-high strength. This versatility is due to two factors:

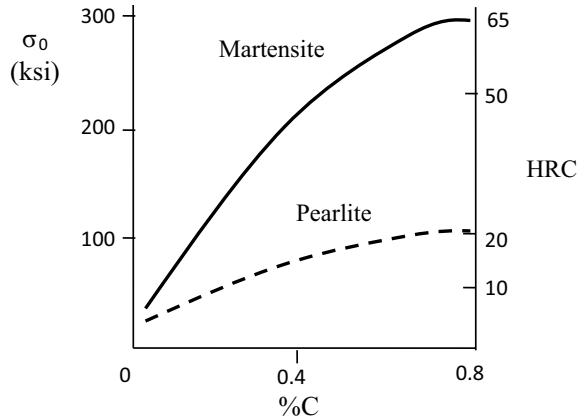
1. The high solubility of substitutional alloying elements, like Cr, Ni, Mo, Mn, and interstitial alloying elements, like C, B, N, in iron (Fe).
2. The phase transformations that the Fe-C system exhibits upon heat treatment.

To better understand this behavior, it is convenient to review the main phase transformations of steel:

- Eutectoid reaction: It is the transformation of austenite into *ferrite* (almost pure Fe) and *cementite* (Fe_3C) that occurs when cooling below 723 °C. Cementite precipitates in the form of plates alternated with ferrite plates, creating a constituent known as *pearlite*. Based on the carbon content, steel can be 100% *ferritic* (less than 0.02% C) up to 100% *pearlitic* (0.8% C).
- Precipitation of fine particles, generally in the form of Cr, Mo, and V carbides.
- Martensitic transformation.

The martensitic transformation is the most important from the technological point of view, because the wide variation of mechanical properties that can produce, as shown in the graph of Fig. 4.34. As it can be seen, martensite can induce as much as three times more hardening than pearlite, depending to the carbon content. Ordinary steel, known as “structural steel” is basically a ferrite-pearlite aggregate. Ferrite is practically pure iron, with a *bcc* structure and very little carbon solubility (0.006% maximum at room temperature), although it can dissolve a great content of alloying elements. Ferrite is soft ($\sigma_o < 70$ MPa, 10 ksi), while cementite is iron carbide (Fe_3C) with a tetragonal structure that has very high hardness values (650 BHN, $\sigma_{\text{UTS}} = 2240$ MPa, 325 ksi). Pearlite is an aggregate of cementite plates alternated with ferrite plates that features mixture strengthening, so the greater the

Fig. 4.34 Variation of yield strength and Rockwell C hardness of steel as a function of carbon content for pearlite and martensite microstructures



pearlite content, the greater the hardness. Martensite, on the other hand, has multiple hardening mechanisms that appear at the same time, all of which provides much more hardening than pearlite, for the same carbon content.

The martensitic transformation in steel is achieved by a heat treatment called *quenching*, which is a rapid cooling from the austenite phase, above 912 °C, followed by a rapid cooling that prevents the diffusion of carbon out of the austenite phase, a necessary condition to form ferrite and cementite, forcing the austenite *fcc* lattice to rearrange into a tetragonal body centered structure, that is implicit in the *fcc* lattice, as shown in Fig. 4.35. This mechanism is called *Bain distortion* and is a bulk transformation that occurs by shear deformation through a process similar to twinning, then being a *difusionless* transformation.

According to the Bain distortion mechanism, martensite is formed from a nuclei that grows inside the austenite grain, accompanied by volume expansion, because of the tetragonal crystalline structure. The shear deformation and volume expansion cause plastic strain in the surrounding untransformed austenite, which results in an additional strain hardening. This process also produces a surface roughness that gives quenched steel pieces a characteristic rough surface aspect, as schematically shown in Fig. 4.36.

The typical appearance of martensite in the metallographic microscope is shown in Fig. 4.37. The microstructure can be visualized as a series of fine plates, similar to twins, alternated with areas of untransformed austenite, known as *retained austenite*. Due to its metastable condition, martensite is easily attacked, showing dark gray tones, whereas retained austenite looks white. The martensitic microstructures are classified as lath, plates or acicular, according to their appearance in a metallographic specimen. Acicular martensite is the hardest and most brittle and therefore is undesirable in most practical applications of quenched steels or in process that involve rapid cooling rates, such as welding.

As martensite has a body centered tetragonal structure with few slip systems, it is quite hard in itself, being this one the main hardening mechanism. In addition, the

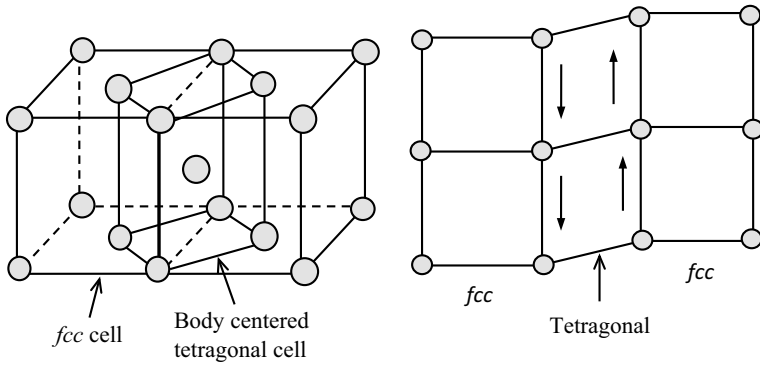


Fig. 4.35 Bain distortion of the austenite forming a body centered tetragonal cell that is implicit in the *fcc* cell. The transformation is diffusionless by shear strain similar to twinning

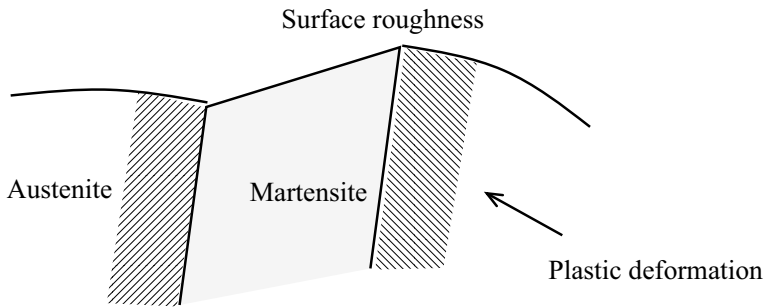


Fig. 4.36 Plastic deformation of austenite and surface roughness, caused by the Bain distortion

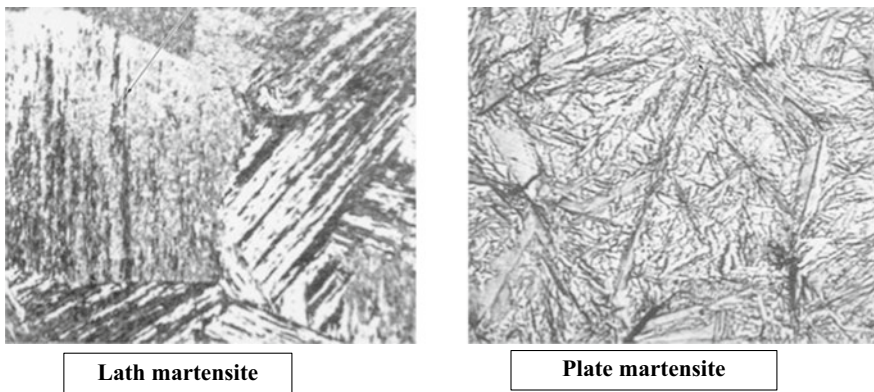


Fig. 4.37 Martensitic microstructures of steel. The light colored regions are retained austenite. Nital attack, bright field

phase boundaries are strong obstacles to the movement of dislocations; carbon, as an interstitial solute, provides solute hardening and there are numerous coherent carbide particles, all of which contribute to the significant hardness levels of martensite. The amount of martensite hardening can be estimated by the following equation:

$$\sigma_o = \sigma_i + \frac{k_o}{\sqrt{d}} + k\sqrt{C} + \alpha Gb\sqrt{\rho_d}$$

where σ_o is the tetragonal cell yield strength, C is the concentration of solute atoms, d is the size of the transformed zone (referred to as the martensite pack size), ρ_d is the dislocation density, G is the shear modulus, b is the Burger's vector and k_o and k are material constants. The first term stands for Pierls stress or lattice friction, the second term is the contribution of the martensite boundaries (similar to grain boundaries), the third term is the solute hardening contribution and the fourth term is the contribution of strain hardening. According to this equation, the finer the martensite plates (small d values) and greater the carbon content, more is the hardening, just as it is known by experience.

Tempering: Since martensite features practically all of the hardening mechanisms, it is expected that, immediately after quench, it should be very hard-and-brittle. For this reason, since ancient times, quenched steels are subject to a further heating treatment, below the austenite transformation temperature (about 715 °C), known as *temper*. The immediate effect of temper heat treatment is the reduction of the dislocation density, which is accompanied by grain refining, and simultaneously, the retained austenite transforms into ferrite and the metastable martensite transforms into ferrite with a dispersion of fine carbides. As the time and temper temperature increase, the martensite and retained austenite transform into the equilibrium phases (ferrite and cementite), and eventually the microstructure will become coarser, reaching low strengths. This reduction of mechanical strength, is compensated by a ductility increment, resulting in a high toughness material. At the microstructural level, tempered martensite is a ferrite matrix, reinforced by fine carbide particle dispersion, which has the metallographic appearance as shown in Fig. 4.38.

The effect of the tempering temperature in the mechanical properties of quenched steel is shown in Fig. 4.39, which is known as “*temper curve*”. Such curves are specific for each steel specification and they are obtained for tempering times of one hour per every 25 mm (1 inch) thickness of material. Temper curves are quite useful, because through them, the mechanical properties of a quench and tempered steel can be chosen, almost at will.

The high degree of fine particle strengthening that occurs in quench and tempered steel has been effectively used for the development of a group of ultra-high strength and high toughness steels named “Maraging”, a term formed by the words “martensitic” and “aging”. The last word referring to the post-quench heat treatment applied to obtain a fine dispersion of precipitates. Maraging steels contain nickel, cobalt, molybdenum and titanium that produce coherent intermetallic precipitates,

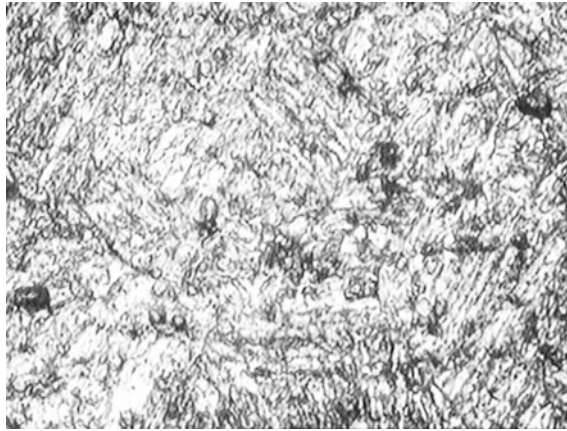


Fig. 4.38 Tempered martensite in steel. Metallographic microscope, Nital attack

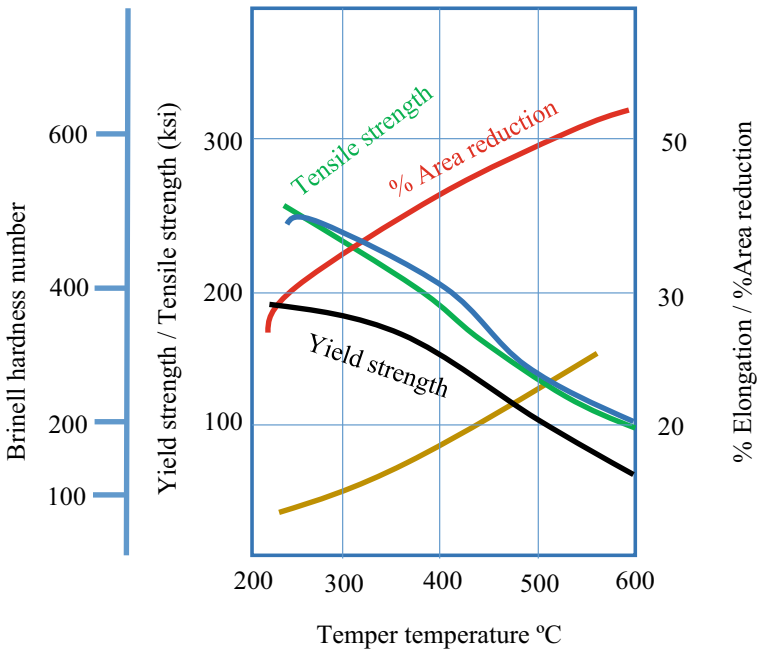


Fig. 4.39 Temper curve, showing the variation of the mechanical properties of quenched steel as a function of tempering temperature (one hour soak time per 1 inch thickness)

which promote precipitation strengthening, in addition to be stable at temperatures up to 400 °C and provide high resistance to corrosion. All of this make Maragin steels an excellent choice for extreme service condition applications.

Chapter 5

Mechanical Behavior of Composites and Polymers



Abstract This chapter is divided in two sections; the first section is an introduction to the nature and classification composite materials, followed with a description of the mechanical behavior of fiber reinforced resin matrix composites, including the concepts of critical fiber length and fiber content to achieve functional composites. The modes of composite materials failure are described along with the most accepted equations to predict the strength of uniaxial fiber reinforced composites. The second section of this chapter deals with the mechanical behavior of thermoplastic polymers, describing the concepts of relaxation and glass transformation, a brief description of the Maxwell and Voight models of stress-strain behavior of thermoplastic polymers and a brief description of the mechanical behavior of elastomers. The chapter finishes with an introduction to the main failure mechanisms of polymers.

5.1 Fundamentals of Fiber Reinforced Composite Materials

Composite materials are macroscopic combinations of two or more materials with different physical properties, produced with the aim of obtaining an intermediate property or rather an improvement of properties, being the most frequently sought: high strength and low density combinations.

A composite material is made up of a matrix which provides shape and volume to the piece to be fabricated, at the same time protects, separates and transmits the applied loads to the reinforcing material. The matrix must not react with the reinforcing material, but has to have good wettability and adherence to it. Usually, the matrix is of low density and should have good formability properties, whereas the reinforcing material should have a high mechanical strength.

Composite materials already exist in nature, such as wood, composed of cellulose fibers in a hemicellulose matrix, or bone, composed of a ceramic called hydroxyapatite reinforced with collagen fibers. Composites are ancient as well, like the straw and mud bricks, used since more than 5000 years ago to construct houses. However, the

first composites considered as engineering materials were developed in the early XX Century, like fiberglass and Bakelite embedded in polymer matrices. Advanced composites may have tensile strengths comparable to those of high alloy steels (up to 2000 MPa, 290 ksi), but their strength-to-density ratio is quite superior, 50–125 [cm^{-1}] for composites, as compared to 5–30 [cm^{-1}] for metals, which makes them really superb materials for advanced applications, such as airplanes and space rockets.

Composites are classified according to the matrix material and type of reinforcement, as shown in Table 5.1. Practically any combination of engineering materials may be used to fabricate composites, but fiber-reinforced composites are the most common type used nowadays, being the most popular those of epoxy resins or other polymers matrix, with tensile strength of 25–200 MPa and density 1.1–1.4 g/cm^3 , reinforced with high strength fibers, such as fiber glass (tensile strength 3100 MPa, density 2.54 g/cm^3) or carbon fiber (tensile strength 3450 MPa, density 1.8 g/cm^3).

In fiber reinforced composites, the fibers should withstand the loads, so they must behave within the elastic regime, whereas the matrix may have an elastic-plastic behavior to some extent. To understand the mechanical behavior of fiber-reinforced composites, consider a single fiber composite stressed under tension, as schematically shown in Fig. 5.1. In this system, the tensile stress is maximum in the middle section of the fiber, while the shear stress is maximum at the ends.

As the load increases, the shear stresses at the composite ends initiate plastic strain of the matrix. On further load increments, plastic deformation spreads out towards the center of the composite, up to the point where the matrix is completely detached of the fiber and the composite fails. Therefore, in order to have an effective composite, the former condition should be avoided by having long enough fibers so the plastic strain of the matrix does not reach the middle section of the fiber. This is called *critical length* (L_c), and is determined by the half fiber length at which the tension force in the fiber equals the shear force in the surface of the fiber, that is:

Table 5.1 Classification of composite materials

Matrix	Metallic	All alloys
	No metallic	Plastics (rigid polymers)
		Resins (solidifying polymers)
		Elastomers (rubbers)
		Ceramics (technical)
	Rocks (ceramic in natural state)	
Form of the reinforcement	Fibers Particles or flakes Plates	Continuum (except particles or flakes) Discontinues

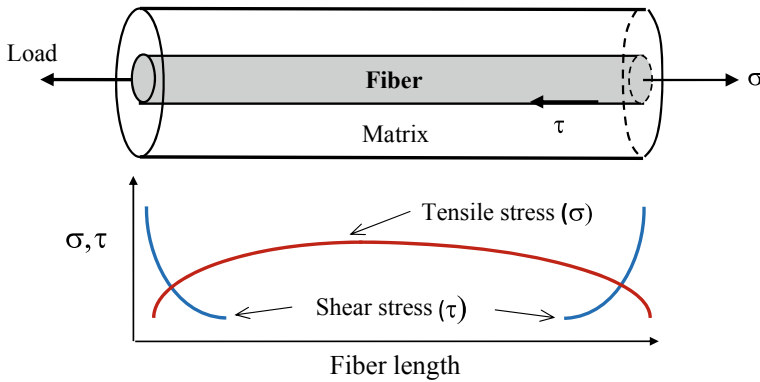


Fig. 5.1 Stress distribution throughout a single fiber composite

$$\sigma_f \left(\frac{\pi d^2}{4} \right) = \left(\frac{\pi d L}{2} \right) \tau_{0M}$$

where σ_f is the fiber tensile strength d is the fiber diameter and τ_{0M} is the shear strength of the fiber-matrix contact area. Solving the above equation for L , gives the fiber critical length:

$$L_c = \frac{\sigma_f d}{2\tau_{0M}}$$

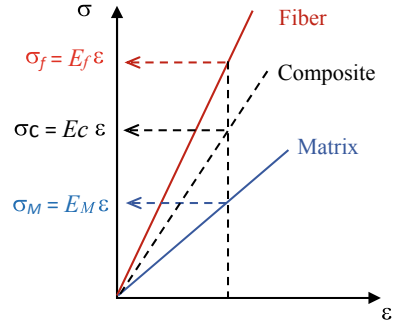
The critical length implies that the fibers may be discontinuous, as long as they are longer than L_c , thus it is not necessary that the fibers be as long as the total length of the piece to be manufactured, which may be a problem in large components, such as aircraft wings or fuselage panels. Since the values of σ_f are tens up to hundreds of times larger than τ_{0M} , and the typical diameter of the fibers should be very small (<0.05 mm), in order to keep the values of L_c in the range of several millimeters. This results in an additional advantage for the fabrication fiber reinforced composites, since it is well known that thin and short fibers are stronger than long and thick ones, and additionally, the moulding and shaping of pieces with thin fibers is a lot easier because they are more flexible. Another advantage of short L_c values is that, if the composite-made component suffers an impact or is cut damaged, the fibers may be still longer than L_c , therefore the composite will be tolerant to the damage.

5.2 Mechanical Behavior of Fiber Reinforced Composites

The stress-strain behavior of fiber reinforced composites, consist of three stages.

Stage I. Elastic deformation in equal strain condition. Under this condition the fiber and the matrix strain equally in elastic regime, so the stress-strain curves

Fig. 5.2 Elastic stress-strain curves of a composite in Stage I (equal elastic strain)



will be as shown in Fig. 5.2. Based on the rule of mixtures, the Young modulus of the composite (E_c) may be calculated as:

$$E_c = E_f f_f + E_M f_M$$

where E_f is fiber Young modulus, E_M is the matrix Young modulus, and f_f and f_M are the volume fraction of the fiber and the matrix respectively. If the composite is made of a single matrix and a single type of fiber: $f_f + f_M = 1$.

Stage II. Quasi-elastic deformation. In this stage, the fibers are elastic, but the matrix begins to yield, persisting equal deformation behavior, hence the composite strength (σ_c) is:

$$\sigma_c = \sigma_f f_f + \sigma'_M f_M$$

where σ_f is the fiber tensile strength and σ'_M is the stress in the matrix when stress in the fiber is equal to the applied stress, that is, when the fibers bear the full load.

Stage III. Composite failure. The failure of the composite can occur by three mechanisms: (1) fiber rupture, which is the desirable condition because it occurs at a stress similar to the tensile strength of the fibers, (2) fiber pull out, the fibers detach of the matrix and they are pulled out, so the strength rapidly drops as more fibers are pulled out; and (3) matrix failure, this is the least desirable mode of failure because it occurs at a stress close to the matrix strength, which is very low. These failure modes depend on the orientation of the fibers with respect to direction of the load, as it will be further described.

As it can be easily understood, the mechanical strength of a composite mainly depends on the fiber content. In order to be useful, a composite should have a mechanical strength greater than that of the matrix. At low loads the composite behaves in Stage I, so the increment of strength as a function of the fiber content is given by the rule of mixtures, but as the loads increase, the mechanical behavior switches to Stage II, and becomes more complex, due to the great difference between the matrix and fiber elastic moduli. In Stage I, the composite's elastic modulus is controlled by the fiber content, as expressed by the following equation:

$$E_C = E_f f_f$$

In Stage II, the composite strength is given by:

$$\sigma_C = \sigma_f f_f + \sigma'_M (1 - f_f)$$

If the fibers are discontinuous, the previous equation is modified as:

$$\sigma_C = \sigma_f f_f \left(1 - \frac{L_C}{2L}\right) + \sigma'_M (1 - f_f)$$

On the other hand, if there are very few fibers, hypothetically, they will fracture rather quickly and the matrix will withstand the full load but with a reduced cross-section area, that is the total cross-section area minus the area occupied by the fibers, therefore the composite's strength is given by:

$$\sigma_C = \sigma_M (1 - f_f)$$

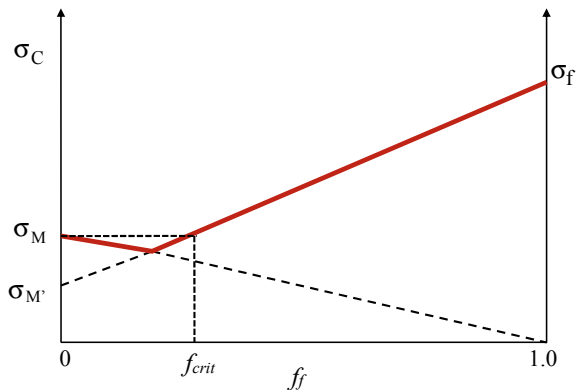
Plotting the σ_C equations for each condition, the curves shown in Fig. 5.3 are obtained.

According to Fig. 5.3, there is a *critical fiber content* (f_{crit}), which is when the composite strength σ_C is equal to the matrix strength σ_M . However, as there is no point on manufacturing a composite if its strength is less than that of the matrix, the fiber content should be higher than f_{crit} . The value of f_{crit} , can be calculated by substituting σ_M into the equation of σ_C for Stage II and solving for f_f , thus:

$$f_{crit} = \frac{\sigma_M - \sigma'_M}{\sigma_f - \sigma'_M}$$

The strength of unidirectional fiber reinforced composites (σ_C), depends on the fiber orientation with respect to the load direction, represented by the symbol ϕ . To

Fig. 5.3 Tensile strength of a composite as a function of fiber content in Stage II for the low fiber content condition (negative slope line) and Stage II behavior (positive slope line)



calculate σ_C as a function of ϕ , consider a composite laminate plate aligned to a coordinate system x, y , let the longitudinal direction of the fibers to be 1 and the direction transverse to the fibers to be 2, as shown in Fig. 5.4.

The of stress and strain transformation equations, from the x, y coordinate system to the 1, 2 coordinate system are:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{Bmatrix} = \begin{bmatrix} m^2 & n^2 & -2mn \\ n^2 & m^2 & mn \\ mn & -mn & m^2 - n^2 \end{bmatrix} \begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{Bmatrix}$$

$$\begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{xy} \end{Bmatrix} = \begin{bmatrix} m^2 & n^2 & -mn \\ n^2 & m^2 & mn \\ 2mn & -2mn & m^2 - n^2 \end{bmatrix} \begin{Bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{12} \end{Bmatrix}$$

where $m = \cos\phi$ and $n = \cos(90 - \phi)$. From the anisotropic theory of elasticity, the constitutive equations are:

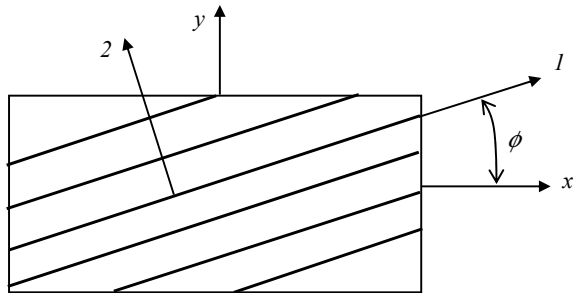
$$\begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ 2\varepsilon_{xy} \end{Bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{16} \\ S_{21} & S_{22} & S_{26} \\ S_{16} & S_{26} & S_{66} \end{bmatrix} \begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{Bmatrix}$$

where S_{ij} is the elastic constant. Here, the short notation of the sub-indexes is used, where: $xx = 1$, $yy = 2$ and $xy = 6$. A similar equation can be written for the longitudinal and transverse fiber directions; to avoid confusions, the elastic constants for these directions is Q_{ij} .

$$\begin{Bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ 2\varepsilon_{12} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{16} \\ Q_{21} & Q_{22} & Q_{26} \\ Q_{16} & Q_{26} & Q_{66} \end{bmatrix} \begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{Bmatrix}$$

From the previous equations, the elastic constants in the longitudinal direction of the laminate (x) can be derived, giving:

Fig. 5.4 Coordinate systems for the analysis of the mechanical behavior of a unidirectional fiber composite laminate, where: x, y arbitrary coordinate system and 1 is the longitudinal direction of the fibers, and 2 the direction transverse to the fibers



$$E_x = \frac{1}{S_{11}} = Q_{11} - \frac{Q_{12}^2}{Q_{22}}$$

$$G_{xy} = Q_{66}$$

$$\nu_{xy} = -\frac{S_{12}}{S_{11}} = \frac{Q_{12}}{Q_{22}}$$

The variation of the elastic constants in the longitudinal direction of the laminate (x) is shown in Fig. 5.5. As expected, the highest value of the Young's modulus corresponds to the fibers parallel to the load direction ($\phi = 0$), and it is equal to the fibers Young's modulus, while at $\phi = 90^\circ$, the Young's modulus is minimum and corresponds to the matrix's value. It is interesting to observe that the Young's modulus drops abruptly (more than 80%) at angles from zero to 20° , indicating that to maintain a high extensional stiffness, the fibers should be as close as possible to the direction of the maximum principal stresses. The shear modulus, on the other hand, is maximum at $\phi = 45^\circ$, but its value is smaller than the isotropic value (given by $G = E/2[1 - \nu]$). This result indicates that the stiffness of a unidirectional fiber composite is largely controlled by the shear modulus of the matrix.

The ultimate tensile strength of unidirectional fiber composites also depends on the ϕ angle. The equations that represent the composite's strength for each failure mode, can be derived from the Mohr's circle shown in Fig. 5.6.

The failure Mode I occurs when the stress component σ_{xx} reaches the ultimate tensile strength of the fibers σ_f ; which, from the Mohr's circle, is:

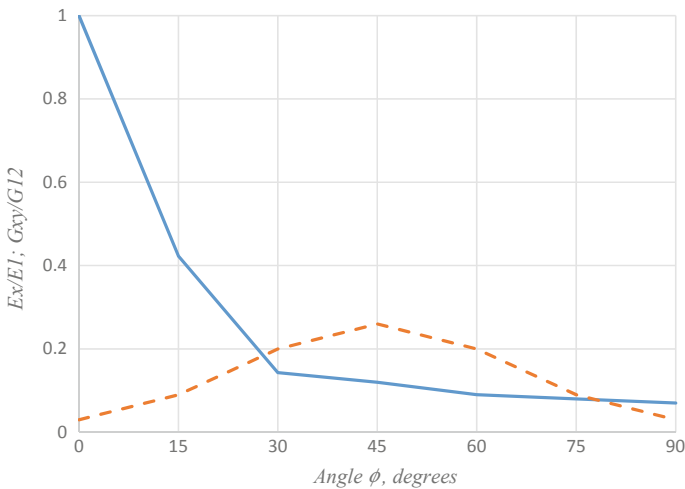
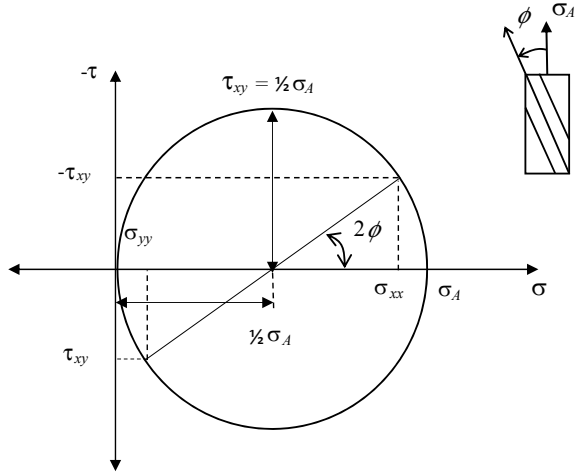


Fig. 5.5 Variation of the Young's modulus E_x (solid line) and the shear modulus G_{xy} (dotted line) with the angle of the fibers of a composite laminate

Fig. 5.6 Mohr's circle of a composite laminate to determine the stress components in the longitudinal, transverse and shear direction of the fibers



$$\sigma_{xx} = \frac{\sigma_A}{2} + \frac{\sigma_A}{2} \cos 2\phi = \frac{\sigma_A}{2} (1 + \cos 2\phi)$$

To simplify the equation, the following identities are used: $\cos 2\phi = 1 - 2\sin^2\phi$; $1 - \sin^2\phi = \cos^2\phi$ and $[\cos^2\phi]^{-1} = \sec^2\phi$. Therefore, if $\sigma_{xx} = \sigma_f$; and solving for σ_A , the composite strength in Mode I is given by:

$$\sigma_{CI} = \sigma_f \sec^2\phi$$

Mode II failure happens when the τ_{xy} stress component equals the cohesion strength of the fiber-matrix interface τ_{fM} , from the Mohr's circle, the shear stress is:

$$\tau_{xy} = \frac{\sigma_A}{2} (\sin 2\phi)$$

By using the identity $(\sin 2\phi)^{-1} = \csc 2\phi$, and solving for σ_A , the Mode II composite's strength is:

$$\sigma_{CII} = 2\tau_{fM} \csc(2\phi)$$

Finally, Mode III failure will occur when the σ_{yy} stress component equals the matrix tensile strength σ_M . From the Mohr's circle, σ_{yy} is given by:

$$\sigma_{yy} = \frac{\sigma_A}{2} - \frac{\sigma_A}{2} \cos 2\phi = \frac{\sigma_A}{2} (1 - \cos 2\phi) = \sigma_A \sin^2\phi$$

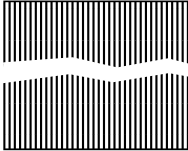
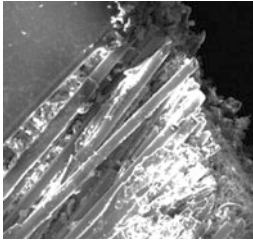
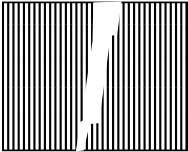
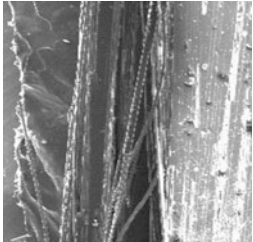
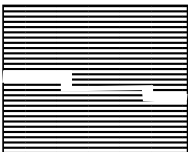
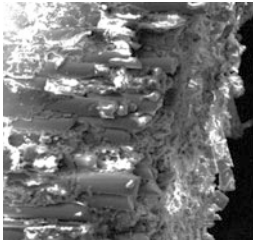
Thus, substituting $\sigma_{yy} = \sigma_M$ and using the identity $[\sin^2\phi]^{-1} = \csc^2\phi$, the composite's strength for Mode III is given by the equation:

$$\sigma_{CIII} = \sigma_M c \sec^2 \phi$$

Table 5.2 shows examples of the three modes of failure of an unidirectional fiber reinforced composite; plotting the equations of Table 5.2, and taking the criterion that composite’s strength is the lowest of each failure mode, the variation of the unidirectional fiber composite strength (σ_C) as a function of the fiber orientation angle (ϕ), is the curve shown in Fig. 5.7. Notice that the composite’s strength is controlled by fiber rupture (Mode I) at fiber angles less than about 8°, while for larger angles the strength is controlled by fiber decohesion modes (II and III), which emphasize the importance of having a very strong matrix-fiber cohesion.

Based on the previous analysis, the design and manufacture of composite materials for structural components under multi axial stresses, require to have enough fibers in the maximum principal stress directions. In practice, this is achieved by placing layers of laminates with fibers in different orientations, obtaining a so called *laminated composite* or sandwich type. Another common practice is to manufacture a fiber weaving, similar to a fabric. Figure 5.8 shows schematically examples of such arrangements.

Table 5.2 Tensile strength of unidirectional fiber reinforced composite and failure modes

Failure mode	Tensile strength	Example
Mode I: Fiber rupture	$\sigma_{CI} = \sigma_f \sec^2 \phi$ 	
Mode II: Fiber decohesion	$\sigma_{CII} = 2\tau_{fM} c \sec(2\phi)$ 	
Mode III: Matrix failure	$\sigma_{CIII} = \sigma_M c \sec^2 \phi$ 	

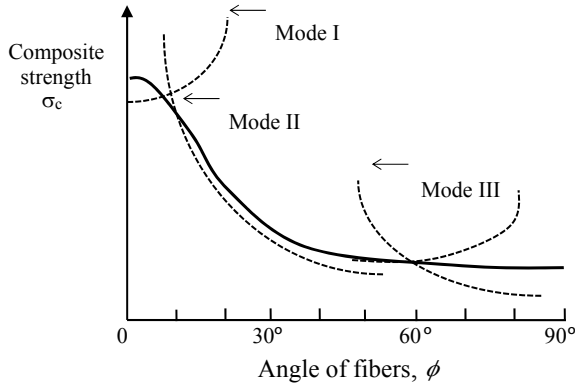


Fig. 5.7 Tensile strength of a composite with unidirectional fibers as a function of the fiber angle with respect to the load axis

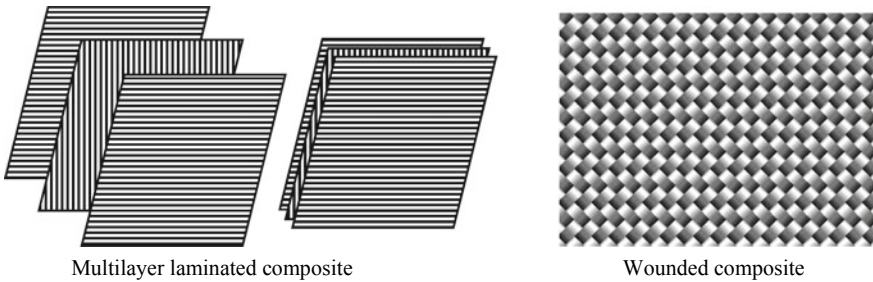


Fig. 5.8 Strategies to ensure having enough fibers in the maximum principal stress direction of a laminated composite

5.3 Mechanical Behavior of Polymers

A *polymer* is a material formed by long molecular chains that are constituted by molecular units called *monomers*. The most widely known polymers are those formed by carbon-hydrogen molecules, but there are also other molecules that make polymers, such as silicon oxide. When monomers link to each other to form a chain, the process is called *polymerization by addition*; like polyethylene, which results from the linkage of ethylene molecules. If monomers of different kinds join together, the process is called *copolymerization*, like polyvinyl chloride (PVC) that comes from the polymerization of vinyl chlorine and vinyl acetate. When double joint links in monomers “open”, forming joint bonds, it is said that polymers are *bifunctional*. Polymerization by *condensation* takes place when monomer units chemically react to form a chain, with a molecular product in each point of reaction.

In polymers, the main chain joints are composed of primary links, so they are strong and stable, whereas the chains are joined to each other by weak links of the

Van der Waals type, cross links, covalent links or hydrogen bridges, giving the polymer its great flexibility and ductility.

According to their response to temperature, polymers are classified as.

Thermoplastic. They soften as the temperature increases and when cooling, they recover their original properties, keeping their molecular form. Typically, they have chains joined by Van Der Waals links that, when subject to stress, can separate, rotate and glide over themselves. An example is polyethylene.

Thermoset. They keep their mechanical strength and molecular structure up to certain temperature, after which they chemically break down. Typically, they have chains joined by tridimensional grids of covalent links. An example is Bakelite (polyphenol-formaldehyde).

Generally, it is assumed that polymers are amorphous materials, but many of them are constituted by a mixture of small crystals (about 10 nm long), surrounded by amorphous zones in an edge-micelle type array, as schematically shown in Fig. 5.9.

The polymer's crystal morphology is typically in form of layers with folded molecular chains, perpendicular to the plane and their size and quantity depend on the cooling rate. In polymers such as nylon and polyethylene, the crystals may group up in structures called *spherulites*, which are visible in the microscope under polarized light, as shown in Fig. 5.10. Spherulites may turn into micelles when the polymer is plastically strained.

The mechanical behavior of polymers is essentially the same as in any solid: the externally applied loads produce stresses, stresses generate strains, that are first elastic, and then plastic and when the material can no longer strain, it fractures. Thermoset polymers usually feature a brittle behavior, whereas thermoplastic polymers, exhibit a non-linear elastic deformation combined with time-dependent plastic deformation. Thermoplastic polymers are not very strong (the tensile strength is up to about 140 MPa, 20 ksi), but they feature very high elongations, of several hundreds of percent prior to rupture. Figure 5.11 shows the typical stress-strain curves of common polymers.

Another characteristic of thermoplastic polymers is that, at a fixed strain, the stress may drop through time; this phenomenon is called *relaxation*. The *Relaxation Modulus (E_r)* is the stress/deformation ratio after a given time. Usually E_r depends on temperature, the degree of crystallization and the type of bond. Figure 5.12 shows the variation of E_r with respect to temperature of polystyrene; it is observed that crystalline and cross link polystyrene have two plateaus, one at low temperatures, called



Fig. 5.9 Schematic representation of the crystalline and amorphous regions in a polymer

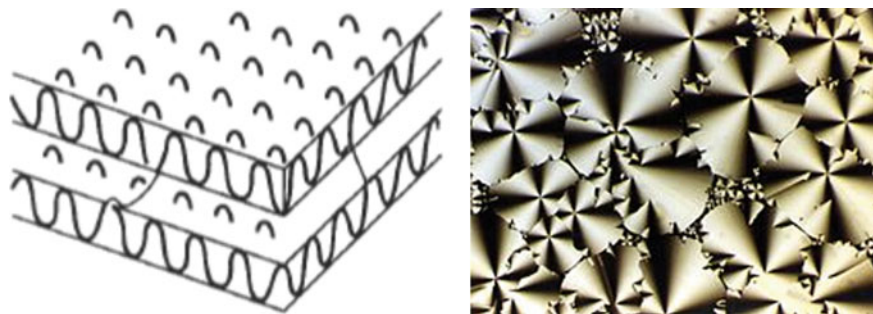
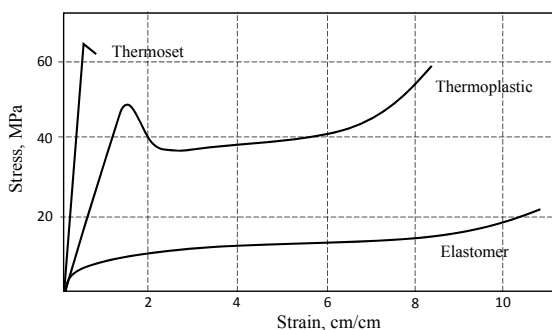


Fig. 5.10 Folded chains in polymeric crystals and spherulite crystals in nylon. Image taken from [https://en.wikipedia.org/wiki/Spherulite\(polymer_physics\)](https://en.wikipedia.org/wiki/Spherulite(polymer_physics))

Fig. 5.11 Schematic stress-strain curves of polymers



glassy, and a second one at intermediate temperatures, called *rubbery*, while amorphous polystyrene is always rubbery. At elevated temperatures, the values of E_r drop sharply because the secondary bonds melt, and the material exhibits a *viscous flow*.

Another important property of polymers is the *glass transition temperature* (T_g), which is the temperature at which the elastic behavior changes into a viscoelastic, and the deformation becomes time-dependent. The value of T_g is determined by the point where the slope of the specific volume V_e versus *Temperature* plot changes, which normally is between 0.5 and 0.75 of the melting temperature, as shown in Fig. 5.13. The glass transition is caused by the weakening of the secondary bonds that allow the molecules to glide and twist over each other.

The stress-strain behavior of polymers in the viscoelastic regime can be represented by mechanical models formed by an array of helical springs and hydraulic dashpots. The springs represent the elastic strain caused by the stretching of “frozen” molecules and the dashpots represent the viscous flow, caused by the glide and rotation of polymer chains. In most polymers, at temperatures lower than T_g , the spring strain dominates, so the deformation is proportional to the load, whereas at intermediate temperatures, the spring elongation is delayed by the dashpot, so the

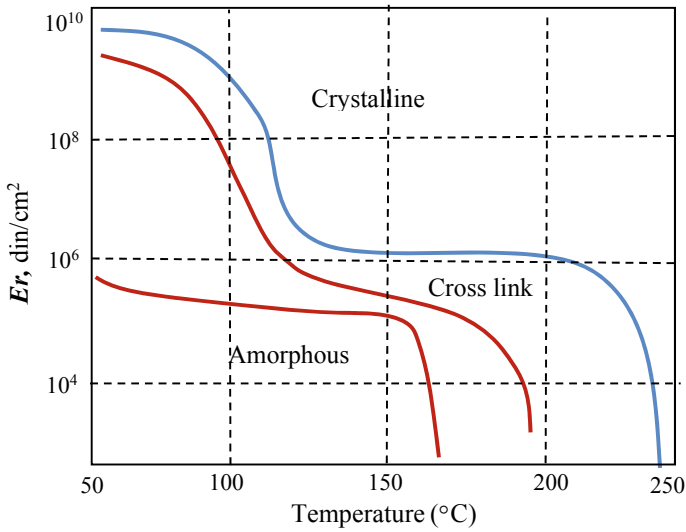
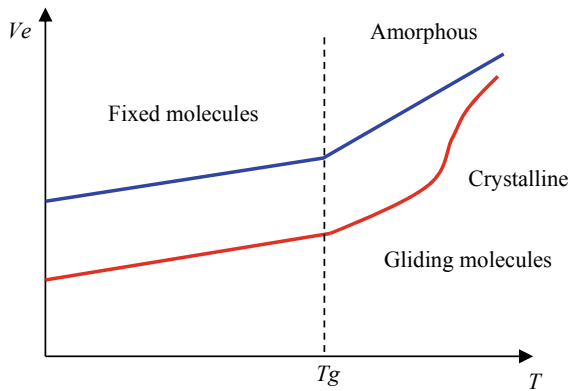


Fig. 5.12 Relaxation modulus of polystyrene after 10 s

Fig. 5.13 Specific volume curves in terms of temperature, showing glass transformation temperature



elastic strain is time dependent and finally, at high temperature, the viscous flow controls the strain, so it can be represented by the dashpot alone.

At low temperatures, where the glassy behavior dominates, the stress-strain behavior is represented by the Maxwell's model, which consist of a spring connected in series with a dashpot, as shown in Fig. 5.14.

If the spring follows Hooke's law, its strain is given by $e_E = \sigma/E$. If the fluid in the dashpot is Newtonian (the viscosity coefficient is independent of the strain rate),



Fig. 5.14 Maxwell's model for glassy behavior; the strain is initially elastic, controlled by the spring, followed by viscous flow, controlled by the hydraulic dashpot

the viscous strain given by $e_V = \sigma t / \eta$ (where t is time and η is viscosity coefficient). Therefore, in Maxwell's model the total strain will be:

$$e = e_E + e_V = \frac{\sigma}{E} + \frac{\sigma t}{\eta}$$

If after a time t_i , the load is removed, polymer remains strained because the dashpot keeps its length, but the stress will drop because the relaxation of the spring, according to the following equation:

$$\sigma = eE \exp\left[-\frac{Et}{\eta}\right]$$

The Maxwell's behavior is schematically depicted in Fig. 5.15. Notice that when $t = E/\eta$, the previous equation is reduced to $0.3679 Ee$ and strain is in a steady state, and the value of t is the *relaxation time*.

At intermediate and high temperatures with respect to the T_g , the behavior is rubbery and the strain is simultaneously elastic and viscous, which is represented by the Voigt's model, consisting of a parallel system of spring and dashpot, as shown in Fig. 5.16.

Again, if the polymer is under a constant stress, the total strain is:

$$\sigma = \left(\frac{\sigma}{E}\right) \left[1 - \exp\left(-\frac{Et}{\eta}\right)\right]$$

And if after a time t_1 , the load is removed, the strain is given by:

$$e = e_1 \exp\left(-\frac{Et}{\eta}\right)$$

The Voigt's model plot is shown in Fig. 5.17.

As it may be foreseen from the previous description, the mechanical behavior of polymeric materials is greatly dependent on the temperature. Figure 5.18 shows the effect of temperature on the mechanical behavior of a thermoplastic polymer. At temperatures below T_g the behavior is similar to a thermoset polymer, which is linear-elastic with brittle fracture. At temperatures closer, but still under T_g , the behavior is elastic-plastic, with ductile tearing failure mode. Finally, at temperatures equal or greater than T_g , the behavior is viscous-elastic, just as described in the previous section.

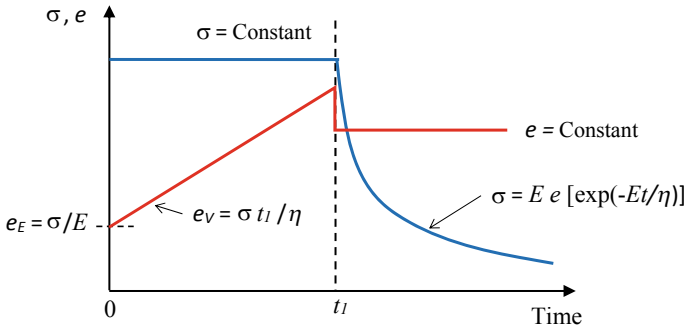


Fig. 5.15 Stress-strain behavior as a function of time for the Maxwell model

Fig. 5.16 Voigt's model for rubbery behavior, where strain is simultaneously elastic and viscous

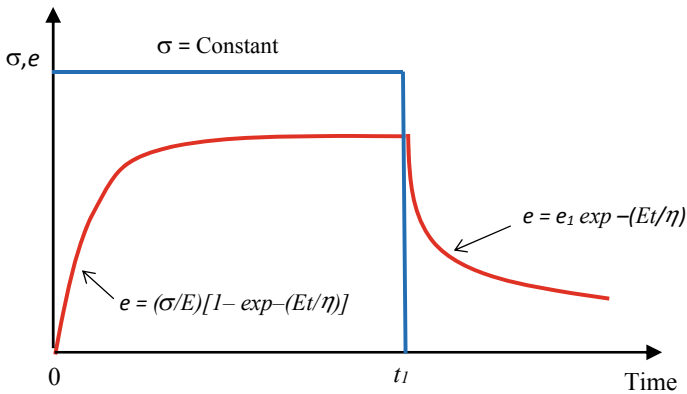
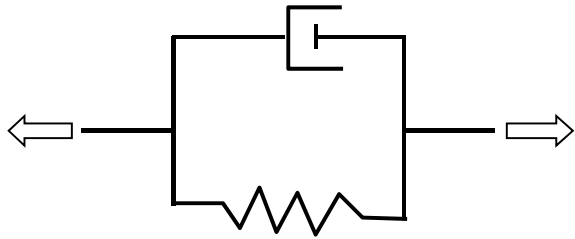
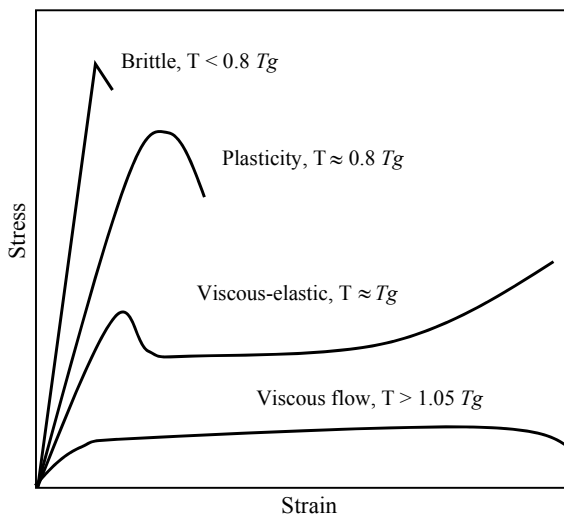


Fig. 5.17 Voigt's model stress-strain behavior for rubbery polymers

Fig. 5.18 Effect of temperature on the stress-strain behavior of a thermoplastic polymer. Adapted from [1]



5.4 Mechanical Behavior of Elastomers

Elastomers are polymeric materials that exhibit large elastic deformations, up to hundreds of percent, but with an elastic linear elongation of no more than 1%, while the rest of the elastic deformation is non-linear; they also show little plastic deformation, so their fracture is, ironically, brittle. Elastomers are usually very soft, they have a low elastic modulus and their stress-strain curve does not show a well-defined yield strength, but it is normally very close to the tensile strength. The tensile strength of commercial elastomers ranges from 5 to 35 MPa (700–5000 psi) and the elongations may be up to 2000%. The stress-strain curve of an elastomer is schematically shown in Fig. 5.19.

The elastomeric behavior of rubber is observed at temperatures higher than T_g and is attributed to the rotation of polymer chains and the folding of monomer joints, which allow the molecular chains to, first align and then stretch, as schematically shown in Fig. 5.20. When an elastomer does not have cross links, it deforms in an elastic-plastic mode, leaving a small plastic deformation when the load is removed. When the elastomer does have cross links, the deformation is totally elastic, so when the load is removed, the material recovers its original form and dimensions.

One of the most widely used elastomers is synthetic rubber, which comes from a natural rubber collected by the ancient Olmec and Maya cultures from Mexico of a tree called “chicle”, for making balls, shoes and water proof textiles. In 1839 Charles Goodyear developed the *vulcanization* process to produce a very strong rubber that was used to make car tires, bands, gloves among many other things. Modern rubber is in most cases, isoprene mixed with chloroprene and isobutylene, which are copolymerized by vulcanization, in which cross joints are formed by the

Fig. 5.19 Schematic stress-strain curve of an elastomer

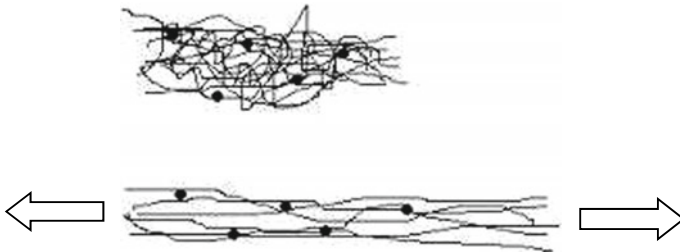
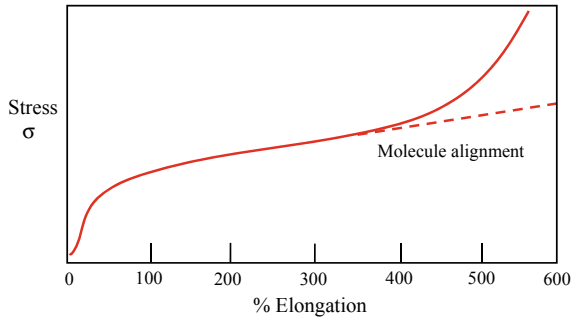


Fig. 5.20 Deformation mechanism of an elastomer

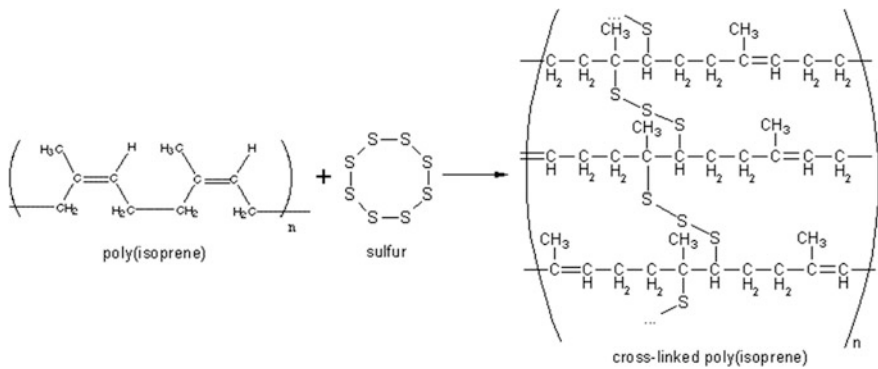


Fig. 5.21 Copolymerization reaction of synthetic rubber by the vulcanization process. Image taken from the article Rubber, www.newworldencyclopedia.org/entry/Rubber

addition of sulfur between the double joints of neighboring chains. The copolymerization reaction to fabricate modern rubber is shown in Fig. 5.21.

The vulcanization of rubber allows to obtain a wide range of physical and mechanical properties, at first by varying the amount of added sulfur, typically from 0.5 to 5wt%. Low sulfur amounts produce a soft and elastic rubber, like elastic bands; the greater the sulfur content, the harder and more brittle rubber becomes,

Table 5.3 Mechanical properties of some elastomers frequently used in the manufacture of mechanical components

Elastomer	Tensile strength (psi)	% Δl	Density (gr/cm^3)	Use
Polysopropene	3000	800	0.93	Tires for light use
Polybutadiene	3500	800	0.94	Automotive rubbers
Polyisobutylene	4000	350	0.92	Pipes, insulation, coatings
Polychloroprene (neoprene)	3500	800	1.24	Hoses, bands, wire insulation
Butadien-estirene (rubber BS or SBR)	3000	2000	1.0	Car and truck tires
Butadiene-acronitrile	700	400	1.0	Gaskets, fuel hoses, O rings
Silicon	1000	700	1.5	Gaskets, seals

such as the belts used in motors and pumps. Table 5.3 presents the typical mechanical properties of elastomers commonly used in industry and daily life.

5.5 Failure Mechanisms of Polymers

The elastic deformation mechanism of elastomers and thermoplastic polymers can be described in simple terms as the unfolding of the tangled molecules by the action of a tensile stress. Since the cross links are directional, they have a “memory” of their original position, thus, when the load is removed, they return to its original position and the body recovers its initial geometry and dimensions. When the temperature is greater than $0.8 T_g$, plastic deformation occurs by molecule sliding, due to the melting of the cross links. When the chains are totally unfolded, the strength is provided by the primary bonds in the chains, so the stress-strain behavior becomes brittle until the point where the material breaks. This behavior is called *cold drawing*, and its typical stress-strain curve is depicted in Fig. 5.22.

The failure of polymer can also occur by a mechanism called *crazing*, which is the formation, growth and interconnection of internal voids, where the voids contain fibrillary bridges of molecular chains that maintain the toughness until they break, enabling the propagation of a macroscopic crack along the direction perpendicular to the maximum principal stress direction. Crazing does not occur in compression, nor at low or high temperatures, with respect to T_g , and depends on the degree of crystallinity. Figure 5.23 shows a scanning electron microscope image of a crazing void in a thermoplastic polymer.

Another failure mechanism of polymers is *shear banding*, which consist on the formation of highly localized shear deformation bands in the directions of the maximum shear stress, so shear banding can occur in materials stressed in tension

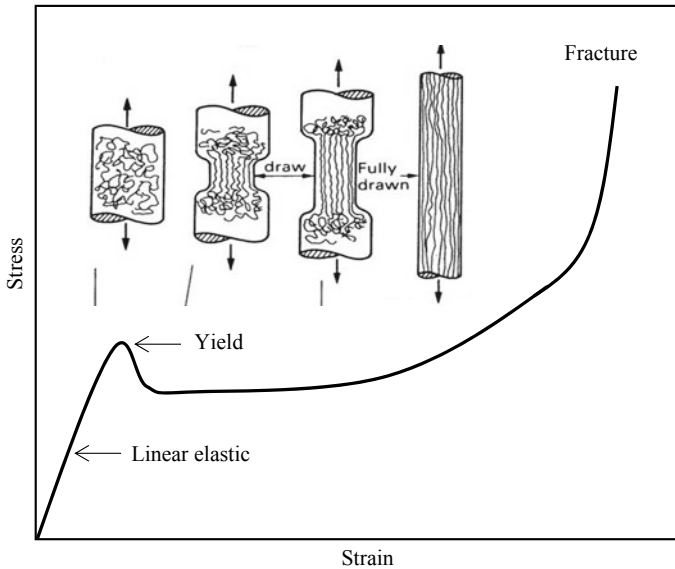


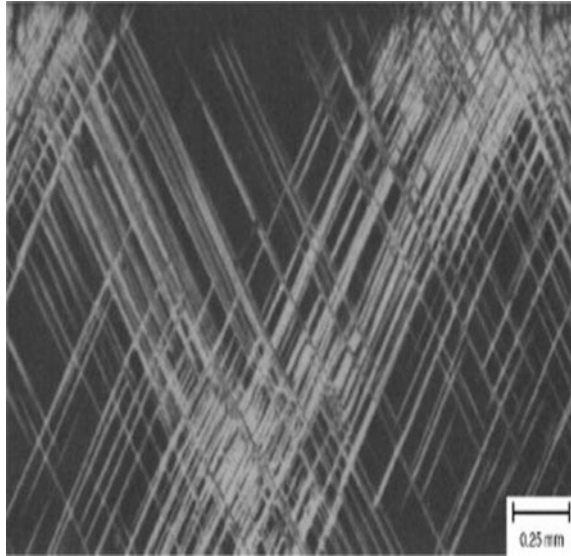
Fig. 5.22 Cold drawn behavior of a thermoplastic polymer up to failure

Fig. 5.23 Craze in a polymer. Image taken from the article Molecular dynamics of fracture in polymers, www.home.iitk.ac.in



or compression, since in both cases, the shear bands lay at 45 degrees of the principal stress direction, as derived from the Mohr’s circle. Under tension strain, shear banding competes with crazing, but in compression, shear banding is the predominant failure mechanism, since crazing occurs only in tension. However in glassy polymers as well as thermoset polymers, shear banding is the dominant deformation and fracture mechanism. In rubbery polymers the deformation of the bands is by drawing of the polymer chains, so the fracture after shear banding is brittle. Figure 5.24 shows an example of shear banding in a polymer.

Fig. 5.24 Shear banding and fracture of polystyrene. Image taken from the article Fracture and failure analyses of plastics and reinforced plastics, Dr. K. Padmanabhan, www.slideshare.net/PadmanabhanKrishnan



Reference

1. Gibson LJ, Ashby MF (1998) Cellular solids-structure and properties. Pergamon Press, Oxford

Chapter 6

Fracture



Abstract This chapter presents a comprehensive introduction of the fracture phenomena as the final stage of the mechanical behavior of engineering materials. It begins with the basic concepts of fracture, the classification of fractures from the engineering point of view and a brief description of the static fracture mechanisms, namely cleavage and micro-void growth and coalescence. The chapter continues with an introduction to fracture mechanics, starting with the analysis of cohesive strength and the Griffith analysis of brittle fracture, to continue with the Irwin's linear elastic fracture mechanic approach, including the concepts of energy release rate, the R-Curve, the stress intensity factor, and fracture toughness, and finishing with an explanation of the Rice's nonlinear fracture mechanics J-Integral and the Wells' CTOD analysis. The previous ideas are used to present an introduction the fundamentals of structural integrity analysis. The final section of this chapter explains the significance and practical aspects of the Charpy impact test.

6.1 Basic Concepts of Fracture

Fracture may be defined as the separation or fragmentation of a solid under the action of stresses, that results in the formation of two new surfaces. The fracture is considered as the culmination of the stress-strain process and it goes through three stages:

- I. Crack initiation or crack nucleation.
- II. Crack propagation.
- III. Final separation.

Depending on the amount of plastic deformation that precedes the fracture, it is classified as *ductile*, if there is a noticeable plastic deformation and *brittle* if there is little or no plastic deformation at all. The drawings in Fig. 6.1 depict the basic features of these fractures. It is important to point out that this classification is valid only from the engineering point of view, because some brittle fractures may actually occur by plastic deformation mechanisms, with plastic deformation

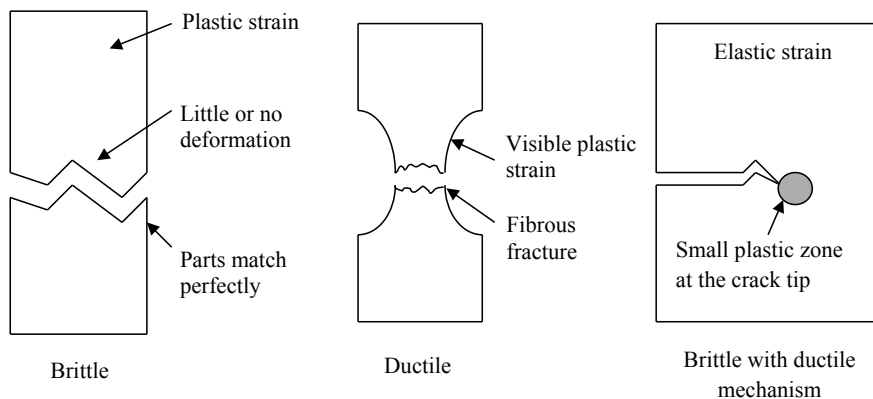


Fig. 6.1 Classification of fracture in terms of the amount of preceding plastic deformation

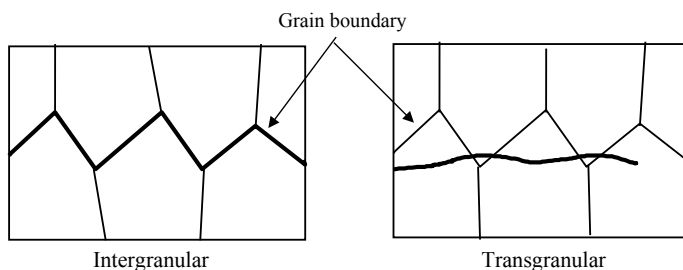


Fig. 6.2 Classification of fracture by the path through the microstructure in polycrystalline materials

concentrated in a narrow region around the crack, thus the fractured piece may have a brittle appearance at macroscopic scale.

The fractures in polycrystalline materials are classified by the path of the crack through the microstructure, as shown in Fig. 6.2, this classification is:

- **Intergranular**, the propagation path is along the grain boundaries.
- **Transgranular**, the crack propagates across the grains, which in turn, is subdivided into: *crystalline*, where the crack path is along a well-defined crystallographic direction and *non-crystalline*, where the crack path follows a path independent of any crystallographic direction. In this last case the fracture plane is usually perpendicular to the direction of the maximum principal stress.

When a fracture occurs under a single load application in a rather short time, it is called *static* or *overload* fracture, and it can be ductile or brittle. On the other hand, fractures may grow at slow rates or occur by increments through time, in such case, it may be by one or a combination of the following mechanisms:

- (1) **Fatigue:** a fracture that results from repetitive or fluctuant load cycles.
- (2) **Stress Corrosion Cracking:** also known as environmental fracture, is a gradual cracking mechanism caused by the synergistic action of a susceptible material under sustained stresses in a corrosive environment.
- (3) **Creep fracture:** it is the propagation of cracks under conditions of creep, which is the deformation through time and under constant stress at high temperatures.

These types of fractures gradually progress until the crack reaches a critical size and then the crack accelerates to high growth rates, typically a fraction of the speed of sound, and the component fails in a very short time (few seconds or less). Therefore, the final stage of fracture is always an overload fracture, which can be either ductile or brittle.

6.2 Static Fracture Mechanisms

Cleavage. The dominant mechanism of brittle fracture is *cleavage*, which is the direct separation of a plane by the rupture of atomic links under a tension stress. The most representative model of cleavage is a pair of parallel atomic rows, being separated by a tensile stress, as depicted in Fig. 6.3. The fracture plane is called *cleavage plane*, which, in crystalline materials is a low-index plane, but it is not the most compact one. Some structures, such as *fcc* do not exhibit cleavage, except in special conditions, such as heavy cold work. The glassy polymers, ceramics and rocks typically fracture by cleavage, while amorphous materials exhibit cleavage in a macroscopic way, with a fracture plane perpendicular to the maximum principal stress direction.

In polycrystalline metals, cleavage produces smooth facets, oriented in different angles, which make the fracture look grainy and shiny. At microscopic level, the main feature of cleavage is a fine topography on the grain facets, known as “river pattern”, with the appearance shown in Fig. 6.4.

The river pattern is formed by the rotation of the cleavage planes when the crack passes from one grain to another. In order to keep continuity, the rotation is done by increments, or “steps”, whose edges form the river pattern, as schematically shown in Fig. 6.5. The name “river pattern” was given because the lines look like several small rivers going downstream and joining to form one large river.

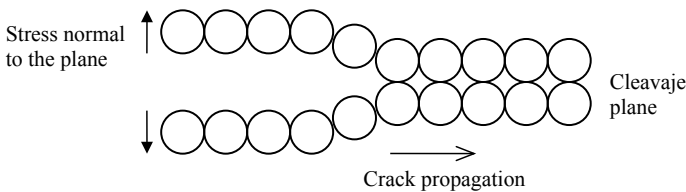


Fig. 6.3 Atomic model of a fracture by cleavage

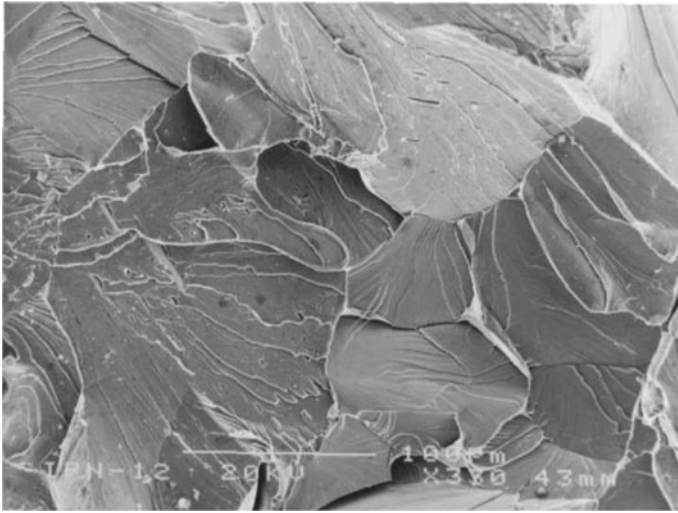


Fig. 6.4 Microscopic appearance of a cleavage fracture in a polycrystal, showing the characteristic river pattern. Scanning electron microscope image

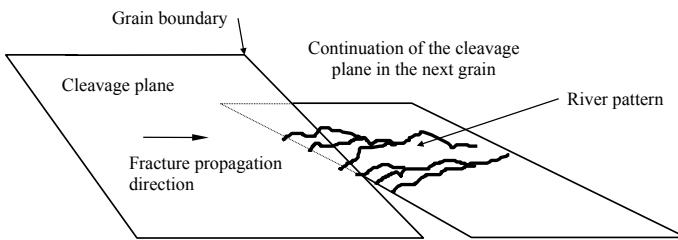


Fig. 6.5 Formation of the river pattern due to the rotation of the cleavage plane from one grain to other

Ductile fracture. The basic mechanism of ductile fracture is the nucleation, growth and coalescence of micro voids. Micro voids are formed at inclusion particles by the triaxial stress that appear once the neck is formed in the tension strained material. This mechanism illustrated in Fig. 6.6. It initiates when the triaxial stress in the neck region causes decohesion of inclusion particles, forming numerous voids that grow by plastic deformation and eventually interconnect forming an internal cavity. As the deformation goes further, the reduction of transversal area increases the stress until the material fails by shear strain.

Since the maximum shear plane is located at about 45° from the tension axis, the final fracture will leave a shear lip that gives a cup-shaped border in one side and a cone in the opposite face, this is why this type of fracture is known as *cup and cone*. At microscopic scale, the central region of the ductile fracture is made up of numerous dimples, which are the halves of the microvoids that coalesced prior to the

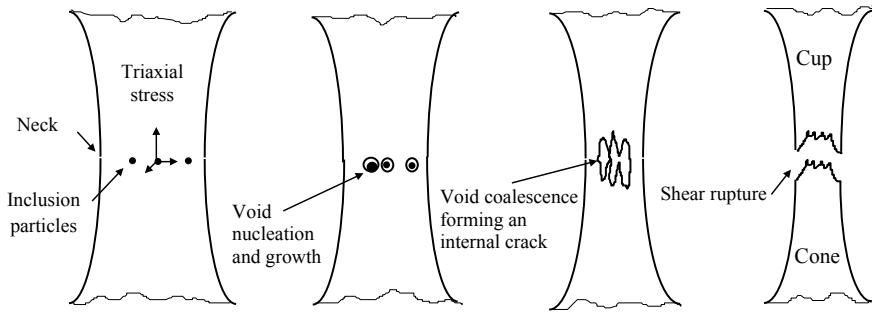


Fig. 6.6 Ductile fracture mechanism in tension by nucleation and coalescence of microvoids

final separation, as seen in Fig. 6.7. The depth of the dimples is proportional to the ductility level, so very ductile materials will exhibit deep dimples and less ductile materials will show shallow dimples. It is common to observe the void initiation particle in the bottom of the dimples; if the particle is not visible it may be in the other side of the fracture.

6.3 Fracture Mechanics

Fracture mechanics is the part of solid mechanics that studies the effect of a crack in the mechanical strength of structural and mechanical components. It also allows to determine the fracture resistance of materials and provides a methodology to study the crack growth processes. Since the fracture of a structural or mechanical component means the end of its service life, or the failure of the fabrication process, the study of fracture mechanics is fundamental to assess the in-service performance and to properly design the fabrication processes that involve plastic deformation.

Fracture mechanics studies the quantitative relations that lead to fracture, considering the following factors:

1. Mechanical properties. Elastic constants, tensile strength, fracture toughness, ductility.
2. Geometry and dimensions. Shape, cross section size, load application points.
3. State of stress. Plane stress, plane strain, pure shear, combined states.
4. Type, orientation and size of cracks.
5. Fracture mechanism: Brittle, ductile, fatigue, stress corrosion cracking, creep.

In practice, fracture mechanics seeks to answer three fundamental questions:

1. What is the load or stress that causes fracture?
2. What is the crack size that produces fracture under a given load?
3. How long does it take a crack to grow?

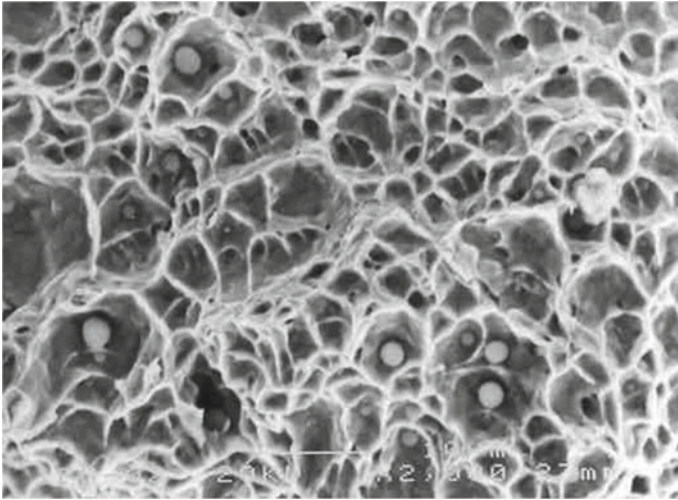


Fig. 6.7 Microscopic appearance of a ductile fracture by nucleation and coalescence of microvoids. Scanning electron microscope image

A description of the basic principles of fracture mechanics, starting with the estimation of the theoretical strength, going through the introduction of the stress intensity factor and fracture toughness and culminating in its application on fracture strength calculation is presented in the next paragraphs.

Cohesive strength. According to the cleavage mechanism, a fracture stress must be sufficient to separate the atomic planes; based on this, the fracture strength is the cohesive strength. The atom bond strength (σ) varies according to the atom separation (x), as shown in Fig. 6.8.

The theoretical cohesive strength (σ^*) is the maximum of the σ versus x curve, reached at $x = a_0/2$, as the atom separation further increases, the atom bond strength is reduced, so the fracture process is irreversible. The variation of the atom bond strength as a function of the atom separation is given by this equation:

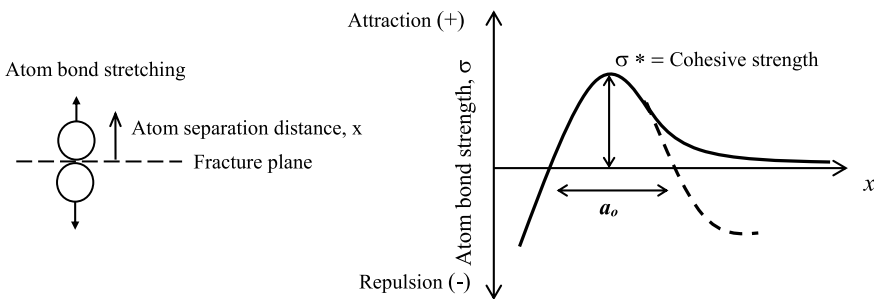


Fig. 6.8 Cohesive strength model

$$\sigma = \sigma^* \operatorname{sen} \left(\frac{2\pi x}{a_0} \right)$$

If the term $2\pi x/a_0$ is close to zero, the next approximation is valid:

$$\sigma = \sigma^* \left(\frac{2\pi x}{a_0} \right)$$

Assuming that the separation of planes produces an elastic deformation, given by the Hooke's law, as to $e = \sigma/E$, where E is Young's modulus, and e is the elastic strain, equal to x/a_0 , thus:

$$\sigma^* \left(\frac{2\pi x}{a_0} \right) = E \frac{x}{a_0}$$

Solving for the cohesive strength:

$$\sigma^* = \frac{E}{2\pi}$$

Substituting typical values of E (for most metals E is in the range of hundreds of GPa) it can be seen that the cohesive strength is much greater than the experimental stresses to produce cleavage, which are in the range of hundreds of MPa, that is 10–100 times larger. However, since cleavage is a real phenomenon, the logical conclusion is that there must be a condition that reduces the fracture stress or that concentrate stress high enough to reach the cohesive strength in solid materials.

The effect of the presence of defects in the strength of a material was analyzed by Inglis¹ in 1913, according to the following procedure. In an infinite plate with a central elliptic hole of length $2c$ and ratio at the tip equal to r , the maximum stress at the tip of the hole is given by:

$$\sigma_{max} = \sigma \left[1 + 2\sqrt{\frac{c}{r}} \right]$$

If r approaches to zero, the previous equation may be reduced to:

$$\sigma_{max} = 2\sigma\sqrt{\frac{c}{r}}$$

Assuming a microscopic hole, similar to crack, of length $c = 10^{-6}$ m (one micron) and $r = 3 \times 10^{-10}$ m (typical length of an inter-atomic space in iron), the stress at the tip of the hole increases more than 115 times, sufficient to reach the cohesive

¹Inglis [1].

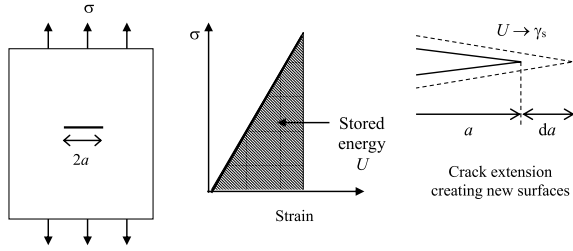


Fig. 6.9 Griffith's model of brittle fracture. Left: thin plate with a central crack. Center: stress-strain record. Right: Conversion of the stored energy U into surface energy γ_s

strength. This reasoning encouraged the idea that fracture was related to the presence of defects in the material.

In 1920 Alan Arnold Griffith made a significant input to the prediction of the fracture stress by stating that fracture is the result of a process of energy transfer. Griffith was an English engineer, graduated from University of Liverpool, who worked in the Royal Aircraft Factory, and later in Rolls Royce. In addition to his contribution to the study of fracture, Griffith designed the axial flow turbo-engines and was pioneer in the technology of vertical takeoff and landing of fixed-wing aircraft.

According to Griffith,² a crack will propagate when the elastic energy stored in the body is released at a rate that equals the creation of surface energy as the crack grows. The demonstration of the Griffith criterion is as follows:

Consider an elastically strained plate with a central crack, as shown in Fig. 6.9. The stored energy is:

$$U = \frac{\pi\sigma^2 a^2}{E}$$

where: σ is the applied stress in the plate, a is the crack size and E is Young's modulus. In a completely brittle fracture (without plastic deformation), the work supplied by the loads is stored as elastic potential energy and is consumed to create two new fracture surfaces. If the stored energy is U and the surface energy is γ_s , the energy balance is:

$$\Delta U = U + 4\gamma_s a$$

As the crack propagates, the conversion rate of stored energy should be equal to the rate of increment of surface energy. Mathematically, this is expressed as:

$$d\Delta U/da = 0$$

²Griffith [2].

Substituting terms and solving for stress, the Griffith's equation for fracture stress is obtained:

$$\sigma = \sqrt{\frac{2E\gamma_s}{\pi a}}$$

In order to verify his theory, Griffith measured the surface tension of silicon glass fibers of 2 in. long and 0.01 in. diameter ($E = 9010$ ksi, $\nu = 0.251$, $\sigma_{uts} = 24.9$ ksi), tested at temperatures ranging from 750 to 1100 °C, out of which he determined the surface energy (0.0031 *lb-in*). Then, he made hollow spheres of the same glass, introducing cracks with a cutter and pressurizing them until rupture. His experimental results varied $\pm 10\%$ with respect to his predictions. Despite the fracture stresses of very brittle materials, such as glass and white iron, calculated by Griffith were very close to the experimental values, the Griffith's equation was rarely applied due to the great difficulty to determine the fracture surface energy of engineering materials.

The energy criterion. Georges R. Irwin, born in El Paso, Texas on February 26, 1907, was a physicist graduated from The Illinois University. He began his career in 1937 at the U.S. Naval Research Lab, working in ballistics and developing, among other, the anti-bullet vest. In 1946 he was in charge of the studies of the Liberty ship fracture problem. Based on Griffith's ideas Irwin developed the Linear Elastic Fracture Mechanics (in honor to Griffith, Irwin used the letter G as symbol of the energy release rate). In 1967 he was appointed "Boeing University Professor" at Lehigh University and in 1972 he moved to Maryland University. He received numerous awards, including the Gold Medal of The American Society for Metals. He died on October 9th, 1998.

Irwin³ restated Griffith's postulate and proposed that the energy release rate in a cracked body must be equal to the demand of fracture work, so as to make a crack propagate spontaneously. Irwin named this condition as *instability* and consequently when the energy release rate is insufficient, the crack does not propagate, so it is said that it is *stable*. Just as in Griffith's criterion, the available energy in a cracked body under load comes from the work done by the loads, which is stored in the body as elastic energy. Each of these contributions is calculated as follows.

Assuming that a perfectly elastic body has a crack of length a , under a load P , the load-displacement record would be as shown in Fig. 6.10. If the crack extends a magnitude Δa and the load is fixed, the crack faces will have an opening displacement equal to ΔV , and the load point will shift from A to B, thus, the energy used by the crack extension is the area within the triangle OAB.

The balance of energy can be expressed as:

$$U = F - W$$

³Irwin published his work for the first time in [3].

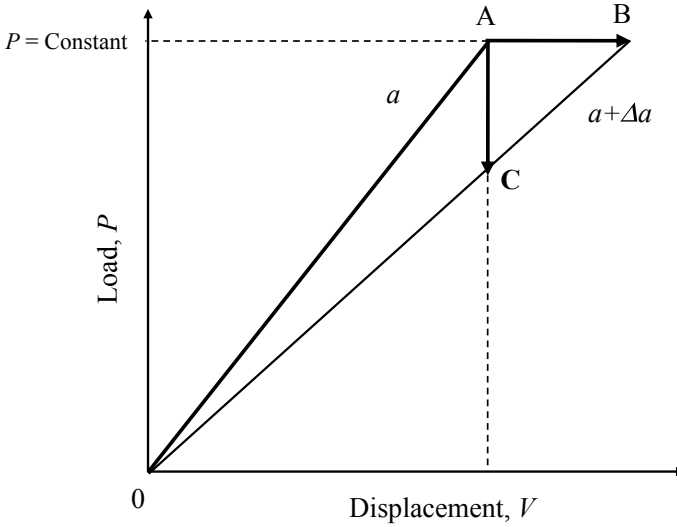


Fig. 6.10 Load-Displacement record of a plate with a crack under tension. If the crack extends under a fixed load, the displacement increases from A to B. If the displacement is kept constant, the load drops from A to C as the crack grows

where U is the stored energy, F is the work supplied by loads and W is the work necessary to make the crack grow. Since the crack growth is produced by the energy conversion as the crack extends. Irwin introduced the concept of *energy release rate* (G), which is expressed mathematically as:

$$G = \frac{d(F - U)}{da}$$

Then, if R is the amount of work needed to cause crack extension, R is the fracture resistance and can be written as:

$$R = \frac{dW}{da}$$

Based on this analysis, Irwin postulated the *Energy Criterion* of fracture, as:

If $G > R$, the crack propagates

So far, Irwin's analysis is identical to that of Griffith's, but Irwin's merit was being able to express the energy release rate as a function of an easy to measure parameter, which is the compliance (C), defined as the inverse of the load-displacement curve slope. The analysis carried out by Irwin in order to experimentally determine the values of G and R is described as follows:

From Fig. 6.10, the area under the straight line OA is $\frac{1}{2}Pv$, which represents the stored energy U . By expressing U in differential form by unit thickness B , it is obtained that:

$$dU = \frac{1}{2} \frac{Pv}{B}$$

On the other hand, the differential work carried out by the load is:

$$dF = \frac{Pv}{B}$$

Substituting into the definition of G :

$$G = \frac{1}{B} \left(P \frac{dv}{da} - \frac{d(1/2Pv)}{da} \right)$$

If $C = v/P$, then $v = CP$ and substituting the previous equation:

$$G = \frac{1}{B} \left(P \frac{dCP}{da} - \frac{1}{2} \frac{d(P^2C)}{da} \right)$$

For constant load conditions:

$$G = \frac{P^2}{B} \left(\frac{dC}{da} - \frac{1}{2} \frac{dC}{da} \right)$$

Which results in:

$$G = \frac{P^2}{2B} \left(\frac{dC}{da} \right)_P \quad \text{Constant load}$$

For conditions of constant displacement, ($dv/da = 0$), therefore:

$$G = \frac{1}{B} \left(\frac{-d(1/2Pv)}{2a} \right)$$

Thus:

$$G = -\frac{P^2}{2B} \left(\frac{dC}{da} \right)_v \quad \text{Constant displacement}$$

Notice that G changes of sign whether the crack extension occurs under constant load (G positive) or constant displacement (G negative) conditions. This has enormous practical consequence, since it means that in a crack propagating under

constant load conditions, the energy release rate increases as the crack grows, therefore, the more it grows, the more energy is available for propagation, so the process will be self-accelerated, leading to crack instability. In opposition, under constant displacement conditions, the energy release rate diminishes as the crack grows, so the propagation will be decelerated until crack stops growing, setting the conditions for crack arrest.

Stress intensity factor. Although the Irwin approach allows the experimental determination of the fracture resistance, its practical application still turns out to be complicated, since the value of G depends on the geometry, crack size and load conditions. To solve this problem Irwin related the values of G and R with a much more convenient parameter which is the *Stress Intensity Factor*, represented by the symbol K .

To define K , Irwin used the solution of the stress distribution around the crack tip of an elastically strained infinite plate obtained by Westergard⁴ a few years earlier. Figure 6.11 shows a coordinate system (x, y) with origin at the crack tip, that defines the stress tensor components of a volume element located at the (θ, r) coordinates from the crack tip.

Westergard found that the stress components around the crack tip are determined by the following equation:

$$\sigma_{ij} = \sigma \sqrt{\frac{a}{2r}} f_{ij}(\theta)$$

Irwin observed that for a given position (r, θ) , the magnitude of the stress depends on the term $\sigma(\pi a)^{1/2}$, which he defined as:

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

where K_I is the Stress Intensity Factor. The subindex I refers to the Mode I, or opening mode, of the displacement of the crack surfaces. Irwin demonstrated that the energy release rate in a cracked body, in linear elastic strain conditions is related to K , according to the following equations:

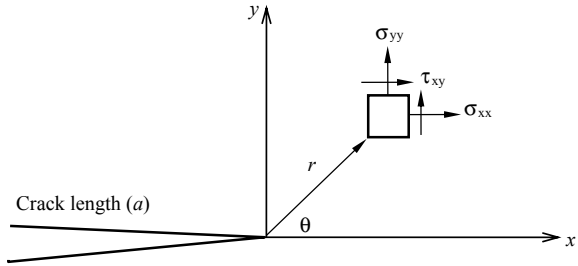
$$G = \frac{K^2}{E} \quad \text{Plane stress}$$

$$G = \frac{K^2}{E} (1 - \nu^2) \quad \text{Plane strain}$$

According to this result, the energy release rate G is proportional to K , so when $G = R$, K will reach a critical value and the crack will propagate. This critical value of K has been termed as *fracture toughness* and it represents the crack extension resistance of a material under conditions of linear elastic strain.

⁴Westergard [4].

Fig. 6.11 Two dimension coordinate system around the crack tip in an infinite plate



Irwin’s analysis is the basis of Linear Elastic Fracture Mechanics, which can be summarized by the following principle of similitude:

Cracks in different bodies, with different loads but the same K , have the same behavior.

The practical applications of this principle consists of determining the crack resistance in a laboratory test, of a cracked specimen with known K . The specimen is tested by an increasing-force machine, and the fracture load and the crack size at the onset of instability are determined, and Kc is determined. Then, with this value, the fracture load of a structure is calculated, provided K and the crack size are known. This procedure is schematically illustrated in Fig. 6.12.

The R-Curve: The energy criterion can be graphically represented by a plot of $[G, R]$ versus $[Crack\ Size]$ known as the *R-Curve*. For convenience, the initial crack size (a_o) is plotted to the left of the origin and the crack extension (Δa) to the right. The R and G functions are drawn on the right side of the graph. The construction of the R-Curve is as follows:

The stress intensity factor of an infinite plate with central crack, under uniform stress is given by:

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

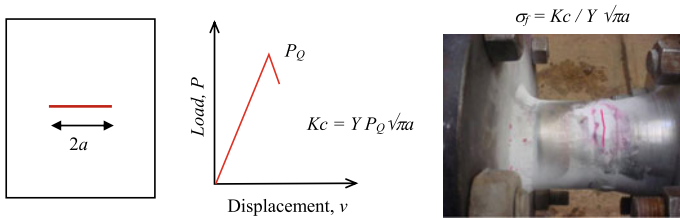


Fig. 6.12 Application of the fracture mechanics similitude principle. The fracture strength, Kc is determined. Kc is used to calculate the failure stress of a component the K and the crack size are known



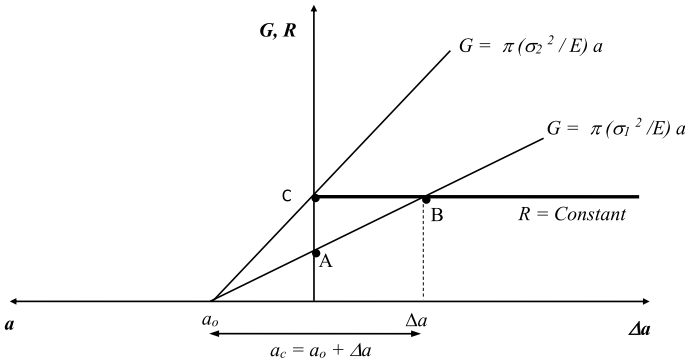


Fig. 6.13 *R*-Curve for an infinite plate with a central crack, *R* constant

Using Irwin’s equation for plane stress:

$$G = \frac{K^2}{E}$$

Substituting *K* into *G*, the following equation is obtained:

$$G = \frac{\pi\sigma^2 a}{E}$$

This equation is plotted, along with the *R* function, as shown in Fig. 6.13. Notice that the *G* is a straight line with slope equal to $\pi\sigma^2/E$. In this case, *R* is independent of *a*, so its plot is a horizontal straight line.

The point A in the *R*-Curve corresponds to the conditions: $\sigma = \sigma_1$, $a = a_0$, and $G < R$, thus the crack will not propagate, and it is in stability. To initiate crack propagation, the crack has to grow an extension Δa , up to the point where $a = a_c$ in order to meet the energy criterion ($G = R$), this is point B. The other option to initiate instability is that the stress increases up to σ_2 , where, once again, $G = R$, point C. Notice that in both cases, as the crack grows further, *G* is increasingly greater than *R*, therefore there is more and more energy release, so the crack propagation accelerates, setting full instability conditions.

In materials that fail after some plastic deformation (ductile fracture), *R* is not constant, but it increases as of the crack grows; so, the *R*-Curve has the shape shown in Fig. 6.14. In this case, for the crack size a_0 under a σ_1 stress, $G < R$, and the crack is stable (no extension). If the stress is increased to σ_2 , then $G > R$, and the crack will grow to some extension Δa , until *G* becomes smaller than *R* and the crack will stop. The crack extension Δa is called “pop-in” and is regarded as a short term stable crack growth. Finally, if the stress increases to σ_3 , $G > R$ and the crack will grow under instability conditions. The tangential point of the *G* and *R* curves indicates the critical crack size a_c . Notice that if the initial crack size is longer, the critical size will be reached at a lower stress, but the stable extension will be larger.



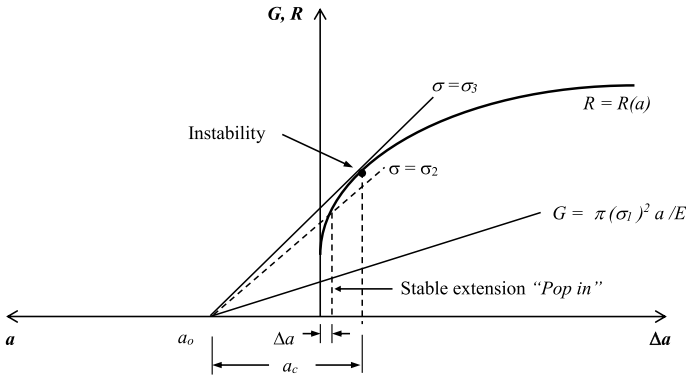


Fig. 6.14 R-Curve for R dependent on the crack size

6.4 Fracture Toughness

The standard test to determine the Plane Strain Fracture Toughness, identified by the symbol K_{IC} , was introduced in 1970 under the designation ASTM E 399, years later the test was grouped along with other fracture toughness tests in the ASTM E 1820 standard. The K_{IC} value is referred as plane strain because, for this condition its value is independent of the thickness and therefore it is a material property. The subindex I refers to Mode I crack displacement (tension opening). The test consists of loading up to fracture a standard specimen containing a fatigue crack emanated from a machined notch. During the test, the load (P) and the crack mouth opening displacement of the specimen (V) are recorded. From this record, the fracture load (P_Q) is determined, while the critical crack size (a_c) is determined by fractographic examination of the fractured specimen. Then, using the K function of the test specimen, the K_{IC} value is determined.

Since at the beginning of the test there is no certainty that the conditions of linear-elasticity and plane strain are achieved, the validity of K_{IC} must be verified. For that, a conditional fracture load P_Q is determined from the P versus V record, as shown in Fig. 6.15. The Type I record is limited plasticity, where P_Q is the load value at the intersection of a straight line with a slope reduced 5% with respect to slope of the OA line of the P versus V curve. This value is represented by the symbol P_5 . The Type II record shows upper yield, in this case, P_Q is the maximum load preceding P_5 . Finally, Type III corresponds to the ideal case (linear record) and $P_Q = P_{max}$.

Once the value of P_Q is known, a conditional value of the fracture toughness K_Q is calculated, then, the following criteria should be verified:

$$(W - a) < 2.5 (K_Q / \sigma_{YS})^2$$

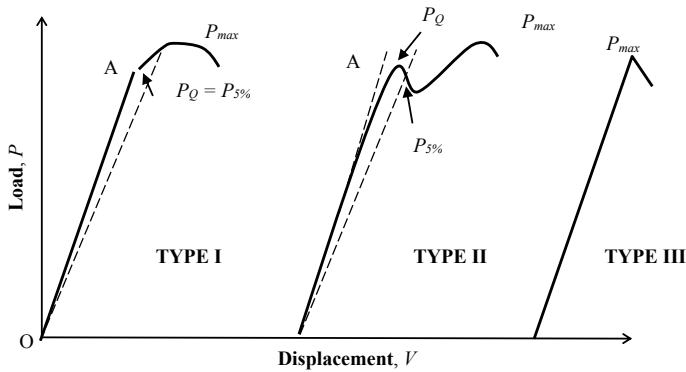


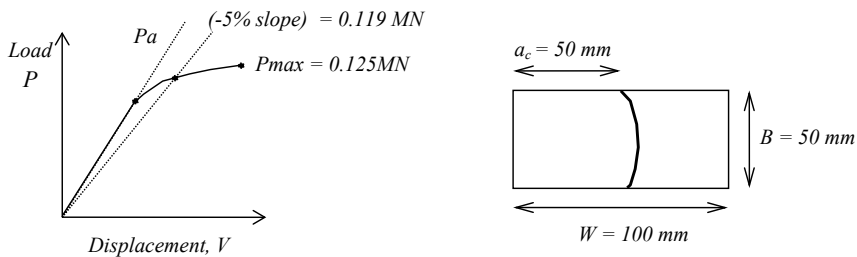
Fig. 6.15 Typical load-displacement records of a fracture toughness test for determining the fracture load P_Q

$$P_{max}/P_Q \leq 1.10$$

where B and W are the width and thickness of the specimen respectively, and σ_{YS} is the 0.2% yield strength in tension. If the preceding criteria are met, then the K_Q is equal to K_{IC} , otherwise test is not valid, and a larger specimen will be required in order to meet the linear-elastic and plane strain requirements.

The following examples illustrate the above procedure:

Example 1 The load-displacement record shown below was obtained from a CT specimen, 50 mm thick and 10 mm wide, of a steel with $\sigma_{YS} = 1000$ MPa. If the critical crack size was 50 mm, determine K_{IC} and verify if the test is valid.



Solution Verification of the linearity condition:

$$P_{max}/P_Q = 0.125/0.12 = 1.05 < 1.1, \text{ the test is valid in terms of linearity.}$$

Calculation of K_Q : the equation of K_I for the CT specimen is:

$$K = P/(BW^{1/2})f(a/W)$$

For $(a/W) = 0.5$; $f(a/W) = 9.66$, substituting data:

$$K_Q = P_Q f(a/W)/(B/\sqrt{W}) = 0.125(9.66)/(0.05\sqrt{0.1}) = 76.4 \text{ MPa}\sqrt{\text{m}}$$

Applying the thickness criterion:

$$2.5(K_Q/\sigma_{YS})^2 = 2.5(76.4/1000)^2 = 0.015 \text{ m} = 15 \text{ mm} < (W - a) = 100 - 50 \text{ mm} \\ = 50 \text{ mm}$$

Since the size of the ligament (uncracked section of the specimen before fracture) is greater than the required ligament size, the test is valid and $K_{IC} = 76.4 \text{ MPa}\sqrt{\text{m}}$.

Example 2 A structural steel has a $\sigma_{YS} = 550 \text{ MPa}$ and an estimated fracture toughness $K_{IC} = 70 \text{ MPa}\sqrt{\text{m}}$. Calculate the width and thickness (W , B) of a CT specimen in order to carry out a valid test.

Solution To meet the plane strain requirement, the width W , assuming a crack size $a = 0.5 W$, should be:

$$(W - a) = 0.5 W > 2.5(K_{IC}/\sigma_{YS})^2 = 2.5(70 \text{ MPa}\sqrt{\text{m}}/550 \text{ MPa})^2 = 0.0405 \text{ m} = 40.5 \text{ mm} \\ W > 2 \times 40.5 \text{ mm} = 81 \text{ mm}$$

For a CT specimen, $B = W/2 = 40.5 \text{ mm}$, so a CT specimen of, at least, 41 mm thick and 81 mm wide meets the requirement.

The thickness and linearity requirements of the Standard ASTM E399 limit the application of the plane strain fracture toughness test to brittle materials. For example, for a forged steel with $\sigma_{YS} = 350 \text{ MPa}$ and $K_{IC} = 150 \text{ MPa}\sqrt{\text{m}}$, the required thickness of a CT specimen is $B = 2.5 (150/350)^2 = 0.459 \text{ m} = 45.9 \text{ cm}$ and thus, $W = 2B = 91.8 \text{ cm}$. A CT specimen with such dimensions would weight about 2000 kg. Additionally, the test machine should have a load capacity of about 10,000 kN. A testing machine of such capacity is expensive and the cost and difficulty to machine and handle a test specimen that size make the test prohibitive. Table 6.1 shows some K_{IC} values for materials commonly used in engineering and the widths recommended for a K_{IC} valid test using CT or SEB specimens.

Elasto-plastic fracture strength. It is clear that the K_{IC} test is valid only when the crack-tip plastic zone is limited to a small zone contained within the ligament, in such a way that the load-displacement behavior is mostly linear. However, when the material is very ductile, the load-displacement curve exhibit large deviations from linearity, as shown in Fig. 6.16, which invalidate the K_{IC} test result, thus a different fracture criterion has to be used.

Table 6.1 Typical yield strength values, fracture toughness and minimum required width of a valid specimen for K_{IC} determination

Material	σ_{YS} (MPa)	K_{IC} (MPa \sqrt{m})	B minimum (mm)
High strength steel	1965	57	2.1
Steel quenched and tempered	1830	47	1.65
Forged steel	350	200	816
Structural steel	240	>220	2100
Titanium Al-V	1100	38.5	3.1
Aluminum 7075 T651	550	30	7.4
Aluminum 2024 T3	390	34	19

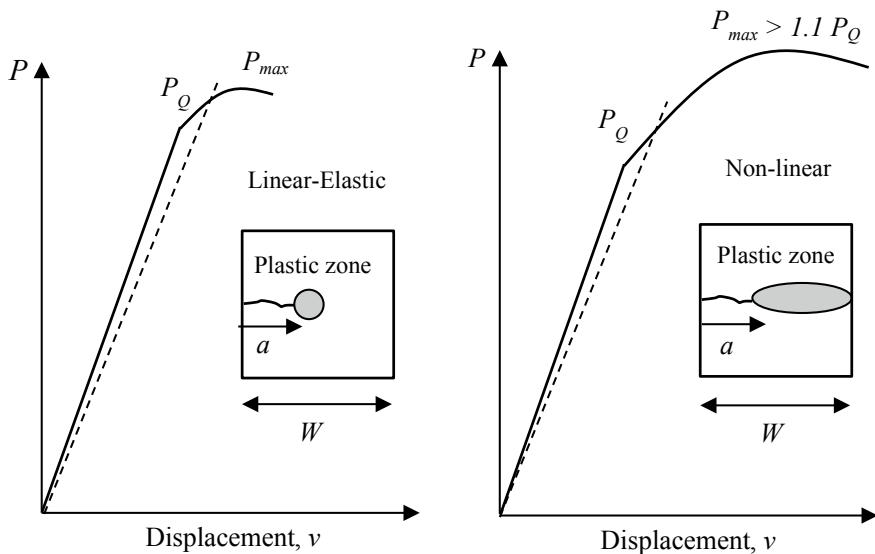


Fig. 6.16 Load versus Displacement records of a cracked body made of linear-elastic material (brittle), and non-linear behavior, ductile material

In 1968, James R Rice⁵ developed the *J-integral* method that became a fundamental parameter for the characterization of crack behavior in non-linear load-strain conditions. From 1964 until 1981, Rice was an assistant professor at Brown University, previously, in 1962 he graduated from Lehigh University, where he took courses with the pioneers of modern fracture mechanics, such as: Fazil Erdogan, George Sih and Paul Paris. He got M.S. and Ph.D. degrees in 1963 and 1964, respectively. Since 1981 he is professor at the Geophysics department of Harvard University. In 1994, he received the Timoshenko Medal, in 1996 he was awarded The Franklin Institute's Francis J. Clamer Medal, and in 2016 he was awarded the ASME Medal.

⁵Rice [5].

From the physical point of view, the *J-integral* is the energy balance around a closed path around the crack-tip, as illustrated in Fig. 6.17. The balance is between the work supplied by the tractions *T* acting over surface elements *ds* of a closed path Γ and the strain energy within the limits of this path.

As the *J-integral* is path-independent, the integration can be done at any distance from the crack-tip, and the energy change due to the crack extension can be calculated whether the stress-strain behavior is linear or non-linear.

Rice showed that when the plastic strain is within a narrow stripe in the ligament, as shown in Fig. 6.18, the *J-integral* value is given by:

$$J = \frac{2A}{B(W - a)}$$

where *A* is the area under the Load-Displacement curve limited by a parallel to the elastic portion of the curve, drawn from the displacement up to where is desired to calculate *J*, as shown in Fig. 6.18.

To include the contribution of the elastic portion, total *J* is given by:

$$J = \frac{K^2}{E} (1 - \nu^2) + \frac{2A}{B(W - a)}$$

The *J-integral* can be used for the analysis of elastic-plastic fracture, because it is equivalent to the energy release rate *G*, which means that *J* = *G*. Rice also demonstrated the equivalence between *K* and *J* by the following equations, that are valid as long as the plasticity is not extensive.

$$J = G = \frac{K^2}{E} \quad \text{Plane stress}$$

$$J = G = \frac{K^2}{E} (1 - \nu^2) \quad \text{Plane strain}$$

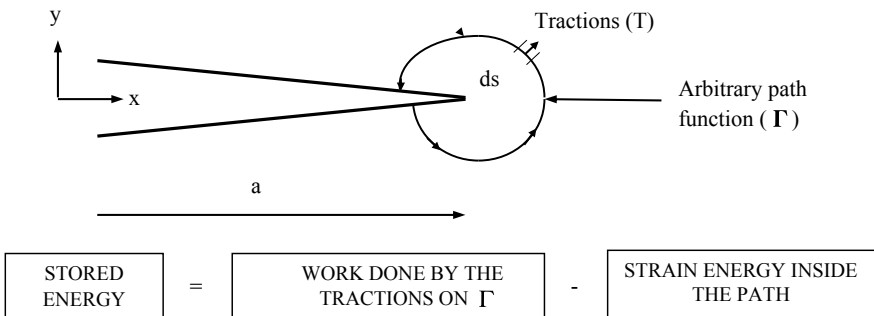


Fig. 6.17 Energy balance of a closed path around a crack



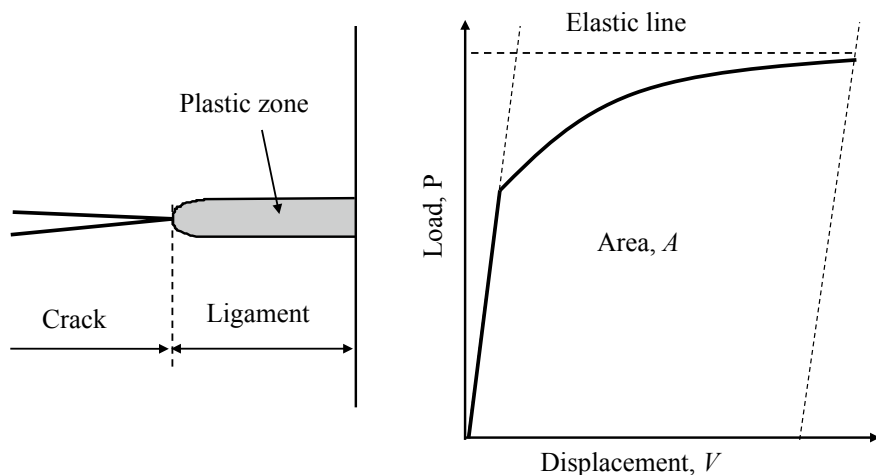


Fig. 6.18 Condition of the plastic zone and area under the Load-Displacement record to calculate J , according to Rice

The J value can be calculated for different geometries and load conditions, through several methods such as: experimental strain measurement, compliance, finite element modeling, etcetera; being all of them complex and time consuming. The critical value of J is designated as J_c and it is a conservative measurement of fracture toughness, as it represents the J value at which a 0.15 mm stable extension of the crack (pop-in) occurs. The procedure to evaluate J_{IC} is described in standard ASTM E-1820 as well as in the British Standard BS-7448. Figure 6.19 summarizes this method.

The crack-tip opening displacement criterion: In 1961, Alan Wells⁶ proposed a failure criterion based on the observation that the fracture initiates after the tension strain at the crack-tip exceeds a critical value, as shown schematically in Fig. 6.20. Wells proposed that the deformation in the direction perpendicular to the fracture plane, named *Crack Tip Opening Displacement*, represented by $CTOD$, may be used as a fracture criterion because it is related to the stress intensity factor by the following equation:

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

The advantage of the use of $CTOD$ as fracture criterion is that it is not limited by linear conditions, such as K_{IC} , nor restricted plasticity, like J_{IC} , so it may be applied in conditions of generalized plastic deformation. Similarly, the $CTOD$ may be

⁶Wells [6].

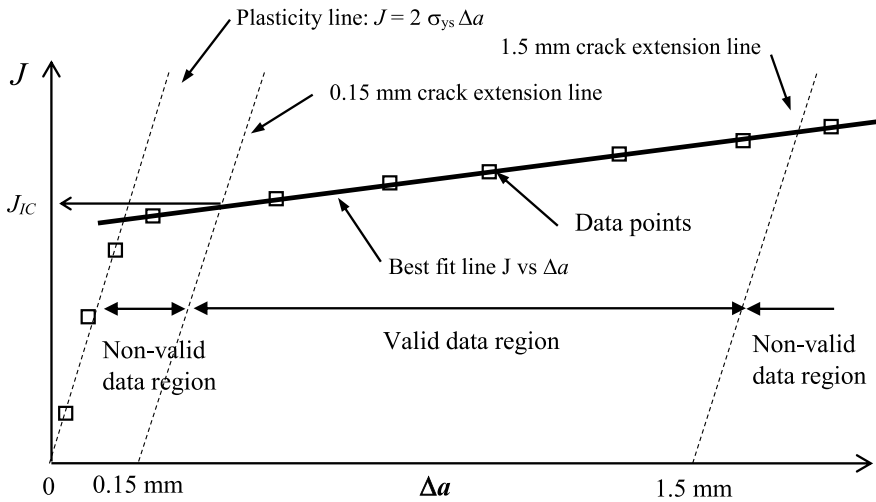


Fig. 6.19 Summary of the standard method to evaluate J_{IC} . Data may be obtained from multiple specimens or one to which several load-unload cycles are applied. The data within the valid region are adjusted to a straight line which intersection with the plasticity line at $\Delta a = 0.15$ mm gives the value of J_{IC}

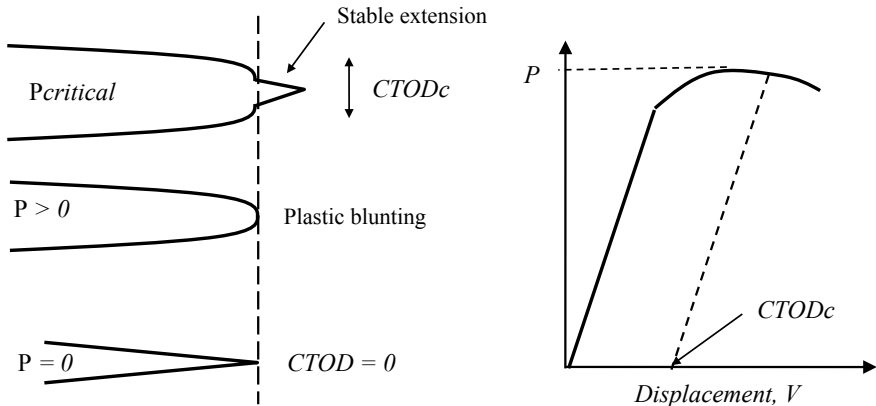


Fig. 6.20 Justification of $CTOD$ as fracture criterion

applied to short cracks and high toughness materials. These reasons make the $CTOD$ a very useful parameter to characterize fracture toughness of welds and high ductility materials. The procedure to evaluate $CTOD_C$ is described in standard ASTM E-1820.

6.5 Structural Integrity

The main contribution of fracture mechanics to structural engineering is that it allows to calculate the fracture load and the crack size of a stressed component containing crack-like flaws. These two values constitute the *remaining strength* and the *critical flaw size*, which are the teoretical foundations of *Structural Integrity*, or *Fitness-For-Service* technology, that merged in the last part of the XX century and nowadays is the main maintenance and reliability strategy of process equipment, building and bridge structures, pipelines, reactors, boilers, pressure vessels and aircrafts, among others, all over the world. The Structural Integrity may be defined as the capability of a mechanical or structural component to bear the imposed loads during a determined time, in a safe way (without failure), while containing defects detected and sized by nondestructive examination.

The remaining strength allows to determine the operating load level that can be applied to a flawed component without risk of failure, which is a necessary information if the Owner/User of a component has to decide of whether or not keep it in-service, or to shut it down and perform a repair. This is very convenient in cases like a submarine pipeline, which due to accessibility limitations cannot be immediately repaired, but the service cannot be interrupted neither. The structural integrity methodology allows to determine if there is a safe operating pressure, along with a maximum allowable flaw size, which are a fraction of the remaining strength and the critical crack size respectively, so the decision of continue operation or shut down can be taken under technical reasons. Furthermore, if the flaw growth rate is known from laboratory tests or field experience, and the operating conditions are reasonably constant, the *remaining life* may be calculated, as the time elapsed from the moment of detection of the crack to the time when it reaches its critical size. This information may be used to establish the detection limits for a non-destructive inspection program as well as the safe operating windows of the flawed component, thus preventing failures, saving money by reducing the number of repairs and ultimately increasing the service life of components that otherwise would be scheduled for substitution.

The remaining strength, being the failure load, can be calculated from the general equation of K , if the component containing a crack-like flaw is working within the elastic regime, which is mostly the case. The equation for K may be written as:

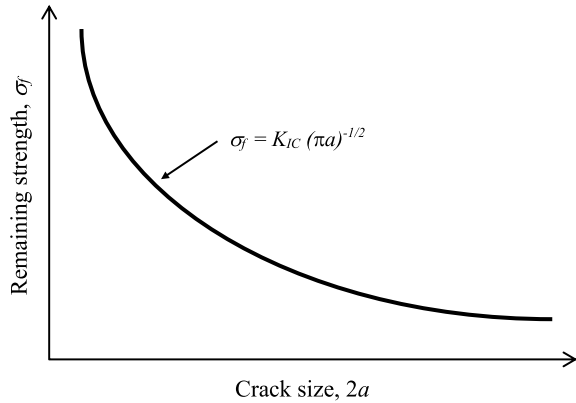
$$K = Y \sigma_A \sqrt{\pi a}$$

where Y is a geometric factor, σ_A is the applied stress and a is the crack size. If $K = K_{IC}$, and the previous equation is solved for P , an equation for the failure stress is obtained.

$$\sigma_f = \frac{K_{IC}}{Y \sqrt{\pi a}}$$

A plot σ_f versus a of the previous equation gives the *Remaining Strength Curve* (RSC), which is depicted in Fig. 6.21. As it can be seen, the RSC goes to infinity as

Fig. 6.21 Remaining Strength Curve calculated by linear elastic fracture mechanics



the crack size approaches to zero and is asymptotic for very large crack sizes. Neither situations are real, because, the strength cannot be higher than the material's strength, and the crack size cannot be larger than the component's width. Therefore, there should be a minimum and maximum crack size for which the linear-elastic fracture mechanics calculations of the remaining strength are valid.

A simple approach to determine the valid limits for the calculation of the remaining strength by linear-elastic fracture mechanics is as follows: consider a rectangular plate of thickness B and width W , which contains a central crack of length $2a$. The applied stress may be calculated as $\sigma_A = P/BW$; if the cracked area is rectangular, the remaining cross-section, referred as the *ligament*, can be calculated as $B(W-2a)$, then the *net stress* σ_N can be written as:

$$\sigma_N = \frac{P}{B(W - 2a)}$$

Multiplying and dividing by W , the term P/BW can be replaced by σ_A to obtain:

$$\sigma_N = \sigma_A \frac{W}{(W - 2a)}$$

Now, taking the criterion that failure by generalized plastic deformation will occur when the stress in the ligament equals the ultimate tensile strength, the applied stress becomes the *net-section* remaining strength, thus:

$$\sigma_f = \sigma_{uts} \left(1 - \frac{2a}{W} \right)$$

Plotting this equation in the RSC, the limits of the applicability of linear elastic fracture mechanics can be determined assuming the criterion that the remaining strength will be the least calculated by each method, as depicted in Fig. 6.22.

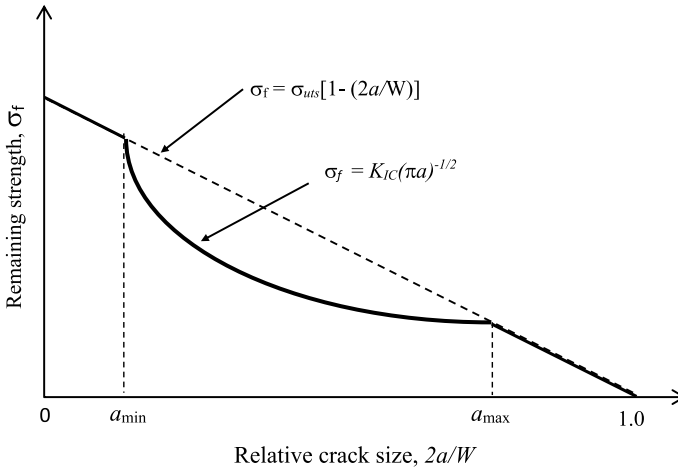


Fig. 6.22 Remaining Strength Curve determined by the combination of linear elastic fracture mechanics and net-section methods

The RSC may be effectively used to perform Fitness-For-Service assessments, according to the following criteria:

- The maximum allowable flaw size is the one that gives a remaining strength equal or greater than the design stress or the proof stress, for example the stress resulting of the hydrostatic test of a pressure vessel or piping. Any crack-like flaw of this size or smaller will be accepted during the quality control inspection of a new component.
- The critical flaw size is the one that gives a remaining strength equal to the applied stress under normal operating conditions or a minimum stress level specified by the user. Any flaw of this size or larger implies an imminent risk of failure, therefore, upon detection the operating load has to be reduced to a safe value, which can be determined from the RSC, until a corrective action is taken.
- The assessment of flaws of sizes between the maximum allowable flaw size and the critical flaw size is done by calculating the safety margin between the normal operation stress and the remaining strength calculated for a given flaw size. The lesser the safety margin the greater severity of the defect, and therefore the urgency of a corrective action. The defect severity levels may be established by the user of in conjunction with structural integrity experts.
- The remaining life can be determined from the RSC if the flaw size is converted into elapsed time, dividing it by the growth rate. It is important to mention that the remaining life calculated in this way should not be taken as a warranty of a time of use, or the time to take a corrective action, but it should be used to determine an appropriate monitoring program. Most Integrity Management Plans recommend that a corrective action, whichever it is, is taken at no later than one half of the remaining life.

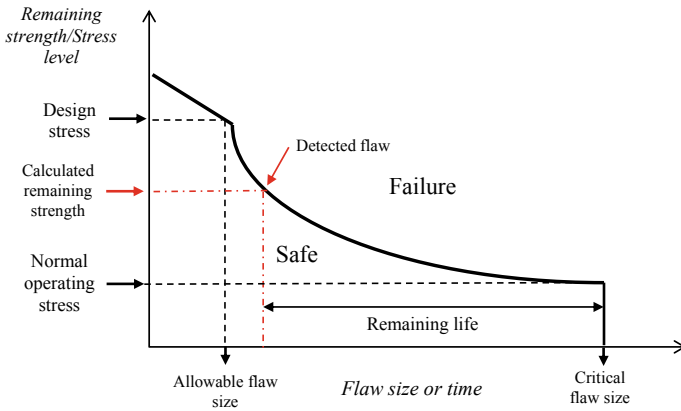


Fig. 6.23 Remaining Strength Curve showing the parameters of a structural integrity assessment of a component containing crack-like flaws

Figure 6.23 depicts the above described parameters in a schematic way. It is worth to mention that a RSC should be specific for a combination of material, environment and service conditions, and it is valid as long as these conditions are reasonably constant. Also it is responsibility of the structural integrity engineer to apply the appropriate safety factors to the material properties, flaw sizes and service loads, as well as, to establish the level of conservatism in the determination of the remaining strength and the critical flaw size.

The ability to obtain the RSC of structural components of strategic importance in the industry motivated extensive investigations to improve the accuracy of the calculations of remaining strength, beyond the limitations of the net-section and linear elastic fracture mechanics methods. As a result in 1975, Dowling and Townley⁷ developed a crack assessment method that combined the linear elastic fracture mechanics criterion for brittle fracture and an elastic-plastic fracture criterion introduced by Doughdale, which is currently known as the *Two-Parameter Criterion*. This method allows to assess brittle, elastic-plastic and plastic collapse failures by using the following equation:

$$Kr = Sr \left[\frac{8}{\pi^2} \ln \sec \left(\frac{\pi}{2} Sr \right) \right]^{-1/2}$$

$$Kr = \frac{K_I}{K_{IC}}$$

$$Sr = \frac{\sigma_A}{\sigma_0}$$

⁷Dowling and Townley [7].

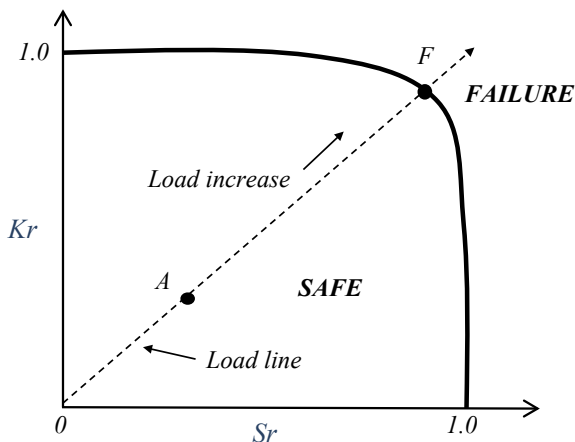
where K_I is the applied stress intensity factor in Mode I, K_{IC} is the plane strain fracture toughness, σ_A is the applied stress and σ_0 is the yield strength. The plot Kr versus Sr is known as the Failure Assessment Diagram (FAD) and has the form depicted in Fig. 6.24. The use of the FAD is quite straight forward; if an assessment point falls (A) within the envelope of the Two Parameter equation curve, the flaw is stable and the component does not fail; if the assessment point falls on the boundary or outside the envelope curve (F), the flaw is unstable and failure is imminent. Additionally, since Kr and Sr both depend linearly on the load, a straight line traced from the origin to the evaluation point will represent the load path; the distance from the origin to the envelope is the failure load, therefore, the length of the segment between the evaluation point and the envelope represents the safety margin in terms of load.

In 1986, Milne, Ainsworth, Dowling and Stewart presented the report CEBG R/H/R6 in England, which would be worldwide recognized as the R6 Code that contained refined expressions for the FAD and greatly improved the accuracy of the fracture mechanics calculations of the remaining strength of structural components containing crack-like flaws. Nowadays the R6 Code is the base of the methods provided by the international recognized standards of structural integrity, such as the API 759-1/ASME FFS-1 "Fitness-for-Service" and the British standard BS 7910 "Standard Methods for Assessing Cracks". The R6 Code is applied in the following way:

1. Define the structural modes of failure as a function of the load line ratio, as follows:

$$\begin{array}{ll} Kr/Sr > 1.8 & \text{Brittle Fracture} \\ 0.2 < Kr/Sr \leq 1.8 & \text{Elastic - plastic Fracture} \\ Kr/Sr \leq 0.2 & \text{Plastic Collapse} \end{array}$$

Fig. 6.24 Failure Assessment Diagram by the Two-Parameter Criterion



- Provides options for the FAD equation, according to the desired level of accuracy:

Option 1. $Kr = (1 - 0.14 Sr^2) [0.3 + 0.7 \exp(-0.65 Sr^6)]$

Option 2. $Kr = (E \varepsilon_{ref} / Sr \sigma_0 + Sr^3 \sigma_0 / 2E \varepsilon_{ref})^{-1/2}$

Option 3. $Kr = (Je/J)^{1/2}$

Where ε_{ref} is the real strain at the applied stress, J is the J -Integral at the applied load and Je is the elastic J -Integral of the structure.

- The right end cut-off of the FAD has been expanded to consider strain hardening behavior, being $Lr = 1.15$ for structural steels, $Lr = 1.55$ for C-Mn steels and $Lr = 1.8$ for stainless steels.
- The parameter Sr is replaced by $Lr = L_A/L_{Lim}$ to facilitate the analysis in conditions where the stress is not uniform or a reference stress cannot be established.

Figure 6.25 shows the FAD as presented in the standard API 579-1/ASME FFS-1 2007.

6.6 The Charpy Impact Test

Brittle fracture is the most dangerous form of fracture, since it occurs very fast and without noticeable plastic deformation that may give a warning indication. Carbon steels are usually very tough at ambient and moderate temperatures, but at low temperatures they become brittle, so many catastrophic failures have been caused by this phenomenon. One of the most popular stories of brittle fracture is the Titanic sink,⁸ but the most important from the engineering point of view was the Liberty Ships case, that occurred just after the World War II. Figure 6.26 shows a Liberty ship that broke into two pieces while docked in calm waters and without load in a cold winter day.

The Liberty ships were the first ones fabricated by electric arc welding, which revolution forever the steel construction industry. With this innovation with respect to the traditional riveted hull technique, the construction time of ships more than 14,000 ton displacement and 135 m overall length was reduced from 244 days to less than 42 days per ship. The USA fabricated around 2710 Liberty ships, a pace greater than the German submarines could sink them (it is said that thanks to this the Allies won the WW II), but more than 400 suffer major structural failures in the hull, and more than 20 broke apart.

⁸The crash of the Titanic against the iceberg made a crack of about 1.5 m long that was not a high risk for the ship's structural integrity, however one theory proposes that the brittleness of the rivets and the plates of the hull led to the formation of a more than 70 m long aperture that eventually caused the sinking.

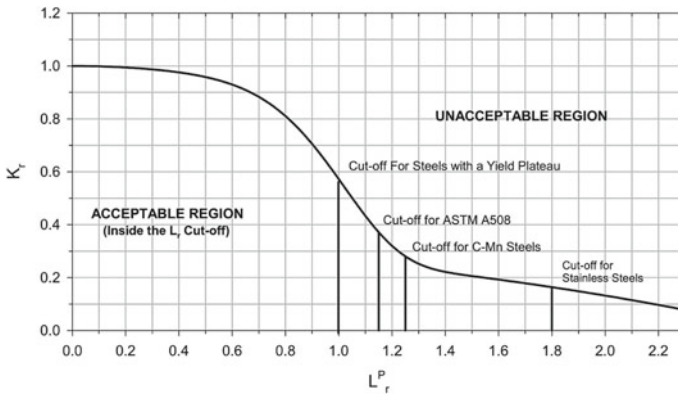


Fig. 6.25 Failure Assessment Diagram as presented in API 579-1/ASME FFS-1 2007



Fig. 6.26 Liberty ship that failed while docked in calm waters during a cold winter day

After the war, in 1945, the US Congress commissioned the Navy Research Center to study the causes of failure of the Liberty ships: one of the researchers was, no one but George R. Irwin, the father of modern fracture mechanics. The investigations performed by the Navy used the Charpy impact test, introduced by the French engineer Georges Charpy in 1901, and determined that the absorbed impact energy C_v of carbon and low alloy steels fall sharply in a narrow range of temperatures, as schematically shown in Fig. 6.27. This phenomenon was named *ductile-brittle transition*, and since then it gets great attention, not only for design, but also as a factor to consider in structural integrity assessments.

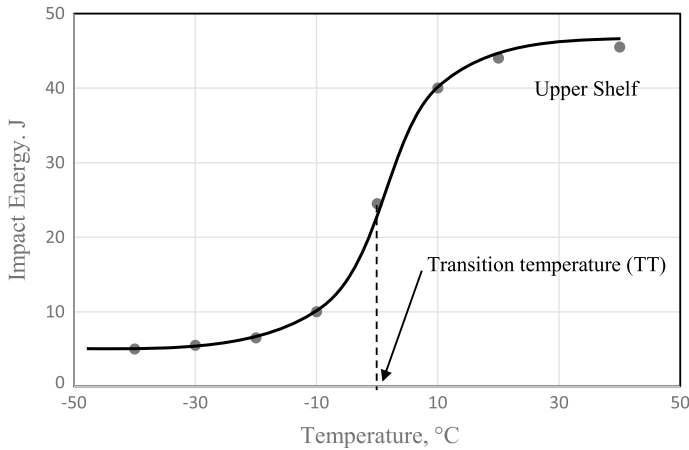


Fig. 6.27 Typical Charpy impact energy as a function of temperature of a ferritic steel

The Charpy impact test, as it is used today is described in the ASTM Standard E23 and ISO 148-1. It consists of breaking a standard specimen with a hitter mounted on a pendulum. The specimen is a square bar of 10×10 mm and 50 mm length, that contains a V notch of 60° and 1 mm depth in the middle. During the test, part of the kinetic energy of the pendulum is absorbed by the fracture process of the specimen, so the absorbed energy is the difference of potential energy between the initial and final height of the pendulum. Figure 6.28 illustrate this principle and a schematic of the apparatus set up for the Charpy test. The Izod test is similar to the Charpy test, the only difference is that the specimen is held in vertical position.

Perhaps the most important effect on the ductile brittle transition of carbon steel is the carbon content. In general, as shown in Fig. 6.29, the higher the carbon content, the lower transition temperature, which is a detrimental effect.

The effect of alloying elements and grain size on the C_v values of steel is shown in Table 6.2. As a consequence, to prevent brittleness and weld cracking, a parameter called Equivalent Carbon (CE) has been introduced by the International Institute of Welding, which is calculated by the equation:

$$CE = C + \frac{(Mn + Si)}{6} + \frac{(Cr + Mo + V)}{5} + \frac{(Ni + Cu)}{15}$$

The element content is in wt%, and typically the value of CE is limited to 0.43 for pressure vessels and hydrocarbon transport pipelines.

Even though the stresses and strains in the Charpy impact test specimen can be measured, the values of C_v are not used to determine the fracture loads in fitness for service assessments, neither for design, mainly because C_v is specific for the test specimen. However, C_v values are widely used as a material specification to

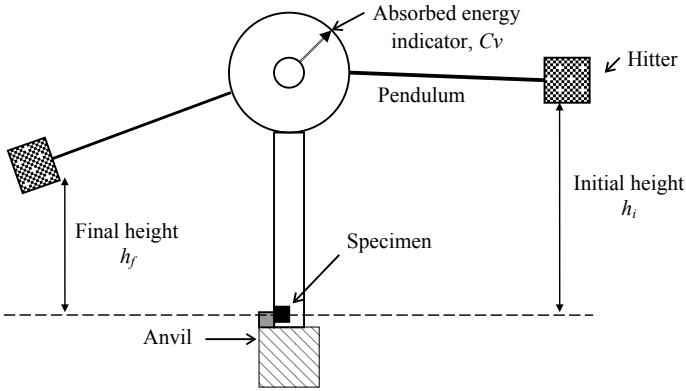


Fig. 6.28 Apparatus and principle of the Charpy impact test. Absorbed Energy $C_v = Mg(h_i - h_f)$, where M is the hitter mass, g is the gravity acceleration and h is the height. The strain rate is 10^3 s^{-1}

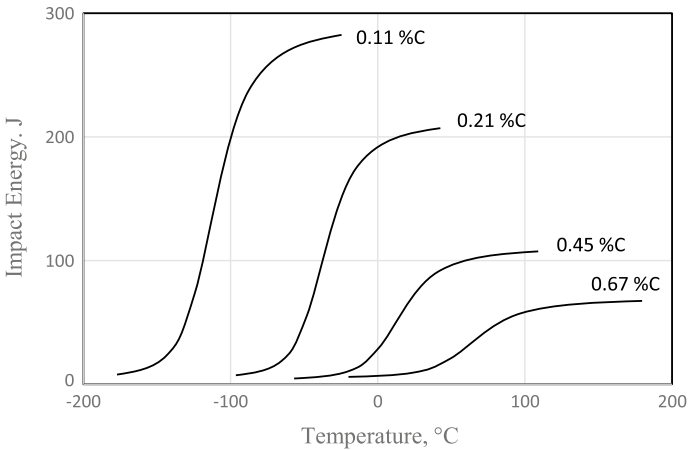


Fig. 6.29 Effect of carbon content on the absorbed impact energy of ferritic Steel

establish a minimum required toughness to prevent brittle fracture, as well as to determine the minimum allowable working temperature as to prevent brittle fracture.

The C_v versus T curve is also useful to compare the performance of materials in terms of their tendency to brittle fracture as schematically depicted in Fig. 6.30. The rule is that a material with higher C_v values at a given temperature will have a better performance, however the performance over a range of temperatures must be examined before making a choice. At room temperature (T_{amb}) material A is better than B because it has higher C_v values, but overall, material B is better because it maintains higher toughness at low temperatures in comparison with material A.

Table 6.2 Effect of the steel alloying elements and grain size on the behavior of Cv

Element	Effect
Mn	Reduces $TT \sim 5\text{ }^\circ\text{C}/0.1\% \text{ Mn}$
P	Increases $TT \sim 7\text{ }^\circ\text{C}/0.01\% \text{ P}$
N	Reduces TT
Ni	Up to $\sim 20\%$ increases Cv
Si	Reduces TT si $\%Si > 0.25$
Mo	Increases $TT \sim 14^\circ\text{C}/0.1\% \text{ Mo}$
Cr	No effect
O, H	Sharply increases TT
Nb, V	Reduces TT below $-30\text{ }^\circ\text{C}$
Grain size	Fine grain increases Cv and reduces TT

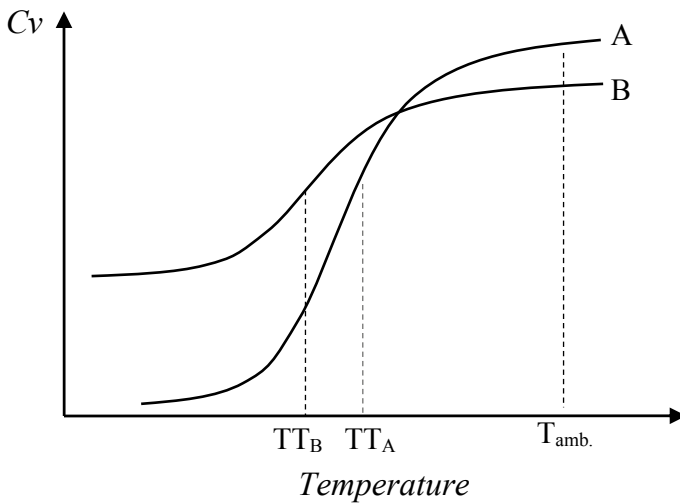


Fig. 6.30 Comparison of the performance of the ductile-brittle transition of two steels

Since Charpy impact tests are relatively easy and cheap to carry out, the Cv values are often used to estimate plane strain fracture toughness K_{IC} , which is a much more difficult and expensive test. The Appendix 9F of the API 579-1/ASME FFS-1 2016 Standard provides the Rolfe-Novak correlation to estimate the fracture toughness in the upper shelf (above the transition temperature), which has the following form:

$$\left(\frac{K_{IC}}{\sigma_{ys}}\right)^2 = 0.64 \left(\frac{Cv}{\sigma_{ys}} - 0.01\right)$$

where K_{IC} is in $\text{MPa}\sqrt{\text{m}}$, σ_{ys} is in MPa and C_V is in J. By the use of these kind of correlations, the structural integrity engineers can obtain the K_{IC} values that they need for their assessments, either from C_V data of the material specifications or from Charpy test specimens extracted from the in-service components, which require less material than the necessary to perform K_{IC} standard tests. It is important to mention that the Rolfe-Novak correlation gives lower bound values of K_{IC} therefore the integrity assessments done with them will be conservative.

References

1. Inglis CE (1913) Stresses in a plate due to the presence of cracks and sharp corners. *Trans Inst Naval Arch* 55:219–230
2. Griffith AA (1920) The phenomena of rupture and flow in solids. *Philos Trans Ser A* 221:163–168
3. Irwin GR (1957) Analysis of stresses and strain near the end of a crack traversing a plate. *J Appl Mech* 24:361–364
4. Westergard HM (1939) Bearing pressures and cracks. *J Appl Mech* 61:A49–A53
5. Rice JR (1968) A path independent integral and the approximate analysis of strain concentration by Notches and Cracks. *J Appl Mech* 35:379–386
6. Wells AA (1961) Unstable crack propagation in metals: cleavage and fast fracture. In: *Proceedings of the crack propagation symposium*, vol 1, Paper 84, Cranfield, UK
7. Dowling AR, Townley CHA (1975) The effect of defects on structural failure: a two-criteria approach. *Int J Press Vessel Pip* 3:77–107

Chapter 7

Fatigue



Abstract This chapter presents a thoroughly description of the fatigue phenomena in engineering materials, starting with an historical synopsis of fatigue failures, the Wöler's concept of fatigue life and a description of the S-N curves. A brief description of the fractographic characteristics of fatigue fracture surfaces and mechanisms is given. The central part of this chapter presents the mechanical methods of fatigue characterization, describes the factors that affect fatigue endurance and explain the most widely accepted methods to estimate the fatigue life of structural components, including the Haigh's diagram, the Weibull's statistical analysis, the Miner's rule and the Manson and Coffin Universal Slopes Method. The fatigue crack growth behavior and the Paris' law are described at the end of the chapter.

7.1 Definition and History of Fatigue

Fatigue is a progressive cracking process under the action of repetitive or fluctuating loads that culminates in the fracture of a material. Since the beginning of the use of materials, fatigue has been a concern for both engineering and science, mainly because fatigue can occur in components under stresses significantly below the yield strength, thus in the early eras of industrialization, fatigue was not considered in design nor predicted for in-service components, but its consequences were often catastrophic. In addition, fatigue does not produce apparent changes either in geometry or microstructure and the fatigue cracks are very fine, so it is difficult to detect during in-service inspections. All of this made fatigue to be considered as the main cause of failure of mechanical and structural components.

The fatigue phenomenon was acquainted early in the nineteenth century. In that time, the brittle-like fractures of machinery and structural components that failed by fatigue was interpreted as a "crystallization"—a way to say embrittlement- of metal caused by frequent use, so it was believed that in-service materials aged until they would become brittle. Thanks to the works of the German engineer August Wöhler, in the second half of the XIX century, it was known that fatigue is a progressive cracking process produced by the action of repeated loads. Wöhler introduced the

formal concept of fatigue and his contributions are still relevant nowadays. It is known that fatigue occurs in practically all engineering materials, including plastics and ceramics, but despite the great scientific and technological advances obtained, fatigue is yet to be thoroughly understood and currently, a great deal of research is carried out in order to better understand its mechanisms, to improve the fatigue resistance of engineering materials, discover new crack detection methods and to increase the accuracy of the methods to predict fatigue life.

For fatigue to take place, it is necessary to comply with three conditions:

1. A tension stress high enough, but below the tensile strength of the material.
2. A stress variation higher than a material property called “fatigue limit”.
3. A sufficient number of accumulated load cycles.

If any of the three conditions is absent, fatigue will not occur, however the identification of a fatigue causing condition requires a careful analysis. As shown in the examples of Table 7.1, there may be cases in which fatigue does not occur, even when the three above mentioned conditions are present. The main reason is that the magnitude of each one of the three factors should be enough to cause the phenomenon.

In engineering design, it is unimportant whether fatigue failures will occur or not, because theoretically fatigue loading occur all the time, but instead the important questions are: How long it will take for it to happen (number of cycles)? And, if such length of time or number of cycles is longer than the expected service time of the component? For this reasons fatigue is assessed in terms of elapsed time or number of cycles to failure, which is termed as *fatigue life*. Unfortunately fatigue life is very difficult to predict, because there are many factors that influence it, among which the most important are:

1. State of stress and stress concentration (Geometry)
2. Mechanical properties (hardness, tensile strength, fracture toughness)
3. Microstructure and thermal treatment
4. Temperature and environment
5. Residual stresses
6. Surface finish.

Before analyzing the characteristics of fatigue and the effect of internal variables (mechanical properties, microstructure, residual stress) and external (environment, stress concentration, surface finish), it is interesting to know the historical events that encouraged fatigue research or that contributed to a better understanding of the phenomenon. Table 7.2 presents a brief review of the paramount events in fatigue history.

In 1860, Wöler carried out the first scientific research on fatigue by cyclic load tests at a constant stress amplitude, presenting his results in graphs of *Stress amplitude vs. logarithm of number of failure cycles*, known as *life curves* or S-N that have the form shown in Fig. 7.1. Thanks to such development, the first discoveries were made in regards with the nature of fatigue, among which outstand:

Table 7.1 Examples of the likelihood of having fatigue failure

Case	Likelihood of fatigue	Analysis of causes
Airplane wings	High	The wings are always under fluctuating stresses, in each takeoff and landing and during the flight. An airplane makes thousands of flights during its service life
Internal combustion motor moving parts	High	These parts are under high amplitude and frequency stresses. A typical car motor works at average 2500 rpm for several thousand hours
Elevator cable	High	The stresses are mainly tensile and there are large load fluctuations and high frequency of use
Truck chassis	Medium	The loads vary over a wide range and repeat large number of times, but designers restrict stresses below the fatigue limit. Often operators overload their trucks
The columns of a metallic structure building	Medium	The column is under quasi-static compression stresses, but earthquakes introduce oscillating stresses that can be high enough to cause fatigue, but the period of return of significant earthquakes can be of several years, depending on the geographical zone
Bolts in machinery, pumps cases and other rotating equipment	Medium	They are designed to endure a large number of load cycles, but overtightening or a corrosive environment may lead to premature fatigue failures
Automobile body	Low	The vibration of the motor is transferred to the car's body and there are variable stresses induced by the wind pressure, however their amplitude is very low to as produce fatigue
Large oil storage tank	Low	The shell is under fluctuating stresses caused by the filling and discharge operations, however this operation is performed few times a month, so there is not enough number of stress cycles
Oil transport pipeline	Low	The stress significantly fluctuate during startups and shutdowns, however, they happen rarely since oil transport pipelines operate under steady and continuous conditions most of the time

1. The number of cycles necessary to produce failure by fatigue increases by a power function of the stress amplitude.
2. There is a stress amplitude value under which fatigue does not occur, termed as *fatigue limit*.

Table 7.2 Historical synopsis of fatigue

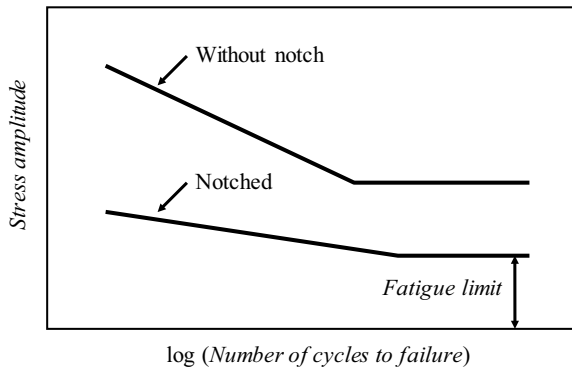
Year	Event
1837	Wilhelm Albert publishes the first article on fatigue
1843	In England it is recognized that railroad wheels and rails can be the result of repetitive stresses. It is wrongly assumed that frequent use “crystalizes” metals, favoring brittle fracture
1850	The term “fatigue” is introduced to describe failures resulting of the frequent use of metallic components
1860	August Wöhler, in Germany carries out the first fatigue tests. He introduced the concept of fatigue life, the S-N curves and the fatigue limit. His studies focused on railroads
1890	Gerber and Goodman in England study the effect of mean stress in fatigue
1919	The 1919 Boston Great Molasses Flood is attributed to a fatigue failure. The ASME Boiler and Pressure Vessel Code is introduced, where fatigue is prevented by limiting allowable stresses
1920	A great slip activity is observed in fatigued metallic components, which leads to the idea that cyclic plastic strain is the mechanism of fatigue. The first textbooks on fatigue are published
1930	Great improvements on the fatigue life of automotive parts are achieved by shoot peening treatment. It is discovered that shoot peening induces compressive stresses that increase the fatigue limit
1945	Miner introduce the cumulative fatigue damage criterion in aircraft design, based on the linear damage hypothesis introduced by Palmgren in 1924
1952	The manufacturer, de Havilland of the United Kingdom introduces, the first commercial passenger jet airliner the DH 106 Comet. In 1954 two Comets crashed after around 3000 flights, due to fatigue of the hull. As a consequence, the structural design is improved to reduce stress concentrators that caused fatigue and the remaining life concept is introduced into the design of airplanes, replacing the design by infinite life assumption
1950	Closed loop servo-hydraulic machines are introduced; they allow the performance of pre-determined load pattern fatigue tests
1951	Weibull introduces a statistical distribution function that can be applied to estimate the probability of fatigue failures
1960	Manson and Coffin publish the Universal Slopes Law
1963	Paul C. Paris demonstrates the relation between the stress intensity factor and the fatigue crack growth rate, initiating the study of fatigue from the fracture mechanics point of view
1967	The collapse of a bridge in West Virginia, USA is caused by a fatigue crack that reached its critical size in a time much shorter than the originally expected. The acceleration of fatigue is attributed to corrosion. The concept of corrosion-fatigue is introduced
1970	An F-111 military aircraft suffers an accident caused by a pre-existing crack in a forged component that grew by fatigue after several hundred hours of flight. The investigation of this failure inaugurated the use of fracture mechanics in the assessment of crack-like flaws detected in-service in aircraft, both military and civil

(continued)

Table 7.2 (continued)

Year	Event
1972	Elber discovers the crack closure phenomenon of fatigue cracks, which explains the effects of average stress, overloads, and other factors in fatigue crack growth
1988	The Aloha Airlines Flight 243 suffered a fuselage fatigue failure while flying at 24,000 feet (7300 m). The pilot was able to land safely at Kahului airport on Maui. This accident proved the success of damage tolerant designs based on fracture mechanics
Current days	Great advances in the study of fatigue mechanisms Fatigue studies in composite and ceramic materials Studies on interactions of fatigue with corrosion and creep Introduction of fatigue assessment methods into the fitness-for-service standards and structural integrity codes

Fig. 7.1 Fatigue life curve, known as “S-N curve”



3. The presence of stress concentrators reduces drastically the number of failure cycles.

Later on, it was observed that fatigue could be divided into three stages which are:

Stage I. Damage accumulation and crack nucleation stage. Also called “*internal damage*”, it happens in absence of stress concentrators. In this stage, cyclic deformation produces dislocation substructure that lead to the formation of surface geometric discontinuities which will further develop as cracks. The extension of this stage depends on the stress concentration. At low stress amplitudes and stress concentration, this stage may represent up to 90% of the fatigue life.

Stage II. Crack growth. Consists on the stable propagation of a fatigue crack, where the crack growth rate depends on the magnitude of the stress intensity factor amplitude, because the crack grows within an elastically deformed body and the crack-tip plastic zone is much smaller than the ligament size. The fatigue mechanism is related to the cyclic strain in the plastic zone and is strongly influenced by the environment.



Stage III. Final fracture. When the crack is about to reach its critical size, fracture mechanism combines cyclic deformation at the crack tip with static fracture mechanisms, such as cleavage or ductile tearing by void coalescence.

7.2 Fatigue Fracture

Fatigue fractures are easily identifiable due to the typical appearance of the fatigue fracture surface, which is schematically shown in Fig. 7.2. The main characteristic of fatigue fracture surfaces is the presence of parallel arc marks in the form of waves, whose center seems to be located in the start zone. Such marks are referred to as *beach marks* for their similarity to the undulations formed by the wind and tide on a sandy beach.

Typical fracture surfaces can be divided into three zones:

1. **Crack initiation zone.** It is the zone where the crack first nucleates and starts growing. The fracture surface at the initiation site is smooth, flat, shiny and with small steps formed by the simultaneous nucleation of several cracks.
2. **Stable growth zone.** The fracture is a relatively flat surface, perpendicularly oriented to the direction of the maximum principal tension stress. It can be either shiny or opaque, depending on the environment where the failed piece was. The beach marks appear in this stage, and they are formed by load amplitude variations, temporary arrest of the crack or by environmental changes.
3. **Final fracture zone.** When the crack is close to reaching its critical size, the high stress concentration leads to a transition of the fracture mechanism that turns the fracture surface rougher and slanted, up to an angle close to 45° with respect to the maximum tensile stress direction, forming a shear lip in the final detachment zone, its final height depends on the material ductility.

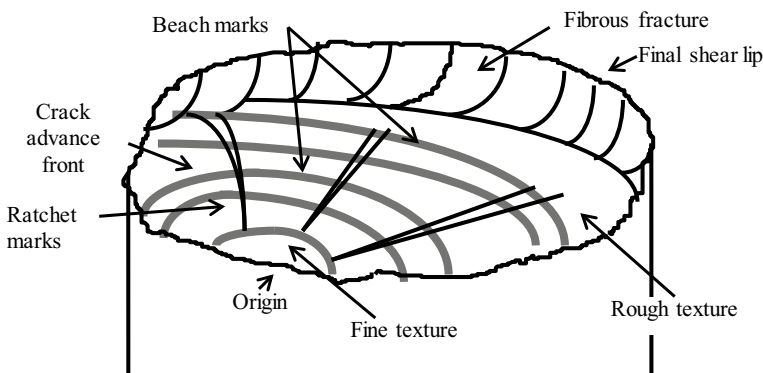


Fig. 7.2 Scheme of a typical fracture surface by fatigue

Generally, fatigue cracks start on a free surface and are often associated to stress concentrators, such as grooves, holes or pre-existing cracks, so it is relatively easy to identify the fracture initiation site. Only in few cases, like casts with severe internal defects such as blows and shrinkage, or materials with sharp shaped precipitates, may have fatigue cracks that initiate in the material's bulk, beneath the free surface. Nonetheless, the characteristics of these fatigue fractures are similar to those observed on fatigue fractures initiated on the free surface. The extent of each fatigue fracture zone depends on the magnitude of stress concentration, load amplitude and fracture toughness of material. The schemes of Fig. 7.3 show some typical configurations of fatigue fracture surfaces in bars under alternating bending stresses.

Fatigue mechanisms. In components with neither pre-existing cracks nor stress concentrators, the fatigue initiating mechanism is by dislocation glide. The most widely known model is the intrusion-extrusion, as proposed by Wood, which is schematically depicted in Fig. 7.4. In this model, the dominating slip plane goes through an alternate gliding process leading to the formation of metal extrusions on planes where dislocations reach a free surface. To keep continuity, the opposite process takes place on a close by plane; that is, the emission of dislocations into the material, thus forming an intrusion. When the intrusion is sufficiently sharp and deep, it turns into a crack. Such mechanism is favored by the planar slip conditions, where plastic deformation along dense slip bands is dominant.

It is an accepted fact that the presence of moderately or severely corrosive environments shortens the initiation stage of fatigue as compared to vacuum or inert environments. An experimental observation is that a gaseous corrosive environment

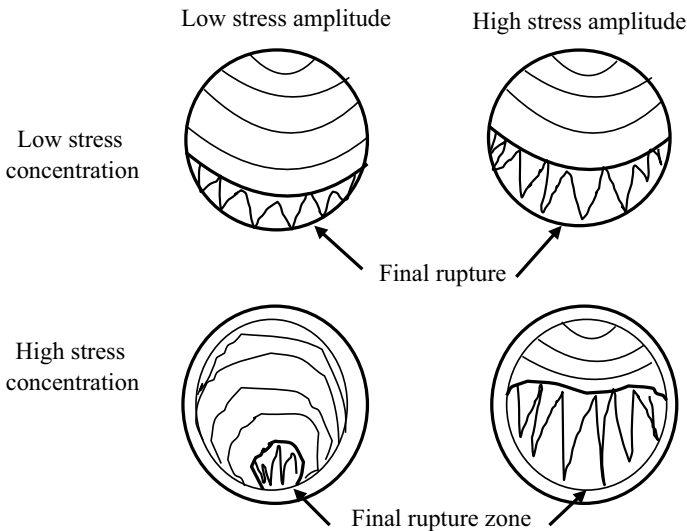


Fig. 7.3 Effect of the stress amplitude and stress concentration on the appearance of alternating bending fatigue fracture surface

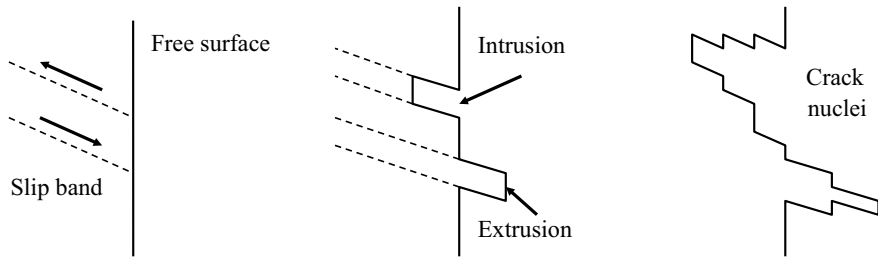


Fig. 7.4 Nucleation of cracks by fatigue due to the formation of intrusions and extrusions on the gliding bands

promotes the absorption of chemical species through the slip bands that further form cavities causing decohesion of the slip bands, as schematically shown in Fig. 7.5.

In strongly oxidant conditions, as it occurs in ovens, heaters and furnaces, the thick layers of brittle oxides fracture under tension, originating microcracks. Such microcracks provide an easy path for the intake of oxidant agents, forming an oxide penetration that cracks again and the process repeats. This mechanism is known as *thermal fatigue* and is shown schematically in Fig. 7.6.

Once the crack has nucleated and reached a macroscopic size (visible at plain sight), fatigue crack growth goes into Stage II, where the crack propagates under linear-elastic stress-strain conditions. The most remarkable characteristic of Stage II

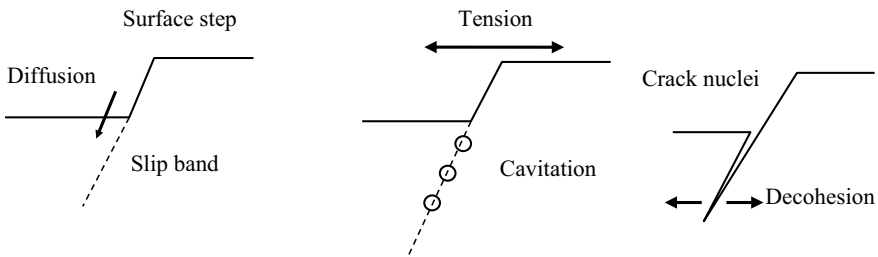


Fig. 7.5 Nucleation of cracks by fatigue due to cavitation of glide bands

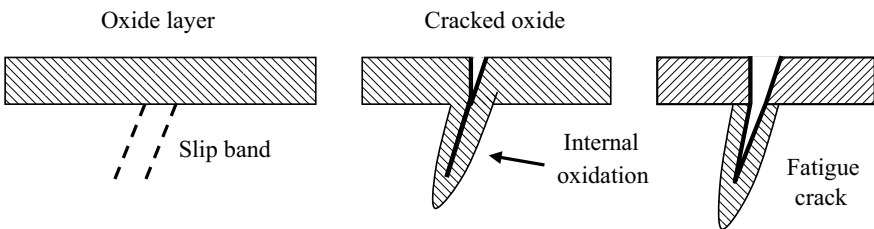
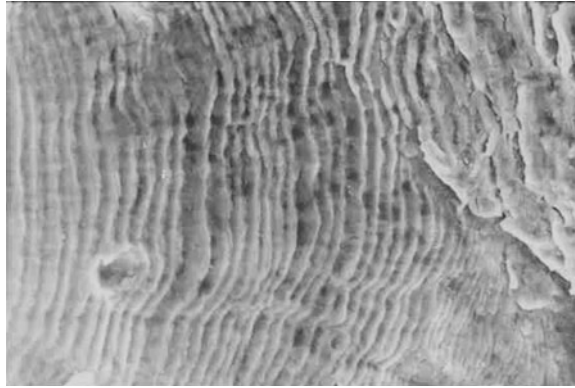


Fig. 7.6 Nucleation of fatigue cracks due to the mechanism known as thermal fatigue

Fig. 7.7 Appearance of microscopic striations on Stage II fatigue fracture surfaces



fatigue fracture surfaces at microscopic level are the *striations*, whose typical aspect is shown in Fig. 7.7. Striations are small parallel grooves with a spacing that matches the macroscopic crack growth rate, therefore, each striation corresponds to one load cycle.

The presence of microscopic striations in a fracture is unmistakable evidence of fatigue, but its absence does not mean that it is not fatigue, because striations are not formed under several conditions. It has been experimentally demonstrated that the formation of striations requires that the following conditions are met:

- (1) Stage II of crack growth, preferably in a high level of ΔK .
- (2) Moderately corrosive atmosphere (humid air).
- (3) High ductility ($\epsilon_f > 20\%$ at the test temperature).
- (4) Characteristics of both multiple slip and cross slip in the process zone.
- (5) Mode I of crack opening displacement or the prevalence of it.

Up to date, a universal mechanism of striation formation has not been recognized, although it is generally accepted that the alternating plastic deformation at the crack tip is the basic mechanism of Stage II fatigue. Pelloux, in 1969, proposed that a Stage II fatigue crack may move forward by the alternate plastic slip at the crack tip, where striations are formed due to the irreversibility of slip caused by surface oxidation and gas adsorption. This model is schematically presented in Fig. 7.8, and it succeed in explaining the existence of striations in the air and their absence in vacuum, although it is questionable why in noble metals, in which the oxide layer is so thin as to exert an effect on the reversibility of slip, perfectly defined striations are observed.

Nix and Flower [1] observed the dislocation substructures under the fracture surfaces aluminum plates fatigued in air. They found alternate bands of high and low dislocation densities, matching the striation spacing. Based on such observations, they assumed that in the up-load half cycle, high density dislocation bands are formed, but at maximum load, the hydrogen formed by decomposition of air moisture is absorbed into the plastic zone, causing a crack extension by cleavage. McEvily and Gonzalez [2] made direct observations of the cracks tips of metals fatigued in air and vacuum. They found that the tip of cracks fatigued in air are

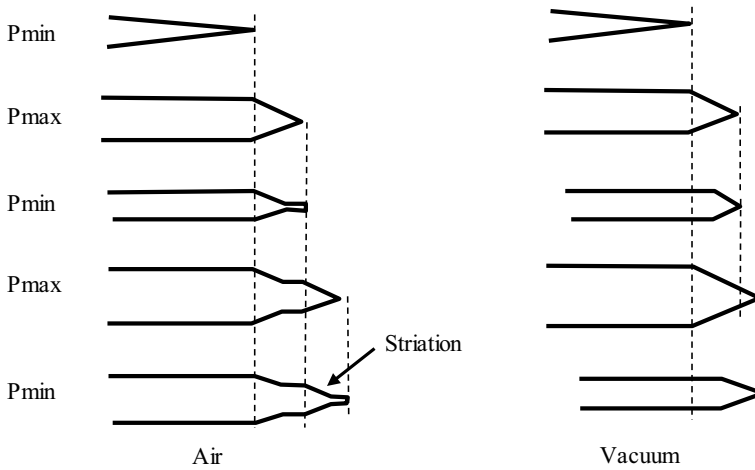


Fig. 7.8 Crack growth by fatigue mechanism through alternate slip as proposed by Pelloux

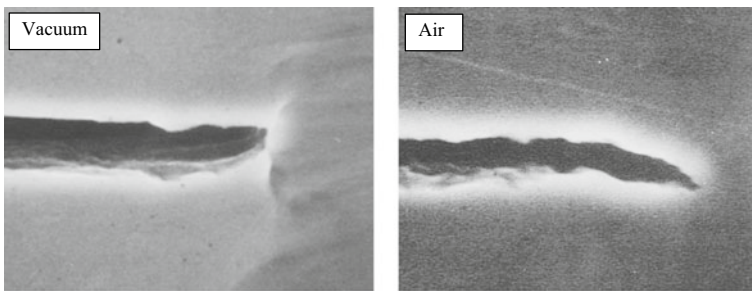


Fig. 7.9 Crack tip deformation of stainless steel fatigued in vacuum and air [2]

sharper than those in vacuum, as shown in Fig. 7.9. They attributed this effect to localization of cyclic plastic deformation in air, caused by a higher dislocation slip irreversibility induced by a thin oxide layer and hydrogen adsorption at the crack tip surface in air. Again, hydrogen is produced by the catalytic decomposition of humidity in the air.

7.3 Mechanical Characterization of Fatigue

Fatigue is characterized by the applied load cycle, which features three main characteristics: (1) amplitude, (2) frequency and (3) mean stress, these are shown in Fig. 7.10. The load cycle amplitude ($\Delta\sigma$) is defined as the difference between maximum (σ_{\max}) and minimum stress (σ_{\min}).

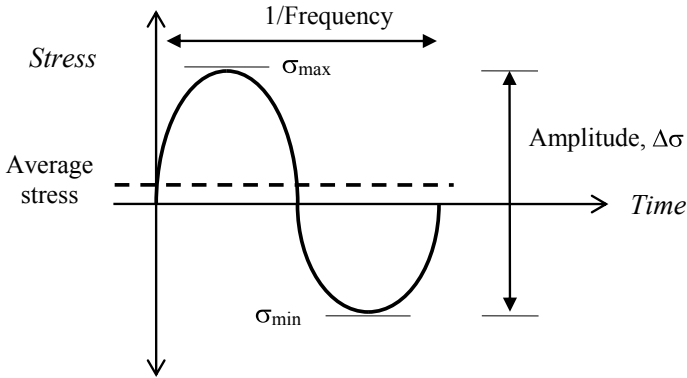


Fig. 7.10 Load cycle parameters

$$\Delta \sigma = \sigma_{\max} - \sigma_{\min}$$

Since $\Delta \sigma$ does not indicate the location of the mean stresses in the cycle, it is necessary to introduce the R ratio, defined as $R = \sigma_{\min} / \sigma_{\max}$. The load cycle as a function of R is shown in Fig. 7.11.

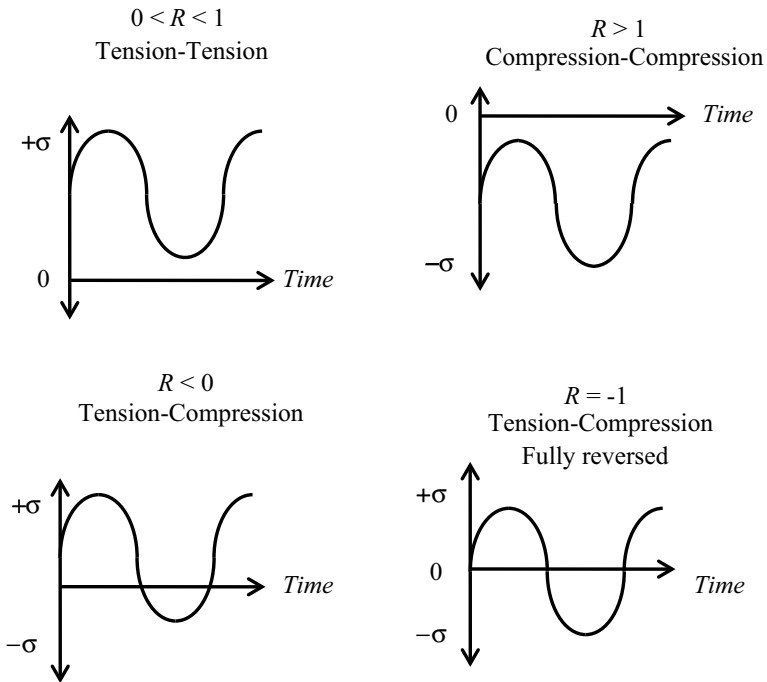


Fig. 7.11 Load cycle patterns as a function of the R value

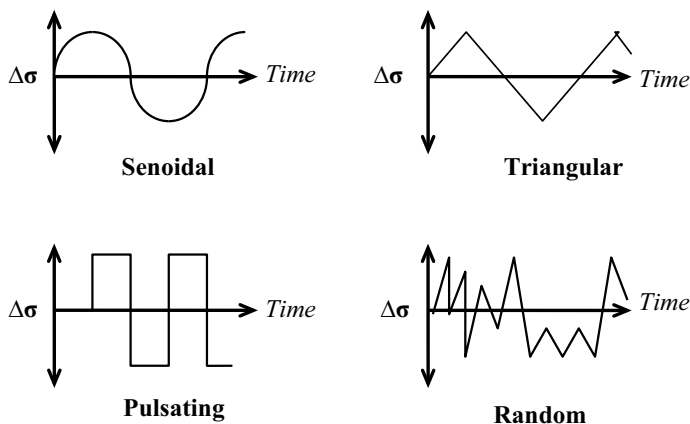


Fig. 7.12 Typical fatigue cycles according to the load ramp shape

The types of fatigue cycles, based on the shape of the load ramp are shown in Fig. 7.12. These are the typical types, although in real life there may be combinations of them.

As it was already mentioned, the aim of a fatigue assessment is to determine the fatigue life or fatigue endurance, which is the number of load cycles that a material withstands until failure at a given stress amplitude. The basic test to determine the fatigue life is carried out by applying on a smooth bar a constant amplitude load wave of predetermined form and frequency and recording the number of cycles to failure. Fatigue tests can be the following types:

- *Tests of in-service components.* They are carried out on the already installed components operating in actual service conditions, applying the necessary instrumentation to record the load pattern and the counting the number of elapsed cycles. They are used when it is desired to assess the performance in real life conditions, but rarely are conducted to failure, because the objective is to verify whether the component can endure a predetermined number of cycles without failure. They are costly and difficult to perform due to the complexity of the instrumentation to record the load cycles.
- *Bench tests on actual components.* In these tests, the component is fitted on a machine that applies a cyclic load pattern, similar to the one that occurs in the actual service. Usually the test stress amplitude is higher than the expected in service to accelerate the test. They are used as final proof of a design and for quality control of critical components such as hydraulic car brake tubing, helicopter rotors, cranes, elevators and etcetera. They are also used for assessing the performance of components that receive high frequency use such as switches, airplane seats, locks, and so forth; in this case the tests are performed on a sample of a production lot.

- *Normalized laboratory tests.* These are standardized tests where the procedure, test specimens, equipment, instrumentation and presentation of results are specified to assure precision and representativeness. The most common standards are those of the American Society for Testing Materials (ASTM) and the International Standards Organization (ISO). The most common laboratory fatigue tests are:
 - *S-N life tests:* Described by ASTM E466-15 “Standard Practice for Conducting Force Controlled Constant Amplitude Axial Fatigue Tests of Metallic Materials”. It consists of a smooth specimen (no grooves nor notched) with a geometry that produces a uniform stress. A constant load amplitude is applied, with fixed R and frequency, under controlled environment conditions. The results are presented in the form of S-N curves. The results are extrapolatable and useful for design, materials selection and Fitness-For-Service assessments. Since the number of failure cycles in these tests is usually high (more than tens of thousands), they are termed as *high-cycle fatigue*.
 - *Low cycle strain-controlled fatigue test:* The test is described in the ASTM E606-12 “Standard Test Method for Strain-Controlled Fatigue Testing”. It uses a specimen of regular geometry, with no precracks or grooves, tested at constant stress amplitude where the maximum stress surpasses the yield strength to produce plastic strain. The stress-strain cycle is recorded. It is used for research since the equipment and instrumentation are costly. As the number of failure cycles in these tests is usually below ten thousand, they are called *low cycle fatigue tests*.
 - *Crack growth tests:* The test is described in the ASTM E647-15e1 “Standard Test Method for Measurement of Fatigue Growth Rates”. In these, a load cycle pattern is applied to a pre-cracked specimen and crack growth is continuously measured, along with the number of cycles. The crack growth rate data (da/dN) as a function of the stress intensity factor amplitude ΔK are recorded to plot the Paris curves. The load amplitude and the specimen dimensions are adjusted so the crack grows in linear-elastic conditions so the ΔK function is valid. They are very precise, and the results can be applied for a variety of purposes such as life assessment, Fitness-For-Service, material selection and in fatigue research.

S-N Life Curves. The results of fatigue life tests are typically presented in the S-N curve, which is a graph on semi-logarithmic paper of the stress amplitude, represented by the symbol $\Delta\sigma$, versus the number of cycles to failure represented by the symbol N_f . The typical S-N curves for metallic materials have the form shown in Fig. 7.13. The main feature of the S-N curve is that N_f exponentially varies with respect to $\Delta\sigma$, indicating that fatigue life is quite sensitive to stress level. The stress level in which the S-N curve becomes asymptotic represents the *fatigue limit*. For practical purposes, the fatigue limit is defined as the stress level in which the material endures more than 10^8 cycles. However, neither all materials nor all load cases exhibit a fatigue limit, as for example, non-ferrous materials such as aluminum. Also, the presence of grooves, residual stresses and corrosive environments can suppress the fatigue limit.

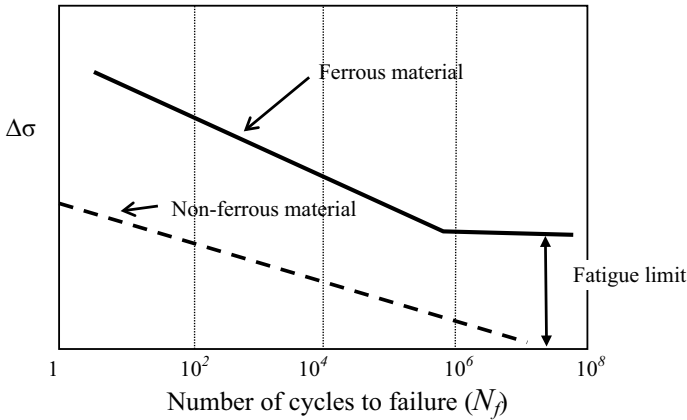


Fig. 7.13 Typical S-N curve for metallic materials

At first, the interpretation of an S-N curve is simple: the further up and to the right the curve is, the better fatigue strength. However, the entire assessment interval has to be considered. For example, in Fig. 7.14, the performance of material B is better at high $\Delta\sigma$ values, but at lower $\Delta\sigma$ values the opposite occurs, material A is better and additionally has a higher fatigue limit.

Another relevant behavior revealed by the S-N curves is that fatigue strength is proportional to tensile strength. Table 7.3 shows the tension properties and fatigue limit for some metallic alloys widely used in engineering application. For steels, it was found that the fatigue limit (σ_{Lim}) is approximately:

$$\sigma_{Lim} = 0.5\sigma_f, \quad \text{for } \sigma_{uts} \leq 200 \text{ ksi}$$

$$\sigma_{Lim} = 100 \text{ ksi}, \quad \text{for } \sigma_{uts} > 200 \text{ ksi}$$

Fig. 7.14 S-N curve of two materials, where material A has a better performance than B at lower stress amplitudes and vice versa

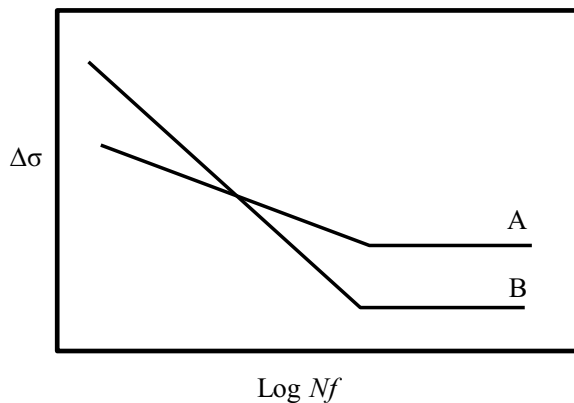


Table 7.3 Tension properties and fatigue limit based on 10^8 cycles of metallic alloys

Material	Condition	σ_{uts} MPa (ksi)		σ_{ys} MPa (ksi)		σ_{Lim} MPa (ksi)	
<i>Steels</i>							
1015	Annealed	455	(66)	275	(40)	240	(35)
1040	Normalized	670	(97)	405	(59)	345	(50)
1040	Quenched	965	(140)	855	(124)	410	(60)
4340	Normalized	745	(108)	475	(69)	340	(49)
4340	Q&T (204 °C)	1950	(283)	1640	(238)	480	(70)
4340	Q&T (427 °C)	1530	(222)	1380	(200)	470	(68)
4340	Q&T (538 °C)	1260	(183)	1170	(170)	670	(97)
HY140	Q&T (538 °C)	1030	(149)	980	(142)	480	(70)
D6AC	Q&T (260 °C)	2000	(290)	1720	(250)	690	(100)
9Ni-4Co-0.25C	Q&T (315 °C)	1930	(280)	1760	(255)	620	(90)
300M	–	2000	(290)	1670	(242)	800	(116)
<i>Aluminum alloys</i>							
1100	Annealed	90	(13)	34	(5)	34	(5)
3014	T6	483	(70)	414	(60)	124	(18)
3024	T3	483	(70)	345	(50)	138	(20)
6061	T6	310	(45)	276	(40)	97	(14)
7075	T6	572	(83)	503	(73)	159	(23)
<i>Titanium alloys</i>							
Ti-6Al-4V	Annealed	1035	(150)	885	(128)	515	(75)
Ti-6Al-2Sn-4Zr-2Mo	Annealed	895	(130)	825	(120)	485	(70)
Ti-5Al-2Sn-2Zr-4Mo-4Cr	Annealed	1185	(172)	1130	(164)	675	(98)
<i>Copper alloys</i>							
80Cu-30Zn	Annealed	524	(76)	435	(63)	145	(21)
90Cu-10Zn	Annealed	420	(61)	370	(54)	160	(23)
<i>Magnesium alloys</i>							
HK31A-T6	T6	215	(31)	110	(16)	62-83	(9-12)
AZ91A	Annealed	235	(34)	160	(23)	69-96	(10-14)

where σ_{uts} is the tensile strength. These correlations are widely accepted, although it must be remembered that only apply for carbon and low alloy steels, but fails for non-ferrous materials, corrosive environments, high temperatures or rough surfaces, where the fatigue limit is severely reduced.

7.4 Factors Affecting Fatigue

The main external factors that affect fatigue are: environment, surface finish and load history, whereas the internal factors are those inherent to the material such as mechanical properties microstructure, and material defects. In the following paragraphs, some of the most important factors in fatigue life will be analyzed.

The most important external factor affecting fatigue is stress concentration, that not only reduces the fatigue life, but it may make the fatigue limit disappear. The main reason is because stress concentration locally increases the stress amplitude, thus exponentially reducing the fatigue endurance. The quantitative effect of stress concentration by semi-elliptical notches on the fully reversed ($R = -1.0$) S-N curves was proposed by Neuber, by the *strength reduction factor* K_f given by the following equation:

$$K_f = 1 + q(K_t - 1)$$

where K_t is the elastic stress concentration factor under fixed load, q is a notch sensitivity parameter, so if $q = 0$, the material is insensitive to notches and if $q = 1.0$ there is full notch sensitivity. Peterson proposed a formula to estimate the value of q , as:

$$q = \frac{1}{1 + \frac{\alpha}{r}}$$

where r is the notch radius, and α is a material's parameter calculated by the following equations:

$$\alpha = 0.0254 \left[\frac{2070}{\sigma_{uts}} \right]^{1.8}; [\text{MPa}, \text{mm}]$$

$$\alpha = 0.001 \left[\frac{300}{\sigma_{uts}} \right]^{1.8}; [\text{ksi}, \text{in}]$$

The effect of notches on fatigue endurance is calculated by multiplying $\Delta\sigma$ by K_f and reading the value of N_f from the S-N curve. It is important to point out that this method only applies for high-cycle fatigue ($N \geq 10^5$). A quick estimate of the effect of stress concentration on the S-N curve can be made by plotting a straight line between the points (σ_{uts} , $N = 1$ cycle) and the value of $\Delta\sigma$ corresponding to 10^6 cycles, in a semi-logarithmic plot, as shown in Fig. 7.15. Then, another straight line is plotted from the same origin to the point valor $\Delta\sigma$ (10^6 cycles) K_f , this line corresponds to the notched S-N curve.

The surface condition is closely related to stress concentration, but at a less degree. As shown in Fig. 7.16, the rougher the surface is, lesser the fatigue strength. The reason for such behavior is that, in rough surfaces the crack initiation stage is

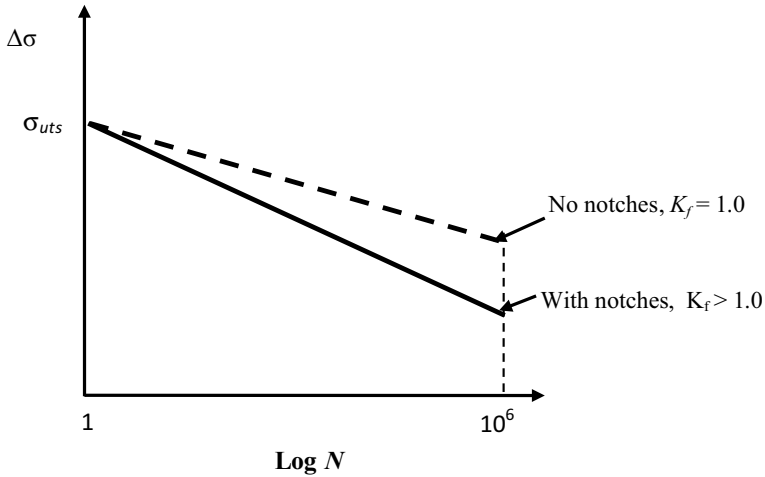
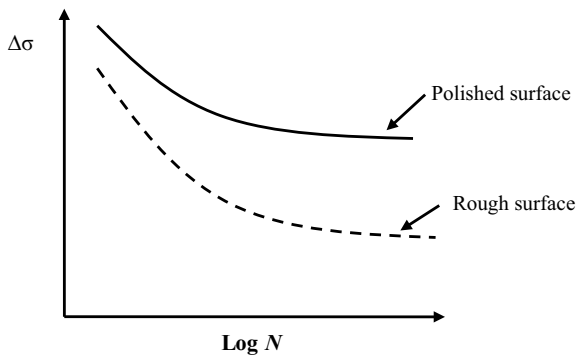


Fig. 7.15 Estimation of the S-N curve of a notched specimen, for fully reversed loading $R = -1$

Fig. 7.16 Effects of surface condition on the S-N curve of a metallic material



practically eliminated, so the cracks begin to grow since the very beginning of the cyclic loading.

The next effect in importance is the environment. It is well known that fatigue can be purely mechanical and occur in inert environments such as vacuum, but the common case for metallic materials is that the service environment is either moderately corrosive, like air, or very corrosive like seawater. In the S-N curve, the effect of environment is similar to that of stress concentration, that means that the more corrosive the environment the lesser the fatigue strength, while highly aggressive environments suppress the fatigue limit disappears, as shown in Fig. 7.17. The effect of the environment corrosiveness in fatigue is attributed to the fact that during crack growth the new fracture surfaces are very active and absorb chemical species that dissolve the material or make it more brittle, thus facilitating crack growth. There is a special condition called fatigue-corrosion, which appears

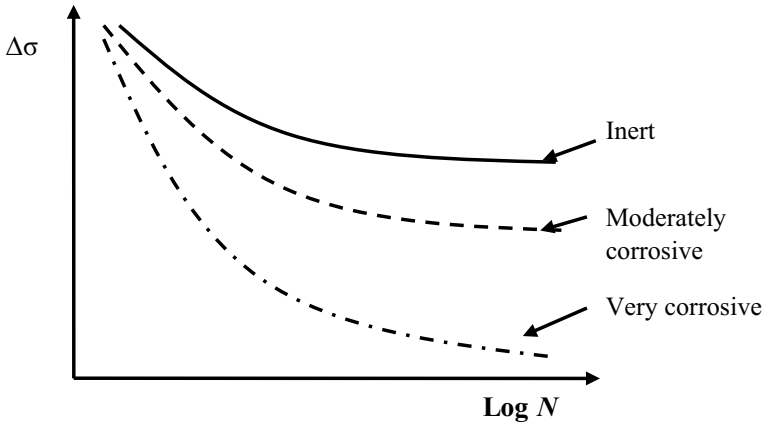


Fig. 7.17 Effect of environment on the S-N curve of a metallic material

when the crack grows due to the combination of stress corrosion cracking and mechanical fatigue.

Finally, the intrinsic factor that most influences fatigue is the microstructure, the main effects are shown schematically in Fig. 7.18. As it is known, microstructure determines mechanical properties, which have a strong effect in fatigue strength, but additionally, the microstructure determines the slip mode, that is, the dislocations behavior, in addition to influencing on the crack path and the fracture mechanism, thus affecting the three stages of fatigue.

As a general rule, the lowest fatigue strength is exhibited by pure metals, probably because they are more susceptible to corrosion and have low mechanical strength. Solid solutions have a higher fatigue strength, mainly due to their higher

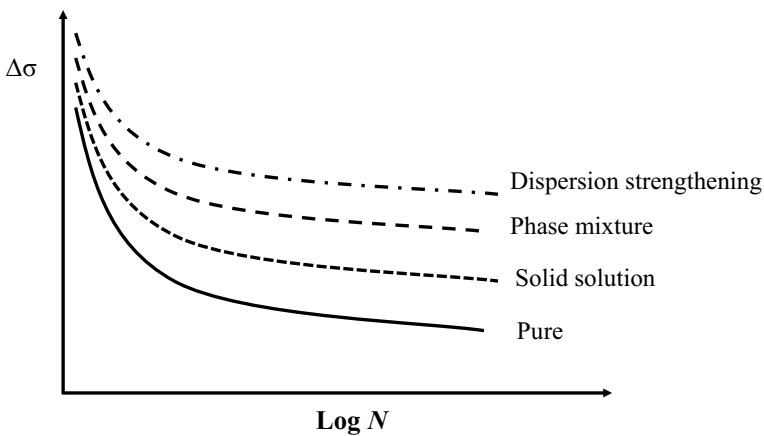


Fig. 7.18 Effect of microstructure on the S-N curve of metallic alloys

mechanical strength. Second phase aggregates are mechanically stronger than solid solutions, but second phases also provide crack closure mechanisms that delay crack growth, which increases fatigue life. Finally, the highest fatigue strengths are observed in materials hardened by particle dispersion, a behavior attributed to the excellent combination of mechanical properties and the delaying effects of the precipitates on the fatigue crack growth mechanism.

7.5 Cyclic Strain Behavior

The invention of closed loop servohydraulic testing machines in the 1950s allowed the study of mechanical behavior of materials under cyclic strain thanks to the capability of these machines to apply predetermined load patterns and instantaneously measure the load-strain response. Figure 7.19 shows a scheme of a closed loop servo hydraulic mechanical testing machine.

A closed loop testing machine consists of a load frame, with a moving head and a fixed one where the specimen is held by a set of grips. A load cell is placed on the fixed head while the moving head is a hydraulic actuator activated by a servo-controlled valve. The system is completed with a linear variable differential transducer (LVDT) that measures the displacement and an extensometer to measure strain. The signal of the transducers is compared with the preset load pattern and the servo-control adjusts the actuator to obtain the pre-set load, displacement or strain. Since the comparison is made at high frequency (MHz), the adjustment is almost instantaneous, so the control is executed on real time.

By monitoring the stress and strain in a servo-hydraulic closed loop machine, a *hysteresis curve* can be obtained, which depending on the strain cycle can be as

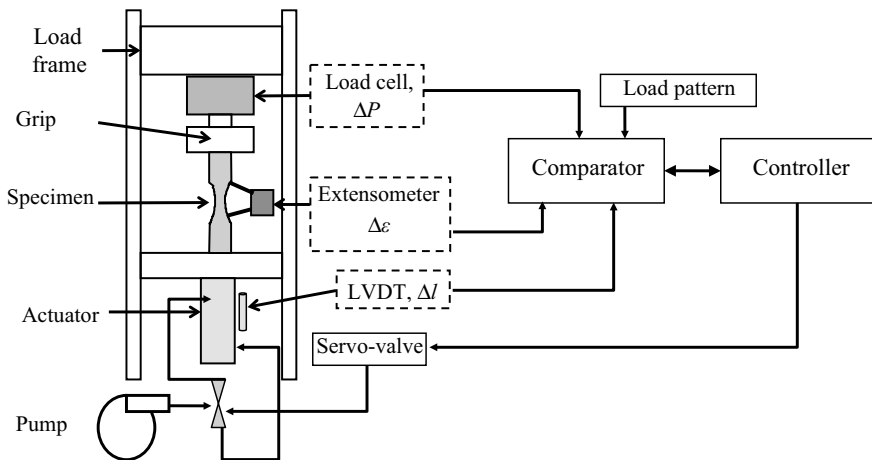


Fig. 7.19 Schematic of a closed loop servo-hydraulic mechanical testing machine

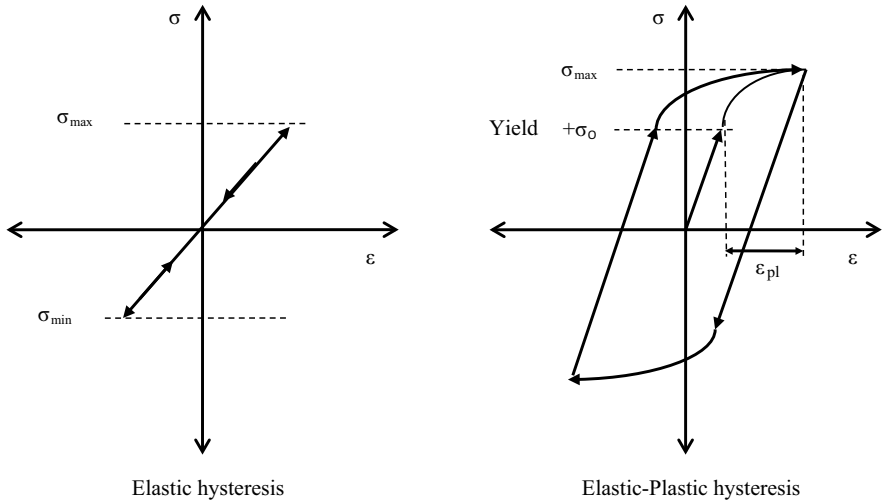


Fig. 7.20 Cyclic load hysteresis curves

shown in Fig. 7.20. Early research showed that *internal damage by fatigue occurs only when there is a cyclic plastic strain component*, all of which led to the study of cyclic strain through the elastic-plastic hysteresis cycle; so this area was termed as *cyclic strain behavior*.

If the hysteresis loop is recorded during several cycles, two types of responses can be observed, as shown in Fig. 7.21. One case is when the amplitude of the hysteresis loop progressively diminishes until reaching a constant amplitude, termed as *saturation*, at this point it is said that the material exhibits *cyclic hardening*. The opposite behavior is when the hysteresis loop amplitude increases up to reaching saturation as well, then it is said that the behavior is *cyclic softening*.

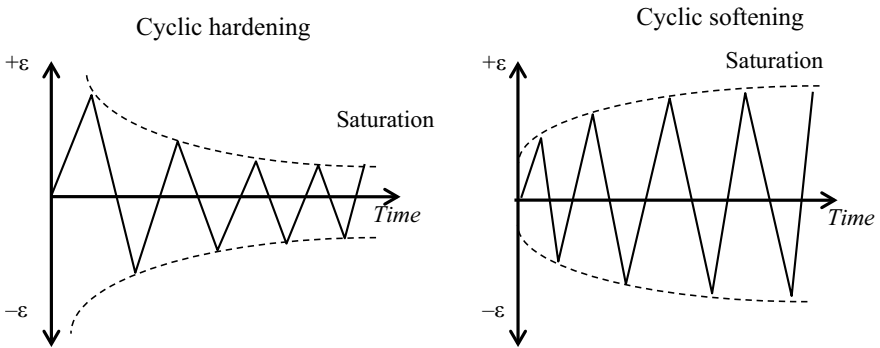


Fig. 7.21 Hysteresis loop response under cyclic loading

Smith et al. [3] observed that cyclic hardening occurs primarily in annealed materials whereas softening is observed in cold worked materials. In other words, in cyclic deformation, strain hardened materials soften and soft annealed materials harden. Manson also found that most engineering alloys, that meet the condition $\sigma_{\text{uts}}/\sigma_0 > 1.4$, cyclically harden, and if $\sigma_{\text{uts}}/\sigma_0 < 1.2$, they soften during cyclic strain; for values in between, the response is variable. Furthermore, it was found that other engineering materials, such as polymers and ceramics, do not exhibit neither cyclic hardening nor softening. If saturation has been reached, and the material is tested in a monotonic uniaxial tension, the stress-strain response may notably differ from the material without previous cyclic strain, as shown in Fig. 7.22.

An equation to calculate the total strain amplitude can be derived as the sum of elastic strain amplitude $\Delta\varepsilon_e$, plus plastic strain amplitude $\Delta\varepsilon_p$, which has the following form:

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p = \frac{\Delta\sigma}{E} + \left(\frac{\Delta\sigma}{K'}\right)^{1/n'}$$

where $\Delta\sigma$ is the stress amplitude, E is the Young's modulus, K' and n' are the Hollomon's equation constants, determined from the real stress-strain curve of the material cyclically strained until saturation. The n' values for most engineering alloys varies from 0.1 to 0.2.

The TEM observation of thin foils prepared from initially annealed metals subject to cyclic strain, show that the dislocation density rapidly increases after the first strain cycles, reaching dislocation densities similar to those of heavily cold worked metals, as shown in the example of Fig. 7.23. As the grains get filled with dislocations, they go through intense interactions mechanisms like those of strain hardening. In previously cold-worked materials that are cyclically strained, the

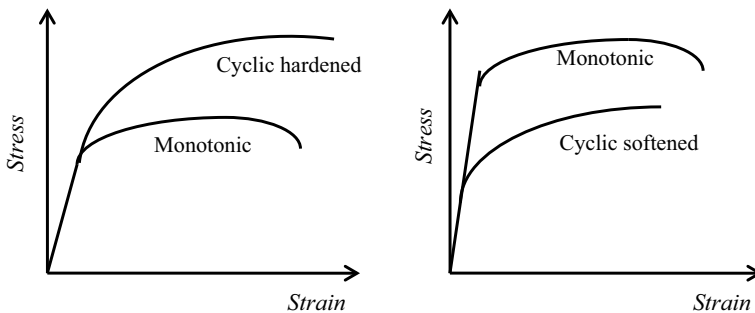


Fig. 7.22 Uniaxial tension response of a metallic material with and without previous cyclic strain

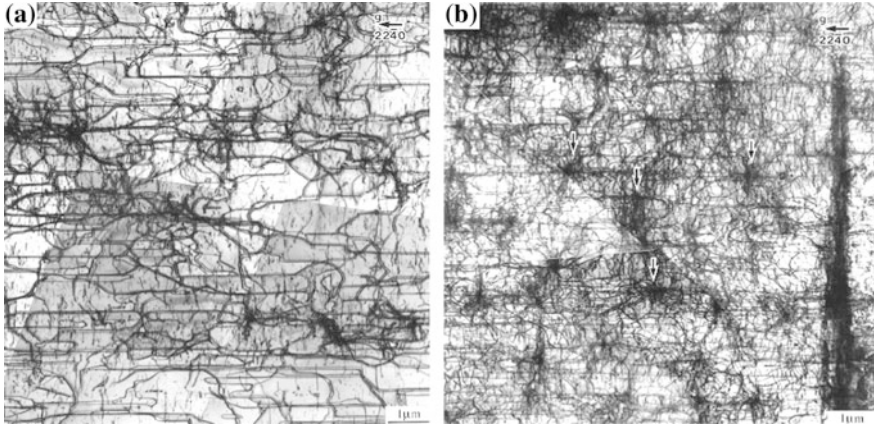


Fig. 7.23 Evolution of dislocation arrays produced by cyclic deformation [5]

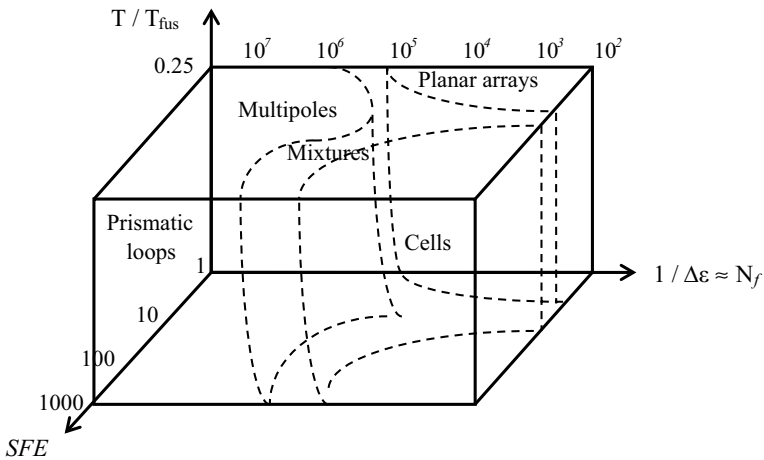


Fig. 7.24 Map of dislocation arrays produced by cyclic deformation in *fcc* metallic materials

initial dislocation arrays are modified. Feltner and Laird [4] elaborated a map of the dislocation substructures produced by cyclic strain in *fcc* materials as a function of temperature, stacking fault energy (*SFE*) and strain amplitude, which is shown in Fig. 7.24. They found that wavy slip materials (high *SFE*) feature the same dislocation substructures, independently of the initial condition, whether annealed or cold worked, whereas in planar slip materials (low *SFE*), the final dislocation array depends on the initial condition.

7.6 Fatigue Life

Since fatigue tests to generate S-N curves are costly and time consuming, several methods have been proposed to estimate fatigue life as a function of stress amplitude (σ_a), mean stress (σ_m), yield strength (σ_{YS}) or tensile strength (σ_{uts}). The most common equations for fully-reversed fatigue are:

$$\frac{\sigma_a}{\sigma_{Lim}} + \frac{\sigma_m}{\sigma_{YS}} = 1 \quad \text{Soderberg (EUA, 1930)}$$

$$\frac{\sigma_a}{\sigma_{Lim}} + \frac{\sigma_m}{\sigma_{uts}} = 1 \quad \text{Goodman (England, 1899)}$$

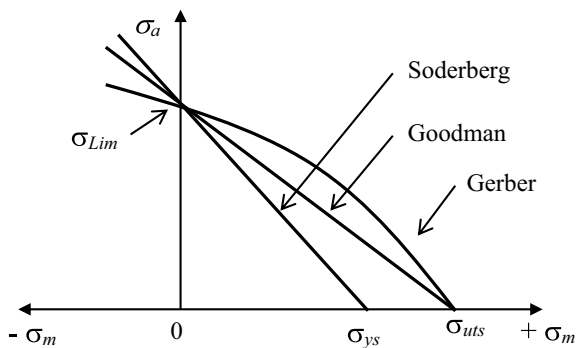
$$\frac{\sigma_a}{\sigma_{Lim}} + \left(\frac{\sigma_m}{\sigma_{uts}} \right)^2 = 1 \quad \text{Gerber (Germany, 1874)}$$

The previous equations are graphically representation in the *Haigh's diagram* shown in Fig. 7.25. Its use is as follows: A point of coordinates (σ_a , σ_m) is located in the Haigh's diagram. If the point falls within the axes and the curve, fatigue life is infinite, but if it falls off the limits, the component will fail prematurely.

In general, Goodman and Gerber's models fit best the experimental data and for that reason they are the mostly used in industry, whereas Soderberg's method is quite conservative and thus not much used. Gerber's model is applied for ductile materials and Goodman's fits better for high strength materials with relatively low ductility.

Statistical analysis of S-N data. If a high enough number of fatigue tests is performed at the same stress level, the S-N data may appear as a Gauss bell distribution, as shown in Fig. 7.26. The area under the bell distribution represents the failure probability, at the middle of the bell, there is a 50% failure probability, that is to mean that 50% of the specimens will fail before the corresponding number of cycles, represented by N_{50} . At three standard deviations to the right of the mean, 99% of the specimens will have failed before enduring N_{99} cycles and at three

Fig. 7.25 Haigh's diagram for determining fatigue life



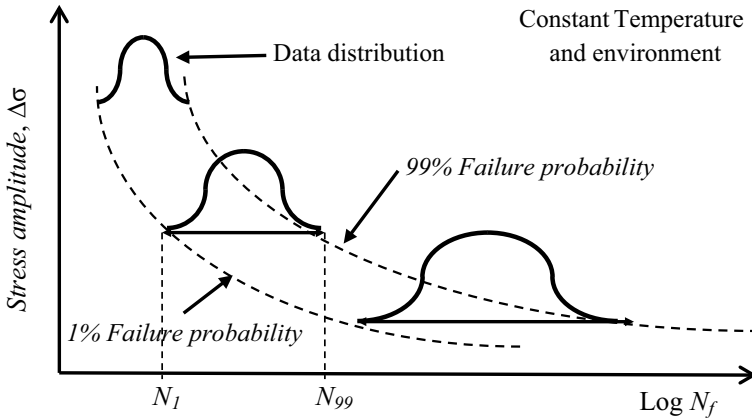


Fig. 7.26 Statistic dispersion of data in an S-N curve and %Probability of failure curves

standard deviations to left of the mean, only 1% of the specimens may have failed after N_1 cycles. By joining the points of the same failure probability for each stress level, a fatigue probability curve may be obtained.

The interpretation of fatigue data as failure probability is more convenient than the absolute numbers. For example, if a steel is meant to fabricate an aircraft landing gear fitting bolt, its design must be made for minimum failure probability, for instance *1% Failure probability*, because it is a critical component. Thus, the number of allowable load cycles before inspection or replacement will be N_1 , at the expected maximum stress amplitude. On the other hand, if the same steel is used for fitting seats inside the cabin, a failure would not be as serious, therefore, the design can be done at, say, *50% Failure probability*, then the inspection or replacement would be carried out after a much higher number of cycles (since the N scale is logarithmic), or else the design would be for a higher stress. The only problem is that a rather high number of tests would be required in order to obtain a representative distribution of the N_f values at a stress level, which means a significant investment in time and cost, that is why statistical analysis is a valuable alternative.

In 1961, Weibull [6] introduced a statistical analysis method to calculate fatigue failure probabilities from a limited number of tests. Such procedure is based on the following probability distribution.

$$F(N) = 1 - \exp \left[- \left(\frac{N}{q} \right)^b \right]$$

where: $F(N)$ is the fraction of population that fails at N cycles and $\Delta\sigma$ constant. The parameter q is the characteristic life at 63.2% of failure probability (which corresponds to ± 1 standard deviation off mean life), b is a parameter related to the

dispersion amplitude. Introducing logarithms, the previous equation can be expressed as:

$$\log \log \left[\frac{1}{1 - F(N)} \right] = b \log N - b \log q$$

The plot of this equation, in loglog versus log scale, is a straight line, with b slope and from which the q value can be obtained. Type loglog versus log graphs are known as *probabilistic graphs* and from them the N_{pp} value can be determined, where pp corresponds to a given percent probability of failure, furthermore this data can be transferred to a probabilistic S-N curve, as shown in Fig. 7.27.

Weibull demonstrated that failure probability $F(N)$ can be estimated from a reduced number of data (n):

$$F(N) = \frac{i}{1 + n}$$

where i is the i esim value of N , given in an increasing order; that is, for n number of data, $i = 1$ for the minimum value of N and $i = n$ for the maximum value of N . The following example illustrates the practical use of the Wiebull analysis, where the N values are given randomly to make the example more explicit.

Example The results of a constant amplitude fatigue life test are listed below. Determine the values of the Weibull constants b and q .

Solution The first step is to assign the i th value of the N data in increasing order and calculate the value of $F(N)$ and then, calculate the values $\log N$ and $\log \log [1 / (1 - F(N))]$. The results are given in the table below:

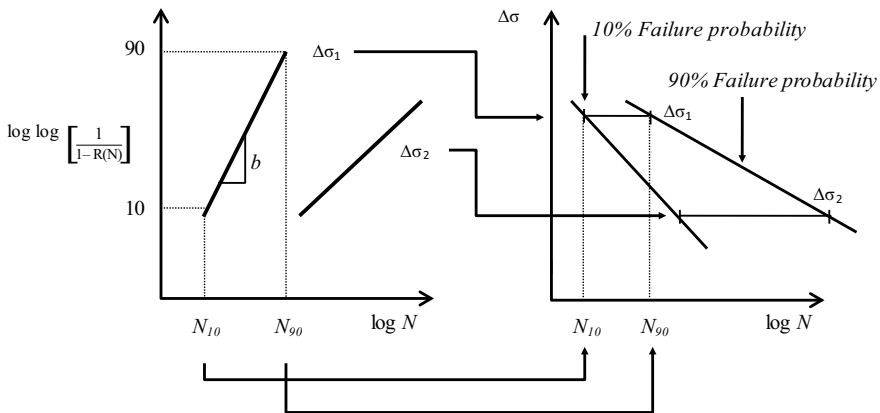
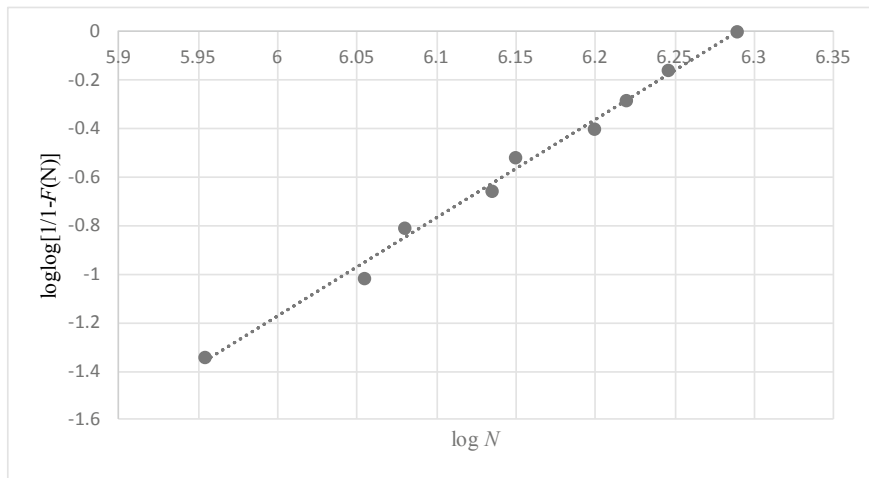


Fig. 7.27 Construction of a probabilistic S-N curve from the Weibull curve



No.	i	$F(N) = i/(1 + n)$	$\log N$	$\log\log[1/(1 - F(N))]$
1,202,264	3	$3/(1 + 9) = 0.3$	6.080	-0.810
900,000	1	0.1	5.954	-1.340
1,412,538	5	0.5	6.150	-0.521
1,364,583	4	0.4	6.135	-0.654
1,659,586	7	0.7	6.220	-0.282
1,945,360	9	0.9	6.289	0.000
1,761,976	8	0.8	6.246	-0.156
1,135,010	2	0.2	6.055	-1.014
1,584,893	6	0.6	6.20	-0.400



The value of b is the slope of the best fit straight line of the $\log\log[1/(1 - F(N))]$ versus $\text{Log}N$ plot, which is 4.0 and the value of q is the corresponding value of N when $F(N) = 63.2\%$, thus $\log\log[1/(1 - F(N))] = 0.362$. So from the above graphic it becomes 6.198 (1,579,103 cycles). The endurance cycles for 99 and 1% probability of failure can also be determined: they are 2,313,396 and 500,034 cycles, respectively.

In the above example is clearly seen that the value of b represents the dispersion of life data. At $b = 3.5$ there is completely random Gaussian data distribution; values of $b \leq 1.0$, indicate a widespread dispersion, meaning that a significant fraction of data will be far below average, termed as “infant mortality” or “premature failure”. On the contrary, if $b > 3.5$ it is beneficial because the life dispersion is low and therefore there is high fatigue life predictability, so the designer can be confident of the expected endurance of the component and the established replacement periods. Finally, if the data plotted on probabilistic paper ($\log\log$ vs. \log) does not follow a straight line, the data are not random and there is a specific cause that deviates them. A non-random situation that is important from the technological point of view appears in some low-alloy steels that exhibit cyclic strain



heating, which leads to premature failure, making a deflection to the left of the Weibul’s graph lower part.

The ASME “Boiler and Pressure Vessel” codes, as well as the API 579-1/ASME FFS-1 “Fitness-For-Service”, use the S-N approach for fatigue design and assessment. The fatigue design curves are S-N curves developed from smooth, base metal bars tested at room temperature. These codes address two curves for most of the ferritic materials used in Class 1 vessels: one for materials with $\sigma_{UTS} < 552 \text{ MPa}$ (80 ksi), shown in Fig. 7.28, the other for materials with 793 MPa (115 ksi) $< \sigma_{UTS} < 892 \text{ MPa}$ (130 ksi). The interpolation for materials with intermediate σ_{UTS} is allowed.

According to the ASME code, fatigue is assessed in terms of the alternating stress intensity (S_a) which is based on the Tresca’s failure criterion, therefore S_a is calculated with the following equation:

$$S_a = \frac{1}{2}(S_{\max} - S_{\min})$$

where S_{\max} and S_{\min} are the maximum and minimum stresses at the assessment zone of the component. The curves allow to determine the endurance cycles at the determined stress amplitude, so the designer can establish a design service life, and if this is insufficient, a new stress amplitude level may be established to accomplish the desired endurance limit, then the design is modified accordingly, may be by increasing the thickness or limiting the allowable stress amplitude. In Fitness-For-Service assessments, the analysis is reverse, that is: if the in-service number of cycles equals or exceeds the value of N_f obtained from the S-N curve, the component is rejected by having a high probability of failure. On the other hand, if N_f at the assessment stress amplitude is greater than the number of in-service stress cycles, the difference will be the remaining life.

Notice that the S-N curve of Fig. 7.28 shows that at 10 cycles, $S_a = 4000 \text{ MPa}$ (580 ksi), which is more than 10 times greater than the typical yield strength (S_{ys}) of

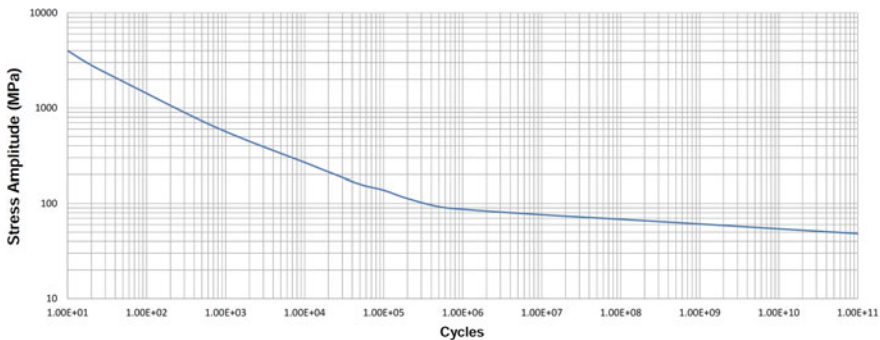


Fig. 7.28 Fatigue curve for ferritic steels for temperatures not exceeding 371 °C and $\sigma_{UTS} \leq 552 \text{ MPa}$ (80 ksi). Source Figure 14B1.M, API 579-1/ASME FFS-1 2016



structural steels (345 MPa, 50 ksi). Since these stress levels are never expected for both, normal and abnormal operating condition, the ASME fatigue analysis approach is not considered for 10 or less cycles. It is interesting also that at one million cycles (10^6), $S_a = 86$ MPa (12.5 ksi), while the maximum allowable stress for the ASME S-VIII and ANSI B31.3 codes is 138 MPa (20 ksi), thus $S_{allow}/S_a = 1.6$, which gives a considerable safety margin for fatigue endurance of pressure vessels and piping designed under these codes.

Effect of variable loads: The load amplitude variation is a frequent condition during the service of mechanical and structural components. Most of the times, the fatigue life under a variable load spectrum can be analyzed by separating the spectrum into a number of blocks of constant load amplitude, as schematically illustrated in Fig. 7.29.

The “Miner’s rule”, proposed by M.A. Miner in 1945, states that the sum of the fraction fatigue lives of each block in the load spectrum equals one, this is:

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

where k is the number of constant amplitude blocks, n_i is the number of cycles in each block and N_f^i is the number of failure cycles at each constant load amplitude block. The following example illustrate the use of the Miner’s rule.

A marine pipeline is subject to alternating deflections by strong storm currents, combined with vortex induced vibration. The deformation amplitude by storm currents is $\Delta\varepsilon_1 = 0.01$, whereas the deformation amplitude by vortex induced vibration is $\Delta\varepsilon_1 = 0.00001$. If the storm frequency is four per year (1.268×10^{-7} cycles/s) and the frequency of vortex induced vibration is 20 cycles/s, what is fatigue life of the pipeline?

Solution The problem can be divided into two blocks, so the Miner’s rule is:

$$n_1/N_f^1 + n_2/N_f^2 = 1$$

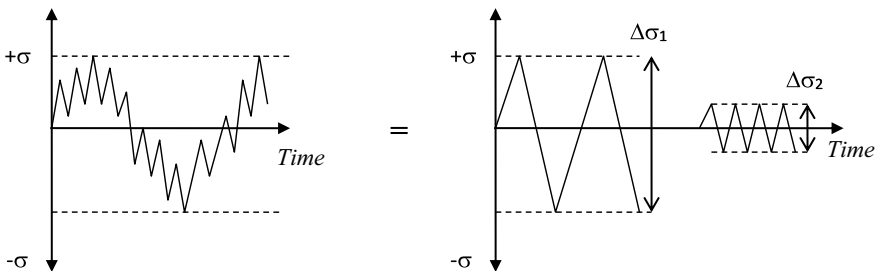


Fig. 7.29 Separation of a variable load pattern (left) into two constant stress amplitude patterns

where N_f^1 is the number of cycles to failure under the storm current deformation amplitude and N_f^2 is the number of cycles to failure due to the vortex induced vibration, which are calculated by solving for N_f the AWS formulas:

For storms: $\Delta\varepsilon \geq 0.002$, thus: $N_f^1 = (0.055/\Delta\varepsilon_1)^{2.5} = (0.055/0.01)^{2.5} = 71$ cycles

For vortex: $\Delta\varepsilon < 0.002$, thus: $N_f^2 = (0.010/\Delta\varepsilon_2)^4 = (0.010/0.00001)^4 = 10^{12}$ cycles

Considering that the failure time (t) is equal to the frequency divided by the number of failure cycles, and solving Miner's equation for t :

$$t = \frac{1}{\frac{f_1}{N_{f1}} + \frac{f_2}{N_{f2}}} = 1.116 \times 10^9 \text{ s} = 35.4 \text{ years}$$

7.7 The Universal Slopes Method of Fatigue

In 1964, Manson and Coffin, researchers from the NASA Lewis Research Center in Cleveland, Ohio, U.S.A., proposed an expression to estimate fatigue life of un-notched components, based on the monotonic uniaxial tension properties. This expression was strongly needed because of the high cost of high cycle fatigue tests, which additionally are time consuming. The principle of this method is based on the fact that fatigue life (N_f) depends on the deformation amplitude, which in turn, depends on the uniaxial tension properties of the material.

As already seen, the cyclic deformation amplitude is:

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

Manson and Coffin found that the elastic component is given by:

$$\Delta\varepsilon_e = \frac{\sigma'_f (2N_f)^b}{E}$$

where σ'_f is the fatigue strength coefficient, defined by the stress value at the hysteresis curve intersection in one load cycle, N_f is the number of cycles to failure, E is Young's modulus and b is the fatigue exponent. Likewise, they found that the plastic component of the cycle is:

$$\Delta\varepsilon_p = \varepsilon'_f (2N_f)^c$$

where ε'_f is the fatigue ductility coefficient, defined by the deformation value at the hysteresis curve intersection in one load cycle and c is the fatigue ductility exponent. Therefore, the Manson-Coffin Law is expressed as:

$$\Delta\varepsilon = \frac{\sigma'_f}{E} (2N_f)^b + \varepsilon'_f (2N_f)^c$$

The cyclic load test procedure for the determination of σ'_f and ε'_f is described in the ASTM E606 standard. The Manson-Coffin Law turned out to be very useful to solve practical problems such as:

- Calculate the maximum allowable deformation amplitude to endure a given number of cycles.
- Calculate the number of cycles to failure at a given deformation amplitude.
- Select the best material for fatigue life (the best material has the highest $\Delta\varepsilon$).

The values of the Manson-Coffin equation constants for several metallic alloys of common commercial use are given in Table 7.4.

The use of the Manson-Coffin Law is easier in logarithmic form, this is:

$$\ln \Delta\varepsilon_e = b \ln(2N_f) + \ln(\sigma'_f/E)$$

$$\ln \Delta\varepsilon_p = c \ln(2N_f) + \ln \varepsilon'_f$$

In a log-log plot, both equations are straight lines, so the life curve is, as shown in Fig. 7.30.

As it can be observed on Table 7.5, b and c vary very little from one material to other, thus the curves for many materials fall within a narrow band. By taking the mean values, a general expression can be obtained, known as *Universal Slopes Law*.

$$\Delta\varepsilon = 3.5 \frac{\sigma_{UTS}}{E} N_f^{-0.12} + \varepsilon_f N_f^{-0.6}$$

Table 7.4 Constants of cyclic deformation for common alloys (stresses in MPa). Data taken from “Deformation and fracture mechanics of Eng. Matls.”. R.W. Hertzberg, John Wiley & Sons

MATERIAL	σ_{ys}/σ'_{ys}	n/n'	$\varepsilon_f/\varepsilon'_f$	σ_f/σ'_f	b	c
SAE 1015 steel annealed	225/240	0.26/0.22	1.14/0.95	725/825	-0.11	-0.64
SAE 1045 steel quenched	1365/825	0.076/0.146	0.72/0.60	2725/2725	-0.081	-0.6
SAE 4340 steel quenched	1370/825	-/0.15	0.48/0.48	1560/2000	-0.091	-0.60
Inox. 304 steel annealed	255/715	-/0.36	1.37/1.02	1570/2415	-0.15	-0.77
Aluminum 2024 T4	305/440	0.20/0.08	0.43/0.21	635/1015	-0.11	-0.52
Aluminum 7075 T6	470/525	0.113/0.146	0.41/0.19	745/1315	-0.126	-0.52

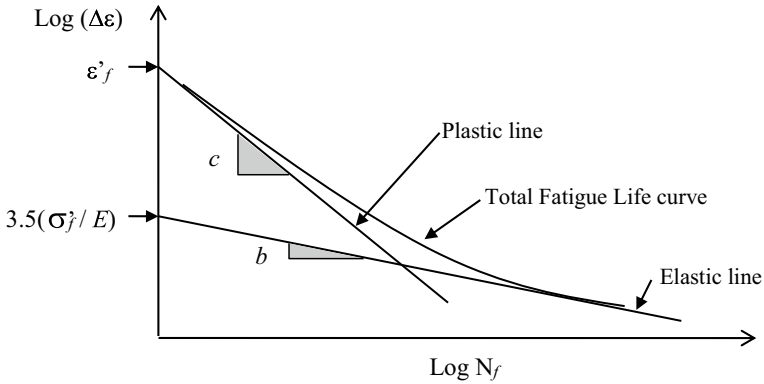


Fig. 7.30 Elastic and plastic lines of the S-N curve to determine the constants of the Manson-Coffin Law

Table 7.5 Paris constants for common engineering materials in air, room temperature and other non-corrosive environments

Material	<i>m</i>	<i>C</i>
Carbon steel	3.0	1.65×10^{-8}
Forged steel	3.0	7.27×10^{-8}
High strength steel, σ_{ys} 552–2026 Mpa	2.25	1.36×10^{-7}
Austenitic stainless steel	3.25	5.61×10^{-9}
Grey cast iron	4.0	8×10^{-9}
Nodular cast iron	3.5	10^{-8}
Aluminum 7021	2.5	10^{-8}
Nickel base alloy	3.3	4×10^{-12}
Titanium base alloy	5.0	10^{-11}
Polycarbonate	4.3	6×10^{-5}
Epoxy	2.0	10^{-4}

For da/dN in mm/cycle and ΔK in $MPa\sqrt{m}$

It is useful to know that for practical purposes, the fatigue ductility coefficient can be approximated by the following equation:

$$\epsilon_f = \ln\left(\frac{100}{100 - \%RA}\right)$$

where $\%RA$ is the percent of area reduction in the uniaxial tension test. This equation allows estimating fatigue life of metallic alloys in terms of their uniaxial tension properties, which are easy to know.

Manson and Ruiz proposed a modification to the universal slopes law so as to express it as a function of the stress amplitude, because frequently it is not possible to know the strain amplitude.



This equation has the form:

$$\Delta\sigma^* = 3.5\sigma_{UTS}N_f^{-0.12} + E\varepsilon_f^{0.6}N_f^{-0.6}$$

where

$$\Delta\sigma^* = \frac{\Delta\sigma\sigma_{UTS}}{2\sigma_{UTS} - \Delta\sigma}$$

Many codes and standards of design have adopted simplified versions of the universal slopes law for the high cycle region, like the American Welding Society (AWS), that proposes the following expressions for welded joints in low carbon steels for structural and pipelines use:

$$\begin{aligned} \Delta\varepsilon &= 0.055 N^{-0.4} \quad \text{for } \Delta\varepsilon \geq 0.002 \\ \Delta\varepsilon &= 0.010 N^{-0.25} \quad \text{for } \Delta\varepsilon < 0.002 \end{aligned}$$

7.8 Fatigue Crack Growth

As it has been stated, fatigue is a process where a crack is nucleated and grows until it causes the final fracture of a component under repeated loads. In the great majority of structural components, the normal operation stresses are below the yield strength, so the crack growth grows in an elastically strained body, therefore linear elastic fracture mechanics may be used to characterize the behavior of fatigue cracks.

In 1961, the American scientist Paris [7] demonstrated that fatigue crack growth rate, defined by da/dN , depends on the magnitude of the amplitude stress at the crack tip, which in linear-elastic conditions depends on the stress intensity factor amplitude ΔK , by the following reasoning:

In the load cycle, the stress intensity factor amplitude ΔK , is defined as:

$$\Delta K = K_{\max} - K_{\min}$$

The general expression for K is:

$$K = P\beta\sqrt{\pi a}$$

where P is the load, β is a geometrical factor and a is the crack size. Since the load amplitude is $\Delta P = P_{\max} - P_{\min}$, the stress intensity factor amplitude can be expressed as:

$$\Delta K = \Delta P \beta \sqrt{\pi a}$$

If ΔP is constant, the load cycle is completely defined with the R ratio:

$$R = P_{\min}/P_{\max}$$

To prove his ideas, Paris did an experiment consisting on applying a constant amplitude cyclic load ΔP onto a pre cracked plate, and measuring the crack growth as a function of the number of elapsed cycles. As predicted by fracture mechanics, the stress intensity factor increases as the crack grows, so Paris regrouped the data, as schematically depicted in the plots of Fig. 7.31.

By plotting the data pairs $(\Delta K, da/dN)$ in a log-log plot Paris found that the graph was linear over most of the intermediate ΔK range. At the low ΔK end, the da/dN values are very small ($<10^{-8}$ mm/cycle), so this defined a threshold value (ΔK_{th}). In the other end, at high ΔK , da/dN sharply increases up to a limit value given by the fracture toughness K_{IC} . These characteristics are shown in Fig. 7.32.

The equation that represent the linear portion of the log-log plot is known as the “Paris Law” and it is:

$$\frac{da}{dN} = C \Delta K^m$$

where C and m are empiric constants, their values for some engineering material are shown on Table 7.5.

The da/dN versus ΔK graph is usually divided into three regions for most engineering materials, these regions are:

- Region I, Near Threshold: The initial values of da/dN are very low, but rapidly increase as ΔK increases. It is assumed that the crack does not propagate below (ΔK_{th}) . This stage is strongly influenced by the stress level and the microstructure.

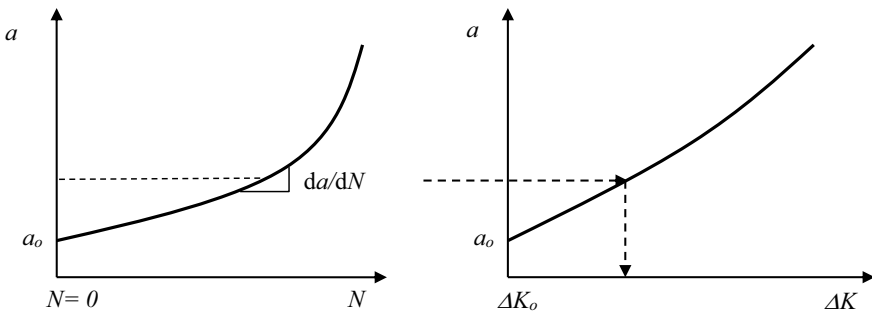


Fig. 7.31 Left, crack size and ΔK variation as a function of elapsed cycles at constant ΔP

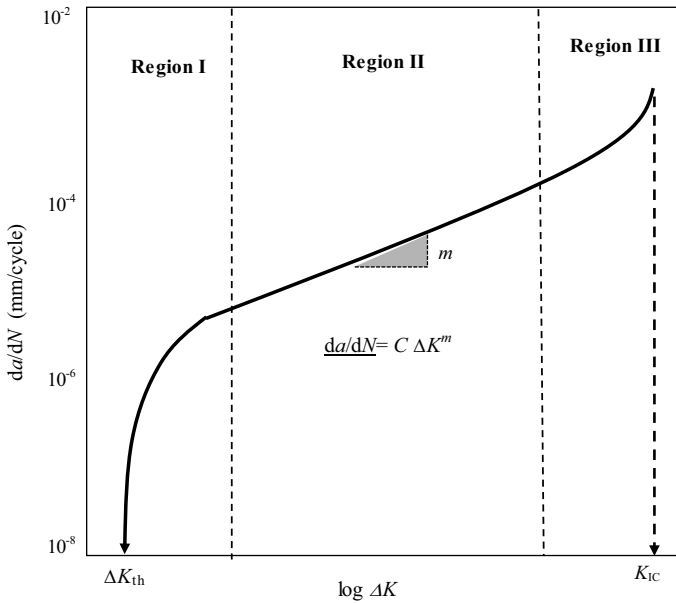


Fig. 7.32 Schematic plot of the fatigue crack growth rate as a function of the stress intensity factor amplitude as found by Paris

- Region II, Paris Region: The da/dN values follow the Paris Law. Normally the microscopic crack extension per cycle is equal to da/dN . It is strongly influenced by the environment.
- Region III, Unstable Crack Propagation: da/dN increases sharply. The fatigue fracture mechanism is combined with static modes of fracture.

Another important aspect of Paris curve is the effect of R . It has been experimentally observed that Paris curves at different R are parallel, that is, they have the same m value, but different C , therefore it is affected by R . For many materials, the dependence of da/dN on R can be described by the equation proposed by Walker, which has the form:

$$\frac{da}{dN} = \frac{C_w}{(1 - R)^n} \Delta K^n$$

where C_w is the value of C when $R = 0$. Since fracture occurs when $K_{max} = K_{IC}$, and according to the definition of R and ΔK , it can be stated that:

$$K_{max} = \frac{\Delta K}{1 - R}$$



So, when $\Delta K = (1 - R)K_{IC}$, the crack growth rate would tend towards its maximum value, which is the right end of Region III in Paris graph. Forman et al. [8] proposed a more general expression to incorporate the effect that R has in the following form:

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

The fatigue analysis based on the fracture mechanics concept, holds two major advantages over the classic S-N curves approach. The first advantage is that fatigue crack growth experimental data does not show the great dispersion observed in the S-N curves and the second one is that the Paris Law equation can be integrated in order to determine the number of cycles to reach a maximum allowable crack size, thus allowing to estimate the remaining life of a fatigued component that already contain cracks. The calculation is valid, as long as linear elastic strain conditions prevail in the component and the K function is known. The procedure is as follows. Starting from the initial crack size a_0 , the Paris Law equation can be integrated by the separation of variables, obtaining:

$$N = \int_{a_0}^{a_c} \frac{da}{c\Delta K^m} = \frac{1}{C(\beta\Delta\sigma)^m \pi^{m/2}} \int_{a_0}^{a_c} \frac{da}{a^{m/2}}$$

where a_c is the critical (or maximum allowable) crack size. Since many engineering alloys feature m values close to 3.0, the integration of Paris law equation can be expressed as:

$$N = \frac{2}{C(\Delta\sigma\sqrt{\pi})^3} \left[\frac{1}{\sqrt{a_0}} - \frac{1}{\sqrt{a_c}} \right]$$

Finally it is important to remember that, when applying the Paris Law equation to assess fatigue crack growth, there has to be great certainty on the applied stress intensity range based on the applied loading, crack configuration and residual stresses, In addition, the Paris Law constants should be according to the material, service environment and R ratio. This is because the results may significantly change on the input data, as seen in the following example:

Example A cylindrical pressure vessel of diameter $D = 1500$ mm and wall thickness $t = 6$ mm made of ferritic steel operates at 21 °C in a non-aggressive environment at a pressure of $P = 1.0$ MPa (145 psi). An in-service inspection detected a semielliptical crack of constant aspect ($a/2c = 0.25$) of $a = 3$ mm depth. Determine the number of full pressure cycles in which the crack will cause a leak, assuming that, when $a = t$, the crack is still stable (leak-before-break condition).

Solution The stress amplitude, according to the ASME B&PV code is:

$$S_a = \frac{1}{2}(S_{\max} - S_{\min})$$

$$S_{\max} = \frac{PD}{2t} = \frac{(1 \text{ MPa})(1500 \text{ mm})}{2(6 \text{ mm})} = 125 \text{ MPa}$$

If $S_{\min} = 0$, $S_a = 62.5 \text{ MPa}$.

For a ferritic steel at 21 °C in a non-aggressive environment: $C = 1.65 \times 10^{-11}$ and $m = 3.0$ (m/cycle, MPa $\sqrt{\text{m}}$). Substituting into the integrated Paris Law equation:

$$N = \frac{2}{C(S_a\sqrt{\pi})^3} \left[\frac{1}{\sqrt{a_o}} - \frac{1}{\sqrt{a_c}} \right]$$

$$= \frac{2}{1.65 \times 10^{-11}(62.5\sqrt{\pi})^3} \left[\frac{1}{\sqrt{0.003}} - \frac{1}{\sqrt{0.006}} \right] = 476,750$$

Notice that the calculation of N is very sensitive to C , because its values are 10^{-11} order of magnitude. The effect of S_a it is very strong too, since it is raised to the 3th power, so an increment of pressure to double will reduce N by eight-fold.

The environment also has a strong effect of the Paris curves. As mentioned before, fatigue can occur in inert environments, so it can be purely mechanical, however, since most engineering materials, especially metals, are susceptible to the environment, the fatigue crack growth under corrosion-fatigue interaction is the general case. The effect of corrosion on the fatigue crack growth curves is observed in two ways: the first one corresponds to a behavior where da/dN increases by the

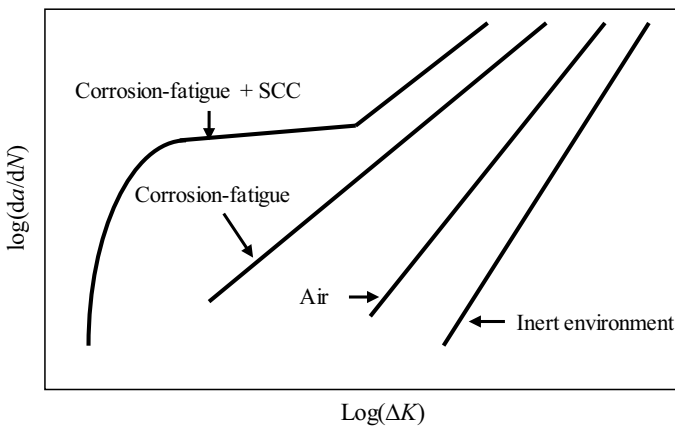


Fig. 7.33 Effect of environment on the Paris curve for metallic materials

effect of the corrosive environment. In this case the Paris plot shifts to the left and upwards, reducing the values of ΔK_{Th} and m (slope of Stage II). The second type of interaction is called “Fatigue + Stress Corrosion Cracking”, where the crack propagation occurs by the combination of cyclic loading and Stress Corrosion Cracking (SCC). This process requires that ΔK is above threshold value identified by K_{ISCC} . When the Fatigue + SCC occur, the Paris curve displays a plateau in the intermediate range of ΔK values. Figure 7.33 shows the schematic Paris graphs for these categories. Finally, under corrosion-fatigue interactions, the effects of R and frequency are stronger.

References

1. Nix KJ and Flowers HM (1982) The micromechanisms of fatigue crack growth in a commercial Al-Zn-Mg-Cu alloy. *Acta Metall* 30:1541
2. McEvily AJ, Gonzalez JL (1992) Fatigue crack tip deformation processes as Influenced by the environment. *Metall Trans A* 23A(8):2211–2221. ISSN 1073-5623
3. Smith RW, Hirschberg MH, Manson SS (1963) NASA TN D-174, NASA
4. Feltner CE, Laird C (1968) *Trans Metall Soc AIME* 242:1253
5. Lukás P, Klensil M (1971) Dislocation structures in fatigued single crystals of CuZn system. *Physica Status Solidi A* 5:247–258
6. Weibull W (1951) A statistical distribution function of wide applicability (PDF) *J Appl Mech-Trans ASME* 18(3):293–297
7. Paris PC, Gomez MP, Anderson WE (1961) A rational analytic theory of fatigue. *Trend Eng* 13:9–14
8. Forman RG, Kearney VE, Engle RM (1967) *J Basic Eng Trans ASME* 89:459

Chapter 8

High Temperature Mechanical Behavior



Abstract This chapter starts with the description of the effects of elevated temperature on the deformation characteristics of materials by means of the Stress-Rupture and the Constant-Stress creep curves. The Creep Power Law for secondary creep is introduced along with the Larson and Miner creep life prediction method. Further, the deformation mechanisms in creep are described, as well as the creep fracture mechanism. Then, the ideas are applied to explain the foundations for the development of high-temperature service materials, known as refractory. The chapter ends with a brief description of the superplastic behavior that results of the combination of strengthening mechanism and high temperature behavior.

8.1 High Temperature Deformation

Some of the most important technological applications of engineering materials are at high temperature service conditions, such as: gas turbines, internal combustion engines, boilers, reactors, heat exchangers, ovens, furnaces and hot forming equipment. All these components are big and costly and their failure usually carries along high consequences (fatalities, injuries and environmental damage) and major economic losses; hence high temperature mechanical behavior is a strategic field for technological and scientific research.

The exposure of engineering materials to high temperatures has several effects, which altogether affect the mechanical behavior. The most important effects are:

1. Yield stress and tensile strength reduction.
2. Increment of ductility by increasing dislocations mobility.
3. Recovery and recrystallization of cold formed materials.
4. Grain growth.
5. Increment of diffusivity.
6. Dissolution and precipitation of second phases.
7. Incipient fusion.
8. Excessive oxidation.

All these processes are thermally activated, therefore, a sufficiently high temperature is required for them to happen, referred as *hot work*. As seen in Chap. 2, the hot work condition is relative to the material's melting point or fusion temperature (T_f). The rule is that at working temperatures of more than $0.4T_f$, where T_f is in Kelvin degrees, the material exhibits high temperature behavior, which is primarily characterized by the strength reduction, suppression of strain hardening, and most important, the plastic strain becomes time-dependent, so the material deforms as if it was a high viscosity fluid. Such phenomenon is called *thermo-fluence* or most commonly *creep*.

From the engineering point of view, creep is defined as plastic deformation through time under a constant stress. After some time, creep culminates in fracture, therefore it is also a failure mechanism. It is currently known that creep is caused by thermally activated diffusive processes, and although these processes occur at any temperature, it is at high temperature ($>0.5T_f$) that it becomes of practical significance, and for that reason creep is considered a high temperature phenomenon.

Perhaps one of the most revealing manifestations about creep is the deformation and distortion of metallic structures affected by fire, as seen in the examples of Fig. 8.1, where the heat of a fire soften columns and beams, causing the plastic flow of the structural components under their own weight.

The general procedure for creep testing is described in the ASTM E139 standard and consists on applying a constant load to a smooth tensile test specimen at constant temperature and measuring the elongation as a function of time. Heating is applied by placing the test specimen inside a heating chamber, usually with a controlled atmosphere to prevent excessive oxidation. The test time may last several hours up to years, so creep tests require specialized laboratory equipment and are very costly.



Fig. 8.1 Example of creep in metallic structures affected by fire. Notice that most of the structural components deformed and collapsed under their own weight

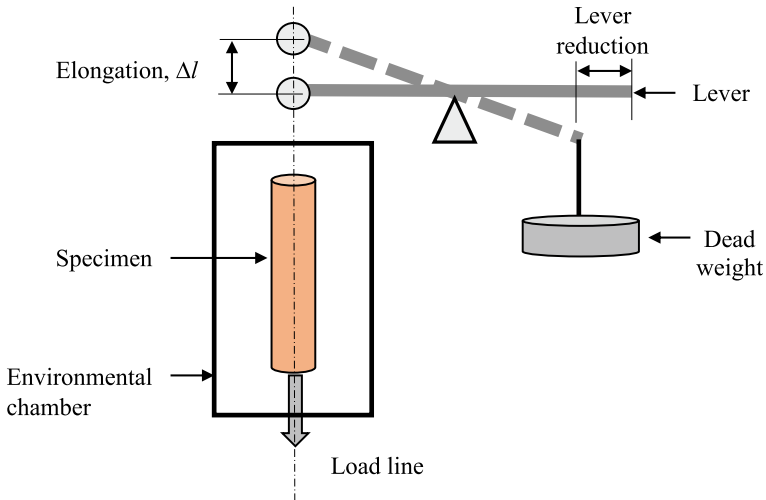


Fig. 8.2 Constant leverage creep test apparatus

In practical engineering creep tests, the load is applied by leverage system with dead weights, designed in such a way that, as the test specimen elongates, the leverage arm shortens, thus reducing the load and compensating the stress increment caused by the reduction of the specimen's cross-sectional, maintaining a fairly constant-stress. The scheme of Fig. 8.2 shows the experimental arrangement of this test.

When the constant-load creep tests for the same material are carried out at different load levels and the rupture time for each load case is recorded and plotted on logarithmic scale, then a *stress-rupture* curve is obtained, which has the shape shown in Fig. 8.3. The main characteristic of the stress-rupture test is that, as both stress and temperature increase, the rupture-time drops in an exponential way. Stress-rupture tests are conducted for periods up to 10,000 h and they are specifically applied to determine the high-temperature strength, and the minimum creep rate of engineering materials. These characteristics make the stress-rupture test useful for design, material selection and quality control purposes.

If the specimen's elongation is recorded and plotted as a function of time, a *creep curve* is obtained, which has the idealized shape shown in Fig. 8.4. The creep curve starts at an instantaneous elongation of the specimen ϵ_0 , followed by a reduction of the *creep rate* ($d\epsilon/dt$), then the creep rate becomes constant for most of the elapsed time and finally the creep rate increases rapidly until rupture occurs.

The constant-stress creep curve has the following characteristics:

Stage I, *transient creep*. It starts at an instantaneous initial deformation ϵ_0 , which is proportional to the applied stress. The creep rate is initially high, but it decreases gradually until reaching a constant value. In this stage, the dislocations substructure is rearranged and some phase transformations may occur, such as dissolution or

Fig. 8.3 Stress-rupture curve of a low alloy Mo-V steel. K. Image taken from: Kimura et al. [1]

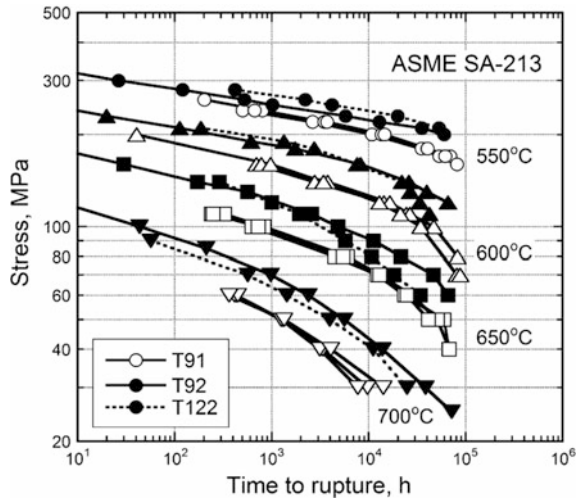
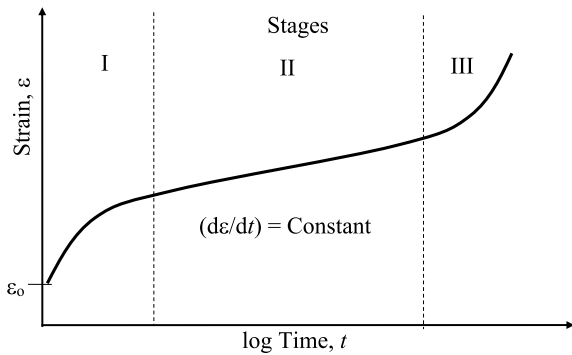


Fig. 8.4 Idealized creep curve at constant stress showing the three stages of creep



precipitation of second-phases. Once the phase transformations, the strain hardening (interaction of dislocations) and recovery processes (dislocation rearrangement and annihilation) reach a dynamic equilibrium, the creep rate becomes constant.

Stage II, secondary creep. In this stage, the dynamic equilibrium between the strain hardening and recovery is maintained. The creep rate in this stage is minimum and is referred to as *steady-state creep*.

Stage III, tertiary creep. In this stage, microstructural changes, such as precipitate coarsening and recrystallization, and the reduction in cross-section area increase the creep rate, along with an increment of vacancy diffusion and grain boundary sliding. These last two processes progress until the interconnection of grain boundary voids, lead to intergranular fractures, typical of creep failures.

Secondary creep is the most important stage, from the technological point of view, as it accounts for the longest creep period, therefore, the majority of research is aimed to estimate the creep rate in this stage.

Andrade carried out the first creep tests in 1914 and found that, most metals follow an empirical equation expressed as:

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

where t stands for time, ε_0 is the instantaneous initial deformation and k and β are experimental constants. It was until 1960, that Garofalo proposed an equation that fits better to the idealized creep curve and it has this form:

$$\varepsilon = \varepsilon_0 + (1 - e^{-rt}) + \left(\frac{d\varepsilon}{dt}\right)_s t$$

where $(d\varepsilon/dt)_s$ is the creep rate in secondary creep and r is the ratio of the transient creep rate divided by the transient creep strain.

8.2 Secondary Creep

In stage II or secondary creep, the creep rate $(d\varepsilon/dt)$, is constant and it is directly related to the applied stress (σ) by the so called *Creep Power Law*, that has the form:

$$\left(\frac{d\varepsilon}{dt}\right)_s = C\sigma^n$$

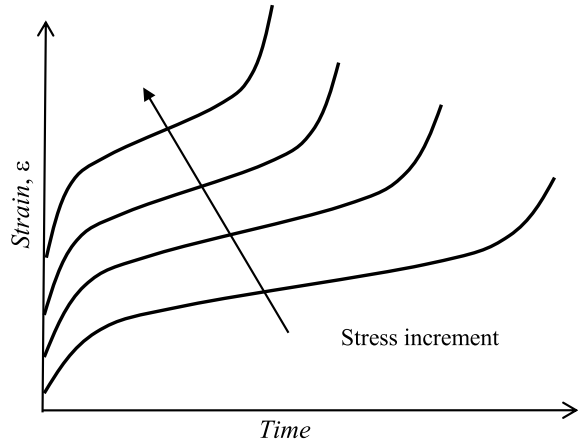
where C is an experimental constant and n is the secondary creep exponent. According to this law, the greater the stress, the greater the creep rate, as shown in Fig. 8.5. This explains why mechanical fasteners such as bolts and screws in high-temperature service get loosen sometime after they are tightened.

The steady creep rate in Stage II has been adopted as the most important design parameter for high temperature service components. The majority of codes set a creep rate limit, in which the applied stress must be adjusted based on the stress-rupture curve. For example, the typical design value for boiler and direct fire heater piping, is defined as to $(d\varepsilon/dt)_s \leq 1\%/100,000 \text{ h}$ or $2.8 \times 10^{-11} \text{ s}^{-1}$, so the pipe's diameter shall not expand more than 1% in 11.5 years at the design stress.

The secondary creep stage involves a process in which thermal activation provides the energy to overcome an energy barrier ΔH , moving the material thermodynamic state, from metastable, to stable. For this reason, the creep rate can be expressed by an Arrhenius type equation:

$$\left(\frac{d\varepsilon}{dt}\right)_s = A e^{\left(\frac{-\Delta H}{RT}\right)}$$

Fig. 8.5 Effect of stress on creep curves at constant temperature



where ΔH is the activation energy, T is the absolute temperature, R is the constant for ideal gases and A is a constant that depends on the material. The determination of ΔH is done by solving the previous equation with two creep rate values obtained at different temperatures in a constant-stress creep test, that is:

$$\left(\frac{d\varepsilon}{dt}\right)_1 e^{\frac{\Delta H}{RT_1}} = \left(\frac{d\varepsilon}{dt}\right)_2 e^{\frac{\Delta H}{RT_2}}$$

From which the activation energy is calculated by:

$$\Delta H = R \left[\frac{T_1 T_2}{T_1 - T_2} \right] \ln \left(\frac{\varepsilon_2}{\varepsilon_1} \right)$$

By this procedure it has been found that ΔH for secondary creep is practically equal to the activation energy for vacancy diffusion, thus, it has been concluded that secondary creep is due to vacancy flow. For the same reason, the Stage II creep mechanism is referred as to *diffusive flow*.

Table 8.1 shows values of ΔH for typical high-temperature materials. It can be observed that ΔH in materials with melting point between 1000 and 1700 °C (Cu, Fe, Ni) is three times higher than that of materials with melting point less than 600 °C (Pb, Al), whereas high melting point materials (Ta, Al₂O₃, W) have much higher ΔH and for such reason they are considered as to *refractory*.

Table 8.1 Activation energies ΔH for secondary creep of engineering materials

Metals	T_f (°C)	ΔH (kJ/mol)	Refractory materials	T_f (°C)	ΔH (kJ/mol)
Lead (Pb)	327	110	Tantalium (Ta)	2996	400
Aluminun (Al)	660	150	Alumina (Al ₂ O ₃)	2072	620
Cooper (Cu)	1083	205	Tungsten (W)	3410	630
Ferrite (Fe α)	1535	240			
Austenite (Fe γ)	1535	280			
Nickel (Ni)	1453	300			

8.3 Creep Life Prediction

As it has been mentioned, the primary practical interest in the study of creep is to predict the rupture time. The stress-rupture test is the practical way to assess the performance of a material at high temperature, however, as mentioned before, these tests are limited by the testing time. The design life of components for high temperature service, is usually established for a minimum of 10,000 h (about 416 days), but they may be in service for 100,000 h (11.4 years) or more. The *creep design life* is defined as a minimum rupture time at the allowable stress and maximum service temperature, so if the rupture allowable stress is greater than the minimum rupture strength for the design life, either the design parameters should be modified, or the material substituted. At first, these criteria obviously requires knowledge of the stress-rupture curve, like the one shown in Fig. 8.3. But, as already mentioned, testing to obtain the rupture strength at very long periods is impractical and costly, consequently the laboratory tests to determine the creep rupture strength are be done under conditions that give rupture times within a reasonable time (generally less than 720 h, 30 days). Obviously, tests conditions that give such rupture times must be extrapolated in order to estimate the expected duration of high-temperature components. The phenomenological creep equations are a means to achieve this objective, but they require sophisticated tests to determine the necessary constants for their use.

In engineering applications, one of the most popular methods to predict creep life is the *Larson-Miller parameter curves*. This method,¹ introduced by F. R. Larson and J. Miller in 1952, is based on an Arrhenius type relation, where r is the creep rate:

$$r = A \exp(-\Delta H/RT)$$

¹Larson and Miller [2]

where A is an experimental constant, ΔH is the creep activation energy, R is the ideal gas constant and T is the temperature. This equation can be re arranged as:

$$\Delta H/R = T(\ln A - \ln r)$$

The rupture time is inversely proportional to r , that is:

$$t_r \propto (1/r)$$

Thus:

$$-\ln t = \ln A - \Delta H/RT$$

Multiplying by T and converting into logarithms base 10, the following equation is obtained.

$$\Delta H/R = T(C + \log t)$$

If the ratio $\Delta H/R$ is independent from stress and temperature, then the material exhibits the same behavior at a given stress level and therefore, the term $T(C + \log t)$ is constant, and is termed as the *Larson-Miller Parameter (LMP)* and it allows to calculate the rupture time L in terms of *LMP*, by the equation shown below:

$$\log_{10} L = \frac{1000 LMP(\sigma)}{T} - C$$

where L is in hours and T is in absolute degrees and $LMP(\sigma)$ is calculated at the evaluation stress level in MPa. Figure 8.6 shows the $LMP(\sigma)$ values for a ferritic steel, L in hours, T is in degrees Kelvin and σ in MPa. The equations to calculate $LMP(\sigma)$ of several materials are provided in the Annex 10B of the API 579-1/ASME FFS-1 2016. For example, the equation for average $LMP(\sigma)$ of Low Carbon Steel:

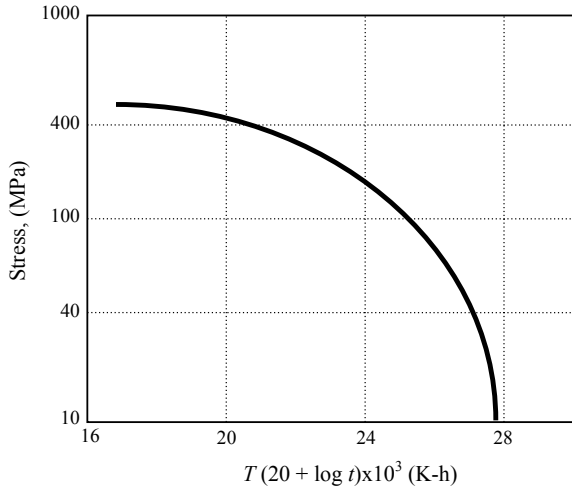
$$LMP(\sigma, ksi) = 39.793713 - 0.15443414\sigma - 2.6260065 \ln \sigma$$

The value of C for each material has to be experimentally determined by at least two sets of time-temperature creep tests at constant stress, by the following procedure.

Assuming that $\Delta H/R$ is a constant at a given stress level, it can be written that:

$$C = (T_2 \log t_2 - T_1 \log t_1)/(T_1 - T_2)$$

Fig. 8.6 Larson-Miller Parameter curve as a function of stress for a ferritic steel



A more precise evaluation can be obtained graphically by the equation:

$$\log t = -C + Cte/T$$

The experimental data $\log(t)$ versus $1/T$ for different stress levels are plotted and the value of C is determined by the intersection of the extrapolation to the origin where $1/T = 0$, as shown in Fig. 8.7. The values of C for some steels commonly used in high-temperature applications are given in Table 8.2. As it can be noticed, the average value of C is 20 for most ferrous alloys, hence LMP is frequently expressed as $T(20 + \log t)$, where T is in degrees Kelvin and t in hours.

The following example illustrates the use of the Larson-Miller Parameter.

Example Determine the rupture time for an S-590 steel at a stress $\sigma = 100$ MPa at 1000 °C. Repeat the calculation for: $T = 700$ °C and $T = 1000$ °C, $\sigma = 50$ MPa.

Fig. 8.7 Determination of the Larson-Miller Parameter from the experimental data $\log tr$ versus $1/T$ obtained at different stress levels and extrapolated at $1/T = 0$

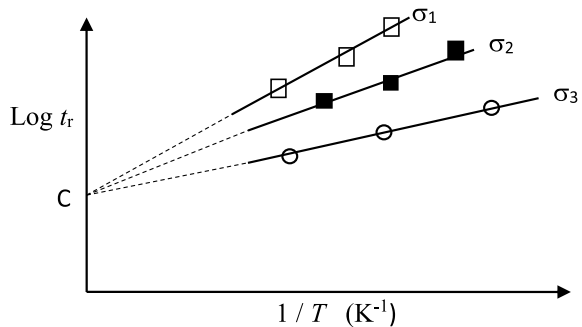


Table 8.2 Constant C for steels (T in degrees Kelvin and t in hours)

Alloy	C (h)
Low carbon steel	18
Stainless steel 18-8	18
Stainless steel 18-8-Mo	17
Steel 2 ¼ Cr-1 Mo	23
Steel S 590	20
Steel Cr-Mo-Ti-B	22

Solution

From Fig. 8.6 at $\sigma = 100$ MPa, $T(20 + \log t) = 25,200$

$$\log t = (25,200/T) - 20 = 25,200/1273 - 20 = -0.2$$

That is, $t = 0.62$ h (37.5 min)

$$\text{For } T = 500^\circ\text{C}, \quad \log t = 25,200/973 - 20 = 5.9$$

That is, $t = 7.93 \times 10^5$ h (90.5 years)

For the stress of 50 MPa, $T(20 + \log t) = 27,200$

$$\text{Clearing : } \log t = (27,200/T) - 20 = 27,200/1273 - 20 = 1.37$$

That is, $t = 23.3$ h.

From the previous example it is clear that temperature has a much greater effect than stress on the creep rupture time. The practical consequence of this is that it is more convenient to prevent overheating than to reduce operation stress to increase the creep rupture time. This may be counterproductive for internal combustion engine and process equipment, as in most of these applications, the higher the temperature, the higher efficiency, hence many efforts in creep research are aimed to the development of materials able to operate at higher temperatures in combination with high strength.

8.4 Deformation Mechanisms in Creep

Deformation mechanisms in creep primarily depend on slip and grain boundary sliding. At relatively low temperatures ($<0.3T_f$) dislocation slip mechanisms prevail, but at high temperatures ($>0.5T_f$), mechanisms based on diffusion and grain boundary glide become predominant, however, the general condition is that these mechanisms may simultaneously occur. The principal creep deformation mechanisms are:

Dislocation climb. When dislocations move through a crystal at high-temperature under high stresses, thermal activation helps to overcome the lattice friction and short-range obstacles located on the slip planes. However, at intermediate and low stress levels, the dislocation glide is assisted by vacancy diffusion by a process called as *dislocation climb*. Dislocation climb is an out-of-plane movement, which occurs by vacancy diffusion, as schematically

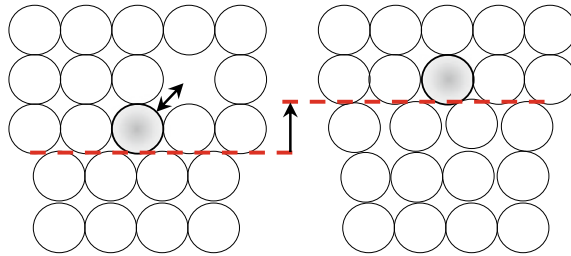


Fig. 8.8 Climb of an edge dislocation by vacancy exchange

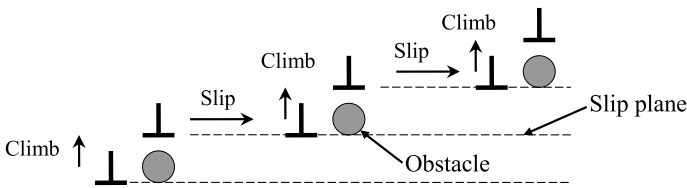


Fig. 8.9 Dislocation climb to overcome non-shearing obstacles

shown in Fig. 8.8. When a vacancy exchanges its place with the atom located on the dislocation line (shaded circle in Fig. 8.8), the segment “climbs” one atomic position in the lattice. If another vacancy moves towards the dislocation line, the process is repeated, until the entire dislocation line climbs. This movement produces shear strain of the crystal, and just like stress-activated glide.

Dislocation climb also assist dislocations overcomes long-range obstacles such as precipitate particles. Once the dislocation overcame the obstacle by climb, it continues gliding until it faces another obstacle, and then the climb process repeats, as schematically in Fig. 8.9.

In 1960, Weertman proposed a model to estimate the strain rate in secondary creep by dislocation movement, which is applicable at temperatures higher than $0.5T_f$, known as the *power-law creep*, represented by the following equation:

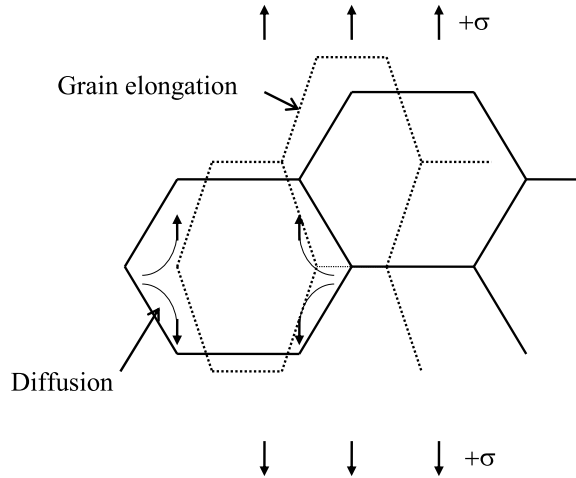
$$\left(\frac{d\varepsilon}{dt}\right)_s = \left[\frac{AD_{eff}Gb}{kT}\right] \left(\frac{\sigma}{G}\right)^n$$

where D_{eff} and A and n are experimental constants.

Creep by diffusive flow. This mechanism was proposed by Nabarro-Herring in 1950, and it involves the grains deformation by massive vacancy flow. The deformation flow is from zones under compression stresses, towards tension stressed zones. Simultaneously, atoms and vacancies flow in the opposite direction, producing elongation, as schematically shown in Fig. 8.10.



Fig. 8.10 Nabarro-Herring model of creep by diffusive flow



The phenomenological equation for the diffusive-flow mechanism is:

$$\left(\frac{d\varepsilon}{dt}\right)_s = \frac{14\sigma b^3 D_v}{kTd^2}$$

where d is the grain size and D_v is the lattice diffusion coefficient. At lower temperatures, close to the limit of the hot work regime, the diffusive-flow mechanism is substituted by a grain boundary sliding mechanism, for which Coble, in 1963, proposed the following model:

$$\left(\frac{d\varepsilon}{dt}\right)_s = \frac{60\sigma b^4 D_{gb}}{kTd^3}$$

The previous equations show that the grain size plays an important role in deformation in creep, since it appears raised at the cube power. This indicates that coarse grain materials have higher creep strength. Such findings have led to the development of coarse grain materials, and even single-crystals in order to improve their performance at high temperatures, as it will be discussed further in this chapter.

8.5 Creep Fracture

Generally, creep fractures are intergranular, since at high temperatures, above the equicohesive temperature (T_{eq}), as described in Sect. 4.3, the grain boundaries are weaker than the grains. The intergranular fracture mechanism in creep results from the combination of grain boundary cavitation and sliding. At macroscopic scale,



Fig. 8.11 Macroscopic (left) and microscopic (right) appearance of a typical creep fracture

creep fractures feature small plastic deformation, minimum neck formation and high surface roughness, whereas at microscopic scale the fracture path is intergranular and the grain facets are covered with numerous dimples. Figure 8.11 shows an example of a typical creep fracture.

At temperatures far above T_{eq} and low stress levels, creep deformation is dominated by grain boundary sliding, which favors intergranular fracture. Intergranular fracture in creep occurs by two basic mechanisms:

1. Grain boundary slip
2. Grain boundary cavitation.

Grain boundary sliding causes decohesion of grain boundary triple joints, forming wedge-shaped cavities, termed as *w-type*, as schematically shown in Fig. 8.12. The *w-type* cavities are formed preferentially in grain boundaries aligned to the maximum shear stress, thus in components under uniaxial tensile loads, they will be at around 45 degrees from the load line.

The condensation of vacancies in the grain boundaries, on the other hand, originates the formation of round cavities termed as *r-type*. When the number and size of the *r-type* cavities is high enough, they coalesce causing an intergranular fracture with little plastic strain associated to it.

The growth rate of *r-type* cavities is controlled by vacancy diffusion and power-law creep, according to the following equation.

$$dr/dt = C D_V r^m \sigma^n$$

where dr/dt is the cavity growth rate, D_V is the vacancy diffusivity coefficient, r is the cavity size, m is an experimental constant, and n is the power-law creep exponent. An interesting aspect of *r-type* grain boundary cavitation is that, according to the previous equation, dr/dt depends on the cavity size, meaning that large cavities grow faster than smaller ones. Additionally, for an *r-type* cavity to grow, the work supplied by stress has to be greater than the void surface energy γ_s . If this condition is not met, small cavities will reduce their size until collapse.

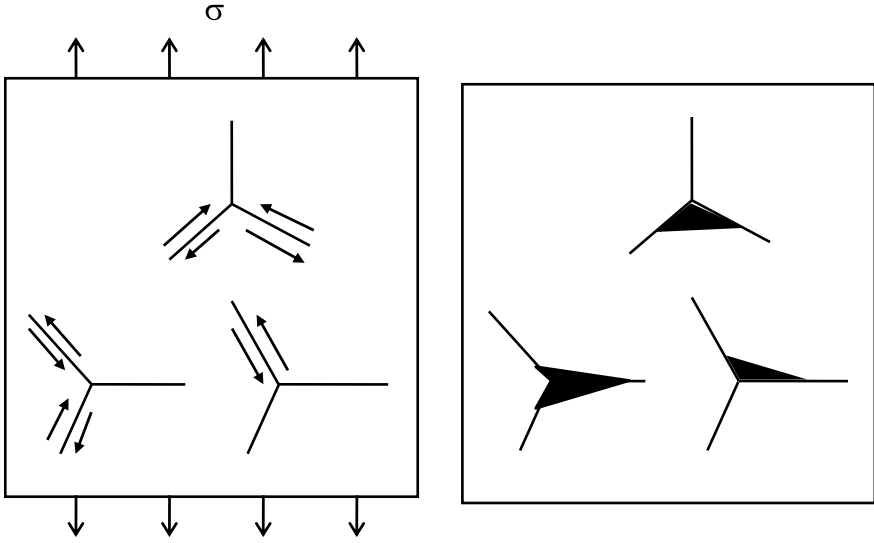


Fig. 8.12 Formation of *w-type* cavities by grain boundary sliding

The cavity minimum size (r_o) in order to maintain a stable growth condition can be calculated by:

$$r_o = 2\gamma_s / \sigma_{GB}$$

where σ_{GB} is the tension stress on the grain boundary. This equation indicates that *r-type* cavitation is favored by high tension stresses and low γ_s values. Since γ_s increases in proportion to the content of solid-solution forming elements, solid solution alloys have better creep strength, because *r-type* voids have to reach larger sizes in order to grow and coalesce, so the intergranular fracture process is delayed.

Another important factor is that *r-type* cavitation preferably occurs in grain boundaries oriented in direction nearly perpendicular to the maximum tension stress, as shown in Fig. 8.13.

Creep crack growth. Creep fracture can also occur locally, at specific locations such as pre-existing cracks or stress concentrators, because, they form local highly stressed plastic zones, so the creep rate increases, and the creep rupture time is drastically reduced. The creep crack growth mechanism is schematically depicted in Fig. 8.14. This is controlled by the cavity growth rate, which in turn, depends on the stress magnitude at the crack tip. According to fracture mechanics, the extension of crack in a strained body is determined by the energy release rate, which results from the balance between the work done by the tractions and the stored strain energy, as described by Rice's *J-integral* (see Sect. 6.4 in this book). But, since creep is a time-dependent process *J* has to be expressed as a function of time, dJ/dt . This derivative is typically represented by the symbol C^* , representing the time-dependent crack energy release rate. It has been experimentally found that:

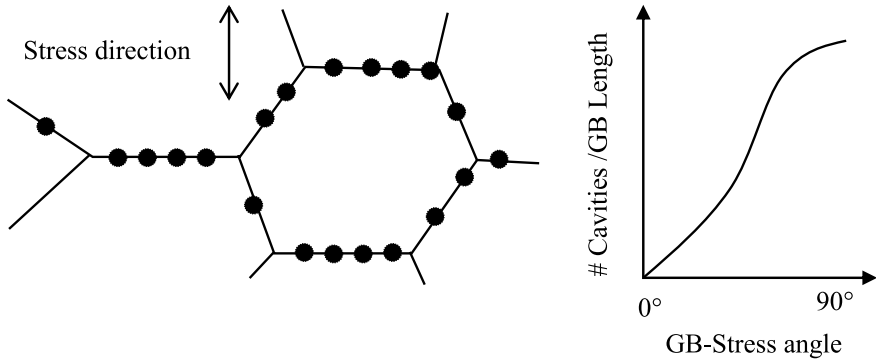


Fig. 8.13 Distribution of cavities in terms of grain boundary orientation with respect to tension stress

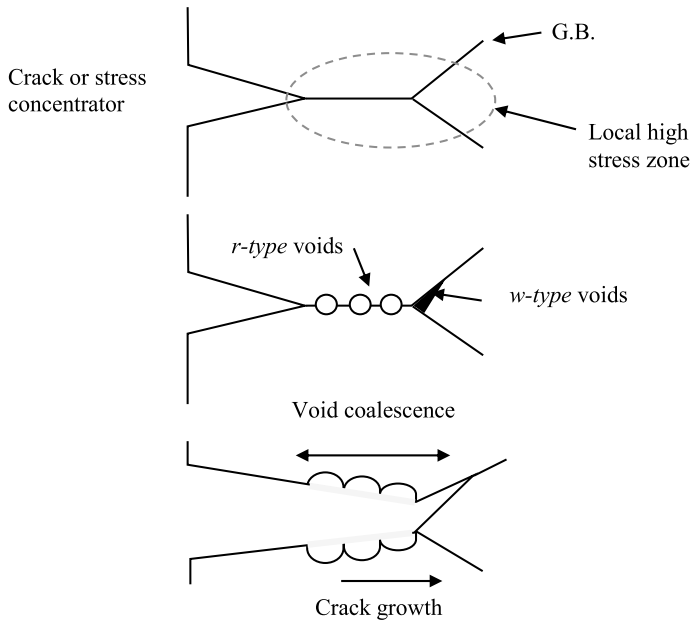


Fig. 8.14 Local creep crack growth mechanism by cavitation of grain boundaries at stress concentrators

$$\frac{da}{dt} = A(C^*)^n$$

where (da/dt) is the crack growth rate, A and n are experimental constants.

8.6 High Temperature Materials

The phenomenological equations along with the deformation mechanisms of creep are useful to predict the effect of variables such as stress, temperature, microstructure and so forth. Research on creep and creep-strength have set the foundations for the development of alloys and materials for high-temperature service, known as *refractory materials*. At first, refractory materials must have a high melting point and high Homologous Temperature, because these two are determining factors on whether deformation is within the creep regime or not. Second, refractory materials should have resistance to oxidation and high temperature corrosion, otherwise their usefulness is limited, such as the case of tungsten, which has high creep strength, but its oxidation strength is so low that when exposed to air at high temperature it oxidizes completely in a matter of few minutes.

In general, the phenomenological equation that describes creep can be generalized as to:

$$\left(\frac{d\varepsilon}{dt}\right)_s = Cte \frac{\sigma D}{TGd^n}$$

where $(d\varepsilon/dt)_s$ is the secondary creep rate, D is the diffusivity coefficient, G is the shear elastic modulus, d is the grain size and Cte is a constant depending on the material. According to this expression, materials with *fcc* structure are more resistant to heat than materials with less-compact structures as they have lower D values. Likewise, materials with high G will have a better performance at high temperatures. Tungsten, alumina, tantalum, as well as, austenitic stainless steels, nickel and cobalt based alloys exhibit these characteristics and therefore they are the main group of materials for high temperature applications.

A term of great importance in the above equation is the grain size (d) because the exponent n varies between 2 and 3, thus moderate grain size increments, significantly increase creep strength. In addition, as mentioned in the previous sections, the grain boundary cavity distribution depends on the orientation with respect to the direction of the applied stress, thus the grains perpendicular to the stress direction have the maximum cavitation, whereas the grain boundaries parallel to the stress, practically do not suffer it. This has led to the design of coarse-columnar grain materials, where the long direction of the grains is oriented parallel to the maximum principal stress direction in order to increase creep strength. Such strategy has gone as far as developing materials without grain boundaries at all, that means,

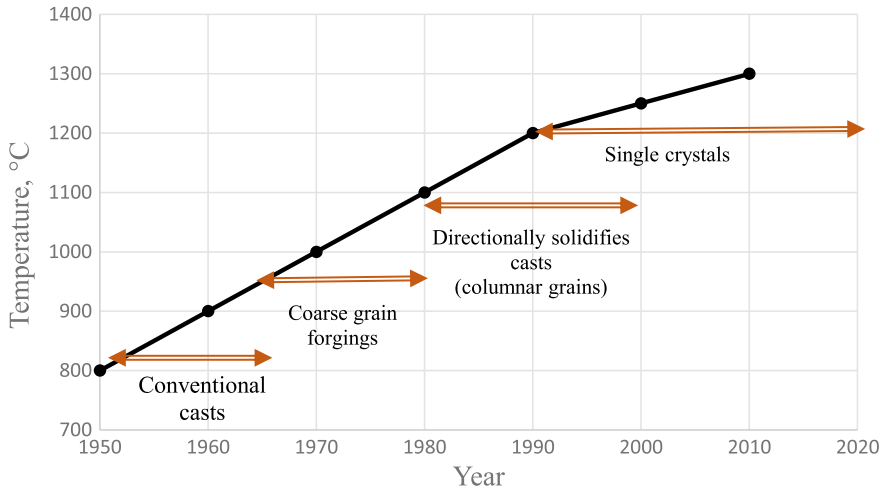


Fig. 8.15 Evolution of creep strength and grain size of materials for blades used in the combustion section of gas turbines and airplane engines

single-crystals. Figure 8.18 shows the evolution of materials for the manufacture of airplane turbine blades in terms of the grain size structure (Fig. 8.15).

The next microstructural characteristic that improves creep strength is size and distribution of second-phase precipitates. At first, the precipitates should be stable and coarse enough to reduce grain boundary sliding; second, they should have a high melting point, and third, they should have interstitial alloying elements. This combination reduces diffusivity, dislocation climb and diffusive flow altogether. These characteristics are featured by nickel-base super alloys, constituted by cuboidal-shaped gamma precipitates, with ordered crystalline structure, which are formulated and specially heat treated to obtain a microstructure like the one shown in Fig. 8.16. Other developments are metallic alloys hardened by alumina or silicon carbide particles, in order to maximize the stability second phase particles. These materials are manufactured by powder metallurgy techniques and exhibit high-temperature strengths as much as 20 times higher than that those of conventional alloys. Table 8.3 lists some of the typical materials for high temperature applications, with their main characteristics and maximum service temperature.

8.7 Superplastic Behavior

Superplastic behavior, also termed *superplasticity* is the ability of some materials to show great elongations in tension, as much as several thousand per cent, so their appearance after a tension test is like the schematic representation of Fig. 8.17. Not long ago, superplasticity was considered as laboratory-curiosity, but currently the need for making complex pieces and reduce manufacturing costs has made

Fig. 8.16 Microstructure of a nickel-base super alloy for high temperature use

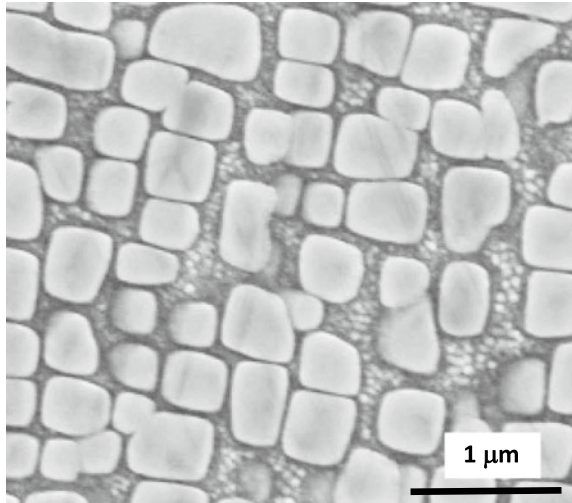


Table 8.3 Typical materials for high temperature applications

Material	Types	Max. service temp. (°C)
High alloy solid solution steels	Austenitic, series 304, 316 y 321 Forged o centrifugal cast	600
Low alloy ferritic steels strengthened by carbides	Up to 4% Cr + (Mo, V) Forged or rolled	650
Nickel Base Superalloys: Ni + (Cr, W, Co)	Supersaturated matrix with ordered precipitates. Directional solidification with columnar grain or single crystal	950
Oxides and refractory carbides: Al ₂ O ₃ , SiC, Si ₃ N ₄	Monolithic, thermal spray or dispersion strengthened mechanically alloyed	1300

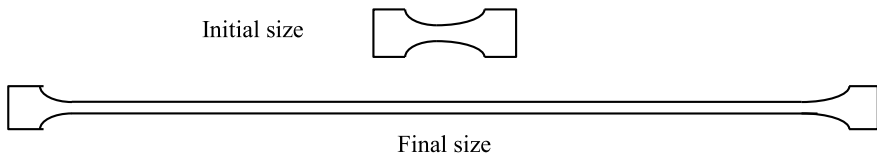


Fig. 8.17 Schematic illustration of a tension specimen after superplastic deformation

superplasticity commercially attractive. To achieve superplastic behavior, it is required that the material does not develop necks during tensile strain, so the mechanism of ductile fracture by nucleation and coalescence of internal voids has to be suppressed (see Chap. 6), therefore the material may continue elongating without rupture. Such condition is obtained if the material has a high strain hardening coefficient, as it will be shown next.

From Garofalo's equation, the equation for secondary creep deformation can be expressed as:

$$\sigma = Ct \left(\frac{d\varepsilon}{dt} \right)^{m'}$$

where σ is the real stress, ε is the real strain, t is time, m' is the dynamic strain hardening coefficient and Ct is an experimental constant. Stress is given by $\sigma = F/A$, where F is the load and A is the cross section area. If the deformation is uniform, $d\varepsilon/dt = dA/dt$, then it can be shown that:

$$-\frac{dA}{dt} = \left(\frac{F}{Ct} \right)^{\frac{1}{m'}} \left(\frac{1}{A^{1-m'}} \right)$$

According to this equation, as m' increases, the dA/dt diminishes, and if m' approaches to 1.0 (linear strain hardening), dA/dt becomes close to zero, which means that necks will not be formed and thus the material will be superplastic. Superplastic behavior has been observed in many alloy systems and it is fundamentally a dynamic strain-induced precipitation phenomenon. Researchers have found that superplastic materials must have m' values higher than 0.5, in addition to complying with the following requirements:

- (1) Low strain rates, from 10^{-4} to 10^{-6} s^{-1}
- (2) Temperature above $0.5T_f$
- (3) Grain size less than 10 microns
- (4) Dynamic second phase precipitation.

There is still controversy about the superplastic deformation mechanism, but it is generally acknowledged that it involves extensive grain boundary sliding, where the deformation of equiaxial is accommodated by grain rotation. During superplastic deformation, the crystalline texture is destroyed, and therefore no cell dislocations nor subgrains are formed, which indicates that the dislocations are quickly annihilated as they appear, thus making strain hardening a dynamic process too.

To determine whether a material can be superplastic, a Stress vs. Strain-rate plot must be constructed from tension tests at several strain rates. The stress-strain rate relation can be rewritten logarithmically as:

$$\log, \sigma = \log Ct + m' \log(d\varepsilon/dt)$$

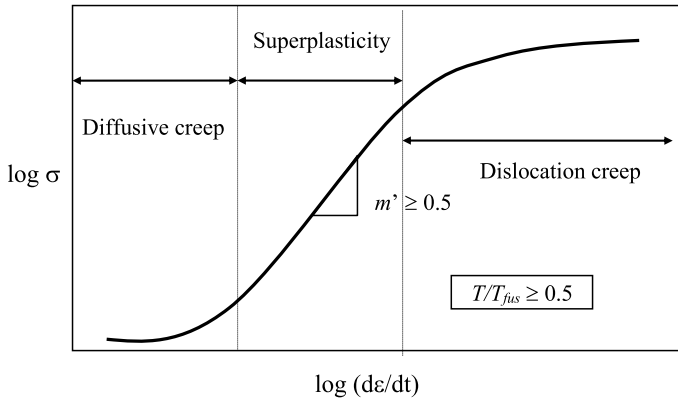


Fig. 8.18 Stress-Strain rate curve to determine the potential for superplastic behavior

Thus, by plotting the experimental data of strain rate and stress in a log-log scale, as shown in Fig. 8.18, the slope of the straight portion of the plot is the coefficient m' , then if $m' > 0.5$ the material may be superplastic. This condition is achieved at intermediate strain rates, since at low and high strain rates, the deformation mechanisms are diffusive creep and dislocation creep, respectively, therefore, the material will not exhibit large elongations, because the aforementioned dynamic processes do not occur.

References

1. Kimura K, Sawada K, Kushima H (2012) Creep ductility of creep strength enhanced ferritic steels. *J Pressure Vessel Technol* 134
2. Larson FR, Miller J (1952) *Transactions ASME* 74:765–771