Practical Engineering Failure Analysis
MECHANICAL ENGINEERING
A Series of Textbooks and Reference Books

Founding Editor
L. L. Faulkner
Columbus Division, Battelle Memorial Institute
and Department of Mechanical Engineering
The Ohio State University
Columbus, Ohio

1. Spring Designer’s Handbook, Harold Carlson
3. Lubrication Fundamentals, J. George Wills
6. Centrifugal Pump Clinic, Igor J. Karassik
10. Vibrations of Shells and Plates, Werner Soedel
11. Flat and Corrugated Diaphragm Design Handbook, Mario Di Giovanni
12. Practical Stress Analysis in Engineering Design, Alexander Blake
13. An Introduction to the Design and Behavior of Bolted Joints, John H. Bickford
15. Spring Manufacturing Handbook, Harold Carlson
17. Gears and Their Vibration: A Basic Approach to Understanding Gear Noise, J. Derek Smith
18. Chains for Power Transmission and Material Handling: Design and Applications Handbook, American Chain Association
20. Gear Drive Systems: Design and Application, Peter Lynwander
22. CAD/CAM Systems Planning and Implementation, Charles S. Knox
24. *Traction Drives: Selection and Application*, Frederick W. Heilich III and Eugene E. Shube
34. *Principles of Automated Drafting*, Daniel L. Ryan
38. *Mechanism Analysis: Simplified Graphical and Analytical Techniques*, Lyndon O. Barton
41. *Design Assurance for Engineers and Managers*, John A. Burgess
42. *Heat Transfer Fluids and Systems for Process and Energy Applications*, Jasbir Singh
45. *Electronically Controlled Proportional Valves: Selection and Application*, Michael J. Tonyan, edited by Tobi Goldoftas
47. *Fabric Filtration for Combustion Sources: Fundamentals and Basic Technology*, R. P. Donovan
51. *Couplings and Joints: Design, Selection, and Application*, Jon R. Mancuso
52. *Shaft Alignment Handbook*, John Piotrowski
55. *Plastics Gearing: Selection and Application*, Clifford E. Adams
56. *Clutches and Brakes: Design and Selection*, William C. Orthwein
57. *Metallurgical Applications of Shock-Wave and High-Strain-Rate Phenomena*, edited by Lawrence E. Murr, Karl P. Staudhammer, and Marc A. Meyers
58. *Magnesium Products Design*, Robert S. Busk
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How to Integrate CAD/CAM Systems: Management and Technology</td>
<td>William D. Engelke</td>
</tr>
<tr>
<td>Cam Design and Manufacture: Second Edition; with cam design software for the IBM PC and compatibles, disk included</td>
<td>Preben W. Jensen</td>
</tr>
<tr>
<td>Solid-State AC Motor Controls: Selection and Application</td>
<td>Sylvester Campbell</td>
</tr>
<tr>
<td>Fundamentals of Robotics, David D. Ardayfio</td>
<td></td>
</tr>
<tr>
<td>Belt Selection and Application for Engineers, edited by</td>
<td>Wallace D. Erickson</td>
</tr>
<tr>
<td>Developing Three-Dimensional CAD Software with the IBM PC,</td>
<td>C. Stan Wei</td>
</tr>
<tr>
<td>Organizing Data for CIM Applications, Charles S. Knox, with contributions by Thomas C. Boos, Ross S. Culverhouse, and Paul F. Muchnicki</td>
<td></td>
</tr>
<tr>
<td>Computer-Aided Simulation in Railway Dynamics, by Rao V. Dukkipati and Joseph R. Amyot</td>
<td></td>
</tr>
<tr>
<td>Fiber-Reinforced Composites: Materials, Manufacturing, and Design,</td>
<td>P. K. Mallick</td>
</tr>
<tr>
<td>Photoelectric Sensors and Controls: Selection and Application,</td>
<td>Scott M. Juds</td>
</tr>
<tr>
<td>Finite Element Analysis with Personal Computers, Edward R. Champion, Jr., and J. Michael Ensminger</td>
<td></td>
</tr>
<tr>
<td>Applied Finite Element Modeling: Practical Problem Solving for Engineers, Jeffrey M. Steele</td>
<td></td>
</tr>
<tr>
<td>Measurement and Instrumentation in Engineering: Principles and Basic Laboratory Experiments, Francis S. Tse and Ivan E. Morse</td>
<td></td>
</tr>
<tr>
<td>Centrifugal Pump Clinic: Second Edition, Revised and Expanded,</td>
<td>Igor J. Karassik</td>
</tr>
<tr>
<td>High Vacuum Technology: A Practical Guide, Marsbed H. Hablanian</td>
<td></td>
</tr>
<tr>
<td>Pressure Sensors: Selection and Application, Duane Tandeske</td>
<td></td>
</tr>
<tr>
<td>Zinc Handbook: Properties, Processing, and Use in Design, Frank Porter</td>
<td></td>
</tr>
<tr>
<td>Thermal Fatigue of Metals, Andrzej Weronski and Tadeusz Hejwowski</td>
<td></td>
</tr>
<tr>
<td>Classical and Modern Mechanisms for Engineers and Inventors,</td>
<td>Preben W. Jensen</td>
</tr>
<tr>
<td>Handbook of Electronic Package Design, edited by Michael Pecht</td>
<td></td>
</tr>
<tr>
<td>Shock-Wave and High-Strain-Rate Phenomena in Materials, edited by</td>
<td>Marc A. Meyers, Lawrence E. Murr, and Karl P. Staudhammer</td>
</tr>
<tr>
<td>Industrial Refrigeration: Principles, Design and Applications, P. C. Koelet</td>
<td></td>
</tr>
<tr>
<td>Applied Combustion, Eugene L. Keating</td>
<td></td>
</tr>
<tr>
<td>Engine Oils and Automotive Lubrication, edited by Wilfried J. Bartz</td>
<td></td>
</tr>
<tr>
<td>Mechanism Analysis: Simplified and Graphical Techniques, Second Edition, Revised and Expanded, Lyndon O. Barton</td>
<td></td>
</tr>
<tr>
<td>Fundamental Fluid Mechanics for the Practicing Engineer,</td>
<td>James W. Murdock</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
86. Vibrations of Shells and Plates: Second Edition, Revised and Expanded, Werner Soedel
89. Finite Elements: Their Design and Performance, Richard H. MacNeal
91. Mechanical Wear Prediction and Prevention, Raymond G. Bayer
92. Mechanical Power Transmission Components, edited by David W. South and Jon R. Mancuso
95. Refractory Linings Thermomechanical Design and Applications, Charles A. Schacht
96. Geometric Dimensioning and Tolerancing: Applications and Techniques for Use in Design, Manufacturing, and Inspection, James D. Meadows
100. Friction Science and Technology, Peter J. Blau
101. Introduction to Plastics and Composites: Mechanical Properties and Engineering Applications, Edward Miller
102. Practical Fracture Mechanics in Design, Alexander Blake
103. Pump Characteristics and Applications, Michael W. Volk
104. Optical Principles and Technology for Engineers, James E. Stewart
105. Optimizing the Shape of Mechanical Elements and Structures, A. A. Seireg and Jorge Rodriguez
106. Kinematics and Dynamics of Machinery, Vladimír Stejskal and Michael Valásek
107. Shaft Seals for Dynamic Applications, Les Horve
108. Reliability-Based Mechanical Design, edited by Thomas A. Cruse
109. Mechanical Fastening, Joining, and Assembly, James A. Speck
110. Turbomachinery Fluid Dynamics and Heat Transfer, edited by Chunill Hah
112. Geometric Dimensioning and Tolerancing: Workbook and Answerbook, James D. Meadows
114. *Handbook of Thermoplastic Piping System Design*, Thomas Sixsmith and Reinhard Hanselka
117. *Fluid Sealing Technology*, Heinz K. Muller and Bernard S. Nau
118. *Friction and Lubrication in Mechanical Design*, A. A. Seireg
119. *Influence Functions and Matrices*, Yuri A. Melnikov
120. *Mechanical Analysis of Electronic Packaging Systems*, Stephen A. McKeown
123. *Gear Noise and Vibration*, J. Derek Smith
127. *Designing for Product Sound Quality*, Richard H. Lyon
128. *Probability Applications in Mechanical Design*, Franklin E. Fisher and Joy R. Fisher
129. *Nickel Alloys*, edited by Ulrich Heubner
133. *Rapid Prototyping Technology: Selection and Application*, Kenneth G. Cooper
134. *Reciprocating Machinery Dynamics: Design and Analysis*, Abdulla S. Rangwala
141. *Practical Guide to Pressure Vessel Manufacturing*, Sunil Pullarcot
144. *Handbook of Machine Tool Analysis*, Ioan D. Marinescu, Constantin Ispas, and Dan Boboc

Copyright © 2004 by Marcel Dekker, Inc.
145. Implementing Concurrent Engineering in Small Companies, Susan Carlson Skalak
146. Practical Guide to the Packaging of Electronics: Thermal and Mechanical Design and Analysis, Ali Jamnia
147. Bearing Design in Machinery: Engineering Tribology and Lubrication, Avraham Harnoy
149. Industrial Boilers and Heat Recovery Steam Generators: Design, Applications, and Calculations, V. Ganapathy
151. Industrial Noise Control and Acoustics, Randall F. Barron
152. Mechanical Properties of Engineered Materials, Wolé Soboyejo
153. Reliability Verification, Testing, and Analysis in Engineering Design, Gary S. Wasserman
155. Intermediate Heat Transfer, Kau-Fui Vincent Wong
156. HVAC Water Chillers and Cooling Towers: Fundamentals, Application, and Operation, Herbert W. Stanford III
159. Piping and Pipeline Engineering: Design, Construction, Maintenance, Integrity, and Repair, George A. Antaki
160. Turbomachinery: Design and Theory, Rama S. R. Gorla and Aijaz Ahmed Khan
162. Fluidized Bed Combustion, Simeon N. Oka
163. Theory of Dimensioning: An Introduction to Parameterizing Geometric Models, Vijay Srinivasan
165. Structural Analysis of Polymeric Composite Materials, Mark E. Tuttle
167. Handbook of Pneumatic Conveying Engineering, David Mills, Mark G. Jones, and Vijay K. Agarwal
171. Vehicle Stability, Dean Karnopp
173. *Liquid Pipeline Hydraulics*, E. Shashi Menon
175. *Mechanical Tolerance Stackup and Analysis*, Bryan R. Fischer
176. *Engineering Design for Wear*, Raymond G. Bayer
180. *Mechanical Alloying and Milling*, C. Suryanarayana
182. *Design of Automatic Machinery*, Stephen J. Derby

**Additional Volumes in Preparation**

*Mechanical Engineering Software*

*Spring Design with an IBM PC*, Al Dietrich

*Mechanical Design Failure Analysis: With Failure Analysis System Software for the IBM PC*, David G. Ullman
Practical Engineering Failure Analysis

Hani M. Tawancy
Anwar Ul-Hamid
Nureddin M. Abbas

King Fahd U. Petroleum/Mineral Research Institute
Dhahran, Saudi Arabia
Preface

Failure of engineering structures in industry can cause loss of life, unscheduled shutdowns, increased maintenance and repair costs, and damaging litigation disputes. To prevent future recurrence of the problem it is essential to carry out an investigation aimed at deterring the root cause of failure. The results of the investigation can also be used as the basis for insurance claims, for marketing purposes, and to develop new materials or improve the properties of existing ones. The treatment of this subject is therefore important in order to develop preventive measures minimizing the incidence of failure and reducing their associated cost to industry.

All engineering failures, without exception, occur at the weakest link in the design–fabrication–performance chain of a product. The ability to identify this weakest link and propose remedial measures is the key for a successful failure analysis investigation. This requires a multidisciplinary approach, which forms the basis of this book. An important feature of this book is that it not only demonstrates the methodology for conducting a successful failure analysis investigation, but also provides the necessary background. Different stages in the life cycle of an engineering product, encompassing design, fabrication and
performance in service, are explained. Factors that can influence the ability of a product to fulfill its functional requirements during service are discussed.

The background needed to conduct a successful failure analysis investigation is developed in this book by introducing important principles of materials science and engineering. Basis principles of fracture mechanics and stress analysis are also elucidated. Various mechanical properties and the tests required to measure them are described. Both macroscopic and microscopic characterization of materials and their relevance to the field of failure analysis are discussed. This knowledge is then used to develop the procedure for a thorough failure analysis investigation. Finally, several case studies covering failures related to processing of materials, design, manufacturing, and service conditions are classified according to the type of industry and included to serve as examples. Within this context, the difference between the mode of failure and cause of failure is emphasized.

It is hoped that this book will provide the reader with the tools required to reach a sound conclusion about the cause of component failure and will provide proper recommendations for preventive measures. Because of the legal implications of failure analysis investigation, the use of proper technical terminology is emphasized throughout the text.

It is our pleasure to acknowledge the support and encouragement received from King Fahd University of Petroleum and Minerals through the Research Committee. The first author is greatly indebted to his wife. Without her unfailing support and encouragement, this project would not have been possible.

\textit{Hani M. Tawancy}
\textit{Anwar Ul-Hamid}
\textit{Nureddin M. Abbas}
Contents

Preface

1 Introduction
  1.1 Engineering Products and Their Performance 1
  1.2 Engineering Properties of Materials 4
  1.3 Classes of the Engineering Alloys 9
  1.4 Structure of Engineering Alloys 10
  1.5 Failure of Engineering Products 11
  1.6 Imperfect vs. Defective Products 12
  1.7 Definition and Objective of Failure Analysis Investigations 14
  1.8 Approach to Failure Analysis Investigations 14
  1.9 Background Requirements of the Failure Analyst: Scope of the Book 16

2 Engineering Design–Fabrication–Performance 18
  2.1 Introduction 18
  2.2 Stages of Engineering Design 20
  2.3 Material Selection 21
  2.4 Fabrication of Engineering Alloys 22
  2.5 Solidification of Ingots 23
2.6 Cold Working 25
2.7 Recrystallization 26
2.8 Thermomechanical Processing 27
2.9 Primary Fabrication Techniques 28
2.10 Secondary Fabrication Techniques 34
2.11 Joining Techniques 37
2.12 Service Performance 41
2.13 Common Causes of Failure 42

3 Principles of Mechanics 44
3.1 Introduction 44
3.2 Concepts of Mechanics 45
3.3 Concept of Mechanical Force 46
3.4 Concepts of Work and Energy 48
3.5 Force and Motion 49
3.6 Conservation of Energy 51
3.7 Concept of Machines 51
3.8 State of Mechanical Equilibrium 52
3.9 Concept of Strain 53
3.10 Concept of Stress 55
3.11 Hook’s Law 59

4 Property Evaluation 61
4.1 Introduction 61
4.2 Nondestructive Tests 61
4.3 Destructive Tests: Measurement of Mechanical Properties 70

5 Stress Analysis 97
5.1 Introduction 97
5.2 Uniaxial State of Stress 98
5.3 Generalized State of Stress 100
5.4 Multiaxial Stress-Strain Relationship 113
5.5 Loading Conditions and Stress 116
5.6 Thermal Stress 120
5.7 Type of Stress Required to Produce Plastic Deformation 121
5.8 Maximum Stresses 122
5.9 Design Stresses 123
## 6 Macroscopic Aspects of Fracture and Fracture Mechanics

6.1 Definition of Fracture  
6.2 Objective of Fracture Mechanics  
6.3 Use of the Terms *Brittle* and *Ductile* in Fracture  
6.4 Crack Loading Modes and Macroscopic Morphology of Fracture Surfaces  
6.5 Crack Propagation Under a Plane Strain Condition  
6.6 Crack Propagation Under a Plane Stress Condition  
6.7 Crack Propagation Under a Mixed State of Plane Strain and Stress  
6.8 Sequence of Events Leading to Fracture  
6.9 Classification of Crack Propagation Modes According to Loading Conditions  
6.10 Variables Affecting Fracture Behavior  
6.11 Basic Principles of Fracture Mechanics  
6.12 Linear Elastic Fracture Mechanics (LEFM)  
6.13 Use of Fracture Mechanics in Design  
6.14 Concept of Allowable Crack Size  
6.15 Use of Fracture Mechanics in Failure Analysis  
6.16 Selection of Materials Resistant to Fracture

## 7 Structure of Engineering Alloys

7.1 Introduction  
7.2 Principles of Thermodynamics  
7.3 Elements of Internal Structure  
7.4 Structure of the Atom  
7.5 Significance of the Electronic Structure of Atom  
7.6 Electronic Structure and Chemical Properties: Classes of Elements  
7.7 Origin of Interatomic Binding Forces  
7.8 Types of Interatomic Binding Forces  
7.9 Bond Strength and Properties of Materials  
7.10 Arrangement of Atoms in Perfect Crystals
7.11 Understanding the Microscopic Plasticity of Perfect Crystals 237
7.12 Crystal Imperfections 240
7.13 Understanding the Microscopic Plasticity of Real Crystals 256
7.14 Alloy Phases and Phase Change 265
7.15 Equilibrium Phase Diagrams 276
7.16 Methods of Strengthening Engineering Alloys 285

8 Materials Characterization 291
8.1 Introduction 291
8.2 Techniques for Microstructural Characterization 291
8.3 Techniques for Chemical Analysis 311
8.4 Microstructure of Engineering Alloys 334

9 Corrosion 344
9.1 Introduction 344
9.2 Low-Temperature Aqueous Corrosion 345
9.3 High-Temperature Corrosion 377

10 Metallurgical Aspects of Fracture and Fractography 397
10.1 Introduction 397
10.2 Microscopic Aspects of Crack Nucleation 398
10.3 Microscopic Mechanisms of Crack Propagation 400
10.4 Fracture Modes and Fractography 409

11 Failure Analysis Procedure 413
11.1 Introduction 413
11.2 Definition of the Problem 414
11.3 Technical Background 415
11.4 Experimental Program and Analysis 415
11.5 Mode of Failure vs. Cause of Failure 418
11.6 Data Interpretation and Terminology 419
11.7 Recommendations 419
11.8 Failure Analysis Reports 420
## Contents

12 Case Studies

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.1 Introduction</td>
<td>422</td>
</tr>
<tr>
<td>12.2 Failure of Engineering Alloys Due to Improper Processing Practice</td>
<td>423</td>
</tr>
<tr>
<td>12.3 Failure of Engineering Products During Manufacturing</td>
<td>450</td>
</tr>
<tr>
<td>12.4 Effect of Variations in Design on Service Performance</td>
<td>457</td>
</tr>
<tr>
<td>12.5 Failure of Engineering Products During Service Because of Unanticipated Service Conditions</td>
<td>466</td>
</tr>
<tr>
<td>12.6 Failure of Engineering Products During Service Because of Improper Material Selection</td>
<td>496</td>
</tr>
<tr>
<td>12.7 Failure of Engineering Products During Service Because of Improper Service Conditions</td>
<td>544</td>
</tr>
</tbody>
</table>

Appendix A Chemical Composition and Classification of Selected Steels     | 581  |

Appendix B Units of Measurements in Mechanics                           | 587  |

Appendix C Moment of Inertia of Selected Cross Sections                | 590  |
1

Introduction

1.1 Engineering Products and Their Performance

Any engineering product ranging from a vacuum cleaner to an aircraft is designed and manufactured to perform specific functions serving human needs over a predetermined service life. Its history includes three main stages: (i) engineering design, (ii) manufacturing, and (iii) service as schematically illustrated in Fig. 1.1.

Engineering design is essentially a decision-making process by which a product can be manufactured at a reasonable cost. Differences in design of a given product, which are reflected on its performance can be related to the extent of engineering analysis preceding any decision made by the designer. Parameters included in such an analysis are (i) geometry and dimensions of different parts, (ii) types of materials used in manufacturing and their specifications, (iii) fabrication and assembling techniques, and (iv) service conditions.

Design is initiated after establishing the market need for a new product or improving the performance of an existing product. Subsequently, the design engineer conceives, plans, and conducts a program to develop complete engineering drawings of the product specifying the above design parameters. A prototype product is then manufactured and tested. If necessary, the design is modified so that the product meets all required specifications. Following this stage, the design is materialized into a commercial product by the manufacturer. Finally, the product is put into service to meet its functional requirements specifying the
procedures of start-up and shutdown as well as the maintenance procedure and its schedule. Figure 1.2 summarizes the above aspects of engineering design.

From our daily experience, we come to realize that various brands of a given product can significantly differ in their performance and durability. When a household item breaks down after a short usage, statements such as “the item is made of a cheap material,” “the item is so cheap,” “the item is poorly made” are often heard to explain the incident. In technical terms, such statements refer to design and manufacturing. Another possible explanation for the incident is that the item is misused, referring to service conditions deviating from manufacturer’s recommendations based upon design parameters.

A direct correlation exists between the cost of any product and the effort spent in engineering analysis to optimize its design, including the quality of materials used in manufacturing as well as the fabrication techniques. One of the
most important aspects of design determining performance is the quality of material selected for the application. Proper selection of material by the designer determines to what extent the user of the product realizes its reliability and satisfactory service life. Quite often, the material dictates the fabrication technique. It then becomes evident that the service performance of any engineering product can be envisioned as a step in a chain of a continuous process beginning with design, as further explained in Chap. 2. Differences in performance and cost of various brands of a given product can therefore be related to variations in (i) design including materials used in manufacturing and (ii) the fabrication techniques used to manufacture the product.

Design engineers are usually more concerned with structural and mechanical functions of the products. Therefore, the reaction of materials to mechanical usage, i.e., how materials respond to applied forces or loads, assumes a primary importance in deciding which material is to be selected. Any engineering product, however, cannot operate in complete isolation fulfilling its mechanical functions without interference from the environment. Various types of environmental effects, e.g., temperature and nature of substances in the environment, can have an important impact on the reliability and functional characteristics of the product.

Almost every design and useful application devised by the engineer is dictated by the quality of materials used in manufacturing, which limits the performance capabilities of the product. For example, the efficiency of an aircraft gas turbine engine is limited by the temperature capabilities of the materials used to manufacture its hottest sections.

Materials used to manufacture engineering products are referred to as engineering or structural materials. To classify a material as an engineering material, it must fulfill two main requirements. First, it must possess desirable qualities or properties rendering it suitable to manufacture certain products. Secondly, it must be readily fabricated in large quantities by standard techniques. Engineering materials can be classified into two main types: (i) metallic and (ii) nonmetallic. Even though the number of nonmetallic materials and their use in manufacturing continue to increase, metallic materials remain indispensable in many industrial applications because of their unique combination of properties.

Since engineering materials are the primary building blocks of any engineering product, material selection is one of the most important decisions made by the designer. It is important that the designer must have a broad knowledge of the meaning of each material property and how it is determined. Such knowledge is essential in terms of identifying the properties relevant to the application and whether their listed values are derived under conditions similar to those encountered in practice. A brief account of the engineering properties of materials is given in Sec. 1.2, and the various classes of engineering metallic materials are described in Sec. 1.3.
To summarize the above discussion, the service performance of a given product is dependent upon the extent to which the various steps of the continuous process of engineering design-manufacturing-performance are correlated and tailored so that the product can effectively perform its functions over its intended service life. Each step in the above process can be considered as a link in a chain connecting design with manufacturing and performance. Depending upon the extent of engineering analysis spent to strengthen these links, various brands of the same product can significantly differ in their performance as well as cost. On the other hand, the above chain can be expected to fall apart at its weakest link, as described in Sec. 1.5.

1.2 Engineering Properties of Materials

Engineering properties are those qualities of structural materials which qualify them for use in manufacturing various products. They can be classified into four main types: (i) mechanical properties, (ii) physical properties, (iii) chemical properties, and (iv) fabrication properties, as summarized in Fig. 1.3. In this section, the various engineering properties are briefly defined. Procedures used to measure these properties and their origin are described in Chap. 4.

1.2.1 Mechanical Properties

Design engineers are primarily concerned with mechanical properties which describe the response of the material to externally applied mechanical loads or forces. A material reacts to an applied force whether it is small or large by distortion or deformation; i.e., it undergoes a change in shape or dimensions, which is known as strain. Simultaneously, the material develops internal forces known as stress to counteract the effect of the applied force. Strain can be measured directly, but stress can only be calculated from knowledge of strain. For a given material, the

Figure 1.3 A summary of important engineering properties of materials.
amount of strain is determined by the magnitude of the applied force. Two types of strain are distinguished: (i) elastic strain and (ii) plastic strain, as shown in Fig. 1.4. When the applied load is relatively small, the strain is elastic. By elastic is meant that when the applied force is removed, the material springs back to its original dimensions; i.e., the elastic strain is reversible or recoverable. In contrast, the plastic strain is irreversible; i.e., a plastically strained material undergoes a permanent change in shape or dimensions. It occurs when the applied load exceeds a certain critical value dependent upon the material. Plastic deformation continues to occur as the applied load is increased until ultimately it is terminated by fracture. Materials which can readily deform in a plastic manner to accommodate the effect of an applied load prior to fracture are said to be ductile.

Plastic deformation of a given material to accommodate the effect of an applied force is a thermally activated time-dependent process, and therefore it is influenced by two important parameters: (i) the rate at which the force is applied or strain rate and (ii) temperature. At a given temperature, if the force is applied relatively slowly in a progressive manner, the material can readily deform; however, the same material can fracture without undergoing an appreciable deformation if the load is applied very rapidly. Increasing the temperature has the same effect as reducing the strain rate.

**Figure 1.4** Schematics illustrating the difference between elastic and plastic deformation. (a) Elastic deformation: When the applied load $P_1$ is relatively small, the deformed part (loaded) springs back to its original shape after the load is removed (unloaded). (b) Plastic deformation: When the load is increased to $P_2$, the deformed part (loaded) undergoes a permanent change in shape after the load is removed (unloaded).
It can be concluded from the above discussion that the response of a material to an applied force or load can be described in terms of three separate phenomena: (i) elastic deformation, (ii) plastic deformation, and (iii) fracture. Usually, designers are primarily concerned with the capacity of the material to carry an applied load without fracturing. In some applications, a certain amount of deformation usually elastic can be tolerated. Frequently, however, the amount of deformation must be limited to a rigidly minimum value, as described below.

An applied load can either be static or dynamic defining various sets of mechanical properties. By definition, a static load such as a dead weight is applied only once. Tensile loads which tend to stretch the material are of primary importance in applications for structural and mechanical functions. Under the influence of a static tensile load, a part made of a given material is stretched and its cross-sectional area is reduced. Two different sets of mechanical properties are defined under static loading condition: (i) tensile properties and (ii) creep properties.

Tensile properties define the response of a material to a progressively increasing tensile load until fracture occurs. They include yield strength, ultimate tensile strength, and tensile ductility. Yield strength is defined as the load per unit cross-sectional area required to initiate plastic deformation on a macroscopic scale. It is a measure of the useful load-carrying ability of the material. Ultimate tensile strength or shortly tensile strength is the load per unit area beyond which fracture occurs. Tensile ductility can be expressed as the percentage elongation at the point of fracture; however, a more meaningful measure is the reduction in cross-sectional area. Up to a temperature corresponding to about one-third the melting point of the material on the absolute scale (degrees Kelvin), the tensile properties are not significantly influenced by the strain rate. However, above that temperature, the effect of strain rate becomes significant.

Since plastic deformation is a thermally activated process, a material can progressively deform in a plastic manner with time at elevated temperatures under the influence of a static load even if it is below the corresponding tensile yield strength at the same temperature. Such a phenomenon is called creep and it becomes significant at temperatures >0.5T_m, where T_m is the melting point on the absolute scale. Plastic deformation resulting from creep is terminated by fracture. Creep properties are expressed in terms of the stress required to produce specific creep strain in a given time at a certain temperature or the stress required to cause fracture.

Dynamic loads are classified into two types: (i) impact loads and (ii) fatigue loads. An impact load resembles a static load in that it can be a single application of a force. However, in contrast with a static load, the rate of deformation is very high, which can localize the stresses in the weak sections of the part leading to fracture. Resistance to fracture under impact loading condition is measured in terms of the energy required to cause fracture, which is known as impact toughness. Fatigue loading is a more common type of a dynamic load. In this case,
a load considerably smaller than the static load corresponding to the yield strength is repeatedly applied to the part in a cyclic manner. Even though the fatigue load is lower than that corresponding to the yield point, it ultimately leads to fracture after a certain number of cycles. Fatigue strength is expressed in terms of the number of cycles required to cause fracture under a given stress.

Resistance to fracture of a given material is expressed in terms of a parameter called the fracture toughness. It relates the stress required to propagate an existing crack with the initial crack length.

Another mechanical property which is frequently used to evaluate the strength of materials is hardness. However, it is not as well defined as other properties such as yield and ultimate strengths because of its dependence on the method of measurement as well as the measurement conditions. Nevertheless, if the hardness value is accurately defined, a direct correlation exists between hardness and strength. A higher hardness reflects higher yield and ultimate strengths and in turn, a lower ductility.

1.2.2 Physical Properties

Among the important physical properties in engineering applications are density and thermal properties. In several applications particularly aerospace, the strength-to-weight ratio of the part is an important design parameter. A high ratio requires that a material of a smaller density has a sufficiently high strength.

When a product is designed to operate at high temperatures, it is essential to check the mechanical properties at the actual service temperature. Frequently, temperature gradients exist in components operating at high temperatures, which are designed to either preserve or minimize this gradient depending upon the particular application. If it is required to maintain a steep temperature gradient, a structural material with low thermal conductivity must be used. Metallic materials are generally classified as good thermal conductors. Thermal properties are expressed in terms of thermal conductivity and coefficient of thermal expansion. By definition, thermal conductivity is the rate of flow of heat energy per unit time across a cube of the material having unit dimensions such that the temperature difference across two parallel faces of the cube is 1°. Linear coefficient of thermal expansion is defined as the increase in length per unit length resulting from an increase in temperature by 1°.

As described in Chap. 5, thermal expansion induces stresses in the part. When the part is subject to thermal cycling as in a gas turbine engine, this causes thermal fatigue, which can lead to failure. It is evident that the susceptibility to thermal fatigue is reduced by lowering the coefficient of thermal expansion, which serves as an example illustrating the interaction among the various properties of materials.
1.2.3 Chemical Properties

A common feature of all engineering products is that they perform their functional requirements in an environment containing other forms of matter ranging from the normal atmosphere to other more aggressive environments. Chemical properties refer to the interaction between a given material and other species in the environment, e.g., oxygen, sulfur, chlorine, carbon, to form other substances. Such interaction, which is known as corrosion, degrades the performance of the material and can lead to failure. Corrosion is classified into two main types: (i) low-temperature aqueous corrosion and (ii) high-temperature corrosion. As its name implies, aqueous corrosion involves chemical reactions in the liquid state. In contrast, high-temperature corrosion refers to reactions occurring in the gaseous state. However, molten substances are involved in a type of high-temperature corrosion known as hot corrosion. When the products of corrosion reactions become porous resulting in metal wastage, the thickness of the part is reduced and in turn its load-carrying ability. Consequently, the designer must be fully aware of the environment in which the product operates. It suffices to mention here that the detrimental effects of corrosion can be minimized by proper selection of materials. Corrosion properties are expressed in terms of rate of metal loss (change in thickness or weight) in specific environments.

1.2.4 Fabrication Properties

Fabrication properties are those qualities which determine the response of a given material to fabrication into useful shapes by various standard techniques. For example, workability or formability is a term used to describe the ability of the material to be formed into various useful shapes by mechanical working operations, which are essentially plastic deformation processes. Evidently, good workability requires that the material has sufficient ductility so that it can readily be formed into useful shapes without cracking or fracture. Another fabrication property is castability. It refers to the ability of the material to be directly formed into useful shapes by casting the molten material into suitable molds and allowing it to solidify. Weldability is another important fabrication property. It is defined as the ability of the material to be joined to the same or another material by different welding techniques. Another related property is brazability.

It can be concluded from the above discussion that although a given set of properties may assume a primary importance in a certain application, other properties can also be important if the interaction among various properties is appreciated. Both the reliability and satisfactory service life can be largely dependent upon the above considerations.
1.3 Classes of the Engineering Alloys

Pure metals are rarely used as engineering materials because of their inadequate properties. However, it is possible to improve certain properties of a given metal by mixing it with another metal or nonmetal to produce an alloy. By definition an alloy consists of a base metal and one or more alloying elements, which can include metals and nonmetals.

Although about two-thirds of the known chemical elements are classified as metals, only a few metals are used as the bases for engineering alloys, particularly aluminum, copper, iron, lead, magnesium, nickel, tin, zinc, and to a lesser extent cobalt. Some of the more common metals used as alloying elements are antimony, beryllium, cobalt, chromium, manganese, molybdenum, titanium, tungsten, vanadium, zirconium, and to a lesser extent rhenium. Carbon, nitrogen, and boron are the most commonly used nonmetallic alloying elements. For example, steel is essentially an alloy of iron (base metal) and carbon (nonmetallic alloying element). However, other metallic elements such as molybdenum, chromium, and nickel are added to steels to improve certain properties. Although it is likely that metallic materials will continue to be the more indispensable engineering materials, the relative importance of nonmetallic materials is rapidly increasing. Depending upon their chemical compositions, engineering alloys can be classified into two main classes: (i) ferrous alloys and (ii) nonferrous alloys.

Ferrous alloys or steels are basically alloys of iron and carbon. Most of the alloys used in engineering applications belong to this category. A commercially produced steel normally contains small concentrations of other elements, particularly manganese, silicon, sulfur, and phosphorus. Provided these elements are present in fractional percentages with the exception of manganese which can reach 1.5 wt%, the product is called plain carbon steel or simply carbon steel. Typically, the maximum amount of sulfur is 0.04 wt% and that of phosphorus is 0.05 wt%, and the silicon content is kept between 0.1 and 0.3 wt%. Low carbon grades or mild steels have a maximum carbon content of 0.23 wt%. Higher carbon grades contain 0.35–0.95 wt% carbon.

Although carbon steels are widely used in manufacturing, their properties are inadequate for more advanced applications. To improve their properties, certain alloying elements are added to carbon steels. Commonly used alloying elements include chromium, nickel, vanadium, molybdenum, tungsten, cobalt, boron, and copper as well as manganese, silicon, phosphorus, and sulfur in concentrations greater than normally present. Depending upon the nature and concentrations of various alloying elements, a large variety of steel grades are produced. When the total amount of alloying elements is 5 wt% or less, the product is called low-alloy steel. Most of the structural steels belong to this category. High-alloy grades, particularly those containing high concentrations of chromium and nickel, are classified as stainless steels.
Various steel grades are classified according to the system developed by the American Iron and Steel Institute (AISI). Each type of steel is assigned a four- or three-digit AISI number. Plain carbon steels and low-alloy steels are assigned four-digit numbers such that the first number refers to the steel grade and the second number indicates the class within the grade. Both the first and second numbers indicate the major alloying element in the steel other than carbon. When the last two numbers are divided by 100, the nominal carbon content is obtained in weight percent, i.e., if the last two numbers are 50, the nominal carbon content is 0.5%. AISI classification of carbon, low-alloy, and stainless steels are summarized in Appendix A. Other grades of steels include (i) tool steels, (ii) free machining steels, and (iii) heat-resistant steels.

Although there are many varieties of nonferrous alloys, their consumption in engineering applications is still a small fraction relative to the ferrous alloys. It is possible to classify nonferrous alloys into five categories: (i) light metals and alloys including aluminum, beryllium, magnesium, and titanium as well as their alloys; (ii) copper and its alloys, particularly brass (alloys of copper and zinc) and bronze (alloys of copper and tin); (iii) nickel-, cobalt-, and iron-based superalloys; (iv) high-temperature alloys, which are also nickel, cobalt, and iron based; (iv) high-performance or exotic alloys, which are mostly nickel based; and (v) miscellaneous metals and alloys such as zinc and its alloys, lead, tin, and antimony.

1.4 Structure of Engineering Alloys

Many of the technologically useful properties of engineering alloys, such as mechanical properties, are determined by their internal structure. To be able to effectively select and use these materials in engineering applications, it is very important to have a basic and broad understanding of the relationship between structure and properties.

Although the fundamental components of all types of materials are the same (protons, neutrons, and electrons), it is so remarkable that so many different types of materials with widely different properties do exist. As explained in Chap. 7, this can be related to (i) the manner in which the fundamental components are combined to form different types of atoms and (ii) the manner in which large numbers of atoms are grouped together to form a material.

It suffices to mention here that the constituent atoms of engineering alloys are arranged in an orderly fashion to form crystals, as schematically illustrated in Figure 1.5a. Therefore, among other types of materials, the engineering alloys are said to be crystalline. A large number of atoms can be grouped in many different ways to form a crystal. Only in a very few applications such as the blades of advanced gas turbine engines, engineering alloys are used in the single-crystal form. In most applications, however, engineering alloys are used
in the polycrystalline form. A polycrystalline material consists of an aggregate of crystals. By virtue of the process leading to the formation of a polycrystalline aggregate, the atoms in each crystal are oriented in a different manner with respect to atoms of a neighboring crystal, as illustrated in Figure 1.5b, giving rise to a bounding surface separating each crystal from its neighbors. Commonly, each crystal in an aggregate is called a grain, and the bounding surface between neighboring grains is called a grain boundary.

1.5 Failure of Engineering Products
When an engineering product ceases to perform one or more of its functions well before its expected service life, it is said to fail. Even though designers and manufacturers continue to strengthen the links among design, manufacturing, and performance, failures still occur and will continue to occur for one reason or another. Failures can cause loss of lives, unscheduled shutdowns of plants, increasing maintenance and repair costs as well as very costly litigation and damages. Any type of failure can be expected to occur at the weakest link in

Figure 1.5  Arrangement of atoms in crystals. (a) A single crystal, (b) an aggregate of crystals (polycrystalline material), (c) arrangement of atoms across a grain boundary.
the chain of the continuous process of engineering design-fabrication-performance as demonstrated by the following examples.

Many of the failures encountered in practice can be related to fatigue loading. One type of fatigue which can be appreciated from our daily experience is the fracture of a paper clip by repeated bending in opposite directions. In an application where a machine part is subjected to fatigue loading, premature failure can readily occur if the geometry of the part is not properly designed for fatigue, even if the part is made of a material known for its best resistance to fatigue. Evidently, in this case, the weakest link in the chain is the part’s geometry. To cite another example, consider a pipe assembled by welding to carry gaseous substances at elevated temperature. Even if the best material is selected for the application in terms of its resistance to heat and environmental effects, premature failure can still occur if the material has a poor weldability. In this case, the weakest link is the incompatibility of the material with the fabrication technique. To ensure that the product undertakes its intended functional requirements, it is equally important that the user follows the design parameters, e.g., service temperature, operating loads, and manufacturer’s recommendations regarding start-up and shut-down procedures and maintenance schedule. Any deviation from such a practice without a prior knowledge of the corresponding effect on performance can lead to premature failure where the weakest link is improper service conditions.

It can be concluded from the above discussion that premature failure can result from weakening any of the links in the chain of the continuous process of engineering design-fabrication-performance. Usually, when a product fails prematurely, it is said to be defective. However, it is extremely important to distinguish between a defective product and an imperfect product as explained below.

1.6 Imperfect vs. Defective Products

In our minds the word perfect is usually connected with the ideal. Although we realize that the ideal does not exist, it is particularly useful to think of the perfect as an upper limit, which cannot be exceeded. To illustrate this concept, consider the functional requirements of the heat engine of an automobile. Its primary function is to convert heat energy resulting from burning the fuel into useful mechanical energy, which drives the automobile. Any heat engine operates between two temperatures $T_1$ and $T_2$, where $T_1 > T_2$. In theory, the ideal efficiency $\eta$ of converting heat into mechanical work is given by $\eta = (T_1 - T_2)/T_1$, where the temperature is measured on the absolute scale (K).

It is evident from the above expression that $\eta = 100\%$ when $T_2 = 0$ K. However, in practice a temperature of 0 K cannot be achieved and therefore 100% efficiency is not possible; i.e., it is not possible to convert all heat energy into useful work. Also, some of the heat input is lost to other processes such as friction. Despite this inherent limitation, the ideal efficiency remains to be a
useful guide for how the real approaches the ideal with further developments in
design and manufacturing. With current technology, there are acceptable levels
of efficiency. When the efficiency drops below those levels for one reason or
another, the engine is said to be inefficient, or defective.

It can be concluded from the above example that any real product can be
envisioned as consisting of imperfections superimposed on a perfect product, i.e.,

\[
\text{Real product} = \text{perfect product} + \text{imperfections}
\]

In practice, manufacturing processes must inevitably introduce some flaws
or discontinuities into any product such as voids, pores, and small cracks. When
such flaws can be accommodated by the product without adversely affecting its
functional requirements over its intended service life, they are classified as imper-
fections; i.e., an imperfection can be defined as a deviation from perfection within
acceptable limits. Conversely, if the flaws responsible for the deviation from
perfection exceed the acceptable limits, they are called defects and in turn the
product is said to be defective, as schematically illustrated in Fig. 1.6.

Having recognized the fact that any real product is imperfect, designers and
manufacturers take the necessary measures to minimize the size and severity
of imperfections so that they lie within acceptable limits. Such measures include
(i) proper selection of materials, (ii) compatible fabrication techniques, and
(iii) nondestructive testing.

A defective product which may fail during testing or during service is cer-
tainly the result of a certain mistake in the continuous process of engineering
design-manufacturing-performance. Learning from mistakes has played a signifi-
cant role in technological advances over the history of mankind. It is evident that

\[ a_1 < a_c \quad \text{(Imperfect (Real Product))} \]
\[ a_2 > a_c \quad \text{(Defective Product)} \]

\[ \text{Acceptable Range} \quad a_c \]

Figure 1.6 Distinction between an imperfect product and a defective product on the
basis of an acceptable flaw or defect size.

Copyright © 2004 by Marcel Dekker, Inc.
before a mistake can be corrected, it must first be identified, which is one of the objectives of failure analysis investigations.

1.7 Definition and Objective of Failure Analysis Investigations

A failure analysis is an investigation carried out to determine the cause of failure of a certain product or equivalently the mistake in the continuous process of engineering design-manufacturing-performance in order to prevent its recurrence in the future. After the investigation is completed, its findings are documented in a report. It is important to realize, however, that failure analysis investigations can serve other useful purposes, as described below.

Results of failure analysis investigations can be extremely useful to designers of the same or similar products. Knowledge of prior incidents of failure and their causes aid the designer in developing a product of improved reliability and durability. Frequently, however, such information never reaches the designer for one reason or another.

Because of the legal aspects of failures, the results of failure analysis investigations can be used as the basis for litigation and insurance claims. Manufacturers can also use the results of failure analysis investigations for marketing purposes to promote a new product with better performance capabilities. Furthermore, by identifying the deficiencies of certain structural materials through failure analysis investigations, it is possible to develop new materials or improve the properties of existing materials.

It is evident from the above discussion that failure analysis reports can be read by a wide audience including the experts and nonexperts. Therefore, it is imperative that the reports must be clearly written and avoid the use of vague language. Since the reports can also be used by a court of law to determine responsibility, it is extremely important to use the proper technical terms with clear definition to avoid any confusion. Because of the importance of this aspect, the use of proper technical terminology is emphasized throughout this text.

1.8 Approach to Failure Analysis Investigations

Having recognized that failure is likely to occur at the weakest link in the chain connecting the various steps of engineering design with manufacturing and performance, the function of the failure analyst is to trace back the history of the product from service to manufacturing and design in order to identify the weakest link, as illustrated in Fig. 1.7; i.e., any failure analysis investigation progresses in a reverse order to that followed until the product is commissioned to service. Evidently, this route covers more than one engineering discipline.
At least, the failure analyst must have a broad basic understanding of the two fields of materials: (i) materials science and (ii) materials engineering. Although these two aspects of materials are interrelated, they are generally regarded as separate fields. Each field is distinguished by the scale over which the behavior of materials is examined, as described below.

Materials science is concerned with the close relationship between the behavior of materials and their internal structure, where materials are viewed on the micrometer and submicrometer scale. Because the performance of materials is a sensitive function of their internal structure, the analyst must have a basic knowledge of materials science and its tools.

In contrast with materials science, materials engineering deals with materials behavior on a continuum basis, i.e., on the centimeter and meter scale requiring a basic knowledge of mechanics. Both stress analysis and fracture mechanics are branches of mechanics, which can be very useful tools in failure analysis investigations. Also, the failure analyst must have a basic knowledge of fabrication techniques of materials, which lies within the spectrum of materials engineering. Distinction between materials engineering and materials science is schematically illustrated in Fig. 1.8.

Since the primary objective of any failure analysis investigation is to prevent future failures, it is extremely important that the analyst must design the investigation accordingly. Frequently, the extent of success of any failure analysis investigation is determined by the foresight used in establishing the initial formulation of the problem being investigated. A well-thought-out initial approach usually results in a more rapid solution of the problem. In principle, the steps to be followed in solving the simplest case should be the same as those followed for the most difficult ones. Such a practice leads to a natural step-by-step case-solving technique. To be an effective failure
analyst, one must clearly define the problem being considered. An uninformed analyst cannot make a contribution to the solution of a problem not understood.

Very frequently metallurgists or materials scientists are called upon to conduct failure analysis investigations. Without an adequate knowledge of materials engineering, however, the analyst can only provide a partial answer to the problem. A similar situation is likely to arise if a materials engineer without an adequate knowledge of materials science is called upon to carry out the investigation.

In conclusion, any failure analysis investigation must be approached from a broad prospective taking into consideration all aspects of the problem. This requires the analyst to have a broad basic understanding of materials engineering and materials science.

1.9 Background Requirements of the Failure Analyst: Scope of the Book

Since this text is designed to provide the necessary background to successfully conduct a failure analysis investigation, its scope parallels the background
requirements of the failure analyst. As pointed out earlier, the failure analyst must have an adequate knowledge of a number of engineering disciplines, particularly materials engineering and materials science. To develop this knowledge step by step, an attempt has been made to balance the various aspects of the two branches of materials and how they are integrated into the process of design-manufacturing-performance.

Chapter 2 emphasizes the continuity of the process of design-manufacturing-performance and its significance to failure analysis investigations. Included in this chapter are the various stages of engineering design as well as the fabrication and assembling techniques used to manufacture a product. Since the field of mechanics is of particular importance to design and failure analysis investigations, its basic principles are covered in Chap. 3. Using these principles, various mechanical properties are defined in Chap. 4 and the tests developed to measure them are described. Methods of nondestructive evaluation are also included in Chap. 4. Having defined these properties, two important fields of mechanics related to mechanical properties are then introduced, namely, (i) stress analysis and (ii) fracture mechanics. Stress analysis is the subject of Chap. 5. Macroscopic aspects of fracture and fracture mechanics are dealt with in Chap. 6.

After developing the above knowledge, the reader’s attention is directed toward the origin of engineering properties which lies in the internal structure of materials. Chapter 7 is devoted to the basics governing the structure of engineering materials. However, before the structure of materials can be correlated with their properties, it must be characterized. To cover this link, Chap. 8 is devoted to materials characterization. Following the introduction of the above two important aspects of materials science, the phenomenon of corrosion which is directly related to the structure of individual atoms is presented in Chap. 9, including low-temperature aqueous corrosion and high-temperature corrosion. Metallurgical aspects of fracture and fractography are the subjects of Chap. 10.

Knowledge developed in Chaps. 2–10 is subsequently used to demonstrate to the reader how to conduct a failure analysis investigation. Chapter 11 deals with the procedure of failure analysis. Selected case studies covering failures related to processing of materials, design, manufacturing, and service conditions are presented in Chap. 12.
2

Engineering Design–Fabrication–Performance

2.1 Introduction

It is recalled that any engineering product is designed and manufactured to perform specific functions serving certain human needs. Once the need for a given product is established, the design engineer conceives, plans, and carries out a program by means of which the product can be manufactured to meet its functional requirements over a predetermined service life. Regardless of the complexity of the product, its service performance is a function of many aspects of design and manufacturing, as well as the adherence of the user to design parameters during service and the maintenance procedure and schedule recommended by the manufacturer. As schematically illustrated in Fig. 2.1, the various aspects of engineering design, manufacturing, and performance are steps in a chain of a continuous process. All these steps must be correlated and tailored to deliver a reliable and durable product. Failure can be expected to occur at the weakest link of the chain process shown in Fig. 2.1, and it is the responsibility of the failure analyst to identify that link in order to be able to prevent future failures.

As can be concluded from the preceding discussion, the cause of premature failure of an engineering product can be traced back to either one or more of three main continuous stages of its history: (i) engineering design, (ii) manufacturing, and (iii) service conditions. Evidently, to achieve the primary objective of a failure analysis investigation cited above, the investigation must be thorough and

Copyright © 2004 by Marcel Dekker, Inc.
complete requiring knowledge of more than one engineering discipline. Failure analysis investigations can vary widely in the extent of their complexity. However, by a careful inspection of the failed component as well as gathering as much information as possible about its history, the analyst can rule out certain possibilities narrowing down the most probable causes of failure as demonstrated later. Another extremely important aspect of failure analysis investigations is determining responsibility for legal proceedings and claims. It is evident from the scheme of Fig. 2.1 that one or more of four main parties must be involved in any case of failure, namely, (i) the designer, (ii) the product manufacturer, (iii) the material manufacturer, and (iv) the end user. Determining exact responsibility requires the use of clear and well-defined technical terms in failure analysis reports as explained later.

It is the objective of this chapter to introduce the earliest stages of history of an engineering product, i.e., its design and manufacturing. Various stages of

Figure 2.1 Schematic illustration of the continuity and main steps of the process of engineering design–manufacturing–performance.
2.2 Stages of Engineering Design

It is recalled from Chap. 1 that engineering design is essentially a decision-making process whose objective is to develop a new product or improve the performance of an existing product. Typically, design begins with an idea or a concept developed to satisfy certain human needs. Before this concept is materialized into a hardware item, engineering design passes through distinct stages, which can be classified into (i) conceptual design, (ii) engineering development, and (iii) production design as shown in Fig. 2.1. After successfully completing these stages, the product becomes ready for large-scale manufacturing. A brief account for each design stage is given below.

Conceptual design is primarily concerned with the functional requirements of the product and how they can be satisfied. At this stage, several concepts or ideas which can satisfy these requirements are studied. Subsequently, a decision is made about the feasibility of the idea. Having determined that the idea is practical and sound, one or more of the conceptual designs are selected for further development. Also, a consideration is given to the availability of structural materials whose properties satisfy the design requirements.

Further modifications of the selected conceptual design(s) result in a fairly matured design during the stage of engineering development. Engineering drawings are developed to specify the various components, their dimensions, assembly procedure, as well as the structural materials to be used in manufacturing. This permits conversion of the design into a prototype model for testing and evaluation, which may call for further modification of the design to improve performance. It is important that all necessary changes particularly in structural materials must be made before the product enters into the production-design stage.

It is possible that the production-design stage reveals that further design changes must be made. Also, this stage requires that all structural materials used in manufacturing must be (i) fully specified, (ii) available in the required quantities, and (iii) readily processed by existing techniques.

Changes in product design after manufacturing progresses can be very costly. Therefore, the required changes including those related to structural materials must be completed during the production-design stage. However, development of new structural materials can justify changes in design after manufacturing if they result in cost reduction and improved performance.

One of the most important aspects of design is the lifetime of the product. Typically, the designer anticipates that failure must not occur before a predetermined lifetime. It is evident, however, that the lifetime of a given product can
vary significantly depending upon variables such as (i) the extent of analysis and
experimental development involved in the various stages of design, (ii) the qual-
ity of materials used to manufacture the product, (iii) manufacturing technology,
and (iv) exact knowledge of service conditions. Obviously, these variables have
a significant effect on the product cost. In practice, the designer compromises
between cost and a predetermined lifetime. However, it must be realized that
an initially lower cost may not be justified if the long-term maintenance and
repair costs as well as the cost of unscheduled shutdowns are considered.

It can be appreciated from the preceding brief discussion that engineering
design involves many decision-making processes. One of the most important of
these decisions is selection of structural materials to perform certain functions
determined by the objectives of the design. Primarily, the designer is concerned
with those qualities of the materials known as properties, which determine their
suitability in certain applications, as described below.

2.3 Material Selection

Both the reliability and durability of any product are largely dependent upon the
qualities of materials used in manufacturing. Material selection is therefore one
of the most important decisions made by design engineers. It is governed by a
number of factors, including (i) properties relevant to the application, (ii) manu-
facturing constraints, (iii) effect of manufacturing procedures on properties,
(iv) prior incidents of failure, (v) maintenance and repair frequency, and (vi) cost.

Proper selection of materials requires a detailed knowledge of all functional
requirements of the material. Depending upon the particular application, these
requirements may include tensile strength, ductility, creep strength, fatigue
strength, resistance to various corrosive environments, fabricability. All antici-
pated service conditions must be considered in preparing a list of functional
requirements. However, it is important to realize that properties derived from
laboratory tests are useful only to the extent to which service conditions duplicate
the test conditions. To a large extent, the manufacturing procedure is dictated by
the material selected for the application. For example, if the material has poor
formability, the part must be fabricated by other techniques such as casting.

When preparing a list of functional requirements for a given application, it
is important to distinguish between absolute and relative requirements. Once this
distinction is made, the list of candidate materials can be narrowed. Absolute
requirements are those which are not subject to compromise. For example, resist-
ance to high-temperature carburization attack as well as creep strength are absol-
ute requirements of a material to be used in constructing reaction tubes of
ethylene cracking process. Relative requirements are those which can tolerate
some compromise and may be satisfied to various extents by different materials.
In the case of the above example, relative requirements may include impact strength and microstructural stability.

Usually, structural materials are provided by the manufacturers in the mill-annealed or heat-treated condition with certain microstructure and corresponding properties. Such initial properties may be of little or no significance in determining actual performance under service conditions. Subsequent fabrication and heat treatment processes involved in manufacturing hardware items can significantly alter the initial properties. A consideration of property changes resulting from manufacturing procedures can be expected to significantly contribute to proper material selection.

Another important factor to be considered in material selection is prior experience related to service performance of candidate materials. Incidents of prior failure and results of failure analysis investigations can be most helpful in selecting the proper material for the application. Equally important is any prior experience concerning some difficulties or unusual service conditions related to a particular material such as a high frequency of maintenance and repair. A due consideration must also be given to the extent to which the material can be standardized. Furthermore, a broad and basic knowledge of the nature and properties of materials is an invaluable aid in proper selection of materials.

Cost is another important factor in material selection. More often, a designer selects a material providing the desired service life at the least possible cost. However, an initially lower cost may not always be justified if the maintenance, repair, and replacement costs as well as the cost of unscheduled shutdowns are considered. To achieve the most economical service life of a given component, it is important to have a better understanding of the factors governing the desired service life of the product and a knowledge of the performance characteristics of various structural materials.

### 2.4 Fabrication of Engineering Alloys

By definition, fabrication is the method or technique used to convert materials into parts and assemblies. Fabrication is closely related to design because a proper selection of materials cannot be made without taking into consideration the techniques required to manufacture the end product. Many of the technologically important properties of engineering alloys particularly mechanical properties are significantly influenced by the fabrication techniques.

As summarized in Fig. 2.1, it is possible to classify the fabrication techniques of engineering alloys into three main classes: (i) primary techniques, (ii) secondary techniques, and (iii) joining techniques. Primary fabrication techniques aim at producing finished or semifinished parts from two types of raw materials: (i) molten metal and (ii) metallic powder. Casting is the technique by which a finished part is directly produced from the molten metal. Wrought
products such as sheets, plates, bars, structural sections, tubes, and wires are produced by first casting the molten metal into an ingot followed by mechanical working techniques. Powder metallurgy is the technique by which a part is fabricated from a powder. Secondary fabrication techniques are the processes required to bring a fabricated part into specified limits of dimensional accuracy, desired surface finish, or a desirable shape. Joining methods are used to assemble parts into a desired component.

Most of the engineering alloys employed in design are used in two forms: (i) cast and (ii) wrought. As pointed out earlier, the material itself dictates the fabrication technique. Typically, the mechanical properties of wrought products are superior to those of cast products of the same alloy. In a few cases, however, a casting may have superior properties. It is therefore important that the designer takes this difference into consideration. Also, it is the responsibility of the design engineer to determine the compatibility between the material and fabrication techniques to achieve the functional requirements of the product. Before describing the various fabrication techniques, it is instructive to review some important related processes particularly solidification, cold working, and recrystallization.

2.5 Solidification of Ingots

Melting of raw materials of an engineering alloy is carried out to thoroughly mix its constituent elements. One of the most remarkable features of the internal structure of solid metallic materials, among others, is their crystalline nature. By crystalline is meant that the constituent atoms are arranged in certain geometrical patterns forming crystals. When a molten metal is poured into a cold mold (see Fig. 2.2a), the liquid metal near the mold wall is the first to solidify. In the liquid state, the atoms move rather freely; however, when heat is extracted from the liquid metal, a few atoms are grouped together to form very small crystals near the mold wall, as schematically illustrated in Fig. 2.2b. Such a process is called nucleation, and the very small crystals initially formed near the mold wall are called nuclei. Since each crystal is nucleated independent of the others, the orientation of the geometric arrangement of atoms differs from one crystal to another. With continued cooling, more atoms are attached to the initially formed nuclei causing them to grow in size, as shown in Fig. 2.2c, and therefore this process is referred to as growth; i.e., solidification consists of two stages: (i) nucleation and (ii) growth.

By virtue of the method of cooling, heat is extracted from the mold along directions perpendicular to its walls as well as along the radial direction. Consequently, the initially formed nuclei grow along the opposite directions of extracting heat toward the center of the mold forming long arms called dendrites, as shown in Fig. 2.2c. Growth of some crystals, however, at the bottom of the mold is obstructed by other crystals, and therefore they maintain a smaller size.
Eventually, the extraction of heat from the central portion of the mold becomes no longer directional particularly if the mold has a large cross section leading to the formation of rather equiaxed crystals at the center of the mold, as shown in Fig. 2.2d. Each crystal in Fig. 2.2d is called a grain, and because the orientation of one crystal differs from that of its neighbors, the continuity of the geometric arrangement of atoms is disrupted when two crystals come into contact. At the contact surface, which is called a grain boundary, the arrangement of the few atoms present deviates from that within the two grains (see Fig. 1.5).

In the solidified condition described above, the alloy is called as-cast or a casting product. Only, in a few cases, the grain structure of the casting (Fig. 2.2d) is visible to the naked eye. Frequently, a low-power microscope is required to reveal such a grain structure. Characterization of the internal structural features of engineering alloys is the subject of Chap. 8.
As explained in Chap. 7, the mechanical properties of engineering alloys are sensitive functions of their internal structure including the grain structure. A coarse, directional, and heterogeneous grain structure such as that shown in Fig. 2.2d leads to poor mechanical properties. Such undesirable features are called defects. Other undesirable defects in as-cast ingots include nonhomogeneous distribution of alloying elements and/or impurities along the ingot cross section. Sometimes, these elements tend to segregate at grain boundaries. Another related defect is called freckling, which is particularly encountered in alloys containing heavy elements such as W, Ta, and Nb. In this case, the heavy element may segregate to an extent where it causes the formation of another substance particularly at ingot surface.

Other inherent defects of solidification include porosity and shrinkage cavity or pipe. Porosity results from the fact that the solubility of gases in a liquid is higher than that in a solid. When gases are liberated from the liquid during solidification, they leave behind round voids or pores, which are sometimes called round blowholes. Shrinkage during solidification is another factor contributing to porosity; however, in this case, the pores are irregular rather than round. It becomes particularly significant near the top of the ingot forming what is known as a shrinkage cavity or pipe. During the final stages of solidification, the remaining liquid metal near the top of the ingot becomes insufficient to completely close the ingot resulting in a cavity assuming the shape of a cylinder or pipe.

To summarize the above discussion, as-cast ingots contain two main types of defects detrimental to mechanical properties: (i) coarse, directional, and rather heterogeneous grain structure and (ii) porosity and shrinkage cavities. Such defects can be eliminated to a large extent by thermomechanical processing involving mechanical working operations and recrystallization.

2.6 Cold Working

It is recalled from Chap. 1 that when a material is acted upon by a progressively increasing tensile force, it first elastically deforms, then it plastically deforms, and ultimately it fractures. Plastic deformation refers to a permanent change in shape or dimensions. As explained later, experiment shows that when a material acquires a given amount of plastic deformation, the force required to induce additional deformation continues to increase. Physically, this means that plastic deformation strengthens the material. Such a technologically important phenomenon is called strain hardening.

Severe plastic deformation such as that encountered during rolling causes the grains of the material to be elongated along the rolling direction as shown in Fig. 2.3. Such a grain structure is maintained and strain hardening persists as long as the material is plastically deformed below a certain temperature as explained in the next section. Under these conditions, the material is said to be cold-worked.

Copyright © 2004 by Marcel Dekker, Inc.
Cold work has the effect of increasing the strength and hardness of the material and reducing its ductility.

### 2.7 Recrystallization

Recrystallization is a process occurring in the solid state whereby a new set of grains relatively free of strain hardening are formed by heating a material which has been previously cold-worked as schematically illustrated in Fig. 2.4. It involves rearrangement of atoms within the old deformed grains to form new grains while the permanent change in shape or dimensions produced by cold work is maintained. During the history of any polycrystalline engineering alloy, recrystallization occurs at one stage or another.

![Figure 2.3](image1.png)

**Figure 2.3** Plastic deformation by rolling.

![Figure 2.4](image2.png)

**Figure 2.4** Schematics illustrating the process of recrystallization. (a) Cold-worked elongated grains. (b) Nucleation of new grains. (c) Growth of new grains. (d) Recrystallized grain structure.
A direct correlation exists between the recrystallization temperature and melting point. Materials of higher melting point are recrystallized at higher temperatures relative to materials of lower melting points. Such a correlation stems from the fact that rearrangement of atoms during recrystallization occurs by a process called diffusion whereby atoms are transported from one position to another under the influence of heat. Also, for a given material, the recrystallization temperature is influenced by the amount of cold work. Increasing the amount of cold work reduces the recrystallization temperature and vice versa.

As explained later, the mechanical properties of a given material are sensitive functions of its grain size. Decreasing the grain size increases the tensile yield strength; however, it reduces the elevated temperature creep strength. Similar to the solidification process described earlier, recrystallization involves two stages: (i) nucleation of new grains and (ii) growth of those grains. Grain size produced by recrystallization is dependent upon a number of variables, particularly the recrystallization temperature, exposure time at that temperature, and the amount of previous cold work. For a given material and certain amount of cold work, there is a minimum recrystallization temperature. If the material is heated to that temperature, the initially formed grains are relatively small. Grain growth occurs with continued exposure at that temperature. Comparatively larger grains are formed if the material is heated to a temperature above the minimum recrystallization temperature. Conversely, for a given recrystallization temperature and time, the grain size is decreased by increasing the amount of cold work.

In opposition to cold work, recrystallization has the effect of reducing the strength and hardness. An equivalent statement is that recrystallization softens the material. Having defined the process of recrystallization, it is possible to redefine cold work as any plastic deformation process carried out at a temperature below the respective recrystallization temperature.

### 2.8 Thermomechanical Processing

Thermomechanical processing refers to operations carried out on as-cast ingots. Its primary objective is to (i) change the undesirable grain structure of the as-cast ingot into a more uniform structure by the process of recrystallization, and (ii) simultaneously change the shape of the ingot into more useful shapes by the process of plastic deformation producing finished or semifinished products. Furthermore, the combined effect of heat and mechanical force eliminates most of the pores and shrinkage cavities produced by solidification.

All mechanical working operations are essentially plastic deformation processes, which can be classified into (i) hot-working operations and (ii) cold-working operations. Hot-working operations are primarily used to improve the structure and properties of as-cast ingots; however, they can also be used to produce finished or semifinished parts. Unlike cold working described earlier,
hot-working operations are carried out above the recrystallization temperature. In contrast with cold-worked parts, hot-worked parts are not strain hardened. Conversely, the objective of all cold-working operations is to produce finished or semifinished products. Usually, hot-working operations used to improve the structure and properties of as-cast ingots are called primary operations, and those operations including cold working are referred to as secondary. However, to avoid confusion with the definition of primary and secondary operations cited earlier, primary and secondary mechanical working operations are referred to in this text as initial and final operations, respectively.

It important to note that cold working does not necessarily imply that plastic deformation is carried out at room temperature. Even if the deformation is carried out at a temperature higher than room temperature but below the recrystallization temperature, it is by definition a cold-working operation. Designers must be aware of the fact that cold-worked parts are hard and less ductile in comparison with hot-worked parts.

2.9 Primary Fabrication Techniques

2.9.1 Casting

Casting is the most direct technique used to produce a finished part. In certain applications, casting is the most economical as well as the only feasible technique, e.g., gas turbine blades. Briefly, casting consists of introducing a liquid material into a mold of a desirable shape where it is allowed to solidify according to a predetermined time and temperature cycle. When solidification is completed, the cast part is removed by opening or destroying the mold. Similar to the case of as-cast ingots described earlier, cast products contain defects such as porosity, shrinkage cavities, and freckles. However, the extent of these defects can be controlled by selecting the proper casting method as described below.

Various casting methods are distinguished by the material of the mold, which can either be (i) sand (sand casting), (ii) ceramic (precision investment casting), or (iii) metal (die casting). Sand casting is the oldest method, which is particularly useful in casting high-melting alloys because the major component of sand (silica: SiO₂) has a higher melting point than the metallic material being cast. Typically, the sand mold is made using a wood pattern of the part to be cast. After removal of the wood pattern, the liquid metal is poured into the mold cavity. In addition to being inexpensive, sand casting is the only practical method for large components and for parts required in small quantities. Figure 2.5 illustrates the configuration of a sand mold for metal casting.

Precision investment casting is particularly suitable for the production of small parts with extremely close dimensional tolerance. In this method, a pattern of the part is made of wax, and the casting ceramic mold is either sprayed onto the
wax pattern or the latter is dipped into a slurry of the ceramic mold. By firing the wax-ceramic assembly at high temperature, the wax pattern is melted away, and the ceramic mold is hardened. Subsequently, the liquid metal is poured into the ceramic cavity. Varieties of this method are used to fabricate products with desirable structures, particularly directionally solidified parts, as well as single-crystal parts as schematically illustrated in Fig. 2.6.

Figure 2.5 A schematic illustrating the sand mold used in metal casting.

Figure 2.6 A schematic illustrating the process of producing directionally solidified and single crystal castings.
Die casting relies upon permanent molds made of metallic materials. Although this method produces parts of closer dimensional tolerance as well as of reduced shrinkage cavities and porosity, it is limited to relatively low-melting materials.

2.9.2 Hot Mechanical Working Operations

As explained earlier, initial hot-working operations are primarily carried out on as-cast ingots to produce more uniform grain structures and eliminate most of the defects created by the solidification process. First, the ingot is heated in a furnace to the desired hot-working temperature and is then plastically deformed in stages to achieve the required thickness. During hot working, the ingot is gradually cooled, and therefore it is important to preheat it before it reaches the temperature where work hardening can occur leading to embrittlement. It is possible to classify these operations into three types: (i) forging, (ii) rolling, and (iii) extrusion. Some of these operations, however, particularly forging, can be used to produce finished or semifinished parts. Each operation is distinguished by the mechanism of plastically deforming the as-cast ingot, as schematically illustrated in Fig. 2.7.

Forging, where the ingot is deformed either by pressing or hammering, is considered to be the most versatile mechanical working operation. Two forging techniques are distinguished by the type of dies used in deformation: (i) open-die forging and (ii) closed- or impression-die forging.

Open-die forging is used to process large parts where an ingot or a work piece is plastically deformed between two simple flat dies. Upsetting of cylindrical shapes is a basic operation of open die forging. As schematically illustrated in Fig. 2.8a, the dies are made to be in contact with the work piece over only a part of its surface. This restricts deformation to that part, which consists of simple compression accompanied by considerable spreading in the lateral direction, as shown in the Fig. 2.8a. Open-die forging is used to improve the structure of as-cast ingots, as well as to produce large parts of simple shapes such as pressure vessels, shafts, and rings.

Closed- or impression-die forging is used to produce finished or semifinished small parts of more complicated shapes. In this case, the ingot or work piece is first prepared by open-die forging, and then is deformed between two dies resembling casting molds in that they contain the impressions of the desired final shape, as schematically illustrated in Fig. 2.8b. Deformation pattern in closed-die forging is more complicated than that in open-die forging.

Forging operations can leave behind a number of defects in the work piece. For example, forging by hammering may limit plastic deformation only to surface layers of as-cast ingots. As a result, the interior of the ingot retains the heterogeneous grain structure characteristic of the as-cast condition.
must be carried out by pressing rather than by hammering. If the work piece is too cold, surface cracking can result from excessive deformation. Internal cracks can also form particularly during upsetting of cylindrical shapes. Other defects include discontinuities resulting from incomplete welding of two surfaces folding against each other during plastic deformation and forging-in of loose scale and lubricant residue on the surface of the part. By proper forging practice, however, it is possible to control the extent of these defects.

Characteristically, forging deformation results in some directionality in the grain structure where the grains and inclusions are oriented parallel to the direction of greatest deformation. This develops what is known as a fiber texture, which is typical of all forged components. Although fiber texture is not considered as a defect, it degrades the mechanical properties of forged components in the transverse direction. Usually, the extent of this inhomogeneity in properties can be controlled by adjusting the amount of forging deformation.

Figure 2.7  Schematics illustrating the main types of mechanical working operations. (a) Rolling. (b) Forging. (c) Extrusion.
Hot extrusion is almost always used to refine the grain structure of as cast ingots. However, it can also be used to produce complex shapes, either solid or hollow. In this technique, the work piece is introduced into a cylindrical container where it is forced to flow through a die orifice. Although the deformation pattern characteristic of extrusion is particularly suitable for eliminating solidification defects, extrusion leaves behind other defects. Because of the nature of plastic deformation during extrusion, inclusions in the workpiece tend to form a ring extending over the rear end of the extruded part. Surface cracking, peeling, and blistering are other extrusion defects.

Rolling is a fabrication technique primarily used to reduce the cross-sectional area of a part. Hot rolling as an initial working operation is used to refine the grain structure of as-cast ingots. Also, it flattens the grains and inclusions, and elongates them in the direction of rolling, resulting in a fiber texture which leads to different properties in the longitudinal (rolling direction) and transverse directions. Blowholes resulting from solidification are either reduced in size or completely closed by hot rolling. In addition, hot rolling is used to produce certain parts such as hot-rolled plates, bars, structural sections, and tubes. Cladding is a special technique of hot rolling where two dissimilar metallic materials are welded together in the solid state by pressure. Some of the typical rolling defects include irregularity in part shape, cracking, and surface defects such as blistering, scratches, and rolled-in dirt.
2.9.3 Cold Mechanical Working Operations

Cold mechanical working operations are strictly used to produce finished or semi-finished parts. Dimensional accuracy and better surface quality as well as improved properties can be achieved by these operations. Usually, the feedstock is a part processed by hot working. Cold rolling and cold extrusion are the most widely used operations; however, cold forging is used to a lesser extent in the production of certain parts.

Cold rolling is primarily used to produce sheet and strip products. Usually, the feedstock is a hot-rolled plate or strip. However, the feedstock for cold rolling can also be as-cast, extruded, or forged part. Cold extrusion is used to produce such parts as seamless pipes and tubing. A closely related technique is that of drawing. It only differs from extrusion in that the workpiece is drawn rather than pushed through the die. Drawing is used to produce rods, wires, and tubes.

2.9.4 Powder Metallurgy

Primary fabrication by the powder metallurgy technique is particularly suitable for metallic components of high melting points such as tungsten, molybdenum, tantalum, and niobium. Also, for some alloys where the component metals are insoluble in the solid state and have limited solubility in the liquid state, the powder technique can be the only feasible fabrication method. Parts produced by this technique are usually porous. Although porosity degrades mechanical properties, it can be useful in certain applications such as bearings. Other advantages of this technique include the economy and ease of fabrication of certain parts, e.g., gears, levers, and cams.

Fabrication by the powder metallurgy technique consists of (i) powder preparation, (ii) mixing, (iii) pressing, compacting, or briquiting, and (iv) sintering. Techniques used to prepare metallic powders include reduction of oxides, vapor condensation, atomizing of liquids by means of gas jets, and electrolytic deposition. Usually, more than one metallic constituent is used to prepare the powder. Therefore, the prepared powder must be thoroughly mixed to ensure homogeneity of the final product. Lubricants are usually added during this stage to facilitate subsequent pressing. Hydraulic or mechanical presses equipped with proper dies are used to compact metallic powders into the desired shape. This is finally followed by sintering, which consists of heating to a high temperature below the melting point, allowing the constituent atoms to rearrange themselves by diffusion and form the final solid product.

A typical defect of parts produced by powder metallurgy is porosity, which degrades mechanical properties. Extent of porosity is dependent upon a number of variables, particularly the compacting pressure, sintering temperature, and time, as well as the powder properties such as grain size, shape, and ductility.
Repressing and resintering are usually employed to minimize the extent of porosity. Infiltration of higher melting powder particles with lower melting powders is another method used to control porosity. During sintering, the low-melting powder melts and fills the voids between particles of the higher-melting powder.

2.10 Secondary Fabrication Techniques

It is the primary objective of secondary fabrication techniques to achieve the required dimensional accuracy and most desired shape of a product. Since these techniques are usually expensive, which can result in a substantial increase in cost, the designer always attempts to minimize the need for secondary fabrication. Therefore, accurate dimensional tolerance and surface finish are only limited to those products where they are absolutely needed.

Any secondary fabrication technique is based upon controlled removal of material from the part surface in the form of chips, which is defined as machining. In general, machining techniques can be classified into two main types: (i) mechanical and (ii) nonmechanical.

2.10.1 Mechanical Machining

Mechanical machining is the conventional and most universally applied technique of secondary fabrication. Basically, it consists of moving the part in a machine relative to a cutting tool, which removes thin surface layers from the part in the form of chips, as schematically illustrated in Fig. 2.9. Depending upon the type of this relative movement, different mechanical machining methods are distinguished, e.g., grinding, polishing, planing, broaching, turning, milling, drilling, sawing, and boring. For example, in turning, the part rotates while the cutting tool moves in a straight line which can either be parallel, perpendicular, or inclined to the surface of the part. In contrast, if the part is machined by drilling, it is kept stationary while the cutting tool rotates along an axis perpendicular to the surface of the part.

![Figure 2.9 A schematic illustrating the mechanical machining process.](image-url)
Since machining processes can have significant effects on properties of the finished part as well as its functional requirements, it is important that in designing a product, consideration must be given to the role of machining. Both the design itself and the material selected for the application determine the technical and economical aspects of machining.

In any mechanical machining technique, the cutting tool must satisfy two requirements. First, it must be harder than the part to be machined. Second, it must maintain a sharp cutting edge. Evidently, the mechanical strength of the part limits the choice of cutting tools. However, even if the cutting tool is adequate, the excessive heat generated at the point of contact between the part and cutting tool can degrade its strength. Ductile materials can also present machining difficulties.

To overcome these problems, some of the engineering alloys, both ferrous and nonferrous, are produced in grades known as free machining. For such grades, certain alloying elements are added to form inclusions, which greatly facilitate the machining process by breaking the chips into small pieces. Sulfur and phosphorus are among the alloying elements used in free-machining grades of carbon steels. Free-machining grades of stainless steels are produced by the addition of selenium and tellurium. Load in another alloying element commonly used to produce free-machining grades of ferrous and nonferrous alloys. It is important to realize that alloying elements used to produce free-machining grades of engineering alloys can degrade other properties, and therefore the designer must be aware of this fact when selecting free-machining alloys for certain applications.

Other important effects produced by mechanical machining include (i) localized surface deformation depending upon the load exerted by the cutting tool, (ii) localized melting, which may result from excessively high cutting rates, blunt cutting tool edge, or improper cooling, and (iii) structural changes in surface layers resulting from excessive heating. All these machining effects can degrade the properties of the part and alter its performance capabilities and, therefore, must be taken into consideration by the designer.

It is evident from the above discussion that machinability is an important property of engineering alloys. Unfortunately, however, it is one of the most difficult properties to clearly define. In practice, machinability is measured with reference to a free-machining alloy grade, which is assigned an arbitrary machinability value of 100. Machinability of other alloys is then rated in comparison with the reference alloy.

2.10.2 Nonmechanical Machining

Although mechanical machining is considered to be the most important secondary fabrication technique, it cannot be applied to parts of complex configurations.
To overcome this limitation, a number of nonmechanical machining techniques have been developed. They include (i) electrical discharge machining (EDM), (ii) electrochemical machining (ECM), (iii) electrochemical grinding (ECG), and (iv) radiant-energy material-removal methods.

Electrical discharge is the most widely used nonmechanical machining technique. It is based upon using a low-voltage, high-frequency direct current to remove thin layers of metallic materials by means of an electrode usually made of brass such that its cross-sectional contours are identical to the cavity to be machined. As schematically illustrated in Fig. 2.10, the electrode and part to be machined are immersed in a dielectric fluid, which must be capable of (i) inhibiting the continuous flow of electric current, (ii) flushing away the metal removed from the part, and (iii) cooling the part. Electric current produced by the resistance-capacitance circuit shown in Fig. 2.10 generates a spark discharge between the electrode and the part removing metal in the form of small segments. Machining by EDM has two important advantages over mechanical machining. First, machinability is not limited by the mechanical strength of the part. Second, irregular and intricate surface contours can easily be obtained because of the dependence of cavity shape only on the electrode configuration, contributing to more accurate dimensional tolerance. One of the disadvantages of EDM is the associated lower fatigue strength in comparison with mechanically machined parts, which contain a work-hardened surface layer. Under extreme conditions, EDM may lead to cracking along grain boundaries. Also, property changes of the part may result from alloying with the electrode material. Further changes in properties of particularly ferrous parts can result from carbon transfer from the hydrocarbon-based dielectric fluid into the part.

Both electrochemical machining and grinding techniques are related to electrical discharge machining. Electrochemical machining (ECM) consists of
passing a direct electric current through a conductive solution, separating the surface of the part from an electrode. In this case, the part acts as an anode, and the electrode behaves as a cathode. Surface layers from the part to be machined are removed by anodic dissolution.

Electrochemical grinding (ECG) is essentially the same as ECM; however, in this case, the electrode is a grinding wheel whose function is twofold. First, as a cathode of an electric circuit, it causes removal of surface layers from the part by anodic dissolution. Second, as a grinding wheel, it contributes to metal removal by grinding. Since the ECM and ECG techniques employ corrosive fluids, the part must be thoroughly cleaned after machining to avoid corrosion damage.

Radiant-energy machining techniques rely upon the use of a high energy, which can either be a laser beam or an electron beam. Since the energy source can be locally focused on a very small surface area of the part, the associated high-energy density can instantly remove surface layers by vaporization. Therefore, the technique is particularly useful for parts of high melting points. Also, it is possible to precisely focus the energy source on selected regions of the part, and therefore remote or otherwise inaccessible regions of the part can readily be machined. It is recalled that mechanical properties of the part are decisive in mechanical machining. In contrast, thermal properties become decisive in machining by radiant energy because this energy is converted into heat at the surface of the part. Relevant thermal properties include the heat of vaporization and thermal conductivity. Evidently, the energy required for machining increases with the heat of vaporization and the thermal conductivity.

2.11 Joining Techniques

Parts produced by the primary and secondary fabrication techniques described above can be assembled into components by a wide variety of joining techniques, which can be classified into (i) welding, (ii) brazing, (iii) soldering, and (iv) riveting. An account for each technique is given below.

2.11.1 Welding Techniques

Welding is by far the most common and important technique for assembling parts into components. Usually, the strongest joint between two parts is produced by welding. However, the actual joint strength is a function of two main variables: (i) the exact welding technique and (ii) the nature of materials to be joined. Improper welding practice is one possible cause of failure during service.

Although there are several welding techniques, all can be classified into two main types: (i) fusion welding where only heat is applied to the surface of parts to be joined and (ii) pressure welding where both heat and pressure are
applied to the surface parts to be joined. Most of the welding techniques used in engineering applications, however, are of the fusion type.

In fusion welding techniques, the metal in the immediate vicinity of the weld is heated to above its melting point. Therefore, the weld joint results from mixing of molten surface layers of the parts to be joined and their subsequent solidification. When a filler metal is used, it is also melted and intermixed with molten layers of the parts. Filler metals must have the same or similar composition to the base metal in order to produce a sound weld joint. Each welding technique is distinguished by the type of heat source used for fusion. It can either be an electric arc, a flame of gas combustion, or the radiant energy of an intense electron beam or laser beam.

In contrast with fusion welding, the heat applied to the surface of parts to be joined by pressure welding does not result in melting. Instead, the temperature becomes sufficiently high to permit diffusion of atoms across the two surfaces to establish the required intimate contact, which is aided by applying a pressure. Although melting occurs in some pressure welding processes, it is confined to much smaller regions in comparison with fusion welding.

Generally, all welding techniques involve the rapid heating of a localized small mass of metal to a very high temperature followed by cooling to ambient temperature. Such a rapid temperature cycling has two important effects on the parts to be welded. First, it results in changes in the internal structure of the weld joint as well as in the region of base metal adjacent to the joint, which is called the weld heat-affected zone. Second, it results in thermal expansion and subsequent contraction of the heated regions.

For fusion welding and pressure welding which involve melting, the structural changes occurring in the weld joint resemble those resulting from solidification. In the as-welded condition, dendritic structures (see Fig. 2.2) are observed in the weld joint. Also, the weld joint retains such defects as porosity and shrinkage cavities. Structural changes similar to those resulting from exposure to elevated temperatures (thermal aging or heat treatment) are observed in the weld heat-affected zone shown in Fig. 2.11. It is evident that these are the only changes encountered in nonmelting pressure techniques of welding. Since

![Figure 2.11](image)

**Figure 2.11** A schematic illustration of a weld fusion zone and heat-affected zone.
the largest mass of metal remains relatively cold during welding, it cannot permit sufficient movement to accommodate the localized shrinkage resulting from the heating-cooling cycles, generating internal stresses. However, these stresses can be relieved by plastic deformation if the material is sufficiently ductile. Conversely, if the material lacks the sufficient ductility, the buildup of these stresses can eventually lead to cracking.

It can be concluded from the above discussion that one of the most important aspects of selecting materials for welded components is the retention of sufficient ductility during and immediately after the heating-cooling cycle. An equivalent statement is that the material should not be embrittled by melting and subsequent rapid cooling, making material selection for fabrication by welding rather critical. Other difficulties are encountered in welding of cold-worked parts and materials which can readily be oxidized. Because of the heating effect, welding of cold-worked parts causes localized loss of strength resulting in inhomogeneous mechanical properties. Oxidation of certain materials during welding, e.g., Al, Ti, and their alloys, can cause severe embrittlement of the welded component; however, this problem can be controlled by welding in protective atmospheres. Furthermore, welding is generally not suitable for joining dissimilar metals because of differences in melting point, oxidation, and thermal expansion characteristics.

Some of the undesirable structural changes in materials resulting from welding can be eliminated by subsequent heat treatment. However, this can be rather difficult if not impossible for large structures, imposing severe limitations on the selection of materials for fabrication by welding. Frequently, it is desired to use the component in the as-welded condition. Therefore, it is important that the designer selects a material whose as-welded properties are compatible with the functional requirements of the product.

2.11.2 Brazing Technique

Brazing is a joining technique similar to welding in that it uses a filler metal. However, it differs from welding in that the filler metal is always different from that of the base metal and also it has a significantly lower melting point. Commonly, the filler metal is referred to as the brazing metal. A brazed joint results from applying the molten filler metal to the surfaces of the parts to be joined. Since the filler metal has a significantly lower melting point, the surfaces of the parts to be joined are not melted, unlike the case of fusion welding. During brazing, only limited diffusion occurs between the brazing metal and the base metal, resulting in a joint of different structure and composition than the base metal and in turn of lower strength than weld joints. However, an adequate strength satisfying the functional requirements of brazed components can
be obtained. Also, joining by brazing overcomes some of welding limitations, particularly the joining of dissimilar metals.

To achieve optimum mechanical properties of brazed components, the clearance between the surfaces of the parts must be small requiring a considerable overlap between the two parts. Aluminum, copper, and silver are commonly used as brazing metals. Ferrous alloys are usually brazed by copper. Also, brazing by copper is particularly useful for joining alloys of high strength and melting point. Joint quality is considerably improved by brazing in a reducing atmosphere to eliminate surface oxide layers, as well as to ensure complete wettability of the mating surfaces by the brazing metal. Brazing by silver is commonly used for some nonferrous alloys, particularly those of relatively low melting point. However, the lower melting point of aluminum alloys increases the difficulty of brazing. Usually, these alloys are brazed by aluminum alloys. Because of the tendency of aluminum to form a persistent surface oxide layer inhibiting the formation of a diffusion bond, special techniques such as ultrasonic vibration are used to break up the oxide layer. Furthermore, the lower recrystallization temperature of aluminum alloys leads to loss of mechanical strength requiring a postbrazing heat treatment.

2.11.3 Soldering Technique

In principle, soldering is very similar to brazing. However, the melting point of the filler metal used in soldering is significantly lower than that used in brazing. Frequently, the joining method is classified as brazing if the melting point of the filler metal is above 500°C, soldering if the melting point is significantly below 500°C. Commonly used filler metals in soldering are tin, lead, antimony, and their alloys. Tin-silver alloys containing up to 5% Ag with higher melting point are also used. Characteristically, the strength of a soldered joint is rather poor. Therefore, the technique is used only when mechanical strength is of a minor importance.

2.11.4 Joining by Adhesive Bonding

Various types of adhesives can be used to form a bond of significant strength between two parts. Adhesives are nonmetallic materials such as epoxy resins and elastomers. Parts joined by adhesives such as thrust bearings can be safely stressed up to the yield point of the base metals.

To join the parts, the mating surfaces are sprayed, brushed, or dipped in a liquid solution of the adhesive, and then the parts are brought into intimate contact. Subsequently, the adhesive material is allowed to solidify by drying, cooling, or by a process called polymerization depending upon the type of adhesive. Various theories have been advanced to explain the origin of bond established by adhesives, e.g., mechanical interlocking, electrostatic attraction,
diffusion, and valence bonding. Adhesives are selected for a given application on the basis of three parameters: (i) mechanical forces applied to the joint, (ii) service environment, and (iii) fabrication requirements.

2.11.5 Joining by Riveting

Joining by riveting differs from the above techniques in that neither a physical bond nor a superficial surface adhesion is established between the two parts to be joined. Instead, the parts are fastened together by means of rivets. Mostly, riveting is used to fasten flat parts such as sheets and plates.

Rivets used to assemble parts are cylindrical rods with one preformed head, which can be hemispherical or conical with flat end. Preformed head of the rivet fits into a bore of one of the parts to be joined. Subsequently, the other head is either cold- or hot-formed into the same shape as the preformed head using a die set or forming tool to complete the joining of the two parts. Various configurations of riveted joints are shown in Fig. 2.12.

2.12 Service Performance

As pointed out earlier, design-fabrication-performance can be visualized as a chain of a continuous process. Such a chain can be expected to fall apart at its weakest link, leading to failure of the product. It is evident that the service performance of any product is a sensitive function of design and fabrication. Both the configuration of the part (shape and dimensions) and the quality of structural materials used in manufacturing are the most important aspects of design influencing the service performance. However, configuration and material quality
are necessary but not sufficient conditions for optimum service performance. Consider, for example, a product designed to withstand fatigue loading conditions during service. Further assume that the design is optimized in terms of geometry and dimensions as well as material selection. Even if the material selected for the application is known to have the best fatigue strength, catastrophic failure during service can occur if flaws or defects are introduced during any of the fabrication processes described earlier.

Extending the above discussion one step further, it can be stated that geometry and dimensions, material quality, and fabrication are all necessary but not sufficient conditions to realize optimum service performance. Even if all these parameters are optimized, catastrophic failure can still occur if the service conditions deviate from the design parameters. Therefore, service conditions are as important as design and fabrication in determining the performance and life of the product.

Generally, it is the responsibility of the operator or user of the product to ensure that the service conditions are compatible with design specifications, e.g., operating temperature, applied forces, environment, and start-up and shutdown procedures. Also, it is equally important that the user adheres to the maintenance procedure and schedule recommended by the manufacturer. Any deviation from such practice during service can lead to failure. In certain instances, however, actual service conditions may deviate from those anticipated during the stage of design. Therefore, the designer must be fully aware of the actual service conditions.

2.13 Common Causes of Failure

Having developed a basic understanding of the continuous process of design-fabrication-performance in the preceding sections, it becomes possible to identify three main common causes of failure: (i) design-related failures, (ii) fabrication-related failures, and (iii) failures related to service conditions. Identifying the cause of failure is not only important in terms of preventing future failures but also in determining responsibility during the course of legal proceedings and claims.

There are two sources of design-related failures: (i) improper configuration of the product and (ii) selection of wrong material. Most of the failures related to improper configuration originate from highly localized stresses, as explained in Chap. 5. Selection of wrong material can result from use of nonrepresentative property data. As pointed out earlier, property data generated in the laboratory must be viewed as useful only to the extent to which service conditions duplicate the test conditions. Generally, materials must be tested under actual service conditions to determine their suitability for a particular application. In this regard, field testing of materials is very useful. Also, it is extremely important that the
designer be aware that fabrication techniques can significantly alter the initial properties of the material.

Fabrication-related failures can result from improper practice on the part of the manufacturer. It is essential that the fabrication technique recommended by the designer must be compatible with the material. For example, a proper material may be selected on the basis of its mechanical strength or environmental resistance. However, if the material has poor weldability, failure is likely to occur, if the product is assembled by welding. As discussed earlier, all fabrication techniques (primary, secondary, and joining) introduce defects into the material. Such defects can lead to failure depending upon their severity and extent. Other than defects, improper heat treatment is another possible source of fabrication-related failure. However, since fabrication is the direct responsibility of the manufacturer, every effort must be made to distinguish between a failure resulting from selecting a fabrication technique incompatible with the material (designer) or improper fabrication practice (manufacturer).

Common causes of failure related to improper service conditions include (i) not following the design parameters during service, (ii) inadequate maintenance, (iii) improper start-up and shutdown procedure, and (iv) unanticipated service conditions. To prevent the incident of failure, it is important on the part of the operator or user of the product to strictly follow the design parameters during service, such as operating temperature, applied forces, and environment. Also, it is equally important to regularly maintain the product according to the procedure and schedule recommended by the manufacturer including the use of proper spare parts. Frequently, during the start-up and shutdowns, certain products may be subject to severe conditions not encountered during normal operation. Therefore, the user must strictly follow the manufacturer recommendations regarding start-up and shutdown procedures. Unanticipated service conditions contributing to failure can result from (i) unawareness of the designer of the actual service conditions at the time of design and (ii) changing some of the process parameters on the part of the user without giving a full consideration to the corresponding effects.

It is recalled from earlier discussions that the physical, mechanical, chemical, and fabrication properties of materials are those qualities which qualify them for use in manufacturing. However, the primary concern of the designer is that the product can safely withstand the applied forces or loads anticipated during service, and therefore mechanical properties are of particular importance in design. Specifically, the designer analyzes the forces generated in the product during service in response to certain applied forces. Subsequently, the configuration of the product and the structural material used in manufacturing are tailored to safely meet the requirements of service conditions. This falls under the branch of physical science called mechanics, which is the subject of Chap. 3.
3

Principles of Mechanics

3.1 Introduction

Mechanics is a branch of physical science dealing with the macroscopic behavior of matter under the influence of a mechanical force. By macroscopic is meant that matter is viewed on a continuum basis with no regard to its constituent atoms or molecules. Generally, mechanics is divided into two main fields: (i) statics and (ii) dynamics. As its name implies, statics is concerned with the cases where no motion is produced by the applied force. Conversely, dynamics deals with motion under the influence of force.

In studying the behavior of a body under the influence of an applied force, two main cases are distinguished: (i) rigid bodies and (ii) extensible or deformable bodies. A rigid body is assumed to be nondeformable under the influence of applied force; i.e., the distance between two points prescribed within the body remains unchanged after applying the force. Important principles of mechanics are developed on the basis of rigid bodies. In practice, however, a body undergoes a change in shape or dimensions, or shortly is deformed in response to the applied force. To study the various mechanical properties of materials, the principles developed to treat rigid bodies in this chapter are extended to cover the mechanics of deformable bodies in Chap. 5.

Other specialized fields of mechanics particularly stress analysis and fracture mechanics are of extreme importance to engineering design and failure analysis. It is the primary objective of stress analysis (Chap. 5) to evaluate
the stresses developed in an engineering product by applying the principles of
mechanics. Knowledge of these stresses allows the designer to select proper
materials permitting safe operation of the product. Depending on the particular
case of failure, it may be required to evaluate the stresses developed in the part
during service and their path to determine the sequence of events leading to
the failure. Fracture mechanics (Chap. 6) aims at identifying the materials prop-
ties that can be used to predict fracture providing another aid to designing
safe products. Application of fracture mechanics to failure analysis investigations
permits the analyst to determine the stress at the point of fracture and its relation-
ship to the design stress.

3.2 Concepts of Mechanics

All branches of physical science, including mechanics, deal primarily with
concepts. Some of these concepts can be easily recognized, and their physical
meanings are well appreciated because of their daily importance to humans.
Such concepts are accepted as being most basic or fundamental; they cannot
be described in terms of anything simpler and therefore are called basic concepts.
Other concepts, however, are rather abstract and their physical meaning can only
be understood by precise definition or derivation from the basic concepts; they
are referred to as derived concepts.

All aspects of mechanics can be described in terms of three basic concepts,
namely, (i) mass \( m \), (ii) length \( L \), and (iii) time \( t \). Any more basic explanation of
these concepts is not possible. Derived concepts are defined in terms of the basic
concepts by the use of mathematical language illustrated by the following
examples.

Area \( A \) is defined as length \( L \) multiplied by length \( L \), which is mathemat-
ically expressed as \( A = L^2 \). Volume \( V \) is defined as area multiplied by length
or length \( L \times L \times L \), i.e., \( V = L^3 \). Velocity \( v \) is also a derived concept;
it is a distance (length) divided by time, i.e., \( v = L/t \). A derived concept related
to velocity is that of acceleration \( a \) defined as velocity divided by time, i.e.,
\( a = v/t = (L/t)/t = L/t^2 \). Another derived concept is momentum; it is defined
as the product of mass \( m \) and velocity \( v \). A system in mechanics is completely
defined if the position and velocity of each mass point are given requiring a
very large number of variables to define a system.

Derived concepts can be divided into (i) scalar quantities and (ii) vector
quantities. If only a magnitude is required to define a concept, it is said to be a sca-
lar, e.g., length, area, and volume. Both a magnitude and direction are required to
define a vector quantity. For example, to completely define the velocity of a mov-
ing object, it is required to specify both its magnitude (speed) and direction. Once
units of measurements are assigned to the basic concepts of mechanics (mass,
length, and time), the units of all derived concepts are defined as summarized in
Appendix B. Mechanics is essentially based upon the derived concept of force described below.

### 3.3 Concept of Mechanical Force

Force is a derived concept, which has both a magnitude and direction; i.e., force is a vector quantity. It is recognized as an external agency in the form of a pull or push action that tends to change the state of motion of a body; i.e., force and motion are so completely interrelated that neither can be defined independent of the other. Both the terms force and load are used interchangeably. However, a force is usually regarded as the effect of a load. Interrelationship between force and motion is expressed in terms of three laws of nature known as Newton’s laws of motion. Similar to basic concepts, such laws must be accepted as most basic or fundamental, which cannot be derived from simpler concepts. Newton’s laws of motion can be stated as follows.

1. Any body at rest or in a state of uniform motion must remain in that state forever unless a resultant force acts on it to change its state of motion.
2. When a body is acted upon by a force, it acquires an acceleration in the direction of the force whose magnitude is proportional to the force.
3. For every force, there is an equal and opposite reaction force.

According to Newton’s second law, if the mass of a body is $m$, the force $F$ required to produce an acceleration $a$ is given by

$$F = Cma$$

where $C$ is a constant. However, by proper choice of the units of force and mass, the constant $C$ becomes unity, and therefore the above equation can be rewritten as

$$F = ma$$  \hspace{1cm} (3.1)

It is important to distinguish between mass and weight. For a given body, its mass $m$ and weight $W$ are related by

$$W = mg$$

where $g$ is the acceleration of gravity associated with the force of gravity exerted on the body. Therefore, weight has the same units of force.

Any vector quantity such as a force can be graphically represented by a directed arrow such that its length is equal to the magnitude of the force and its direction coincides with that of the force, as shown in Fig. 3.1a. It is possible to resolve a vector quantity into components acting along reference directions such as those of a two-dimensional coordinate system of $x$ and $y$ axes shown...
in Fig. 3.1b. If the force $F$ is inclined to the $x$ and $y$ axes by the angles $\alpha$ and $\beta$, respectively ($\alpha + \beta = 90^\circ$), its components are given by

$$
F_x = F \cos \alpha \\
F_y = F \cos \beta
$$

(3.2)

where $F_x$ and $F_y$ are the components of the force along the $x$ and $y$ axes, respectively. Using the rule of the right triangle, the force $F$ is given by

$$
F^2 = F_x^2 + F_y^2 = F^2 \cos^2 \alpha + F^2 \cos^2 \beta
$$

where

$$
\cos^2 \alpha + \cos^2 \beta = l^2 + m^2 = 1
$$

(3.3)

Commonly, $\cos \alpha$ and $\cos \beta$ are called the direction cosines, and are given the symbols $l$ and $m$ respectively, i.e.,

$$
l = \cos \alpha \\
m = \cos \beta
$$

(3.4)

Therefore, Eqs. (3.2) can be rewritten as

$$
F_x = Fl \\
F_y = Fm
$$

(3.5)

In three dimensions defined by the $x$, $y$, and $z$ axes where the direction of force is inclined to the $x$, $y$, and $z$ axes by the angles $\alpha$, $\beta$, and $\gamma$, valid.
Eqs. (3.2) become:

\[ F_x = F \cos \alpha = Fl \]
\[ F_y = F \cos \beta = Fm \quad (3.6) \]
\[ F_z = F \cos \gamma = Fn \]

where \( n = \cos \gamma \), and Eq. (3.3) is modified to

\[ l^2 + m^2 + n^2 = 1 \quad (3.7) \]

Two types of mechanical forces are distinguished: (i) a conservative force and (ii) a nonconservative or dissipative force. A force is said to be conservative when its magnitude does not depend on velocity but rather on position such as the force of gravity and the force of a spring. Conversely, the magnitude of a nonconservative or dissipative force does not depend on position but rather on velocity, e.g., the force of friction.

### 3.4 Concepts of Work and Energy

A derived concept closely related to force is that of work. When a body is displaced a distance \( L \) along the direction of an applied force \( F \), the product of force and displacement is called work. If the force \( F \) remains constant during the displacement \( L \), the work \( W \) is given by

\[ W = FL \quad (3.8) \]

Mechanical work as defined by the above equation is a technical term related to a derived concept of mechanics called energy. By definition, energy \( E \) is the capacity of a body to do work. It has the same units as work, i.e., energy is the product of force and length.

Energy can take several forms including kinetic and potential energy. Kinetic energy \( (KE) \) is energy acquired by motion, which is defined by the relationship

\[ KE = \frac{1}{2} mv^2 \]

where \( m \) is the mass of a moving body and \( v \) is the magnitude of its velocity. Energy stored in a body because of its position is called potential energy \( (PE) \) and is defined by

\[ PE = mgL \]

where \( m \) is the mass of the body, \( g \) is the acceleration of gravity, and \( L \) is the elevation of the body with respect to some reference level.
Generally, only changes in energy can be calculated; i.e., it is not possible to evaluate absolute energy values. However, it turns out that only changes in energy are of significance rather than absolute values of energy.

3.5 Force and Motion

A force can produce two types of motion: (i) translational and (ii) rotational. During translational motion, all straight lines prescribed within the body remain parallel. When a body is in a state of rotational motion, every point within the body revolves in a circular path about some point not necessarily in the body, which serves as the axis of rotation.

In terms of motion, if several forces are simultaneously acting on a body, they can either be concurrent or nonconcurrent. By definition, concurrent forces, which produce translational motion, act on the same point of a body. An equivalent statement is that the lines of action of concurrent forces intersect at a common point, as shown in Fig. 3.2a. Conversely, nonconcurrent forces (Fig. 3.2b), whose lines of action do not intersect at a common point, can produce rotational motion. For example, consider a body that has an axis of rotation such as the axle of a wheel. A force $F$ perpendicular to the axis such that its line of

![Figure 3.2](https://example.com/figure32.png)

**Figure 3.2** Types of motion produced by forces acting on a body. (a) Concurrent forces; translational motion. (b) Nonconcurrent forces; translational and rotational motion. (c) Torque produced by a force causing translational motion.
action does not pass through it, as shown in Fig. 3.2c, tends to produce a clockwise or counterclockwise rotation about that axis. Such a tendency for rotational motion is directly proportional to the force $F$ as well as the perpendicular distance $L$ between the axis and the line of action of the force. By definition, the product of the force $F$ and distance $L$ is called the torque $T$, or the moment of the force, i.e.,

$$T = FL$$  \hspace{1cm} (3.9)

By comparing Eqs. (3.8) and (3.9), it can be seen that the expression of the torque is the same as that of work.

If the magnitude of velocity $v$ of a moving body remains constant, however, its direction changes at a constant rate typifying a curvilinear or circular motion, and the acceleration $a$ is given by $a = v^2/r$. In this case, the force $F$ acting on the body toward the center of the circular path is called centripetal force and is given by

$$F = \frac{mv^2}{r}$$  \hspace{1cm} (3.10)

According to Newton’s third law, the reaction to this force that acts in the opposite direction is called centrifugal force.

It is possible to make an analogy between translational and rotational motion such that a one-to-one correspondence exists between the concepts describing both types of motion, as schematically illustrated in Fig. 3.3.

As can be seen, a translational displacement $L$ along the $x$ axis corresponds to an angular displacement $\theta$ along a circular path. Similarly, the angular velocity $w$ is defined as the rate of change of angular displacement with time. The average angular velocity $\bar{\omega}$ is given by

$$\bar{\omega} = \frac{\phi}{t} = v = \frac{L}{t}$$

Figure 3.3  Schematics illustrating the one-to-one correspondence between rotational motion (left) and translational motion (right) of an object.
where $\phi$ is the change in angular position $\theta$ in time $t$, that is, $\phi = \theta - \theta_0$. Also, the angular acceleration $\alpha$ is defined as the rate of change of angular velocity with time. It is recalled from Eq. (3.9) that the torque or moment of force $T$ is defined in terms of translational concepts as

$$T = FL$$

To rewrite the above equation in terms of rotary concepts, the force is replaced by moment of force, and the linear acceleration is replaced by the angular acceleration, i.e.,

$$T = (FL)(L) = (m\alpha L)(L) = mL^2 \alpha$$

In terms of rotary concepts the product of mass $m$ and the square of distance $L$ from any axis is a concept called the moment of inertia $I$. Therefore, in terms of rotary concepts, the above equation can be written as

$$T = I \alpha$$

By definition, the moment of inertia is a function of the geometry of cross section of the body. Appendix C summarizes the moment of inertia for various geometries of cross sections.

### 3.6 Conservation of Energy

Although energy can be transformed from one form to another, e.g., from potential energy into kinetic energy or vice versa, it is realized from daily experience that during any process, including mechanical processes, energy is conserved. This is known as the principle of conservation of energy.

In all mechanical processes, the sum of potential and kinetic energy of the system remains constant, i.e.,

$$PE + KE = \text{constant}$$

However, since the force of friction is dissipative, i.e., nonconservative, it is not included in the above summation. An equivalent statement is that excluding nonconservative forces, the gain in kinetic energy of a system must be equal to the loss of potential energy during any process.

### 3.7 Concept of Machines

Mechanical work as defined by Eq. (3.8) is a compound concept because it is the product of force and length. It follows that a given amount of work $W$ can be defined in a variety of ways. A small force multiplied by large displacement is
equivalent to a large force multiplied by a small displacement, i.e.,

\[ F_1 L_1 = F_2 L_2 \]  \hspace{1cm} (3.13)

where \( F_1 \) < \( F_2 \) and \( L_1 \) > \( L_2 \).

Equation (3.13) can be rewritten as:

\[ \frac{F_1}{F_2} = \frac{L_2}{L_1} \]  \hspace{1cm} (3.14)

Physically, Eq. (3.14) means that when a given amount of work is done, a “force advantage” becomes possible if a “displacement disadvantage” can be tolerated. This is the basic principle involved in constructing machines. By definition, a machine is a device used to multiply force at the expense of a displacement or to change the direction of a force. According to the principle of energy conservation, a machine cannot do more work than is done on it. In practice, however, as a result of the forces of friction introduced by different working parts of the machine, a corresponding amount of the work done on the machine becomes unavailable to be done by the machine, i.e., the machine always does less work than done on it, leading to the idea of efficiency in engineering design. By definition, the efficiency of any machine is expressed as

\[
\text{Efficiency of a machine} = \frac{\text{work output}}{\text{work input}}
\]

Engineering products are designed such that they remain in a state of “mechanical equilibrium” during service. From the definitions of the various concepts of mechanics described above, the state of mechanical equilibrium is established by satisfying certain conditions as described below.

### 3.8 State of Mechanical Equilibrium

A body is said to be in a state of mechanical equilibrium if two conditions are satisfied. First, the sum of forces acting along any direction within the body must vanish; i.e., Eqs. (3.6) become

\[
\Sigma F_x = 0 \\
\Sigma F_y = 0 \\
\Sigma F_z = 0
\]  \hspace{1cm} (3.15)

where \( \Sigma F_x, \Sigma F_y, \) and \( \Sigma F_z \) are the sum of all components of forces acting on the body along the \( x, y, \) and \( z \) directions, respectively. Second, the sum of all torques acting on a body must also vanish; i.e., Eq. (3.9) becomes

\[ \Sigma L = 0 \]  \hspace{1cm} (3.16)

It is noted that if the forces acting on the body are of the concurrent type, the condition of mechanical equilibrium is satisfied only by Eqs. (3.15). However,
if the forces are nonconcurrent, the condition of mechanical equilibrium is met by simultaneously satisfying Eqs. (3.15) and (3.16).

In the preceding discussion, a body acted upon by a system of forces is assumed to be rigid or nondeformable. When a product is designed for mechanical or structural functions, the designer is primarily concerned with mechanical properties of the materials to be used in manufacturing. Specifically, the designer wishes to know the effects of an applied force on the material. To accommodate the effect of an applied force, whether it is small or large, the material is deformed or is said to be strained. Simultaneously, the material develops internal forces or stresses to counteract the effect of an applied force. Therefore, the effect of an applied force can be analyzed in terms of the concepts of strain and stress defined below.

### 3.9 Concept of Strain

When a body is acted upon by an external force, it undergoes some type of distortion or deformation, which can be a change in shape, volume, or both. All types of materials are characterized by a relative tendency to recover from such distortion upon the removal of the applied force; i.e., the body restores its original shape or volume. In effect, materials tend to resist and recover from deformation. Such a tendency is called elasticity (see Fig. 1.4).

Figure 3.4 schematically illustrates the three main types of deformation, namely, (i) simple one-dimensional stretch (tension) or compression (Fig. 3.4a), (ii) a shearing or twisting type of deformation (Fig. 3.4b), and (iii) overall fractional change in volume or dilation, which can be either expansion or contraction (Fig. 3.4c).

Deformation is expressed in terms of a derived concept called strain. It is defined as the fractional change in dimension or volume. In the case of one-dimensional stretch or compression shown in Fig. 3.4a, the axial strain $e$ is defined as

$$e = \frac{(L - L_0)}{L_0} = \frac{\Delta L}{L_0}$$  \hspace{1cm} (3.17)

where $L$ is the length of the body corresponding to an applied tensile force $F$, and $L_0$ is the initial length. Since $L > L_0$, the tensile strain $e$ is positive. Conversely, for compression, where $L < L_0$, the compressive strain $e$ is negative. Axial strain such as that shown in Fig. 3.4a is accompanied by a transverse strain. As can be seen, when the body is stretched along its axis, it is simultaneously contracted along the perpendicular direction. In this case, the transverse strain $e_{\text{tr}}$ is compressive and therefore negative. For a given material, the ratio of the transverse strain to axial strain is a constant called Poisson’s ratio $\nu$, i.e.,

$$\nu = -\frac{e_{\text{tr}}}{e}$$  \hspace{1cm} (3.18)
Since $e_{ir}$ is negative, the above ratio is preceded by a minus sign because $v$ is positive. For most metallic materials $v$ ranges from 0.25 to 0.45.

When deformation is expressed in terms of the original dimension of the part as in Eq. (3.17), the corresponding strain is called nominal or engineering strain, which is useful in engineering applications. As demonstrated in Chap. 4, however, a better understanding of the sequence of events taking place during deformation of a part requires taking into consideration the change in dimensions of the part with each increment of strain leading to the definition of actual or true strain. To determine the true strain $\varepsilon$, it can be considered to consist of the sum of increments of engineering strains as follows:

$$\varepsilon = e_1 + e_2 + e_3 + \cdots$$

$$\varepsilon = \Sigma (L_1 - L_0)/L_0 + (L_2 - L_1)/L_1 + (L_3 - L_2)/L_2 + \cdots$$

Figure 3.4 Schematics illustrating the three main types of deformation. (a) One-dimensional deformation (tension and compression). (b) Shear deformation. (c) Dilation (tension and compression).
or

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

Since \( e = (L - L_0)/L_0 \) according to Eq. (3.17), \( L/L_0 = 1 + e \), and the true strain \( \varepsilon \) is given by

\[ \varepsilon = \ln (1 + e) \]  
(3.19)

By comparing Eqs. (3.17) and (3.19), it can be seen that for a very small amount of deformation, \( e = \varepsilon \), however, as the amount of deformation increases, the true strain \( \varepsilon \) becomes progressively smaller than the engineering strain \( e \).

A shear or twisting strain \( \gamma \) such as that shown in Fig. 3.4b is defined as

\[ \gamma = \frac{x}{L} = \tan \theta \]  
(3.20)

where \( x \) is the shear displacement over a distance \( L \), and \( \theta \) is the shear or twisting angle.

A volumetric strain or dilation \( \Delta \), such as that shown in Fig. 3.4c, is defined as

\[ \Delta = \frac{V - V_0}{V_0} \]  
(3.21)

where \( V \) is the final volume and \( V_0 \) is the initial volume. In the case of expansion, \( \Delta \) is positive, and it becomes negative for contraction.

### 3.10 Concept of Stress

As a result of strain caused by an external force, counteracting internal forces are developed in the body. Such internal forces can be determined from the conditions of mechanical equilibrium. A simple case to consider is that of uniaxial loading. As demonstrated below, stress is the intensity of internal forces or force per unit area developed within a part in response to an external force.

Consider a rectangular part of an initial cross-sectional \( A_0 \), such as that shown in Fig. 3.5a, acted upon by two forces \( F \) of opposite sense and with their lines of action coinciding with the geometric axis of the part; i.e., the forces are of the concurrent type. It is evident that the net force acting on the part = \( F - F = 0 \) and therefore, according to Eqs. (3.15), it is in a state of mechanical equilibrium.

To determine the internal forces developed by the body, it must be cut at some location such as the cross section \( A-A \) (Fig. 3.5a). Since the entire body is in a state of mechanical equilibrium, any part of it, such as that shown in Fig. 3.5b, must also be in a state of mechanical equilibrium. According to Eqs. (3.15), an internal force \( F_n \) must be developed normal to the cross section \( A-A \) such that it is of the same magnitude as the force \( F \) but of opposite sense. By definition, a force normal to the cross section is called a normal force, and a force parallel to the cross section is called a shearing force.
is called a shear force. Since \( F_n = F \) is a normal force, its shear component \( F_s = 0 \) (Fig. 3.5c); i.e., the cross section is acted upon by

\[
F_n = F \\
F_s = 0
\]

As shown in Fig. 3.5b, the internal force \( F_n \) can be considered as the sum of unit forces \( \Delta F \)'s, each of which is acting on a unit area \( \Delta A \) such that

\[
F_n = F = \sum \Delta F_1 + \Delta F_2 + \Delta F_3 + \cdots \quad (3.22) \\
A_0 = \sum \Delta A_1 + \Delta A_2 + \Delta A_3 + \cdots \quad (3.23)
\]

Physically, Eqs. (3.22) and (3.23) mean that each unit area \( \Delta A \) resists a force \( \Delta F \). By definition, the force per unit area is called stress \( \sigma \), i.e.,

\[
\sigma = F/A_0 \quad (3.24)
\]

Since the stress \( \sigma \) is normal to the cross section, it is called a normal stress. It is evident that the shear stress \( \tau \) acting on the cross section is

\[
\tau = F_s/A_0 = 0
\]
Similar to the case of strain, when the stress is calculated with reference to the original cross-sectional area as in Eq. (3.24), it is called nominal or engineering stress. A normal stress can either be tensile (positive) or compressive (negative).

Corresponding to the true strain defined in the previous section, there exists a true stress $\sigma_t$. Referring to Fig. 3.5, if the instantaneous cross-sectional area corresponding to a tensile force $F$ is $A$, the true stress $\sigma_t$ is given by

$$\sigma_t = F/A$$

Since the volume of the part remains constant during deformation, its initial and instantaneous dimensions are related by

$$A_0 L_0 = AL$$

By combining the above two equations, $\sigma_t$ can be expressed as

$$\sigma_t = F/(A_0 L_0/L) = F/A_0 (L/L_0)$$

From Eqs. (3.24) and (3.17) ($F/A_0 = \sigma$, and $L/L_0 = 1 + e$, respectively), the true stress $\sigma_t$ is given by

$$\sigma_t = \sigma (1 + e) \quad (3.25)$$

It can be concluded from the preceding equation that when the amount of deformation is very small, $\sigma_t = \sigma$, however, as the amount of deformation increases, the true stress becomes progressively larger than the engineering stress. Although the cross section normal to the applied force is free of shear stress, this is not the case for an arbitrary section such as $B$–$B$ in Fig. 3.6, as described below.

If the angle between the force $F$ and the normal to the plane $n$ is $\alpha$ (Fig. 3.6b), the force $F$ can be resolved into two components: a force normal to the section $F_n$, and another force $F_s$ which lies in the plane of the section as shown in Fig 3.6b, where

$$F_n = F \cos \alpha \quad (3.26)$$

$$F_s = F \cos(90 - \alpha) = F \sin \alpha \quad (3.27)$$

If the initial cross-sectional area of the body is $A_0$, the area of the inclined section $A$ is given by

$$A = A_0 / \cos \alpha$$

From the definition of engineering stress [Eq. (3.24)], the normal stress $\sigma'$ acting on the inclined section is given by

$$\sigma' = F_n / A = (F \cos \alpha) / (A_0 / \cos \alpha) = F / A_0 \cos^2 \alpha = \sigma \cos^2 \alpha$$
where \( l \) is the direction cosine according to Eq. (3.4). By definition, the force \( F_s \) in the plane of the section is a shear force, which produces a shear strain such as that shown in Fig. 3.4b, and the shear force per unit area is called shear stress \( \tau \), i.e.,

\[
\tau = F_s/(A_0/\cos \alpha) = F/A_0 \sin \alpha \cos \alpha = \sigma/2 \sin 2\alpha \quad (3.29)
\]

It follows from the preceding equation that the maximum possible value of shear stress \( \tau \) developed by a normal stress \( \sigma \) corresponds to \( \alpha = 45^\circ \), i.e.,

\[
\tau_{\text{max}} = \sigma/2 \quad (3.30)
\]

and it acts along plane whose normal is inclined to the direction of \( \sigma \) by an angle of 45°.

To summarize the preceding discussion, a normal stress is perpendicular to the area upon which it is acting, while a shear stress is parallel to the area. Both normal and shear stresses can either be positive or negative. A normal tensile stress which tends to stretch the body is considered to be positive. Conversely, a compressive stress, which tends to compress the body, is taken to be negative. In terms of motion, a normal stress induces translational motion. Unlike the normal stress, a shear stress tends to slide, twist, or slip layers of the body relative to each other; i.e. it induces a rotational motion. If the rotational motion is in
the clockwise rotation, the shear stress is considered to be positive, and it is considered to be negative if it produces counterclockwise rotation.

It is evident from the preceding discussion that a normal force produces both normal and shear stresses. As a result, the distortion produced in the body results from a combination of translational and rotational motion, as illustrated in Fig. 3.7. Similarly, if a body is acted upon by a shear force, it produces both normal and shear stresses.

### 3.11 Hook’s Law

Experiment shows that within the elastic range of any material, a linear relationship exists between stress and strain, which is known as Hook’s law, i.e.,

\[
\text{Stress} = \text{constant} \times \text{strain} \tag{3.31}
\]

where the proportionality constant is called the modulus of elasticity. Three different moduli are associated with the three types of strain shown in Fig. 3.4. In the case of a normal strain such as a simple stretch, Hook’s law is expressed as

\[
\sigma = E\epsilon \tag{3.32}
\]

where \(\sigma\) is normal stress (tensile or compressive), \(\epsilon\) is the corresponding engineering strain, and \(E\) is called Young’s modulus. For a shear strain,

![Figure 3.7](image)

**Figure 3.7** Translational and rotational motions produced by a normal tensile force. (a) Before deformation; the normal and shear components of force acting on an arbitrary section are indicated. (b) After deformation; the normal force has produced a translational displacement and the shear force has produced a rotational displacement.

Copyright © 2004 by Marcel Dekker, Inc.
Hook’s law becomes
\[ \tau = G \gamma \]  \hspace{1cm} (3.33)
where \( \tau \) is shear stress, \( \gamma \) is the corresponding shear strain, and \( G \) is the shear modulus or modulus of rigidity. If the strain corresponds to contraction or expansion (dilation), a bulk modulus \( B \) is defined such that
\[ \sigma = B(V - V_0/V_0) \]  \hspace{1cm} (3.34)
where in this case the stress \( \sigma \) is called hydrostatic stress as explained later, \( V \) is the corresponding volume, \( V_0 \) is the initial volume, and \( B \) is the bulk modulus.

Young’s modulus \( E \), the shear modulus \( G \), the bulk modulus \( B \), and Poisson’s ratio \( \nu \), defined earlier, are called the elastic constants. They are characteristic properties of a given material and can be shown to be related by the following equations:
\[ K = E/3(1 - 2\nu) \]
\[ G = E/2(1 + \nu) \]  \hspace{1cm} (3.35)
\[ \nu = E/2G - 1 \]

Other related constants include the coefficient of rigidity \( n \) and the coefficient of compressibility \( C \). By definition,
\[ n = \tau/\theta \]  \hspace{1cm} (3.36)
where \( \tau \) is the shear stress and \( \theta \) is the shear or twist angle (see Fig. 3.4b), and:
\[ C = 1/B \]  \hspace{1cm} (3.37)
4

Property Evaluation

4.1 Introduction

It is possible to evaluate the properties of a failed component by two main types of tests: (i) nondestructive and (ii) destructive. By definition, a nondestructive test does not degrade the integrity of the component being tested. However, a given test can be either nondestructive or destructive depending on how it is performed. For example, it is possible to conduct a hardness test on a given component in the field. In this case, the test is nondestructive. A hardness test conducted on a polished specimen in the laboratory is destructive. Obviously, metallographic examination in the laboratory is destructive; however, field metallography is a nondestructive test. Also, stress analysis of a failed component is nondestructive. Although there is no sharp distinction between nondestructive and destructive tests, it is a common practice to consider some specific tests as nondestructive and others as destructive.

Nondestructive tests are primarily carried out to detect flaws and control dimensions. Destructive tests are used to determine certain properties, which are mostly of mechanical nature, such as tensile strength, creep strength, and fatigue strength. Both types of tests are covered in this chapter.

4.2 Nondestructive Tests

4.2.1 Liquid Dye Penetrant Test

A dye penetrant assists the naked eye in detecting cracks at the surface of the failed component. It is based upon penetrating the surface defect by a dye, removing...
excess penetrant, development of the indication, and finally inspection. A defect is marked by a bright color.

Prior to applying the penetrant, the part must be free of grease and oil. After cleaning, the penetrant is applied by spraying, dipping, or painting. Sufficient penetration time must be allowed. Longer penetration time up to 30 min may be required to detect fine or tight discontinuities. Subsequently, the part is drained to remove excess penetrant, rinsed in hot water, and dried. A developer is then applied such that the surface is covered evenly with a thin layer. After allowing the developer to dry, defects become marked by a bright color visible to the naked eye.

A dye penetrant test is rather cheap and the results can be interpreted in a straightforward manner. However, internal flaws or inaccessible surface cannot be examined.

### 4.2.2 Magnetic Particle Inspection Test

This test is based upon the effect of surface flaws on the lines of magnetic flux induced in a magnetic material; i.e., the test is limited to ferromagnetic materials such as Fe, Co, Ni, and some of their alloys. Before conducting the test, the part to be inspected is demagnetized and cleaned. This is followed by heavily magnetizing the part using various methods, such as using (i) a permanent magnet or (ii) passing an electric current either directly through the part or through a coil around the part. More frequently, however, a magnetic field is applied by passing an electric current through a coil around the part. Either direct or alternating current can be used. Direct current is preferred because it travels deeper into the part, which permits subsurface flaws to be detected. Following the magnetization process, the part is dusted with a magnetic powder or painted with a suspension of a magnetic powder revealing displacements of magnetic flux lines at the locations of surface flaws or irregularities, as schematically illustrated in Fig. 4.1. Discontinuities are outlined by magnetically attached particles of the powder, which also indicate the size, shape, and extent of the discontinuity. By combining the magnetic powder with a fluorescent material, it becomes possible to visually detect surface flaws using ultraviolet light.

Since a surface flaw affects the lines of magnetic flux only if it crosses them, it is always recommended to conduct two tests at right angles. Advantages of the test include its simplicity and easy interpretation of the results. Apart from its limited application to ferromagnetic materials, only surface and subsurface flaws at accessible locations can be examined.

### 4.2.3 Radiography

Radiography is defined as the formation of images of an object on a fluorescent screen or other photographic material by means of X rays or gamma rays.
Because of the high penetrating power of these radiations, they can easily pass through quite thick pieces of metallic materials. X rays are produced when a beam of electrons generated from a hot filament (cathode) is accelerated toward a metal target (anode) by a potential difference applied between the cathode and anode in an evacuated glass tube called the X ray tube. Gamma rays which are also part of the electromagnetic spectrum are generated from radioactive sources such as cobalt, cesium, iridium, and thallium. Although both X rays and gamma rays have similar penetrating power, the use of gamma rays is limited to remote or restricted locations.

When the part to be inspected is placed in the path of a sufficiently energetic X ray beam, the intensity of an image formed by the transmitted beam varies with the opacity of the material. For example, if the part being inspected contains cracks or holes, the beam transmitted through this region becomes more intense in comparison with more sound sections free of these defects. Usually, the image of the part is formed by using a photographic plate. Therefore, by placing an X ray tube on one side of the part, and a photographic plate on the other side, a shadow image of the part or a radiograph is obtained where the positions of flaws can be located. Usually, radiography is used to evaluate the quality of welds and castings. Some of its advantages include examination of entire sections and obtaining a permanent record of the findings. One limitation of radiography is that only cracks running parallel to the X ray beam can be detected. Cracks on planes normal to the X ray beam cannot be detected. Also, in most cases cracks assume irregular paths precluding detection. Therefore, radiography is not considered to be a positive means for detecting cracks.

Figure 4.1  Schematics illustrating the effect of a flaw on the lines of magnetic flux. (a) Magnetic lines in the absence of a flaw. (b) Magnetic lines in the presence of a flaw.
4.2.4 Ultrasonic Inspection

Ultrasonic inspection tests are primarily used in flaw detection and thickness measurement. Such tests are based upon the effect of various types of planar defects or boundaries on the transmission of very high frequency sound waves through a metallic material. When a sound wave travels through a metallic material, it is reflected at any type of boundary such as an external surface and a crack surface. Although sound waves of very high frequencies in the megacycle range are capable of detecting small irregularities, they can be easily absorbed if the material has a large grain size. In practice, a frequency of about 2–3 megacycle/sec is usually used.

Sound waves of very high frequency are transmitted into the part being examined by means of a crystal acting as a transducer using an alternating electric current. Usually, the crystal is coupled to the part by an oil or water medium to minimize energy losses resulting from air gaps. Either an oil film is maintained between the transducer and the part or both the part and transducer are immersed in a tank of water.

Although there are a variety of ultrasonic tests, they are all based upon the pulse-echo method schematically illustrated in Fig. 4.2. In this method, sound waves are generated in the transducer by a pulse generator. Subsequently, the

![Figure 4.2](image-url)
transducer transmits the waves into the part being examined, and then receives the reflected waves. Typically, the waves are transmitted into the part in pulses at a rate of about 50 pulses/sec. An oscilloscope with a screen monitoring the speed of the transmitted waves is used to measure the time taken by a reflected sound wave to return from the part into the transducer. By means of a time base control system connected to the X plate of the oscilloscope, the speed of the spot across the screen representing the pulse can be adjusted. Instant of pulse transmission is indicated by the peak $T$ on the screen of the oscilloscope (see Fig. 4.2) when a potential difference is applied to the $Y$ plate of the oscilloscope. Waves returning to the transducer from the part are converted into an alternating current, and after amplification and rectification, it is supplied to the $Y$ plate. Once the $Y$ plates are energized by the current representing the reflected waves, the spot on the screen travels a given distance producing peaks at some locations, e.g., $F$ and $S$, on the screen as shown in Fig. 4.2. If the position of the stronger peak $S$ represents reflection from the surface of the part, the position of the weaker peak $F$ represents a discontinuity or a flaw.

Similar to radiography, an entire section can be examined by the ultrasonic test. However, difficulties may arise in interpreting the results if the surface of the part being examined is rather irregular, producing spurious reflections. Also, it is difficult to distinguish between reflecting boundaries representing deleterious and nondeleterious flaws.

### 4.2.5 Eddy-Current Test

By definition, eddy currents are electric currents generated by induction, e.g., as a result of a magnetic flux or a nearby electric charge. Basically, an eddy-current test system designed to detect flaws consists of two electric coils, an exciter coil and a search coil, as schematically illustrated in Fig. 4.3. When an alternating current passes through the exciter coil, a magnetic flux is generated, which in turn

![Figure 4.3](image-url) A schematic illustration of the Eddy current test.
induces an alternating current in the search coil whose magnitude is dependent upon that of the magnetic flux.

Detection of surface flaws is based upon (i) eddy currents generated in the part being examined upon its insertion into the coil system (Fig. 4.3) and (ii) the corresponding effect on the magnetic flux produced by the exciter coil. If a sample of the same material of the part free of defects is inserted in the coil system, it produces a given effect on the magnetic flux. However, this effect is altered by surface flaws and discontinuities affecting the electrical conductivity of the material along the direction of eddy current flow. By measuring the potential difference across the search coil, the change in current flow produced by the part can be determined. Flaws are detected by comparing the potential differences corresponding to the part and a sample free of defects. Usually, the current frequency employed in the test ranges from 50 to 200,000 cycles/sec. Higher frequencies are used when examining ferromagnetic materials to minimize the countereffect produced by magnetic hysteresis. In this case, however, only surface flaws can be detected because only high-frequency current flows at the surface.

An important advantage of the eddy-current test is its simplicity for repetitive routine inspection. However, without prior correlation, interpretation of the results can be rather difficult.

4.2.6 Field Metallography

It is the primary objective of field metallography to nondestructively evaluate the overall or gross microstructural features of the part being examined. Usually, the technique is used to determine grain size, morphology of grain boundaries and fracture, as well as the general characteristics of major phases present. Although field metallography can provide some valuable information, it cannot substitute a detailed microstructural characterization of samples removed from the failed component.

Field metallography is based upon producing a replica of the surface to be examined. Prior surface preparation methods are essentially the same as those used for metallographic examination in the laboratory (see Chap. 8). First, the surface of the part is ground with successively finer grits of silicon nitride starting with 80-grit followed by 180-grit, 320-grit, and finally 600-grit. Second, the surface is intermediately polished using a 6 μm and 1 μm diamond paste. Final polishing is carried out using a suspended solution of 0.05 μm gamma alumina. A proper etchant is then applied to the surface to outline the gross microstructural features.

To produce a replica of the etched surface as schematically shown in Fig. 4.4, a standard cellulose acetate replication film about 50 μm in thickness is softened by a solvent preferred to be methyl acetate because it minimizes the formation of bubbles when the replica is dried. After softening the replication

Copyright © 2004 by Marcel Dekker, Inc.
film, it is applied to the etched surface and allowed to dry. When the replica becomes completely dry, it is removed from the surface, mounted on a glass slide, and stored. It is possible to evaluate the replica using a field microscope if available; however, more detailed characterization is carried out when the replica is returned to the laboratory.

4.2.7 Measurement of Residual Stresses

It is recalled from Sec. 1.2.1 that when a metallic material is subjected to a tensile load, it responds in three different ways as the applied load increases gradually: (i) Initially the material is deformed elastically; i.e., when the applied stress is removed, the material regains its original dimensions. (ii) When the applied stress exceeds a certain value, the material is deformed plastically, i.e., when the applied stress is removed, only the elastic portion of the strain is recovered, and the material becomes permanently deformed. (iii) Finally, when the stress exceeds the inherent strength of the material, fracture occurs. Stresses are classified into two main types: (i) applied and (ii) residual. An applied stress persists only as long as the part is acted upon by an external force. Once the force is removed, the applied stress disappears. In contrast, the residual stress remains even if the external force, e.g., an overtightened bolt, is removed.
Figure 4.5 summarizes the basis for measuring residual stresses by the technique of X ray diffraction. A uniaxially loaded part of original length $L_0$ and cross-sectional area $A_0$ is shown in Fig. 4.5a. From an atomic point of view, elastic strains correspond to a change in the spacing between atomic planes in the crystals of the material along the direction of applied stress, i.e., along the $z$ axis (Fig. 4.5b). In this case, the spacing of atomic planes normal to the $z$ axis is changed from $d_0$ to $d$; and therefore, the tensile elongation $e_z$ defined in Chap. 3 can be expressed as

$$e_z = (d - d_0)/d_0$$

(4.1)
Within the elastic range of a given material, the stress $\sigma_z$ is linearly proportional to strain such that

$$\sigma_z = E e_z$$  \hspace{1cm} (4.2)

where $E$ is a material property called the elastic modulus or Young’s modulus. Combining Eqs (4.1) and (4.2) results in

$$\sigma_z = E(d - d_0)/d_0$$  \hspace{1cm} (4.3)

Therefore, by measuring the change in lattice spacing, and from knowledge of the elastic modulus of the material, the stress $\sigma_z$ can be calculated from Eq. (4.3). Lattice spacing can be measured by the technique of X ray diffraction as described in more detail in Chap. 8. However, it suffices to mention here that the technique involves bombardment of the specimen by a beam of X rays, as shown in Fig. 4.5a. As a result, the specimen emits diffracted X ray beams, which can be analyzed to determine the $d$ spacings, as described in Chap. 8. From an experimental point of view, measurement of lattice spacing along the $z$ axis (Fig. 4.5) requires that the X ray beam must be directed normal to the axis of the bar, which is physically impossible. Instead, the lattice spacings are measured for crystal planes nearly parallel to the $z$ axis. Since the volume of the part remains constant after applying the load $P$, the transverse strain along the $x$ and $y$ directions corresponding to the reduction in cross-sectional area is compressive and can be expressed as

$$e_x = e_y = (A - A_0)/A_0$$  \hspace{1cm} (4.4)

It is evident that the transverse strain has a negative value because $A_0 > A$. Similar to the case of the longitudinal strain described above, the transverse strain $e_x$ can be expressed as

$$e_x = (d' - d'_0)/d'_0$$  \hspace{1cm} (4.5)

where $d'$ and $d'_0$ are the spacing of atomic planes nearly parallel to the $z$ axis after and before applying the stress respectively. It is recalled from Chap. 3 that the ratio of transverse strain $e_x$ or $e_y$ and longitudinal strain $e_z$ is a material property called Poisson’s ratio $\nu$, i.e.,

$$e_x/e_z = e_y/e_z = -\nu$$  \hspace{1cm} (4.6)

where the minus sign accounts for the fact that the transverse strain has a negative value as described above. For most metallic materials, the value of $\nu$ ranges from about 0.25 to 0.45. Combining the above equations results in

$$\sigma_z = -E/\nu (d' - d'_0)d'$$  \hspace{1cm} (4.7)
X-ray diffractometers modified for stress measurements can be used to determine the above lattice spacing, and from knowledge of $E$ and $v$ for the material, the longitudinal stress $\sigma_z$ can be calculated.

In practice, the stressed condition of parts is more complicated than the simple uniaxial state described above. Generally, the part can be subjected to two or three components of normal stress $\sigma$ at right angles to each other, as explained in Chapter 5. For a biaxial state of stress where $\sigma_x = 0$, i.e., the stress normal to the surface of the part in Fig. 4.5 is 0. In this case, the strain $e_x$ is given by

$$e_x = \frac{d'' - d'_0}{d'_0} \quad (4.8)$$

where $d''$ and $d'_0$ are the spacing of crystal planes parallel to the surface of the bar after and before applying the stress, respectively. It can readily be shown that $e_x$ is given by

$$e_x = -v/E(\sigma_z + \sigma_y) \quad (4.9)$$

Thus, the sum of the stresses $\sigma_z$ and $\sigma_y$ can be calculated using Eq. (4.9).

### 4.3 Destructive Tests: Measurement of Mechanical Properties

It is recalled that in selecting a material for a given engineering application, the primary concern of the designer is its properties. Usually, designers are particularly concerned with mechanical properties, which describe the response of the material to applied forces encountered during service. For better use and selection of materials, the designer must have a knowledge of both the meaning of each property and the method or type of test used to measure that property. When using the results of these tests in design, it is extremely important that the designer must be aware of the fact that they apply only to the test conditions. Also, the usefulness of these results is limited to the extent to which test conditions duplicate actual service conditions. In general, materials must be tested under actual service conditions to determine their suitability for a given application. Various mechanical properties measured by standardized tests are expressed in terms of the concepts of strain and stress defined in Chapter 3.

Various tests have been devised to reveal the mechanical properties of materials under two main types of loading conditions, namely, (i) static loading and (ii) dynamic loading. A static load is applied only once; it induces strain in the material very slowly and gradually and remains constant throughout the service life of the component. Tension, compression, hardness, and creep tests are
used to reveal mechanical properties under a static loading condition. Dynamic loads can be classified into (i) impact loads and (ii) fatigue loads. An impact load resembles a static load in that it is applied only once. However, it differs from a static load in that it induces strain in the material very rapidly. Charpy impact test is devised to measure the resistance of materials to such loads. A fatigue load is a more common type of a dynamic load, and it is applied several times in a cyclic manner. Fatigue test is exclusively used to determine mechanical properties under cyclic loading condition. Another related test is the fracture toughness test.

Design for structural and mechanical functions is based on the useful strength or allowable stress of engineering materials. Usually, in such applications, materials are selected to operate within their elastic range. Sometimes, however, machine parts and structures are operated at stresses exceeding their elastic limit. Also, to guard against catastrophic failure, it is taken into account that the material should plastically deform rather than fracture in case of a sudden overload condition. During service, engineering products are usually subjected to complex systems of forces.

It follows from the discussion above that designing a safe product requires knowledge of (i) the allowable stress or the maximum stress, which can be accommodated safely by the product, and (ii) the onset of plastic deformation under service conditions. Usually, tensile stresses resulting from tensile loading condition play the most important role in structural applications of materials. Provided the material is capable of being plastically deformed, i.e., ductile, a tensile load results in a gradual shrinkage of the load-carrying cross-sectional area and in turn a continuous increase in true stress until fracture occurs. Tensile stresses become more critical if the ability of the material to plastically deform is rather limited, i.e., brittle. In this case, small imperfections or discontinuities result in local weakening by reducing the load-carrying cross sectional area and in turn increasing the local stress, which may lead to sudden fracture even though failure is not predicted on the basis of the nominal cross section. It is therefore evident that tensile properties of materials derived from standard tension tests provide an invaluable aid in designing safe products.

Tension, hardness, creep, impact toughness, and fatigue tests have long been used to evaluate the mechanical properties of engineering materials. More recently, the fracture toughness test has emerged as another important test. Compression is a less common mechanical test. Another test rarely used to specify the mechanical properties of materials is the torsion test. As described below, the uniaxial stress-strain relationship determined from the tension test reveals a number of important mechanical properties of the material. This relationship is extended in Chap. 5 to cover more complex stress conditions.
4.3.1 Tension Test

A standard tension test is considered to be the most versatile mechanical test used to reveal the mechanical properties of a given material. To carry out the test, a standard specimen is machined from the part. It can either be a sheet or a round bar. A sheet specimen is held in the testing machine by means of pins inserted into the holes at both ends, while a round bar is threaded into the machine at both ends. All measurements are carried out across a prescribed gauge length, e.g., 50.8 mm (2 in.), as schematically shown in Fig. 4.6.

It is the function of the testing machine to stretch the specimen by a progressively increasing tensile load acting along its geometric axis until it fractures, as shown in Fig. 4.7. During the course of the test, the load applied to the specimen and corresponding elongation across the gauge length are recorded.

From knowledge of the original dimensions of the specimen, the load-elongation data is converted into stress-strain data using the relationships:

\[ \sigma = \frac{P}{A_0} \]  \hspace{1cm} (4.10)
\[ e = \frac{L - L_0}{L_0} \]  \hspace{1cm} (4.11)

where \( \sigma \) is the engineering stress, \( P \) is the applied load, \( A_0 \) is the original cross-sectional area of the specimen, \( e \) is the engineering strain, \( L \) is the length of the specimen corresponding to load \( P \), and \( L_0 \) is the original length of the specimen.

Since the tensile test involves the application of a single stress acting along the geometric axis of the specimen, it is said to be a uniaxial test. A plot of the

![Figure 4.6](image)

**Figure 4.6** A Schematic illustration of standard tensile specimens. (a) Sheet. (b) Round bar.
stress vs. strain resulting in a stress-strain diagram reveals the uniaxial stress-strain relationship of the material being tested. Two types of tensile stress-strain diagrams can be distinguished: (i) engineering stress-strain diagram and (ii) true stress-strain diagram as described below.

Figure 4.8 illustrates the typical shape of a tensile specimen after removal of the load during the course of the test. Initially, the specimen is elastically deformed when the applied load is relatively small; i.e., it regains its original dimensions after the load is removed, as shown in Fig. 4.8a,b. With continued increase in the applied load, the specimen is plastically deformed; i.e., it acquires a permanent change in its dimensions manifested by an increase in length and reduction in cross-sectional area. During the early stages of plastic deformation, the strain is uniformly distributed across the gauge length (Fig. 4.8c); i.e., the plastic deformation is homogeneous. Subsequently, however, the strain becomes localized, leading to the development of a “neck” as shown in Fig. 4.8d; i.e., the plastic deformation becomes nonhomogeneous. Eventually fracture occurs at the necked section as the applied load continues to increase (Fig. 4.8e).

Depending on the nature of material being tested, the transition from elastic to plastic behavior can either be gradual or abrupt, as illustrated in Fig. 4.9. As can be seen from Fig. 4.9a, the transition from elastic to plastic behavior is gradual, which is typical of nonferrous engineering alloys. This transition becomes...
never, it is possible to divide the diagram into three main regions as described below.

Region $OA$ represented by a straight line in Fig. 4.9a,b corresponds to the elastic range of the material where Hook’s law is applicable; i.e., the stress is linearly related to strain. Once the applied stress exceeds a certain value, the material begins to deform both elastically and plastically. Plastic deformation is homogeneously distributed across the gauge length within region $AB$ of Fig. 4.9a and region $A'B$ of Fig. 4.9b, where the stress-strain relationship is no longer linear, i.e., Hook’s law is not valid. At any point such as $x$ within regions $AB$ and $A'B$, the total engineering strain $\varepsilon_t$, corresponding to an applied stress $\sigma_s$ is

Figure 4.8 Schematics illustrating the change in shape of a tensile specimen after removing the applied load during the course of tension test. (a) Initial specimen shape. (b) Shape of specimen after removing the load within the elastic range. (c) Shape of specimen after removing the load within the uniform plastic range. (d) Shape of specimen after removing the load within the non-uniform plastic range. (e) Shape of specimen after fracture.
made up of two components such that

$$e_s = e_e + e_p$$  \hspace{1cm} (4.12)$$

where $e_e$ is the elastic component and $e_p$ is the plastic component. If the applied load is removed, only the elastic component $e_e$ is recovered, and the specimen acquires a permanent strain $e_p$. As shown in Fig. 4.9a, the total strain $e_s$ is defined by the point resulting from the intersection of a line drawn from point $x$ normal to the strain axis. Each strain component is defined by the intersection with the strain axis of a line drawn from point $x$ parallel to the elastic line $OA$, as shown in Fig. 4.9a.

By definition the stress corresponding to the onset of plastic deformation is called the yield stress or strength $\sigma_y$, which is an important design parameter. It is evident from Fig. 4.9a that the yield strength is not well defined. In this case, it is considered to be the stress corresponding to 0.2% plastic strain, and therefore, it is commonly called the 0.2% yield strength. In contrast, two yield points are well defined by the stress-strain diagram shown in Fig. 4.9b. At the end of the elastic region an upper yield point corresponding to an upper yield strength is defined by point $A$ beyond which the stress is dropped to point $A'$, which is known as the lower yield point. In practice, the yield strength is considered to be the stress corresponding to the upper yield point. It is possible to explain the yield point phenomenon shown in Fig. 4.9b in terms of the microscopic aspects of plastic deformation, as discussed in Chap. 7.
A distinguishing feature of regions $AB$ and $A'B$ in Fig. 4.9a,b is that the stress required to continue plastic deformation increases with the amount of plastic strain. It is recalled from Chap. 3 that this technologically important phenomenon is called strain hardening. Similar to the yield point phenomenon, its origin can be explained in terms of the microscopic aspects of plasticity (Chap. 7).

According to the engineering stress-strain diagrams of Fig. 4.9, strain hardening continues up to point $B$. Beyond this point, the engineering stress-strain diagram indicates that the material ceases to strain harden. Within region $BC$, the strain becomes highly localized in a certain section of the specimen leading to the phenomenon of necking (see Fig. 4.8d). Correspondingly, the engineering stress decreases rapidly, and ultimately the specimen is fractured at point $C$. A summary of the information which can be obtained from a tensile stress-strain diagram is given below.

By definition the slope of the elastic line $OA$ in Fig. 4.9a,b corresponds to the elastic, or Young’s, modulus $E$. As pointed out above, the yield strength $\sigma_y$ corresponding to the onset of plastic deformation is an important design parameter. It is a measure of the load-carrying ability of the material. For both the diagrams shown in Fig. 4.9, the stress corresponding to point $B$, beyond which the stress appears to decrease with strain, is called the ultimate tensile strength $\sigma_{uts}$. It can be used as a measure of the resistance of the material to fracture.

Total elongation to fracture expressed as an engineering strain is sometimes taken as a measure of ductility; a property determining the ability of the material to plastically deform prior to fracture. However, it is important to realize that the elongation to fracture is a function of gauge length. Therefore, when elongation is used as a measure of ductility, the corresponding gauge length must be specified. A more meaningful measure of ductility is the reduction in cross-sectional area of the specimen at the point of fracture expressed as a percentage, i.e., reduction in cross-sectional area $= (A - A_0)/A_0 \times 100$, where $A_0$ is the original cross-sectional area of the specimen and $A$ is the cross-sectional area at the fracture point. Since for most structural applications, the material must exhibit a minimum of 10% tensile elongation prior to fracture in a uniaxial tensile test, it is a common practice to classify materials having $>10\%$ elongation as ductile. A material exhibiting less than 10% elongation is said to be brittle.

It is noted that the area under the stress-strain diagram corresponds to energy per unit volume, i.e., stress $\times$ strain $= (\text{force/area}) \times (\text{length/length}) = (\text{force} \times \text{length})/(\text{volume}) = \text{energy per unit volume}$. Two properties can be derived from areas under the stress-strain diagram. As can be seen from Fig. 4.10a, the area enclosed by the triangle $OAO'$ represents the strain energy $U_e$ required to stress the material from zero stress up to the yield point, which

Copyright © 2004 by Marcel Dekker, Inc.
is given by

\[ U_e = \frac{1}{2} \sigma_y e \]  

(4.13)

Physically, \( U_e \) represents the ability of the material to absorb elastic strain energy, which is called resilience. Combining the above equation with Hook’s law,

\[ \sigma_y = Ee \]

results in

\[ U_e = \frac{\sigma_y^2}{2E} \]  

(4.14)

where \( E \) is the Young’s modulus of the material.

Another property related to the area under the uniaxial stress-strain diagram is called toughness. It corresponds to the total area enclosed by the stress-strain diagram (Fig. 4.10b), and therefore it can be used as a measure of the energy per unit volume absorbed by the material before it fractures. Evidently, toughness is a property dependent upon both the yield strength and ductility and can be considered to represent the resistance of the material to fracture. Various expressions can be used to approximate the total area \( A \) under the stress-strain diagram depending upon whether the material is ductile or brittle. For ductile materials,
A is approximated by

\[ A = \sigma_{uts} e_f \]  \hspace{1cm} (4.15)

or

\[ A = (\sigma_y + \sigma_{uts})/2 e_f \]  \hspace{1cm} (4.16)

and for brittle materials, \( A \) is given by:

\[ A = \frac{2}{3} \sigma_{uts} e_f \]  \hspace{1cm} (4.17)

where \( \sigma_{uts} \) is the ultimate tensile strength, \( e_f \) is the total engineering strain at failure, and \( \sigma_y \) is the yield strength.

According to the engineering stress-strain diagrams of Fig. 4.9, the material ceases to strain harden beyond the point of ultimate tensile strength. Such an apparent behavior, however, results from calculating the stress and strain with reference to the original dimensions of the specimen. In reality, when the true stress and true strain are used to plot the stress-strain diagram, the stress is found to continuously increase with strain till the point of fracture, as shown in Fig. 4.11; i.e., strain hardening continues to occur till the point of fracture.

Discrepancy between the engineering and true stress-strain diagrams in relation to the strain hardening behavior beyond the point of ultimate tensile

![Figure 4.11](image-url)  

Figure 4.11 A schematic illustration of a true tensile stress-strain diagram.
strength can be related to calculating the stress and strain with reference to the original dimensions of the test specimen. This implies that the plastic strain is uniformly distributed over the gauge length from the onset of plastic deformation till the point of fracture. In contrast, from a consideration of the microscopic aspects of plasticity (Chap. 7), it can readily be shown that plastic deformation is a nonhomogeneous process. Although on the microscopic scale, plastic deformation occurs nonhomogeneously from the onset of deformation, its macroscopic effect becomes pronounced during the later stages as manifested by phenomenon of necking encountered in tensile tests.

To explain the difference in shape between the engineering and true stress-strain diagrams, it must be recognized that the stress required to continue plastic deformation is determined by two opposing factors: (i) strain hardening leading to an increase in stress and (ii) reduction in cross-sectional area leading to a decrease in stress. During the earlier stages of deformation where the macroscopic strain is rather uniformly distributed over the gauge length, the effect of strain hardening dominates the effect of reducing the cross-sectional area leading to an overall increase in stress with strain as exhibited by both the engineering stress-strain diagram and true stress-strain diagram (regions AB and A′B in Fig. 4.9a,b).

Eventually, the opposing effects of strain hardening and reduction in cross-sectional area are balanced at point B corresponding to the ultimate tensile strength. Beyond that point, deformation becomes highly localized in the necked section rather than uniformly distributed over the entire gauge length. Therefore, calculating the stress and strain with reference to the original dimension leads to the apparent reduction of stress with strain within region BC of the engineering stress-strain diagram; i.e., the effect of strain hardening is dominated by the effect of reducing the cross-sectional area. Conversely, if the stress and strain are referred to the instantaneous dimensions, the stress continues to increase with strain as shown in the true stress-strain diagram. Such a deviation from the engineering stress-strain diagram can be understood with reference to Eqs. (3.19) and (3.25), defining the true strain and stress, respectively. For a relatively large elongation, the true strain becomes significantly smaller than the engineering strain. Conversely, the corresponding true stress becomes significantly larger than the engineering stress leading to the shift of the true stress-strain diagram into higher stress and lower strain levels, as schematically illustrated in Fig. 4.12.

For many engineering alloys, the relationship between true stress $\sigma_t$ and true plastic strain $\varepsilon$ can be expressed as

$$\sigma_t = K \varepsilon^n$$  \hspace{1cm} (4.18)

where $K$ is a constant called the strength coefficient, and $n$ is another constant known as the strain-hardening exponent.
For a given material, the shape of the tensile stress-strain diagram and in turn the tensile properties derived from that test are influenced by a number of variables, the most important of which are (i) grain size, (ii) temperature, and (iii) strain rate.

Experiment shows that for a given material, the yield strength increases with decreasing grain size according to the relationship

\[ \sigma_y = \sigma_0 + K_{yd}d^{-1/2} \]  

\[ (4.19) \]

where \( \sigma_y \) is the tensile yield strength, \( \sigma_0 \) is a material constant called the friction stress, \( K \) is an empirical constant, and \( d \) is the grain diameter. Also, decreasing the grain size increases the rate of strain hardening. Both effects of grain size can be explained in terms of the microscopic aspects of plasticity (Chap. 7).

Since plastic deformation is a thermally activated process, it is time dependent; i.e., time is required to accommodate the effect of an applied stress by plastic deformation. Evidently, as the rate of applying the load or strain rate is increased at a given test temperature, the time required to plastically deform is shortened; i.e., a lag between stress and strain is created. As a result, the yield strength increases with strain rate. A similar effect is produced if the test temperature is reduced; i.e., the yield strength increases with decreasing the test temperature at a given strain rate. Therefore, the effect produced by increasing the strain rate is equivalent to that produced by reducing the test temperature.

It can be concluded from the discussion above that increasing the strain rate at a given test temperature shifts the tensile stress-strain diagram into higher stress levels and lower strain levels. For a given strain rate, increasing the test temperature shifts the tensile stress-strain diagram into lower stress levels and higher strain levels.
Because of the interconnection between the effects of strain rate and temperature on the shape of the tensile stress-strain diagram as described above, the actual effect of strain rate on the shape of the stress-strain diagram is linked to the dependency of yield strength of the material on temperature, which is a function of its structure. Some of the engineering alloys exhibit a slight dependency of yield strength on temperature. In this case, strain rate has a slight effect on the shape of the tensile stress-strain diagram. For other alloys, however, which exhibit a strong dependency of yield strength on temperature, the shape of the tensile stress-strain diagram is strongly influenced by the strain rate. Such alloys are said to be strain-rate sensitive.

Generally, for most engineering alloys, the stress-strain diagrams derived from a tensile test carried out at room temperature is changed by only a few percent if the strain rate is changed by an order of magnitude up to a temperature corresponding to about one-third of the melting point on the absolute temperature scale. Above that temperature, however, the dependence of stress-strain diagram on temperature becomes significant, requiring the use of standard strain rates to ensure reproducibility of the results.

4.3.2 Compression Test

In contrast with the case of tensile loading described above, a compressive loading condition results in a gradual increase in the load-carrying cross-sectional area. As a result, the phenomenon of necking is not encountered during the course of a compression test. Therefore, materials which behave in a brittle manner under the influence of tensile stresses become stronger in compression.

A compression test is carried out in the same testing machine used in the tension test. It is evident from Fig. 4.7 that reversing the direction of the moving crosshead subjects the specimen to a compressive stress. The test is carried out by fixing a standard specimen between two anvils and then a gradually increasing compressive load is applied, as shown in Fig. 4.13. Frictional forces developed along the specimen-anvil interface prevent surface layers of the specimen from being deformed. Therefore, the amount of deformation varies from 0 at the surface to a maximum at the center of the specimen resulting in the barrel shape shown in Fig. 4.13. Ductile materials such as the engineering alloys are seldom tested in compression to reveal properties relevant to design. However, the test is of particular importance in determining the stress required to cause plastic deformation or flow stress in metalworking operations, e.g., rolling, extrusion, and forging.

4.3.3 Creep and Stress Rupture Tests

As explained in Sec. 4.3.1, a tensile test specimen is plastically deformed as the applied stress is progressively increased. Since plastic deformation is a thermally
activated process, it can occur with time at a constant stress. This phenomenon is called creep. At temperatures below about \(0.5T_m\), where \(T_m\) is the melting point of the material on the absolute scale; creep deformation is negligible. Above that temperature, however, the application of a given stress well below the yield strength derived from a tensile test carried out at the same temperature can cause appreciable time-dependent plastic strain, which can eventually lead to fracture. Therefore, resistance to creep deformation assumes a primary importance in high temperature structural applications.

Typically, the plastic creep strain acquired by the material at the point of fracture is only a small fraction of the elongation to fracture observed in a tensile test carried out at the same temperature. Because the plastic strain accumulated in the material as a result of creep at a given temperature is dependent upon the applied stress, engineering structures intended for use at elevated temperatures are designed on the basis of a finite lifetime.

For engineering purposes, creep tests are conducted at constant load corresponding to a given engineering stress rather than at a constant true stress. Usually, the creep test is not continued to the point of fracture. Design criteria for high-temperature structural applications are typically based upon the time required to produce a given plastic strain under a specified applied stress, e.g.,
time to 0.5% or 1% creep strain. Also, it is a common practice to conduct creep tests for specified periods of time at given temperature and stress, e.g., 1000 or 10,000 hr, and determine the plastic strain accumulated in the specimen. Another related test is called the stress-rupture test. It is an accelerated creep test where an engineering stress greater than that employed in a conventional creep test is applied to the specimen until it fractures. In this case, both the time and elongation to rupture are used as design parameters.

The significance of creep as a design limiting factor in high-temperature structural applications can be understood in terms of the relationship between maximum operating temperature in a power cycle and efficiency of energy conversion. In a power cycle, the lower temperature is always limited by the ambient temperature. However, the maximum temperature is limited by the properties of materials used to manufacture the hottest components of the energy conversion system, e.g., boiler tubes of a steam power plant and turbine sections of gas turbine engines. Increasing the maximum temperature leads to an increase in efficiency resulting in more power at a lower cost.

Creep test data are recorded as plastic strain vs. time constructing what is known as a creep curve such as that shown in Fig. 4.14. A typical creep curve consists of three distinct regions: I, II, and III. Region I is referred to as the stage of transient or primary creep. During primary creep, the strain initially increases rapidly with time, and then decreases toward the end of this stage. Subsequently, during region II, which is known as the stage of steady-state creep, the strain rate becomes constant. At the onset of region III known as tertiary creep,
the strain rate increases rapidly until the specimen is fractured. Usually, the stage of steady-state creep has the longest duration, however, for a given material, this may significantly be influenced by both temperature and applied stress as described later in this section.

Microscopic aspects of creep deformation are discussed in Chap. 7. It suffices to mention here that during the stage of primary creep, the strain rate is dependent upon the initial structure of the material. As the material continues to plastically deform, it strain-hardens similar to the case of a tensile test. Strain hardening decelerates the creep rate towards the end of the primary stage. It is recalled from Chap. 3 that recrystallization results in softening or reducing the strength of a strain-hardened or cold-worked material. Another related process is called recovery, which results in partial softening of the material by relieving some of the effects of strain hardening without forming new grains. During steady-state creep the opposing effects of strain hardening and recovery are balanced resulting in a constant strain rate. Toward the end of steady-state creep, voids are formed in the material by various processes, causing deformation to be localized in the weakest section similar to the case of necking in a tensile test. This accelerates the creep rate during the tertiary stage, which is terminated by fracture.

Among the three creep stages described above, the stage of steady-state creep is of particular practical significance. Typically, a material selected to resist creep conditions is required to maintain specified dimensional tolerances after extended exposure at elevated temperatures. As an example, a part is designed such that the allowable increase in length is only 0.1% after 100,000 hr of service corresponding to a creep rate not in excess of $10^{-8}$ per hour. However, running a creep test for 100,000 hr is costly as well as time consuming. In practice, creep measurements are conducted up to a certain time and then the results are extrapolated to predict the behavior after extended periods of time. As shown in Fig. 4.15,

Figure 4.15 Extrapolation of creep data.
the creep test is discontinued at point \( x \) within the steady-state stage. Since this stage is represented by a straight line, the results are extrapolated to time \( t \), and the creep strain after time \( t \) (\( et \)) can be expressed as:

\[
et = e_0 + mt
\]  

(4.20)

where \( e_0 \) is the intercept of the steady-state line with the strain axis (Fig. 4.15) and \( m \) is the slope of the line corresponding to the steady-state creep rate. Both \( e_0 \) and \( m \) can be obtained from the short-term creep test, and the strain acquired after an extended time \( t \) is calculated from the above equation.

Increasing the test temperature at a given stress is found to produce three important effects. First, it shifts the entire creep curve into a higher strain level and corresponding shorter time, as schematically illustrated in Fig. 4.16. Second, it progressively reduces the duration of the primary and steady-state creep stages. Eventually, the creep rate is accelerated at the beginning of the test until the specimen is fractured. A similar effect to that described above is obtained if the temperature is held constant and the stress is increased.

For engineering applications, creep data can also be represented in terms of what is known as a Larson-Miller plot. To construct this plot, creep data such as time to 0.5 and 1% creep strain as well as time to rupture are obtained as functions of stress and temperature. A Larson-Miller parameter \( P \) is defined by

\[
P = T(\log t = 20)
\]  

(4.21)

where \( T \) is the test temperature in absolute degree (K), and \( t \) is the time to a specified creep strain or rupture. For each test temperature \( T \), the parameter \( P \) is plotted

\[\begin{align*}
\text{Strain} & \quad T_5 \to T_4 \to T_3 \to T_2 \to T_1 \\
\text{Time} & \quad (\text{Constant Stress})
\end{align*}\]

\[\begin{align*}
\text{Strain} & \quad \sigma_5 \to \sigma_4 \to \sigma_3 \to \sigma_2 \to \sigma_1 \\
\text{Time} & \quad (\text{Constant Temperature})
\end{align*}\]

Figure 4.16 Effect of temperature and stress on the shape of creep curve. (a) Temperature effect. (b) Stress effect.
against the stress. Larson-Miller plots are particularly useful in comparing the creep resistance of different alloys, as illustrated in Fig. 4.17.

4.3.4 Hardness Test

Although hardness is a rather complex property depending upon other properties of the material, it can roughly be defined as the resistance of the material to indentation. It can be used as a guide in evaluating the mechanical strength of materials. Because a hardness test is rather simple and quick, it is frequently used as an important means for quality control. Other applications of a hardness test include (i) evaluation of the thermal-mechanical history of the material, (ii) a rough estimate of the tensile strength of materials particularly steels, and (iii) detection of hardening effects produced by various processes such as phases changes, and carbon and nitrogen pickup. However, due to its complexity as well as ill definition, the results of hardness tests can be easily misused. Therefore, it is extremely important to evaluate hardness test data in terms of the particular method of measurement and also the exact measuring conditions.
Hardness tests can be broadly classified into two main types: (i) surface hardness tests and (ii) microhardness tests. For all these tests, however, hardness is measured from the local penetration of the material being tested by an indenter of a given material, size, and shape. In the surface hardness test, the indenter is placed at the surface of a polished specimen, and a given load is applied for a specific time. Then, the load is removed, and the area of indentation is measured. Hardness is expressed as load per unit area. A microhardness test is based upon the same principle; however, in this case the indenter is placed on a specific microstructural features viewed by means of a light optical microscope.

Various types of hardness tests include (i) Brinell test, (ii) Vickers or diamond ball test, (iii) Knoop test, and (iv) Rockwell test. Figure 4.18 illustrates the shape of indentation produced by the Brinell, Vickers, and Knoop tests. A Brinell test uses a steel ball usually 10 mm in diameter as an indenter. Other standard ball diameters can also be used. From a theoretical point of view, any load can be used to conduct the test. However, for metallic materials, a standard load of 3000 kg is usually used. After removing the load, the indentation diameter (Fig. 4.18a) is measured in at least two mutually perpendicular directions using a low-power light optical microscope. Tables are available for determining the indentation area from the measured average diameter. If such tables are inaccessible, however, the value of Brinell hardness number (HB) in units of kg/mm$^2$ can be calculated from

$$
HB = \frac{2P}{\pi D} \left[ D - (D^2 - d^2)^{1/2} \right]
$$

where $P$ is the load in kilogram, $D$ is the diameter of the indenter ball in millimeter, and $d$ is the average indentation diameter in millimeters.

Since a steel ball is used in the Brinell test, the values of HB are reliable only for up to 450. Above this value, the steel ball itself can be deformed. Limitations of this test include (i) variability of the depth of indentation leading to inaccuracy in measuring the hardness value and (ii) possible deformation of

![Figure 4.18](image)

**Figure 4.18** Shape of indentation produced by various hardness tests. (a) Brinell. (b) Vickers. (c) Knoop.
the steel ball by relatively harder materials. These limitations can be overcome by the Vickers (HV) or diamond pyramid hardness test (DPH).

In the DPH test, a diamond indenter having the shape of a square pyramid with face angles of 136° is used. By measuring the two diagonals of indentation (Fig. 4.18b) using a light optical microscope usually built in the system, the indentation area is calculated from the average diagonal, and the value of HV or DPH in kg/mm² is given by

$$HV = 1.8544P/D^2$$  \hspace{1cm} (4.23)

where $D$ is the average diagonal of indentation. Since a diamond indenter is used in the DPH test, the hardness of any material can be measured, and therefore the test is more versatile than the Brinell test. By means of very small loads and a high-power light optical microscope, the DPH test can also be used as a micro-hardness test.

A Knoop hardness test is similar to the Vickers test described above. However, it is almost exclusively used as a microhardness test. In this case, the diamond indenter is rhomboid, and therefore it produces an elongated indentation (Fig. 4.18c). To obtain the Knoop hardness number KHN, the longer diameter of the indentation $l$ is measured, and the hardness value is given by

$$KHN = 14.2P/l^2$$  \hspace{1cm} (4.24)

where $P$ is the load. Usually, the Knoop test is used to measure the hardness of thin surface layers such as plated surface layers and carburized layers.

A common drawback of the Brinell, Vickers, and Knoop tests is that the test procedure is rather lengthy and cumbersome, requiring precise measurement of indentation followed by calculation of the hardness from tables. These difficulties are eliminated by using the Rockwell hardness test, characterized by being a faster test providing a direct and an efficient method for accurately measuring the hardness value as described below.

Multiple indenters and standardized test loads are used in the Rockwell test in addition to a preload minimizing the effect of surface roughness and enabling the hardness value to be directly measured and read on a dial gauge. Each combination of an indenter and load has a characteristic hardness scale of an arbitrarily set maximum value corresponding to zero penetration of the specimen by the indenter. Such scales are identified as $A$, $B$, and $C$ and the corresponding hardness values read $R_A$, $R_B$, and $R_C$. Figure 4.19 schematically illustrates the procedure of a Rockwell hardness test. First, the indenter is placed in contact with the specimen surface (Fig. 4.19a). A relatively small preload, usually 10 kg, is then applied to provide a well-defined reference position $h_1$ for measurement of the depth of indentation (Fig. 4.19b). Then, the major load is applied, causing a
penetration $h_2$ corresponding to the hardness of the material (Fig. 4.19c). Finally, the main load is removed to eliminate the portion of elastic deformation receding the indenter to a depth $h_3$ (Fig. 4.19d). From the depth difference $h_3 - h_1$, the dial gauge registers the value of Rockwell hardness.

### 4.3.5 Impact Toughness Test

As pointed out earlier, the area enclosed by the tensile stress-strain diagram represents the energy absorbed by the test specimen before it fractures, which is a measure of toughness—a property involving both strength and ductility. However, toughness is closely related to the impact resistance of the material. In practice, materials can have various types of stress raisers such as internal cracks, grain boundaries, and second phase particles. For this reason, impact toughness tests are conducted on standard notched specimens, and therefore they are also called notch toughness tests. The Charpy test described below is the most commonly used to measure toughness.

A standard specimen with a notch cut at midsection as shown in Fig. 4.20a is used in the Charpy test. First, the specimen is supported at both ends. Then, a pendulum of known mass, usually 27.2 kg, is released from a given height and is made to strike the specimen at midsection opposite to the notch at a certain speed typically 5.3 m/sec. As a result, the specimen is fractured, and the pendulum swings upward as shown in Fig. 4.20b. By definition, the work done $W$ by the pendulum in fracturing the specimen is the energy absorbed by the material.
Figure 4.20 Charpy impact toughness test. (a) Impact toughness test specimen. (b) Typical impact toughness test equipment. (c) Impact toughness test illustrating the ductile-to-brittle transition temperature.
before it fractures or its toughness. It is given by

\[ W = m(h_0 - h_1) \] (4.25)

where \( m \) is the mass of the pendulum, \( h_1 \) is the final height of the pendulum, and \( h_0 \) is the initial height.

An important characteristic of impact toughness is its sensitivity to temperature. Many of the engineering materials exhibit a pronounced change in impact toughness when the temperature is lowered, as schematically shown in Fig. 4.20c. Although toughness reflects the notch sensitivity of the material resulting from stress raisers, it is not directly used as a design parameter like the yield strength. However, the transition temperature shown in Fig. 4.20c is an extremely useful guide in the selection of materials performing under low-temperature conditions.

### 4.3.6 Fatigue Test

It can be concluded from a tensile test that a given material should perform safely if it is subjected to a static load producing a stress below its yield strength. However, repeated application of a relatively smaller stress in a cyclic manner, as shown in Fig. 4.21 where the maximum stress is normally safe if it is applied only once, can lead to failure. This behavior is known as fatigue. Since fatigue failure resulting from relatively small stresses occurs after a very large number of cycles exceeding \( 10^6 \), it is called low-stress high-cycle fatigue or shortly high-cycle fatigue. It is the most common mode of failure of rotating machine parts, e.g., rotating drive shaft and torsion bar.

Another type of fatigue results when the maximum stress is relatively high producing macroscopic strain in each cycle. In this case fatigue failure occurs after a smaller number of cycles less than \( 10^5 \). This type of fatigue, which is called high-stress low-cycle fatigue or shortly low-cycle fatigue is simply demonstrated by repeated bending of a paper clip. Fatigue can also result from thermal stresses generated by rapid temperature cycling or temperature gradients as occur in steam and gas turbine engines, steam boilers, and nuclear equipment.

In a typical fatigue test a standard specimen is loaded in pure bending as shown in Fig. 4.21a to produce tensile and compressive stresses. When the specimen is made to rotate, it becomes subject to alternating tensile and compressive stresses as shown in Figure 4.21b. The Moore test setup illustrated in Fig. 4.21c is commonly used to measure fatigue strength. If the specimen is rotated under the effect of a bending load, its circumference passes from a state of tension into a state of compression upon completion of each revolution constituting the stress cycle shown in Fig. 4.21b. Each cycle is repeated several thousand times per minute until the specimen is fractured. It is the objective of the test to determine the
number of cycles $N$ to failure as a function of stress amplitude $\sigma$. This is known as the fatigue life. Therefore, the test is repeated for different stress amplitudes to obtain a set of $N$ and $\sigma$ values. A plot is then constructed illustrating the effect of the stress amplitude $\sigma$ on the number of cycles to failure $N$. It is usually called the $\sigma$-$N$ curve where $\sigma$ refers to stress and $N$ refers to number of cycles. A typical $\sigma$-$N$ plot is schematically illustrated in Fig. 4.22. As can be seen, the number of cycles to failure increases as the stress decreases until a stress is reached below which the number of cycles to failure becomes infinite or larger than $10^8$. This critical stress is called the ‘endurance limit’. In a low-cycle fatigue test, the strain rather than the stress is plotted against the number of cycles to failure, and in this case, the plot is called the e-$N$ curve.

Typically, the $\sigma$-$N$ and e-$N$ curves are influenced by a number of variables, which can lead to a considerable scatter in the test data. Among these variables are surface finish, surface hardness, nature of the environment, and presence of residual stresses.

Copyright © 2004 by Marcel Dekker, Inc.
4.3.7 Fracture Toughness Test

A fracture toughness test is used to measure the resistance of the material to fracture under specified conditions in terms of a material property known as fracture toughness. Two fracture conditions are distinguished depending upon the dimensions of the specimen: (i) fracture under plane strain condition and (ii) fracture under plane stress condition defining two fracture toughness parameters, as discussed in Chap. 6. It suffices to mention here that fracture under plane stress condition occurs when the specimen is relatively thin. In this case, the applied stress during the test develops a state of biaxial stress in front of the crack. However, if the specimen is rather thick, a state of plane strain is developed ahead of the crack. As explained in Chap. 6, a stress intensity factor $K$ is defined in terms of the applied stress and crack length. Under plane stress condition, fracture occurs when $K = K_c$, where $K_c$ is the critical stress intensity factor or fracture toughness under plane stress condition. If the specimen is sufficiently thick to develop a state of plane strain, fracture occurs when $K = K_{lc}$, where $K_{lc}$ is the fracture toughness under plane strain condition. It is the objective of the fracture toughness test to determine the fracture toughness parameter usually under plane strain condition, i.e., $K_{lc}$.

Figure 4.23 illustrates the geometry of a compact tensile specimen commonly used in fracture toughness tests. Prior to the test, a crack is introduced into the specimen by fatigue loading. Pins inserted through holes bored into the specimen are used to apply a tensile load in the same testing machine used in the tension test. After completion of the test, the length of the preexisting fatigue crack is measured at different positions. Different expressions are developed for the fracture toughness parameter.
to relate the stress intensity factor $K$ with the applied load $P$, the crack length $c$, and the specimen geometry. For the compact tensile test specimen shown in Figure 4.23, $K$ is given by

$$K = \frac{P}{BW^{1/2}}[16.7(c/W)^{1/2} - 104.7(c/W)^{3/2} + 369.9(c/W)^{5/2} - 573.8(c/W)^{7/2} + 360.8(c/W)^{9/2}]$$

where $B$ is the specimen thickness and $W$ is the distance from the center of holes and specimen edge as shown in Fig. 4.23.

### 4.3.8 Torsion Test

Similar to the compression test, the torsion test is not as widely used as the tension test to determine the mechanical properties of engineering alloys. Its objective is to determine the shear strength of the material such as the modulus of elasticity in shear, the torsional yield strength, and the modulus of rupture. In addition to test specimens, the torsion test can be carried out on full size parts such as shafts and axles.

During the course of a torsion test, a cylindrical specimen is subjected to a twisting moment or torque in a special machine, as illustrated in Fig. 4.24a. As a result, a shear stress is developed in the specimen, which varies linearly with radial distance from the center of the cylinder, i.e., it varies from zero at the center of the cylinder to a maximum value at the surface (Fig. 4.24b). The shear stress $\tau$ as a function of radial distance $r$ can be expressed as

$$\frac{\tau}{r} = \frac{\tau_{\text{max}}}{R}$$

or

$$\tau = \tau_{\text{max}} \left(\frac{r}{R}\right)$$

(4.27)

Copyright © 2004 by Marcel Dekker, Inc.
where $\tau_{\text{max}}$ is the maximum shear stress developed at the surface and $R$ is the radius of test specimen.

As shown in Fig. 4.24b, the shear stress $\tau$ acts on an elemental area $dA$ corresponding to an elemental force $dF$, which is given by

$$dF = \tau dA$$

(4.28)

Combining Eqs. (4.27) and (4.28) results in

$$dF = \tau_{\text{max}} \left( \frac{r}{R} \right) dA$$

(4.29)

It is recalled from Eq. (3.9) that a torque $T$ is the product of force and the perpendicular distance between the axis of rotation and the line of action of the force. Therefore, the elemental torque $dT$ corresponding to $dF$ can be expressed as

$$dT = dF \cdot r = \tau_{\text{max}} \frac{r^2}{R} dA$$

(4.30)

and the total torque $T$ can be obtained by integrating $dT$ across the entire cross section, i.e.,

$$T = \int_A dT = \int_A \tau_{\text{max}} \frac{r^2}{R} dA$$

Because $\tau_{\text{max}}$ and $R$ are constants, the above equation can be rewritten as

$$T = \tau_{\text{max}} / R \int_A r^2 dA$$
By definition, the term $\int r^2 dA$ is called the moment of inertia of the cross section $I$, and therefore, the above equation can be rewritten as

$$\tau_{\text{max}} = TR/I$$

(4.31)

For a solid cylinder, $I = \pi d^4/32$ where $d$ is the cross section diameter. Substituting the value of $I$ into Eq. (4.22) results in

$$\tau_{\text{max}}(\text{solid cylinder}) = T d/2(32/d^4)$$

or

$$\tau_{\text{max}} = 16T/\pi d^3$$

(4.32)

For a hollow cylinder of an external diameter $d_1$ and internal diameter $d_2$, $I = \pi/32 (d_1^4 - d_2^4)$, and $\tau_{\text{max}}$ is given by

$$\tau_{\text{max}}(\text{hollow cylinder}) = 16Td_1/\pi(d_1^4 - d_2^4)$$

(4.33)
Stress Analysis

5.1 Introduction

Stress analysis is a branch of mechanics dealing with evaluating the stresses developed in a part subjected to mechanical forces. It has important applications in engineering design, as well as in failure analysis investigations. By means of stress analysis, the designer can evaluate the stress pattern developed by applied mechanical forces. On the basis of this evaluation, proper materials with specific mechanical properties can be selected so that the product can safely function during service. In relation to failure analysis investigations, it is possible to evaluate the stresses in the failed component at the time of failure. A comparison of these stresses with the allowable or design stress provides important clues about the possible causes of failure.

In general, stress analysis is carried out on the basis of mechanical equilibrium conditions defined in Chap. 3. It is recalled that the first condition of mechanical equilibrium requires that the sum of forces acting along any given direction must vanish [Eq. (3.15)]. This is the only condition of mechanical equilibrium, which must be satisfied if the forces are of the concurrent type. Once the forces are determined, the stresses can be calculated by dividing the forces by the respective areas upon which they are acting. If the forces are of the nonconcurrent type, the preceding condition must be supplemented by the condition that the sum of torques acting on the body must also vanish [Eq. (3.16)].

It is the objective of this chapter to develop the principles governing stress analysis. From these principles, it is shown that once the state of stress at any
point within a part is known, it is possible to determine the state of stress at any other point. A uniaxial state of stress is the simplest case of stress analysis to be considered as described below.

5.2 Uniaxial State of Stress

It is recalled from Sec. 3.10 that if a part is acted upon by a tensile force \( F \) along its geometric axis, any arbitrary section of the part is acted upon by a tensile stress \( \sigma \) and a shear \( \tau \) given by

\[
\sigma = \sigma^2 \quad (5.1)
\]

where \( \sigma \) is the stress acting on the cross-sectional area normal to the applied force and \( l \) is the direction cosine defined as the cosine of the angle \( \alpha \) between the normals to the cross-sectional area and an arbitrary section, and

\[
\tau = \sigma/2 \sin 2\alpha \quad (5.2)
\]

When the angle \( \alpha = 0 \) corresponding to the cross-sectional area, the preceding equations become

\[
\sigma' = \sigma \quad \text{and} \quad \tau = 0
\]

Figure 5.1 illustrates the uniaxial state of stress. It is noted that there is only one set of parallel planes acted upon by only a normal stress \( \sigma \); i.e., the shear stress acting on these planes is \( \tau = 0 \). Also, the normal stress acting on these planes is the maximum possible normal stress developed in the part. To distinguish the planes free of shear stress, they are called the principal planes, and the normal stress acting on these planes is called the principal stress.

Any plane within the part other than the principal plane is acted upon by both a normal stress and a shear stress. It is evident from Eq. (5.2) that the shear stress reaches a maximum value \( \tau_{\max} \), given by

\[
\tau_{\max} = \sigma/2 \quad (5.3)
\]

corresponding to \( \alpha = 45^\circ \). On these planes, the normal stress \( \sigma' \) is given by

\[
\sigma' = \sigma/2 \quad (5.4)
\]

A stress defined by the equation stress = force/area represents an average stress over the area. If this average stress is constant over the area, it is said to be uniform. Generally, however, the stress may not be uniform. In this case, the area can be made infinitesimally small, approaching a point to define a uniform stress at a point. It is rather more convenient to describe the stress at a point in terms of a reference coordinate system such that the \( x \) axis is normal to the plane passing through the point and defined by the \( y \) and \( z \) axes, as shown in Fig. 5.1. In this
In the uniaxial case, the normal stress lies along the $x$ axis, and the shear stress $\tau$ lying in the $y$-$z$ plane can be resolved into two components along the $y$ and $z$ axes.

Conventionally, the normal stress component is identified by one subscript corresponding to the coordinate of the axis along which it is acting; i.e., the normal stress $\sigma$ in Fig. 5.1d is labeled $\sigma_x$. In contrast, a shear stress component is identified by two subscripts; the first subscript corresponds to the coordinate axis normal to the shear plane, and the second subscript identifies the coordinate axis coinciding with the direction of the shear stress. Therefore, the two shear stress components in Fig. 5.1 are labeled $\tau_{xy}$ and $\tau_{xz}$, where

$$\tau^2 = \tau_{xy}^2 + \tau_{xz}^2$$  \hspace{1cm} (5.5)

Replacing $\sigma'$ in Eq. (5.1) with $\sigma_x$ results in

$$\sigma_x = \hat{P} \sigma_1$$  \hspace{1cm} (5.6)
Squaring Eq. (5.2) and combining it with Eqs. (5.5) and (5.6), recalling that \( l = \cos \alpha \) and that \( \cos^2 \alpha + \sin^2 \alpha = 1 \), results in

\[
\tau^2 = \tau_{xy}^2 + \tau_{xz}^2 = l^2 \sigma_1^2 - \sigma_x^2
\]  
(5.7)

It is evident from Eqs. (5.6) and (5.7) that if the state of stress at any point within the part is known, i.e., \( \sigma_x, \tau_{xy}, \) and \( \tau_{xz} \) are given, it is possible to determine the principal stress \( \sigma_1 \) as well as its orientation. Conversely, if the principal stress \( \sigma_1 \) and its orientation are known, the state of stress at any given point within the part can be determined.

Some important conclusions can be drawn from the above analysis of uniaxial state of stress, which can be summarized as follows:

1. It is evident from Fig. 5.1 that there is only one set of parallel planes on which the maximum normal \( \sigma_1 = P/A_0 \) acts, and on these planes, the shear stress \( \tau = 0 \). Such planes are uniquely defined within the body, and to distinguish them from other planes, they are called principal planes. Also, the stress \( \sigma \) is called principal stress; i.e., the principal stress is normal to the principal plane.

2. If the principal stress is \( \sigma_1 \), the plane oriented at 45° from principal stress is acted upon by (i) the maximum possible shear stress \( (\tau_{\max}) \) developed in the body where \( \tau_{\max} = \sigma_1/2 \) and (ii) a normal stress \( \sigma' = \sigma_1/2 \).

To summarize the preceding discussion, the uniaxial state of stress at any point within the part is completely defined in terms of one normal stress component \( \sigma_x \) and two shear stress components \( \tau_{xy} \) and \( \tau_{xz} \). If the uniaxial state of stress at any point is completely defined, it is possible to determine both the magnitude and orientation of the principal stress. Conversely, if the magnitude and orientation of the principal stress are known, it is possible to determine the state of stress at any point. As described below, this analysis can be extended to cover more complex states of stress.

### 5.3 Generalized State of Stress

During service, engineering structures and machine parts are usually subjected to more complex states of stress than the uniaxial state described in the previous section. It is therefore required to develop means by which more complicated states of stress can be defined. Specifically, the primary objective is to be able to evaluate the stresses developed in the part from knowledge of the state of stress at any given point.

If the part is subjected to more than one force acting along different directions, it is possible to extend the argument used for the uniaxial state of stress to
describe the most generalized triaxial or three-dimensional state of stress. In this case, the stress at a point is defined in terms of all stress components acting on three mutually perpendicular planes of an elemental cube enclosing the point, as shown in Fig. 5.2. As can be seen, the stress components are

\[ \sigma_x, \tau_{xy}, \text{ and } \tau_{xz} \] acting on the plane defined by the y and z axes

\[ \sigma_y, \tau_{yx}, \text{ and } \tau_{yz} \] acting on the plane defined by the x and z axes

\[ \sigma_z, \tau_{zx}, \text{ and } \tau_{zy} \] acting on the plane defined by the z and y axes.

Since the part is in a state of mechanical equilibrium, both Eqs. (3.15) and (3.16) must be satisfied. By applying Eq. (3.15), stresses of identical magnitude and opposite directions must be acting on parallel faces, as shown in Fig. 5.2. Application of Eq. (3.16) results in

\[ \tau_{xy} = \tau_{yx} \]

\[ \tau_{xz} = \tau_{zx} \]

\[ \tau_{yz} = \tau_{zy} \] \hspace{1cm} (5.8)

To summarize, the most generalized triaxial state of stress at any point within the part is completely defined in terms of six stress components:

a. Three normal components \( \sigma_x, \sigma_y, \text{ and } \sigma_z \)
b. Three shear components \( \tau_{xy}, \tau_{xz}, \text{ and } \tau_{yz} \)

Since the uniaxial state of stress is defined in terms of one principal stress \( \sigma_1 \) and one set of principal planes, the more generalized triaxial state of stress can

**Figure 5.2** Illustration of the generalized or three-dimensional state of stress.
be defined in terms of three mutually perpendicular normal stresses $\sigma_1$, $\sigma_2$, and $\sigma_3$, and three mutually perpendicular sets of principal planes. By convention $\sigma_1 > \sigma_2 > \sigma_3$; i.e., $\sigma_1$ is the maximum normal stress developed in the part, and $\sigma_3$ is the minimum normal stress taking into consideration the sign of the stress (a tensile stress is positive, and a compressive stress is negative). As demonstrated in Sec. 5.2 for the uniaxial state of stress, the three sets of principal planes are uniquely identified, where each set is characterized by the condition that the shear stress component $\tau = 0$.

Different states of stress can be distinguished by the magnitudes of the three principal stresses, as summarized in Fig. 5.3. When two of the three principal stresses vanish, the state of stress is reduced to the uniaxial state described in Sec. 5.2. In this case, the state of stress is defined in terms of one principal stress, which can either be positive (tensile) or negative (compressive). Using the convention $\sigma_1 > \sigma_2 > \sigma_3$, if the stress is positive, it corresponds to the maximum stress $\sigma_1$. Conversely, if it is negative, it corresponds to the minimum stress.

Figure 5.3 The three possible states of stress. (a) Uniaxial. (b) Biaxial. (c) Triaxial.
Two principal stresses are required to define a biaxial state of stress, where three different cases can be encountered: (i) If the two stresses are positive (tensile), the larger stress corresponds to \( \sigma_1 \), the smaller stress corresponds to \( \sigma_2 \), and \( \sigma_3 = 0 \); (ii) if one stress is positive, and the other stress is negative, the positive stress corresponds to \( \sigma_1 \), the negative stress corresponds to \( \sigma_3 \), and \( \sigma_2 = 0 \); and (iii) if the two stresses are negative, the maximum stress is \( \sigma_1 = 0 \), the less negative stress corresponds to \( \sigma_2 \), and the more negative stress corresponds to \( \sigma_3 \). Such distinction must be made in order to correctly calculate the maximum shear stress, as demonstrated later. For a triaxial state of stress, the maximum and minimum stresses become evident taking their signs into consideration. In the extreme case where \( \sigma_1 = \sigma_2 = \sigma_3 \), a state of hydrostatic stress is defined.

Having identified the stress components required to define the most generalized three-dimensional or triaxial state of stress at a given point within a body, it is possible to determine the state of stress at any other point either analytically or graphically. Frequently, structures and machine parts encountered in practice have one dimension shorter than the others. In this case, the principal stress normal to the shorter dimension, e.g., thickness of the part, is relaxed; i.e., it approaches 0, which reduces the state of stress into a biaxial (Fig. 5.3b). Such a state of stress is known as plane stress. An account for both the analytical and graphical methods of stress analysis is given below.

### 5.3.1 Analytical Method

If the six stress components at any point within the part are \( \sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{xz}, \) and \( \tau_{yz} \), it can readily be shown that the principal stresses can be calculated from the following cubic equation:

\[
\sigma^3 - \sigma^2(\sigma_x + \sigma_y + \sigma_z) + \sigma(\sigma_x \sigma_y + \sigma_y \sigma_z + \sigma_z \sigma_x - \tau_{xy}^2 - \tau_{xz}^2 - \tau_{yz}^2) - (\sigma_x \sigma_y \sigma_z + 2\tau_{xy} \tau_{xz} - \sigma_x \tau_{yz}^2 - \sigma_y \tau_{xz}^2 - \sigma_z \tau_{xy}^2) = 0
\]

(5.9)

It is evident that there are three values of \( \sigma \), or three roots of the above equation, corresponding to the three principal stresses. As demonstrated for the uniaxial state of stress (Sec. 5.2), the orientation of the principal plane within the part is uniquely defined. Also, for the most generalized triaxial state of stress described by Eq. (5.9), the orientation of the principal planes are uniquely identified regardless of which coordinate system is selected to describe the state of stress. As indicated earlier, many parts can be analyzed in terms of a biaxial or plane state of stress. In this case, the state of stress at any point within the part is completely defined in terms of three stress components \( \sigma_x, \sigma_y, \) and \( \tau_{xy} \), noting that \( \tau_{xy} = \tau_{yx} \), and the remainder of the stress components vanish, reducing the
cubic stress equation to
\[ \sigma^3 - \sigma^2(\sigma_x + \sigma_y) + \sigma(\sigma_x\sigma_y - \tau_{xy}^2) = 0 \]

or
\[ \sigma[\sigma^2 - \sigma(\sigma_x + \sigma_y) + (\sigma_x\sigma_y - \tau_{xy}^2)] = 0 \quad (5.10) \]

Since \( \sigma = 0 \) satisfies the preceding equation, it is one of its roots consistent with the biaxial state of stress. It is clear that the other two roots corresponding to the two nonzero principal stresses can be determined by solving the quadratic equation:
\[ \sigma^2 - \sigma(\sigma_x + \sigma_y) + (\sigma_x\sigma_y - \tau_{xy}^2) = 0 \quad (5.11) \]

Since the principal stresses and planes are uniquely identified, they are independent of the coordinate system selected to describe the state of stress at any point including the principal coordinate system. Mathematically, this means that the coefficients of Eq. (5.11) are constants, i.e.,
\[ \sigma_x + \sigma_y = \sigma_x' + \sigma_y' = \ldots = \sigma_1 + \sigma_2 \quad (5.12) \]

and noting that the shear stress component \( \tau = 0 \) on the principal planes,
\[ \sigma_x\sigma_y = \sigma_x'\sigma_y' = \sigma_1\sigma_2 \quad (5.13) \]

Equation (5.11) is a quadratic equation of the general type
\[ ax^2 + bx + c = 0 \]

which has two roots given by:
\[ x = \left(-b \pm \sqrt{b^2 - 4ac}\right)/2a \quad (5.14) \]

By comparing the coefficients of Eqs. (5.11) and (5.13), the two principal stresses defining a biaxial state of stress are given by
\[ \sigma = \frac{1}{2}(\sigma_x + \sigma_y) \pm \sqrt{\left[\frac{1}{2}(\sigma_x - \sigma_y)\right]^2 + \tau_{xy}^2} \frac{1}{2} \]

or
\[ \sigma' = \frac{1}{2}(\sigma_x + \sigma_y) + \sqrt{\left[\frac{1}{2}(\sigma_x - \sigma_y)\right]^2 + \tau_{xy}^2} \frac{1}{2} \]
\[ \sigma'' = \frac{1}{2}(\sigma_x + \sigma_y) - \sqrt{\left[\frac{1}{2}(\sigma_x - \sigma_y)\right]^2 + \tau_{xy}^2} \frac{1}{2} \quad (5.15) \]

where \( \sigma' > \sigma'' \).
It is recalled from Sec. 5.2 that the maximum shear $\tau_{\text{max}}$ corresponding to a uniaxial state of stress defined by the principal stress $\sigma_1$ is given by $\tau_{\text{max}} = \frac{1}{2}\sigma_1$ and that the normals to the planes of $\tau_{\text{max}}$ are oriented at $45^\circ$ from the direction of the principal stress or principal axis. A similar expression can be derived for the biaxial state of stress as described below.

Consider an element of a part subjected to a biaxial state of stress defined by the principal stresses $\sigma_1$ and $\sigma_2$, where $\sigma_1 > \sigma_2$ as shown in Fig. 5.4a. Assume that the element has a unit thickness along the $z$ axis, and its dimensions along the $x$ and $y$ axes are $dx$ and $dy$, respectively, and it is required to determine the shear stress $\tau$ developed parallel to an arbitrary plane within the element inclined at an angle $\theta$ to the $x$-$y$ plane (Fig. 5.4a). It is recalled that the procedure of stress analysis is based upon balancing the forces acting on the part along certain directions. In the following analysis, it is assumed that $\sigma_1$ is the maximum stress and $\sigma_2$ is the minimum stress.

Since the stress $\sigma_1$ acts along a direction normal to an area given by $(dy)$ (1) as shown in Fig. 5.4a, the corresponding tensile force $F_{1t}$ is given by $F_{1t} = \sigma_1 (\text{area}) = \sigma_1 dy$. Similarly, the tensile force $F_{2t}$ corresponding to the stress $\sigma_2$ is given by $\sigma_2 dx$. From the geometry of Fig. 5.4b, the shear force component of $F_{1t}$ parallel to the arbitrary plane is $F_{1t} \cos(90^\circ - \theta) = F_{1t} \sin \theta$, and that of $F_{2t}$ acting in opposite direction is $F_{2t} \cos \theta$. Therefore, the net shear force $F_s$ is given by

$$F_s = F_{1t} \sin \theta - F_{2t} \cos \theta = \sigma_1 dy \sin \theta - \sigma_2 dx \cos \theta$$

**Figure 5.4** Determination of shear stress corresponding to a biaxial state of stress. (a) An element of a part subjected to a biaxial state of stress. (b) Analysis of stresses acting on an arbitrary plane.

Copyright © 2004 by Marcel Dekker, Inc.
Since the above shear force acts parallel to an area $dA$, the shear stress $\tau = F_3/dA$, or

$$\tau = (\sigma_1 \, dy \sin \theta - \sigma_2 \, dx \cos \theta)/dA$$

(5.16)

From the geometry of Fig. 5.4a, $dA$ is given by

$$dA = ds (l) = ds$$

and from Fig. 5.4b,

$$\sin \theta = dx/ds,$$

$$\cos \theta = dy/ds$$

Therefore, Eq. (5.16) becomes:

$$\tau = (\sigma_1 \sin \theta \cos \theta \, ds - \sigma_2 \sin \theta \cos \theta \, ds)/ds = (\sigma_1 - \sigma_2)(\sin \theta \cos \theta)$$

By definition, $\sin 2\theta = 2 \sin \theta \cos \theta$, i.e., $\sin \theta \cos \theta = \sin 2\theta/2$. Substituting into the above equation results in

$$\tau = [(\sigma_1 - \sigma_2)/2] \sin 2\theta$$

(5.17)

Since $\sigma_1$ and $\sigma_2$ are constants, it is evident from the preceding equation that the shear stress $\tau$ is a sinusoidal function of the angle $\theta$. When $\theta = 0$, $\tau = 0$, corresponding to the orientation of a principal plane, which by definition is free of shear stress. A maximum value of $\tau$ is reached on planes oriented at $45^\circ$ from the principal stress. In this case, $\sin 2\theta = \sin 90 = 1$, and the maximum shear stress $\tau_{\text{max}}$ is given by

$$\tau_{\text{max}} = (\sigma_{\text{max}} - \sigma_{\text{min}})/2$$

(5.18)

To correctly calculate the maximum shear stress using Eqs. (5.15) and (5.18), it is essential to identify the maximum and minimum principal stresses, i.e., $\sigma_1$ and $\sigma_3$, remembering that by convention $\sigma_1 > \sigma_2 > \sigma_3$. As explained earlier, three cases can be distinguished:

1. $\sigma'$ is positive (tensile) and $\sigma''$ is negative (compressive). In this case, $\sigma_1 = \sigma'$, $\sigma_2 = 0$, $\sigma_3 = \sigma''$, and therefore, from Eq. (5.15), $\tau_{\text{max}}$ is given by

$$\tau_{\text{max}} = (\sigma' - \sigma'')/2 = \pm \left[\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right]^{1/2}$$

(5.19)

2. $\sigma'$ is positive (tensile) and $\sigma''$ is positive (tensile). In this case, $\sigma_1 = \sigma'$, $\sigma_2 = \sigma'$, and $\sigma_3 = 0$, and therefore $\tau_{\text{max}}$ is given by

$$\tau_{\text{max}} = (\sigma' - 0)/2 = \sigma_1/2$$

(5.20)
3. $\sigma'$ is negative (compressive) and $\sigma''$ is negative (compressive), noting that $\sigma' > \sigma''$, i.e., $\sigma'$ is less negative than $\sigma''$. In this case, it is evident that $\sigma = 0$ is the maximum stress, i.e., $\sigma_1 = 0$, and $\sigma''$ is the minimum stress, i.e., $\sigma'' = \sigma_3$, and therefore $\tau_{\text{max}}$ is given by

$$\tau_{\text{max}} = \frac{1}{2} \left[ 0 - (-\sigma_3) \right] = \sigma_3 / 2$$  \hspace{1cm} (5.21)

Having determined the principal stresses and maximum shear stress, it remains to be determined their orientations within the part. It can readily be shown that for the most generalized triaxial state of stress, the direction cosines $l$, $m$, and $n$ of each principal stress $\sigma$ relative to a selected $x$-$y$-$z$ coordinate system are given by

$$l(\sigma - \sigma_x) - m\tau_{xy} - n\tau_{xz} = 0$$
$$-l\tau_{xy} + m(\sigma - \sigma_y) - m\tau_{yz} = 0$$
$$-l\tau_{xz} - m\tau_{yz} + n(\sigma - \sigma_z) = 0$$  \hspace{1cm} (5.22)

For a biaxial state of stress defined by $\sigma_x$, $\sigma_y$, and $\tau_{xy}$ ($\sigma_z = \tau_{xz} = \tau_{yz} = 0$), the above equations become

$$l(\sigma - \sigma_x) = 0$$
$$-l\tau_{xy} + m(\sigma - \sigma_y) = 0$$
$$n(\sigma) = 0$$  \hspace{1cm} (5.23)

By combining the preceding equations and noting that $l^2 + m^2 + n^2 = 1$, the values of $l$, $m$, and $n$ can be calculated. Once the directions of principal stresses, and in turn the orientations of the principal planes, are determined, the orientations of the planes of maximum shear stress are fixed by the condition that their normals are oriented at 45° from the directions of principal stresses. It is recalled from Sec. 5.2 that for a uniaxial state of stress, the normal stress $\sigma$ acting on the planes of maximum shear stress is $\sigma_1 / 2$; this is modified to

$$\sigma = \frac{1}{2} (\sigma_1 + \sigma_2) = \frac{1}{2} (\sigma_x + \sigma_y)$$  \hspace{1cm} (5.24)

for the biaxial state of stress. Also, Eqs. (5.6) and (5.7), which determine the normal and shear stresses at any point within a part subjected to a uniaxial state of stress, become

$$\sigma = l^2 \sigma_1 + m^2 \sigma_2 + n^2 \sigma_3$$
$$\tau^2 = l^2 \sigma_1^2 + m^2 \sigma_2^2 + n^2 \sigma_3^2 - \sigma^2$$  \hspace{1cm} (5.25)

where $l$, $m$, and $n$ are the direction cosines of the plane passing through the point relative to the principal axes. For a biaxial state of stress, the above equations are
reduced to

\[ \sigma = l^2 \sigma_1 + m^2 \sigma_2 \]
\[ \tau^2 = l^2 \sigma_1^2 + m^2 \sigma_2^2 - \sigma^2 \]

(5.26)

It follows from Eq. (5.18) that for a triaxial state of stress defined by the normal principal stresses \( \sigma_1, \sigma_2, \) and \( \sigma_3, \) the corresponding maximum shear stresses are given by

\[ \tau_1 = (\sigma_1 - \sigma_3)/2 \]
\[ \tau_2 = (\sigma_1 - \sigma_2)/2 \]
\[ \tau_3 = (\sigma_2 - \sigma_3)/2 \]

(5.27)

where \( \tau_1 > \tau_2 > \tau_3. \) Since the shear stress can either be positive or negative, the planes of maximum shear stresses bisect the angle between the two respective principal stresses; i.e., the angle between normal to a plane of maximum shear stress and the principal stress is 45°, as shown in Fig. 5.5.

5.3.2 Graphical Method: Mohr’s Circle

To avoid some of the mathematical difficulties encountered with the analytical method described above, Mohr’s circle is developed to graphically represent the state of stress. It allows rapid and exact determination of (i) the maximum

---

Figure 5.5 A schematic illustrating the relative orientations of principal planes and planes of maximum shear stress.
and minimum principal stresses ($\sigma_1$ and $\sigma_3$), (ii) the maximum shear stress, (iii) orientations of the principal stresses and maximum shear stresses, (iv) the normal stress acting on the plane of maximum shear stress, and (v) the state of stress at any point within the part being analyzed.

By definition, Mohr’s circle is a two-dimensional plot representing the state of stress at any point within the part being analyzed. As described for the analytical method, a biaxial or plane state of stress is defined by $\sigma_x$, $\sigma_y$, $\tau_{xy}$, and $\tau_{yx}$ ($\tau_{xy} = -\tau_{yx}$). To construct the corresponding Mohr’s circle, two perpendicular axes are selected to represent the normal stress $\sigma$ and shear stress $\tau$. Conventionally, the horizontal axis represents the normal stress and the vertical axis represents the shear stress, as shown in Fig. 5.6. Positive normal stresses (tensile) and negative normal stresses (compressive) are plotted on the right- and left-hand side of the $\sigma$ axis, respectively. Shear stresses producing clockwise rotation are considered positive and plotted on the upper part of the $\tau$ axis, and negative

![Figure 5.6 An illustration of Mohr’s circle.](image-url)
shear stresses causing counterclockwise rotation are plotted on the lower part of the $\tau$ axis.

A state of stress defined by $\sigma_x$, $\sigma_y$, $\tau_{xy}$, and $\tau_{yx}$ can be graphically represented by two points, as illustrated in Fig. 5.6: Point $A$ whose coordinates are $\sigma_x$ and $\tau_{xy}$, and point $B$ whose coordinates are $\sigma_y$ and $-\tau_{xy}$. The two points are then connected by a straight line $AB$ whose intersection with the horizontal $\sigma$ axis defines the center $C$ of Mohr’s circle. As can be seen, the coordinates of point $C$ are

$$\frac{1}{2}(\sigma_x - \sigma_y) + \sigma_y = \frac{1}{2}(\sigma_x + \sigma_y) \quad \text{and} \quad \tau = 0$$

where $\frac{1}{2}(\sigma_x + \sigma_y) = \sigma_{avg}$ is the average of the normal stresses. By definition of Mohr’s circle construction, points $A$ and $B$ must lie on the circle, and since $\tau_{xy} = \tau_{yx}$, $CA = CB = \text{radius of the circle}$, which corresponds to the maximum shear stress $\tau_{max}$.

Since lines $CA$ and $CB$ coincide with the $x$ and $y$ axes, respectively, it is evident that any angle measured from the Mohr’s circle is twice the true angle. As a general rule, the radius of the circle drawn to any point on its circumference corresponds to the axis normal to the plane whose stress components are given by the coordinates of that point. Also, the angle between two radii of the circle drawn to selected points on its circumference is twice the angle between the normals to the two planes whose stress components are represented by the coordinates of the two points.

Intersections of Mohr’s circle with the horizontal $\sigma$ axis determine the magnitudes of two principal stresses $\sigma'$ and $\sigma''$, where $\sigma' > \sigma''$. It is noted that in the example of Fig. 5.6, $\sigma'$ is positive and $\sigma''$ is negative. Therefore, $\sigma' = \sigma_1$, $\sigma'' = \sigma_3$, and $\sigma_2 = 0$. As explained earlier if both $\sigma'$ and $\sigma''$ are positive, $\sigma_1 = \sigma'$, $\sigma_2 = \sigma''$, and $\sigma_3 = 0$ is the minimum principal stress, and if both $\sigma'$ and $\sigma''$ are negative, $\sigma_1 = 0$, $\sigma_2 = \sigma'$, and $\sigma_3 = \sigma''$. As illustrated in Fig. 5.6, if the angle between the $\sigma$-axis and $x$-axis is $\alpha$, the true angle between the principal stress $\sigma_1$ and the $x$ axis is $\alpha/2$.

Noting that $ADC$ in Fig. 5.6 is a right triangle, $DC = 1/2(\sigma_x - \sigma_y)$, and $DA = \tau_{xy}$, the radius of the circle $CA$ is given by

$$CA = \left[\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right]^{1/2}$$

From the geometry of Fig. 5.6, and noting that the stress is measured from point $O$, the magnitudes of the principal stresses $\sigma_1$ and $\sigma_2$ are given by

$$\sigma_1 = \sigma' = \sigma_{avg} + \text{radius of the circle } CA$$

and

$$\sigma_3 = \sigma'' = \sigma_{avg} - \text{radius of the circle } CA$$
that is,

$$\sigma_1 = \frac{1}{2}(\sigma_x + \sigma_y) + \left\{\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right\}^{1/2}$$

and

$$\sigma_3 = \frac{1}{2}(\sigma_x + \sigma_y) - \left\{\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right\}^{1/2}$$

which is the same result obtained by the analytical method [see Eq. (5.15)]. Since in the case of Fig. 5.6, \(\sigma_2 = 0\), the maximum shear stress \(\tau_{max} = (\sigma_1 - \sigma_3)/2 = \left\{\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right\}^{1/2}\), which corresponds to the radius of Mohr’s circle (see Fig. 5.6). Therefore, three circles must be drawn, as shown in Fig. 5.7. One circle defines \(\sigma_1\) and \(\sigma_2\), a second circle defines \(\sigma_2\) and \(\sigma_3\), and a third circle defines \(\sigma_1\) and \(\sigma_3\).

When the state of stress is such that both \(\sigma_x\) and \(\sigma_y\) are positive (tensile stresses), \(\sigma_1 = \sigma'\), \(\sigma_2 = \sigma''\), and \(\sigma_3 = \sigma''\). In this case, the state of stress is

\[\begin{align*}
\sigma_1 &= \sigma' \\
\sigma_2 &= \sigma'' \\
\sigma_3 &= \sigma''
\end{align*}\]

\[\begin{align*}
\sigma_1 &= \frac{1}{2}(\sigma_x + \sigma_y) + \left\{\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right\}^{1/2} \\
\sigma_3 &= \frac{1}{2}(\sigma_x + \sigma_y) - \left\{\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right\}^{1/2}
\end{align*}\]
represented by the three circles shown in Fig. 5.8a. Figure 5.8b illustrates the state of stress corresponding to \( \sigma_x \) and \( \sigma_y \) negative (compressive), i.e. \( \sigma_1 = 0, \sigma_2 = \sigma', \) and \( \sigma_3 = \sigma'' \).

If Mohr’s circle is drawn to scale, it is possible to directly read the principal and maximum shear stresses as well as their orientations. However, it is important to correctly label the principal stresses identifying the maximum stress \( \sigma_1 \) and minimum stress \( \sigma_3 \) so that the maximum shear stress can be determined correctly.

To summarize the above discussion, the most generalized triaxial state of stress is completely defined in terms of three mutually perpendicular normal stresses \( \sigma_1, \sigma_2, \) and \( \sigma_3 \) called the principal stresses, which can either be tensile, compressive, or any combination of these. Such stresses act on three mutually perpendicular sets of planes whose orientations within the part are uniquely identified. Each principal plane is distinguished by the condition that its shear stress component \( \tau = 0. \) Since most engineering structures and machine parts have one of their dimensions shorter than the others, the principal stress normal to that dimension is relaxed, i.e., it approaches zero, which reduces the state of stress into a biaxial or plane stress. In this case, the state of stress at any point is defined in terms of one normal stress component \( \sigma_x \) and two shear stress components \( \tau_{xy} \) and \( \tau_{xz} \).

Having defined the concepts of strain and stress as basic tools in analyzing the response of an engineering structure or a machine part to external applied force, it is required to establish the relationship between strain and stress. Knowledge of such a relationship is important in two aspects of design. First,
designs for mechanical functions are essentially based upon mechanical property data generated from uniaxial tensile tests, where the stress is determined as a function of strain. Second, stress is not a directly measurable quantity; however, strain can directly be measured by various techniques such as electric strain gages. Therefore, if the relationship between stress and strain is known, the stress can be calculated from the measured strain as described below.

5.4 Multiaxial Stress-Strain Relationship

It is recalled from Chap. 4 that standard tensile tests are used to reveal the uniaxial relationship. In addition to providing very useful mechanical property data such as the elastic range, yield strength, ultimate tensile strength, and ductility, the results of the uniaxial tensile test can be used to predict the onset of plastic deformation when the part is subjected to a more complex state of stress as shown later. Also, the uniaxial stress-strain relationship can be extended to determine the multiaxial stress-strain relationship within the elastic range, as described below.

Consider the most generalized triaxial state of stress defined by the three principal stresses \( s_1, s_2, \) and \( s_3 \) acting along the \( x, y, \) and \( z \) axes, as shown in Fig. 5.9a. To determine the elastic strains developed by these stresses, it is possible to visualize that the three stresses \( s_1, s_2, \) and \( s_3 \) are applied independently.

![Figure 5.9](image-url)
one at a time. Subsequently, the three independent strains are combined to determine the net strain of the part assuming that its material is isotropic; i.e., its properties are independent of the direction along which it is measured. This assumption is valid for many engineering materials.

When the stress \( \sigma_1 \) is applied along the \( x \) axis as shown in Fig. 5.9b, it produces a positive engineering strain \( e_1 \) given by

\[
e_1 = \frac{\sigma_1}{E}
\]

where \( E \) is the elastic modulus (see Sec. 3.11). Simultaneously, the stress \( \sigma_1 \) produces lateral negative strains along the \( y \) and \( z \) axes given by

\[
e_y = -\nu e_2 \quad \text{and} \quad e_z = -\nu e_3
\]

where \( \nu \) is Poisson’s ratio as described in Sec. 3.9 and \( e_2 \) and \( e_3 \) are the strains produced by the stresses \( \sigma_2 \) and \( \sigma_3 \) along the \( y \) and \( z \) axes, respectively. Therefore, the net strain \( e_1 \) along the \( x \) axis produced by the stress \( \sigma_1 \) is given by

\[
e_1 = \frac{\sigma_1}{E} - \nu e_2 - \nu e_3
\]

Since \( e_2 = \frac{\sigma_2}{E} \) and \( e_3 = \frac{\sigma_3}{E} \), the preceding equation can be rewritten as

\[
e_1 = \frac{\sigma_1}{E} - \nu \sigma_2/E - \nu \sigma_3/E
\]

(5.28)

Similarly, the net strains along the \( y \) and \( z \) axes are given by

\[
e_2 = \frac{\sigma_2}{E} - \nu \sigma_1/E - \nu \sigma_3/E
\]

(5.29)

\[
e_3 = \frac{\sigma_3}{E} - \nu \sigma_1/E - \nu \sigma_2/E
\]

(5.30)

Eliminating \( E \) from the preceding three equation results in

\[
e_1/\sigma_1 - \nu (\sigma_2 + \sigma_3) = e_2/\sigma_2 - \nu (\sigma_3 + \sigma_1) = e_3/\sigma_3 - \nu (\sigma_1 + \sigma_2)
\]

(5.31)

Similar to stresses, the strains \( e_1, e_2, \) and \( e_3 \) produced along the directions of \( \sigma_1, \sigma_2, \) and \( \sigma_3 \) are called the principal strains. It can be shown that regardless of the state of stress, the change in volume of the part per unit volume, i.e., \( \Delta V/V_0 \), where \( V_0 \) is the initial volume, corresponding to the triaxial state of stress is given by

\[
\Delta V/V_0 = e_1 + e_2 + e_3
\]

(5.32)

When \( \sigma_1 = \sigma_2 = \sigma_3 \), the state of stress is referred to as hydrostatic. In this case, \( \Delta V/V_0 = 3e \), and the elastic relationship between stress and strain is given by

\[
\sigma = K(DV/V_0) = K(3e)
\]
or

\[ K = \frac{1}{3} (\sigma/e) \]  \hspace{1cm} (5.33)

where \( K \) is the bulk modulus. It can readily be shown that \( K \) is related to the tensile elastic modulus by the relationship

\[ K = E/3(1 - 2\nu) \]  \hspace{1cm} (5.34)

where \( \nu \) is Poisson’s ratio. For a biaxial state of stress, Eq. (5.32) becomes

\[ \Delta V/V_0 = e_1 + e_2 \]  \hspace{1cm} (5.35)

and for a uniaxial state of stress, the change in volume is given by

\[ \Delta V/V_0 = e_1 \]  \hspace{1cm} (5.36)

It is recalled from Sec. 5.3.1 that the three principal stresses \( \sigma_1, \sigma_2, \) and \( \sigma_3 \) produce maximum shear stresses oriented at 45° from the respective principal stresses. Such shear stresses produce shear strains along the same directions. Combining Eqs. (5.27) and (3.33) \((\tau = G\gamma)\) results in

\[ \gamma_1 = (\sigma_1 - \sigma_3)/2G \]
\[ \gamma_2 = (\sigma_1 - \sigma_2)/2G \]
\[ \gamma_3 = (\sigma_2 - \sigma_3)/2G \]  \hspace{1cm} (5.37)

where \( \gamma_1, \gamma_2, \) and \( \gamma_3 \) are the principal shear strains. Since \( E = 2G(1 + \nu) \), the principal shear strains can be related to the principal normal strains Eq. (5.31) by

\[ \gamma_1 = e_1 - e_3 \]
\[ \gamma_2 = e_1 - e_2 \]
\[ \gamma_3 = e_2 - e_3 \]  \hspace{1cm} (5.38)

To derive the multiaxial stress-strain relationships within the plastic range, certain simplifying assumptions are made, including: (i) The change in volume accompanying plastic deformation is negligible; (ii) similar to the elastic case described in the previous section, the principal strains remain parallel to the principal stresses; (iii) the ratio of principal shear strains to principal shear stresses is constant; (iv) the elastic component of the total strain is negligible.
Based upon the assumptions above, the true principal strains $\varepsilon_1$, $\varepsilon_2$, and $\varepsilon_3$ are related to the principal true stresses $\sigma_1$, $\sigma_2$, and $\sigma_3$ by the following equations

$$
\varepsilon_1 = (\sigma_1/k)(1/n[\alpha^2 + \beta^2 + 1 - \alpha\beta - \alpha - \beta](1 - n)/2n[1 - \alpha/2 - \beta/2])
$$

$$
\varepsilon_2 = (\sigma_2/k)(1/n[\alpha^2 + \beta^2 + 1 - \alpha\beta - \alpha - \beta](1 - n)/2n[\alpha - \beta/2 - 1/2])
$$

$$
\varepsilon_3 = (\sigma_3/k)(1/n[\alpha^2 + \beta^2 + 1 - \alpha\beta - \alpha - \beta](1 - n)/2n[\beta - \alpha/2 - 1/2])
$$

where $k$ is the strength coefficient and $n$ is the strain hardening exponent of the material. Noting that $\sigma_1 > \sigma_2 > \sigma_3$, the parameters $\alpha$ and $\beta$ in the preceding equation are given by

$$
\alpha = \sigma_2/\sigma_1 \quad \text{(5.40)}
$$

$$
\beta = \sigma_3/\sigma_1 \quad \text{(5.41)}
$$

Different states of stress defined in the preceding discussions can be mechanically developed by one or more types of three loading conditions: (i) axial, (ii) flexural, and (iii) torsional. Stresses can also be produced by changing the temperature of the part (thermal stresses). In the next section, the axial loading is distinguished from the flexural and torsional loading in terms of the stresses produced.

### 5.5 Loading Conditions and Stress

From a mechanical point of view, a part may be subjected to an axial load, flexural load, torsional load, or any combination of these during service. To safely design a part, it is essential to (i) define and evaluate the stresses developed by each type of loading and (ii) evaluate the stresses developed by a combined loading condition. In this section, the stresses developed by each type of loading condition are distinguished from each other. Evaluation of stresses developed by a combined loading condition is explained with solved problems at the end of this chapter in Sec. 5.14.

By definition an axial load acts along a geometric axis of the part and can be either uniaxial, biaxial, or triaxial. In general, the stress produced by an axial load whether it is normal or shear is given by

$$
\text{Axial stress} = \text{load/area acted upon by the load} \quad \text{(5.42)}
$$

When one part is resting on another and transfers a force to it, a bearing or compressive stress is developed at the surface of contact. Similar to direct compressive stress, the bearing stress is a measure of the tendency of the applied force to crush the supporting part, and is given by

$$
\sigma_{\text{bearing}} = \text{applied force/bearing area} \quad \text{(5.43)}
$$
For flat surfaces in contact, the bearing area is simply the area over which the force is transferred from one part to another. If the two parts have different areas, the smaller area is used in the calculation. Bearing stresses are developed in such parts as riveted joints.

To define the stresses produced by a flexural load, consider a part fixed at both ends and is acted upon by a force \( P \) normal to its long axis, as shown in Fig. 5.10a, corresponding to pure bending. To be in a state of mechanical equilibrium, the part must be acted upon at both ends by two forces each of which equals \( P/2 \) and both forces are acting along a direction opposite to that of \( P \) (Fig. 5.10a). Such a loading condition produces a bending moment. It is recalled from Chap. 3 that the moment of an applied force is defined as its tendency to

\[\text{Figure 5.10} \quad \text{General characteristics of a flexural loading condition. (a) A part subjected to a flexural load. (b) Determination of the bending moment produced by the load. (c) Normal stresses produced by the load.}\]
rotate the body and is measured by multiplying the magnitude of the force by the perpendicular distance from the line of action of the force to the axis of rotation. A bending moment about any point is the algebraic sum of the moments of all vertical forces to one side of that point.

The bending moment $M$ can be calculated by cutting the part at some location. If the section is cut along $X-X$ as shown in Fig. 5.10a, b, the moment $M_1$ is given by $(P/2)l_1$ according to the above definition. Along another section taken at a distance $l_2$, the corresponding moment $M_2 = P/2 \times l_2$, and so on. Therefore, the bending moment $M$ produced by the force $P$ is given by

$$M = \sum_{0=0}^{l_2} P/2l_1 + P/2l_2 + \cdots$$

or

$$M = P/2(l)$$

(5.44)

where $l$ is the length of the part. When the part is loaded in bending as shown in Fig. 5.10a, its upper part assumes a convex shape resulting from stretching or tensile stresses. Because the lower part is compressed, it assumes a concave shape. Figure 5.10c illustrates the stress distribution along a cross section of the part. As can be seen, the tensile positive stress has a maximum value at the surface, and it decreases gradually until it vanishes at the geometric axis of the part, which is called the neutral axis. Below this axis, the stress becomes compressive, and it increases gradually with depth reaching a maximum value at the bottom of the same magnitude as the maximum tensile stress. Evidently, it is important to calculate the maximum stress $\sigma_{\text{max}}$ developed in the part. This is given by

$$\sigma_{\text{max}} = Mh/I$$

(5.45)

where $h$ is the distance from the neutral axis to the surface of the part (see Fig. 5.10a), and $I$ is the moment of inertia of the part cross section with respect to the neutral axis. Appendix B summarizes the moment of inertia for different geometries of part cross section where the neutral axis is indicated by the dotted line.

Similar to the case of an axial load, a flexural load also produces a shear stress. As shown in Fig. 5.10b, an internal force $P/2$ is developed along an arbitrary section to balance the external force. Since the line of action of the internal force $P/2$ is parallel to the cross-sectional area, it produces a shear stress $\tau$ given by

$$\tau = P/2A_0$$

(5.46)

where $A_0$ is the cross-sectional area. It is to be noted that the direction of shear stress is dependent upon the location within the part, as illustrated in Fig. 5.11a. As can be seen, the shear stress developed at $X'_1-X'_1$ section is negative. A positive shear stress
of the same magnitude is developed along the $X_2-X_2$ section. At the middle section of the part where the bending force acts, the shear stress vanishes, as illustrated in Fig. 5.11b.

It is recalled from Sec. 4.3.8 that loading in torsion produces a shear stress reaching a maximum value at the surface of the part. At any cross section of a cylinder loaded in torsion, the shear stress varies from zero at the center of the cylinder to a maximum value at the surface. Torsional shear stress is given by an equation similar to Eq. (5.45). In the case of a solid cylinder of diameter $d$, the bending moment $M$ is replaced by the torque $T$, and $h$ is replaced by $r/2$, where $r$ is the radius ($r/2 = d/4$); therefore, Eq. (5.45) becomes

$$\tau_{\text{max}} = \frac{Td}{4I}$$
From Appendix B, the moment of inertia $I$ of a circular cross section is $\pi d^4/64$. Substituting this value into the preceding equation results in

$$\tau_{\text{max}} = (Td/4)/(\pi d^4/64)$$

and therefore

$$\tau_{\text{max}} = 16T/\pi d^3$$

which is the same as Eq. (4.32). For a hollow cylinder (tube) of an external diameter $d_1$ and internal diameter $d_2$, the shear stress $\tau$ is given by

$$\tau = (16Td_1)/\pi (d_1^4 - d_2^4)$$

Just as axial and flexural loads produce normal and shear stresses, a torsional load also produces a normal stress. Since the angle between the normals to the planes of maximum normal stress and maximum shear stress is always 45°, the maximum normal stress in a cylinder loaded in torsion is developed on planes oriented at 45° from the cross section. As shown in the solved problems in Sec. 5.14, cylinders loaded in torsion fail along the planes of maximum normal stress.

### 5.6 Thermal Stress

It is well known that when a part is subjected to changes in temperature, it either expands (heating) or contracts (cooling). Considering the case of uniaxial deformation as in a tensile test, if a part of an initial length $L_0$ is subjected to a temperature increase of $\Delta T$, the corresponding elongation $\Delta L$ is given by

$$\Delta L = \alpha L_0 \Delta T$$

where $\alpha$ is a material property called the coefficient of linear thermal expansion. It’s units are length/length degree, i.e., $\alpha$ is expressed in degree$^{-1}$. Since by definition the engineering strain $e = \Delta L/L_0$, the above equation can be rewritten as

$$e = \Delta L/L_0 = \alpha \Delta T$$

In three dimensions corresponding to volumetric strain $\Delta V/V_0$ where $\Delta V$ is the change in volume and $V_0$ is the initial volume, the above equation becomes

$$\Delta V/V_0 = 3\alpha \Delta T$$

If the part is not free to change its dimensions as a result of temperature changes, e.g., it is restrained by external forces, the tendency to change dimensions is accommodated by building up stresses in the part. It is recalled from Chap. 3 that within the elastic range the normal stress $\sigma$ is given by

$$\sigma = Ee$$
where $E$ is the elastic modulus and $e$ is the engineering strain. Combining the above equation with Eq. (5.50) results in

$$\sigma = \alpha E \Delta T \quad (5.52)$$

In practice, the resistance of structural materials to thermally induced stresses is measured in terms of what is known as the thermal stress index $I$, expressed as

$$I = \frac{s_{uts}}{k} \frac{k}{\alpha E} \quad (5.53)$$

where $\sigma_{uts}$ is the ultimate tensile strength, and $k$ is the thermal conductivity of the material. As the value of $I$ becomes higher, the material is said to be more resistant to thermal stresses.

If during service the part is subjected to a rapid change in temperature, e.g., a thermal shock, the index $I$ given by Eq. (5.53) can be useful in assessing the resistance of the material to fracture. Under thermal shock conditions, the temperature capability of the material is given by

$$T_{\text{max}} = \frac{s_y}{(1 - \nu)} \frac{1}{\alpha E} \quad (5.54)$$

where $T_{\text{max}}$ is the maximum temperature the material can withstand and $\nu$ is Poisson’s ratio. Another effect related to changes in temperature is thermal fatigue, resulting from temperature cycling, as discussed later in this chapter.

It can be appreciated from the preceding discussion that regardless of the external effect producing stresses in the part, there are only two basic types of stress: (i) normal stress (either tensile or compressive) and (ii) shear stress. Since engineering alloys are usually selected to operate within their elastic range, it is essential to identify and be able to evaluate the stress required to produce plastic deformation.

### 5.7 Type of Stress Required to Produce Plastic Deformation

Usually engineering alloys for mechanical functions are selected to operate within the elastic range. However, to guard against catastrophic failure, the designer takes into account that the part should plastically deform rather than fracture in case of an unexpected overloading condition. Also, in some cases, machine parts and structures are operated within the plastic range. It is therefore important to identify the stress which can produce plastic deformation.

When the three principal stresses defining a triaxial state of stress become equal, i.e., $\sigma_1 = \sigma_2 = \sigma_3$, defining a hydrostatic state of stress, no shear stresses can be developed according to Eq. (5.27), i.e., $\tau_1 = \tau_2 = \tau_3 = 0$. Experiment shows that for this special case of hydrostatic state of stress, no plastic
deformation can occur. Evidently, only a shear stress can produce plastic deformation consistent with the prediction of microscopic aspects of plasticity outlined in Chap. 4. Also, experiment shows that plastic deformation is suppressed when the state of stress is changed from uniaxial \( \tau_{\text{max}} = \sigma_1/2 \) to biaxial, where \( \tau_{\text{max}} = (\sigma_1 - \sigma_3)/2 \) as a result of reducing the shear stress.

To summarize, plastic deformation can only occur by the shear component of an applied stress. Therefore, to predict the onset of plastic deformation in a part, it is required to determine the maximum shear stresses developed. As explained earlier, the maximum shear stresses can be evaluated from knowledge of the principal stresses. Furthermore, knowledge of the principal stresses is used to predict failure, as shown later.

5.8 Maximum Stresses

Knowledge of the maximum stresses developed in a part in comparison with the strength of material used in manufacturing is essential to ensure safety during service. Both the maximum normal and shear stresses must be determined requiring knowledge of the principal stresses. As demonstrated earlier, it is possible to completely define the state of stress in a part in terms of the principal stresses \( \sigma_1, \sigma_2, \text{ and } \sigma_3 \), where \( \sigma_1 > \sigma_2 > \sigma_3 \), i.e., \( \sigma_1 \) is the maximum possible normal stress developed in the part and \( \sigma_3 \) is the minimum normal stress. Based upon previous discussions, it is possible to summarize the maximum normal and shear stresses for each of the three states of stress as follows:

1. **Uniaxial state of stress defined by the principal stress \( \sigma_1 \) \( (\sigma_2 = \sigma_3 = 0) \): \( \sigma_1 \) is the maximum normal stress (either tensile or compressive) and \( \tau = \sigma_1/2 \) is the maximum shear stress.

2. **Biaxial state of stress defined by two principal stresses: (i) If the two stresses are positive, \( \sigma_3 = 0 \), and the maximum normal stress is \( \sigma_1 \), and \( \tau = (\sigma_1 - \sigma_2)/2 \) is the maximum shear stress. (ii) If one stress is positive (tensile) and the other is negative (compressive), \( \sigma_2 = 0 \) and the positive stress is \( \sigma_1 \), the negative stress is \( \sigma_3 \), and the maximum shear stress \( \tau_{\text{max}} = |\sigma_1 - (-\sigma_3)|/2 = (\sigma_1 + \sigma_3)/2 \). (iii) If the two stresses are negative (compressive), \( \sigma_1 = 0 \) is the maximum normal stress, the less negative stress is \( \sigma_2 \), the more negative stress is \( \sigma_3 \), and the maximum shear stress \( \tau = |0 - (-\sigma_3)|/2 = \sigma_3/2 \).

3. **Triaxial state of stress defined by the principal stresses \( \sigma_1, \sigma_2, \text{ and } \sigma_3 \) \( (\sigma_1 > \sigma_2 > \sigma_3), \text{ taking the sign of the stress into consideration} \): \( \sigma_1 \) is the maximum normal stress and \( \tau = (\sigma_1 - \sigma_3)/2 \) is the maximum shear stress. A hydrostatic state of stress is a special case where the three principal stresses \( \sigma_1, \sigma_2, \text{ and } \sigma_3 \) have the same magnitude and sign and the maximum shear \( \tau = 0 \).
5.9 Design Stresses

Most designs for structural applications are based on mechanical property data obtained from tensile tests on smooth bars particularly the yield strength and ultimate tensile strength. However, since the safe lifetime of an engineering structure cannot be predicted from such simple tests, it is a common practice to use a “factor of safety” to guard against unanticipated fracture during service. A factor of safety is a number \( N \) divided into either the yield or ultimate tensile strength obtained from the tensile test to obtain what is known as the allowable stress. Whether the yield strength or ultimate strength is used as the basis for design is dependent upon (i) type of loading and (ii) ductility of the material. It is possible to identify three main types of loading: (i) static, (ii) cyclic, dynamic, or repeated, and (iii) impact or shock. Generally, for brittle materials where the yield strength approaches the ultimate strength, the ultimate strength is used as the basis for design, and therefore the design or allowable stress \( \sigma_d \) is given by

\[
\sigma_d \text{(brittle materials)} = \frac{\sigma_{uts}}{N} \quad (5.55)
\]

For ductile materials, the yield strength is used as the basis for design in the case of a static loading, i.e.,

\[
\sigma_d \text{(static loading, ductile material)} = \frac{\sigma_y}{N} \quad (5.56)
\]

However, in the case of cyclic or impact loading, the ultimate strength is used as the basis for design using ductile materials, i.e.,

\[
\sigma_d \text{(cyclic loading or impact loading, ductile material)} = \frac{\sigma_{uts}}{N} \quad (5.57)
\]

Many parts and components are covered by codes and specifications determining the design or allowable stress, e.g., pressure vessels, bolts, and welded components. For parts not covered by codes and specifications, the factors of safety listed in Table 5.1 are usually used to obtain the design stress.

When parts are subjected to shear stresses during service such as riveted joints and springs, a design shear stress \( \tau_d \) is used. However, values of yield strength in shear for different metallic materials are not as commonly reported as yield strength. Usually, the design shear stress is obtained from tensile strength data. It is recalled that a tensile stress \( \sigma \) develops a maximum shear stress \( \tau = \frac{1}{2} \sigma \).

<table>
<thead>
<tr>
<th>Type of loading</th>
<th>Ductile materials</th>
<th>Brittle materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Yield strength basis: ( N )</td>
<td>Ultimate strength basis: ( N )</td>
</tr>
<tr>
<td>Cyclic</td>
<td>Ultimate strength basis: ( N )</td>
<td>Ultimate strength basis: ( N )</td>
</tr>
<tr>
<td>Impact (shock)</td>
<td>Ultimate strength basis: ( N )</td>
<td>Ultimate strength basis: ( N )</td>
</tr>
</tbody>
</table>

Table 5.1 Factors of Safety for Parts Not Covered by Codes and Specifications

Copyright © 2004 by Marcel Dekker, Inc.
Therefore, the yield strength in shear $\tau_y = \frac{1}{2} \sigma_y$, and the design shear stress $\tau_d$ is given by

$$\tau_d = \sigma_y / 2N$$  \hspace{1cm} (5.58)

where $N$ is the factor of safety. Usually, the design bearing stress is taken as $0.9 \sigma_y$.

Even though most designs are based on yield strength requirements, structural components are merely required to withstand the applied load without fracture or more practically without plastic deformation and a minimum of elastic deformation. Therefore, selection of materials for structural applications may not reside so much in meeting certain yield strength requirements but in producing resistance to fracture at low strains. More recently, another material property $KIC$ known as the plane strain fracture toughness, to be described later in this chapter, has become a part of materials procurement specifications and is used as a quantitative design parameter in some cases.

It is evident from the preceding discussion that if the design parameters are strictly followed during service, the part should only elastically deform. However, in some cases, machine parts and structures are operated beyond the yield point where plastic deformation occurs. Also, as pointed out earlier, prediction of the onset of plastic deformation under a complex state of stress is important in terms of preventing catastrophic failure. Therefore, it is required to develop both the elastic and plastic stress-strain relationships corresponding to complex states of stress, as described below.

5.10 Criterion for the Onset of Plastic Deformation (Yielding)

In a simple tensile test, macroscopic plastic deformation is initiated at a stress corresponding to the tensile yield strength of the material $\sigma_y$. As demonstrated earlier, plastic deformation occurs by a shear stress. It is recalled that the maximum possible shear stress $\tau_{\text{max}}$ produced by a triaxial state of stress defined by the principal stress $\sigma_1$, $\sigma_2$, and $\sigma_3$ ($\sigma_1 > \sigma_2 > \sigma_3$) is given by

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

Since in the case of uniaxial tensile test, $\sigma_y = \frac{1}{2} \tau_y$, the simplest criterion for yielding corresponding to a triaxial state of stress known as the Tresca’s criterion is

$$(\sigma_1 - \sigma_2) > \sigma_y$$  \hspace{1cm} (5.59)

However, a more satisfactory criterion for yielding is that of Von Mises, which states that yielding occurs when

$$\sigma_y = \left[\left((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right)/2\right]^{1/2}$$  \hspace{1cm} (5.60)
5.11 Stress Concentration

Many engineering structures and machine parts are designed such that they contain sudden changes in shape or geometry. When a part containing a geometrical discontinuity is subjected to an external force, the local stress at the discontinuity can be considerably greater than the average stress. This is referred to as stress concentration, which can be a primary cause of failure. Some of the typically encountered sources of stress concentrations are (i) notches, (ii) fillets, (iii) rough surface finish, (iv) press fits, (v) inclusions, and (vi) residual stresses. Since most failures start at the surface of the material, it is of extreme importance to consider the surface condition of the part giving rise to stress concentration, as demonstrated by the following examples.

Consider a part containing a hole such as that shown in Fig. 5.12 subjected to a tensile force $P$. For cross sections such as $A$-$A$ at a considerable distance from the hole, the stress $\sigma$ is uniformly distributed over the cross-sectional $A_0$, and is given by $P/A_0$. However, as the hole is progressively approached, the stress becomes more and more nonuniformly distributed. It has been shown that at a cross section passing through the hole ($B$-$B$ in Fig. 5.12), the stress reaches a

![Figure 5.12](image)

**Figure 5.12** A schematic illustration of stress concentration. (a) A uniaxial tensile load $P$ applied to a part containing a hole. (b) Stress distribution across the hole.
maximum value $\sigma_{\text{max}}$ given by

$$\sigma_{\text{max}} = K\sigma$$  \hspace{1cm} (5.61)

where $K$ is a stress concentration or intensity factor and $\sigma$ is the stress at the cross section $A-A$, which is usually referred to as the nominal stress. It is evident that the value of $K$ is dependent on the particular geometry of the part. For the example shown in Fig. 5.12, it has been shown that the approximate value of $K = (3w/d)/(w/d + 1)$, where $w$ is the width of the bar or its diameter and $d$ is the diameter of the hole. Stress concentration can also change the local state of stress, as described below.

When a part is uniaxially loaded in tension as shown in Fig. 5.13a, it is in a uniaxial state of stress. However, if the part contains a V notch as shown in Fig. 5.13b, the state of stress at the notch becomes triaxial. Because of the reduced cross section at the notch, the local stress can reach the yield point before other sections of the part tending to stretch the part along the direction of the applied stress and shrink it in the transverse direction. Since the other sections of the part cannot plastically deform, a constraint is placed upon the notched section creating a state of triaxial stress—the applied tensile stress and two mutually perpendicular induced tensile stresses.

![Figure 5.13 Effect of part geometry on the state of stress. (a) A smooth part uniaxially loaded in tension (uniaxial state of stress). (b) Development of a triaxial state of stress by introducing a notch.](https://www.marceldekker.com/books/MaterialsProcessing/1050272797/8808501007/5.13.png)
Since there are many sources for stress concentration, the value of $K$ is the product of a number of parameters, including a surface factor, size factor, reliability factor, temperature factor, shape factor, and a miscellaneous factor taking into account the effects of corrosive environments, surface layer, residual stresses, etc. All these parameters can either be obtained from available charts or calculated. Stress concentration related to a shape factor, which is a function of part geometry, is accounted for by the theoretical stress concentration factor $K_t$, defined as

\[ K_t = \frac{\text{actual maximum stress in the part}}{\text{applied nominal stress}} \]  

Values of $K_t$ can be calculated for simple shapes; however, for complex parts experimental stress analysis methods must be used. Tables and charts are available in the literature which give values of $K_t$ for many types of geometry, as shown in the examples of Fig. 5.14 where the part is uniaxially loaded in tension.

**Figure 5.14** Stress concentration factor $K_t$ for selected geometries.
5.12 Criteria for Mechanical Failure

Prediction of failure is an important part of design to ensure safety as well as reliability. Various criteria for mechanical failure are developed for the most generalized triaxial state of stress. Among the most commonly used criteria are those of the maximum normal stress and maximum shear stress.

When the material is brittle, the criterion of maximum normal stress is particularly useful. Since a brittle material does not undergo significant macroscopic plastic deformation prior to fracture in a uniaxial tensile test, the tensile yield strength is used as the basis to predict failure. According to this criterion, failure occurs when the maximum principal \( \sigma_1 \) becomes equal to or exceeds the tensile yield strength, i.e.,

\[
\sigma_1 \geq \sigma_y
\]  

(5.63)

For ductile materials, the criterion of maximum shear stress is used to predict failure. In this case, failure occurs when the maximum shear stress \( \tau_{\text{max}} \) developed in the part reaches the shear stress corresponding to yielding in a uniaxial tensile test. It is recalled that the shear stress at yielding \( \tau_y = \frac{1}{2} \sigma_y \), and therefore for a triaxial state of stress, failure occurs when

\[
(\sigma_1 - \sigma_3)/2 > \frac{1}{2} \sigma_y
\]

or

\[
(\sigma_1 - \sigma_3) > \sigma_y
\]  

(5.64)

5.13 Applications: Analysis of Stresses in Specific Components

5.13.1 Thin Spherical Pressure Vessel

A spherical pressure vessel may contain a liquid or gas exerting a pressure on its walls. It is the objective of stress analysis to determine stresses developed in the walls as a result of the internal pressure. As shown in Fig. 5.15a, the internal pressure \( p \) acts along a direction perpendicular to the wall of the sphere. Figure 5.15b illustrates the area upon which the pressure \( p \) is acting, which is \( \pi d_i^2/4 \), where \( d_i \) is the internal diameter of the sphere. By definition, the corresponding force is given by

\[
F = p \pi d_i^2/4
\]

From a consideration of the state of mechanical equilibrium, the above force is balanced by a tensile force \( F \) developed within the walls of the sphere, as shown in Fig. 5.15a. However, this force acts on the cross-sectional area of
the sphere given by $\pi/4(d_e^2 - d_i^2)$, where $d_e$ and $d_i$ are the external and internal diameters, respectively (Fig. 5.15b). For a thin-walled sphere where the ratio of internal diameter $d_i$ to wall thickness $t(d_i/t)$ is $>10-15$, the cross-sectional area can be approximated by a rectangular area whose length is $\pi d_i t$ (circumference of the circle whose diameter is $d_i$), and width is $t$. If the corresponding tensile stress is $\sigma$, $F$ is given by

$$F = \sigma \pi d_i t/4$$
Combining the above two equations results in

\[ p \pi d^2 / 4 = \sigma \pi d_i t \]

or

\[ \sigma = pd_i / 4t \]  \hspace{1cm} (5.65)

To determine the relationship of the stress \( s \) to the principal stresses as well as its orientation, a suitable system of coordinate axes is selected such that its origin \( O \) coincides with the center of the sphere. If the \( x \) and \( y \) axes are selected to lie on the surface of the sphere, the \( z \) axis becomes normal to the surface, i.e., normal to the thickness of the sphere. An elemental cube at any point is extracted from the sphere, as shown in Fig. 5.15c. In view of the applied force \( F \), the only non zero stress component on the elemental cube is \( \sigma_x \), reducing the cubic stress Eq. (5.9) to

\[ \sigma^3 - \sigma^2 (\sigma_x) = 0 \quad \text{or} \quad \sigma \times \sigma (\sigma - \sigma_x) \]

Therefore, two of the principal stresses equal 0 and the third principal stress \( \sigma = \sigma_x = \sigma_1 \). Also, Eq. (5.23), determining the orientations of the principal planes, becomes, respectively,

\[ l(\sigma - \sigma_x) - 0 - 0 = 0 \]
\[ -0 + m(\sigma) - 0 = 0 \]
\[ -0 - 0 + n(\sigma) = 0 \]

Since \( \sigma = \sigma_x (\neq 0) \), all three values of \( l, m, \) and \( n \) cannot be zero, and \( l^2 + m^2 + n^2 = 1 \), the possible values of \( l, m, \) and \( n \) are \( l = \pm 1, \ m = 0, \ n = 0 \). To summarize, the maximum normal stress \( \sigma_1 \) developed in a thin-walled sphere as a result of the internal pressure is given by

\[ \sigma_1 = pd_i / 4t \]

and is oriented along the \( x \) axis. It then follows that the maximum shear stress \( \tau_{max} = \frac{1}{2} \sigma_1 \) and oriented at 45° from \( \sigma_1 \), is given by

\[ \tau_{max} = pd_i / 8t \]  \hspace{1cm} (5.66)

To examine the fracture behavior of the sphere, a set of parallel lines following the direction of \( \sigma_1 \) is drawn on the surface, as shown in Fig. 5.16a. Such lines are called the normal stress trajectories. Another set of lines is drawn perpendicular to the normal stress trajectories. If the material of the sphere is brittle, the fracture path...
follows the direction of these lines normal to $\sigma_1$ (Fig. 5.16b). Recalling that the maximum shear stress is oriented at $45^\circ$ from the principal stress, shear stress trajectories are drawn as shown in Fig. 5.16c where only one set is shown. If the material is ductile, the fracture path follows these trajectories as shown in Fig. 5.16d.
5.13.2 Thin Cylindrical Pressure Vessel (Thin-Walled Tube)

Similar to the thin-walled spherical pressure vessel described above, a thin-walled tube carries a fluid (gas or liquid) exerting pressure $p$ on its walls. Both longitudinal and circumferential (hoop) tensile stresses are developed in the cylinder, as illustrated in Fig. 5.17a.

From a consideration of mechanical equilibrium, both stresses can be determined as illustrated in Figs. 5.17b,c. As can be seen from Fig. 5.17b a longitudinal tensile force $F_l$ is developed to balance the force resulting from the internal pressure $p$ acting on an area given by $\pi d_i^2/4$, where $d_i$ is the internal diameter, i.e.,

$$F_l = p(\pi d_i^2/4)$$

A longitudinal tensile stress is developed by this force and acts on an area given by $\pi d_i^2/4(d_e^2 - d_i^2)$ where $d_e$ and $d_i$ are the external and internal diameters, respectively. This area can be approximated by $\pi dt$ for a thin-walled cylinder where $d_i/t > 10–15$ and $t$ is the thickness of the cylinder. In this case, the stress is nearly uniform across the thickness, i.e.,

$$F_l = \sigma_l \pi d_i t$$

Combining the above two equations results in

$$p(\pi d_i^2/4) = \sigma_l \pi d_i t \quad \text{or} \quad \sigma_l = pd_i/4t$$

As shown in Fig. 5.17c, a circumferential force $F_c$ is developed to balance the force resulting from the internal pressure acting on an area given by $d_iL$. A circumferential stress is developed by this force and acts on area given by $2tL$; thus,

$$F_c = p(d_iL) = \sigma_c(2tL) \quad \text{or} \quad \sigma_c = pd_i/2t$$

It is evident from the above analysis that

$$\sigma_c = 2\sigma_l$$

To determine the relationship of $\sigma_l$ and $\sigma_c$ to the principal stresses, a coordinate system with an origin at point $O$ is selected such that the $x$ axis is normal to the axis of the cylinder, the $y$ axis is parallel to the tube axis, and the $z$ axis is normal to the surface (Fig. 5.17a). In this case, the only nonzero stresses developed by the internal pressure are the circumferential or hoop stress acting along a direction normal to the tube axis and the longitudinal stress along the tube axis, as shown in Fig. 5.17a; i.e., the only nonzero stress components in the cubic stress Eq. (5.9) are $\sigma_x$ and $\sigma_y$, and therefore this equation is reduced to

$$\sigma^3 - \sigma^2(\sigma_x + \sigma_y) + \sigma(\sigma_x\sigma_y) = 0$$

Copyright © 2004 by Marcel Dekker, Inc.
Figure 5.17  Stress analysis of a thin-walled cylinder. (a) Longitudinal and circumferential stresses developed in the cylinder. (b) Basis for determining the longitudinal stress. (c) Basis for determining the circumferential stress.
or

\[ \sigma [\sigma^2 - \sigma (\sigma_x + \sigma_y) + (\sigma_x \sigma_y)] = 0 \]

Dividing the above equation by \( \sigma \) results in

\[ \sigma^2 - \sigma (\sigma_x + \sigma_y) + \sigma_x \sigma_y = 0 \]

and the above quadratic equation has two roots, giving two values of \( \sigma \) as follows:

\[ \sigma_1 = \left( \sigma_x + \sigma_y \right) + \left[ (\sigma_x + \sigma_y)^2 - 4\sigma_x \sigma_y \right]^{1/2} / 2 = \sigma_x \]

\[ \sigma_2 = \left( \sigma_x + \sigma_y \right) - \left[ (\sigma_x + \sigma_y)^2 - 4\sigma_x \sigma_y \right]^{1/2} / 2 = \sigma_y \]

and the orientations of the principal planes with respect to the \( x, y, \) and \( z \) axes are determined from Eq. (5.23), which becomes

\[ l(\sigma - \sigma_x) = 0 \]

\[ m(\sigma - \sigma_y) = 0 \]

\[ n(\sigma) = 0 \]

Since \( \sigma \) does not equal 0, \( n = 0 \). Noting that \( l^2 + m^2 + n^2 = 1 \) for \( \sigma_1 = \sigma_x \), \( l = 1 \) or 0, and therefore \( m = 0 \) for \( l = 1 \), and \( m = 1 \) for \( l = 0 \). When \( \sigma_2 = \sigma_y \), for \( n = 0 \), \( m = 1 \) or 0, and \( l = 0 \) for \( m = 1 \) and \( l = 1 \) for \( m = 0 \). As demonstrated earlier, \( \sigma_c \) is the maximum normal stress developed in the cylinder and \( \sigma_z = 0 \) is the minimum stress, i.e,

\[ \sigma_1 = \sigma_c = \frac{pd_i}{2t} \]

\[ \sigma_2 = \sigma_l = \frac{pd_i}{4t} \]

\[ \sigma_3 = 0 \]

(5.67)

and therefore, the maximum shear stress \( \tau_{\text{max}} \) is given by

\[ \tau_{\text{max}} = \frac{(\sigma_1 - \sigma_3)}{2} = \frac{(pd/2t)}{2} = \frac{pd}{4t} \]

(5.68)

Figure 5.18 summarizes the fracture behavior of a thin-walled cylinder. Normal and shear stress trajectories are shown in Fig. 5.18a. Figure 5.18b illustrates a common mode of failure where longitudinal cracks parallel to the direction of \( \sigma_l(\sigma_x) \) are developed by the circumferential tensile stress \( (\sigma_l = \sigma_2) \).

### 5.13.3 Thick-Walled Cylinder

When the ratio of the internal diameter of the cylinder to thickness is less than 10–15, the assumption of a uniform stress across the thickness of the cylinder is no longer valid. If the cylinder is of medium thickness, the circumferential...
tensile stress $s_1$ (maximum tensile stress) is given by

$$s_1 = \frac{p d_e}{2t}$$

where $p$ is the internal pressure, $d_e$ is the external diameter, and $t$ is the thickness. However, if the cylinder is extremely thick, the stress $s_1$ is given by

$$s_c = \frac{p (r_e^2 + r_i^2)}{(r_e^2 - r_i^2)}$$

\[ \text{Figure 5.18 Fracture behavior of a thin-walled cylinder. (a) Normal and shear stress trajectories. (b) Common failure mode; longitudinal cracks are developed by the circumferential stress } \sigma_c. \]
where \( p \) is the internal pressure, \( r_e \) is the external radius of the cylinder, and \( r_i \) is its internal radius.

### 5.13.4 Riveted Joints

Riveted connections consist of perforated plates joined together by rivets. They are used in pressure vessels, as well as in structural work, to maintain rigid and leak proof connections. Strength of riveted joints is limited by the smaller of (i) the capacity of the rivets to transmit load between the joined plates and (ii) the tearing resistance of the plates. To prevent creation of frictional forces, the joined plates are not clamped tightly, and therefore the rivets are beared on the holes. Generally, three types of stresses can develop in riveted connections: (i) shear, (ii) tensile, and (iii) bearing. Each of these stresses can be calculated as follows.

Three basic configurations of riveted joints are distinguished as (i) lap joint, (ii) butt joint with one cover plate, and (iii) butt joint with two cover plates. A single-riveted lap joint such as that shown in Fig. 5.19a is the simplest joint to consider. Such a joint is said to fail if any relative displacement or shear occurs between the two joined plates under the influence of an applied tensile load \( P \), as illustrated in Fig. 5.19b. Since only one cross section of the rivet is subjected to shear stress (Fig. 5.19a), the joint is said to be of the single-shear mode. If the diameter of the hole is \( d \), the area \( A \) subjected to a shear stress \( t \) is \( \pi d^2/4 \), and therefore

\[
\tau = P/(\pi d^2/4) = 4P/\pi d^2
\]

Butt-riveted joints are illustrated in Fig. 5.20. In general, the shear strength of a riveted joint is dependent upon two parameters: (i) shear mode and (ii) the pitch. Riveted joints can be of a single- or higher-order shear mode. A riveted butt joint with one cover plate (Fig. 5.20a) is also of the single shear mode. However, a butt joint with two cover plates (Fig. 5.20b) is of a double shear mode because, as can be seen, two sections are subjected to shear.

A rivet pitch is defined as the largest distance from the center of one rivet to the center of adjacent rivet in the same row, as illustrated in Fig. 5.21. It is shown that a row of rivets is defined as a line of rivets in a direction normal to the line of action of the applied load. To determine the shear stress developed in a generalized riveted joint, the number of rivets in the pitch must be determined. In the example of Fig. 5.21, row 1 contributes 1 rivet and row 2 contributes \( \frac{1}{2} + \frac{1}{2} = 1 \) rivet, i.e., the total number of rivets in the pitch = 2. In general, the shear stress \( \tau \) developed in a riveted joint can be expressed as:

\[
\tau = P_s/nA_s \text{ (order of shear mode)}
\]
or

\[ \tau = \frac{P_s}{n(\pi d^2/4)} \]  (order of shear mode)

where \( P_s \) is the applied shear load, \( n \) is the number of rivets per pitch, and \( A_s \) is the cross-sectional area of the rivet subjected to shear = \( \pi d^2/4 \) (\( d \) is the hole diameter).

Another mode of failure of riveted joints is tearing of the plate resulting from a tensile stress, as shown in Fig. 5.22. As can be seen, tearing occurs along a section through the rivet hole whose area \( A \) is given by \( wt - dt = \)

Figure 5.19 Shear of single riveted lap joint. (a) A schematic illustration of a single riveted lap joint. (b) Failure of the joint by shear across the shaded area.
where \( w \) is the width of the plate, \( t \) is its thickness, and \( d \) is the diameter of the hole. In this case, the tensile stress \( \sigma_t \) is given by:

\[
\sigma_t = \frac{P}{(w - d) t}
\]

Bearing is also a possible mode of failure of riveted joints. Excessive bearing pressure can cause plastic deformation resulting in enlargement of the rivet.
hole, as shown in Fig. 5.23. Since the area subjected to bearing stress is \(td\), the bearing stress \(\sigma_b\) is given by
\[
\sigma_b = \frac{P}{td}
\]

5.13.5 Welded Components

Two main types of welds are distinguished: (i) butt weld and (ii) fillet weld. A butt weld can either be single or double, as illustrated in Fig. 5.24a,b. Fillet welds

**Figure 5.22**  Failure of a riveted joint by tearing.

**Figure 5.23**  Failure of a rivet by excessive bearing pressure.
include the side type and the transverse, or end, type. Figure 5.24c,d illustrates these types of welds.

For a butt weld, the weld area is given by $Lt$, where $L$ is the length of the weld and $t$ is the thickness of the thinner plate of the joint. If the weld
is supporting a load $P$ (tensile or compressive), it develops a normal stress $\sigma$ given by

$$\sigma = \frac{P}{Lt}$$

In general, butt welds are as strong as the base material. However, it is recalled from Chap. 2 that the welding process can lower the strength of the base material adjacent to the weld. Therefore, the allowable tensile stress for the base material should be somewhat lower to account for possible weakening by the welding process.

As shown in Fig. 5.24, the direction of applied load $P$ is normal to the transverse fillet weld and parallel to a side fillet weld. Mechanical strength of both welds is determined by shear stresses developed in what is known as the throat of the weld regardless of the direction of the applied force. Figure 5.25 illustrates the throat at the triangular shaped fillet. In this case, the area subjected to shear stress is given by the throat $t$ of the weld multiplied by its length $L$. From the geometry of Fig. 5.25, $t = h \sin 45^\circ$ where $h$ is the height or leg of the weld. If the applied load is $P$, the shear stress $\tau$ developed at the throat is given by

$$\tau = \frac{P}{Lh \sin 45^\circ}$$

### 5.14 Solved Problems

The primary objective of stress analysis in failure investigations is to determine whether the failed component is properly designed so that the stresses developed during service are within the design or allowable limit. In this section some
examples are given to illustrate the applications of the relationships developed in previous sections.

**Problem 1**

The state of stress at a given point in a part is determined by: $\sigma_x = 50 \text{ MPa}$, $\sigma_y = 0$, $\sigma_z = 0$, $\tau_{xy} = 30 \text{ MPa}$, $\tau_{xz} = 0$, $\tau_{yz} = 0$. Determine the principal stresses and their orientations relative to the $x$, $y$, and $z$ axes, and the maximum shear stress both graphically and analytically.

**Solution:** To solve the problem graphically, a Mohr’s circle is constructed as shown in Fig. 5.26. Points A and B with coordinates of $(\sigma_x, \tau_{xy})$ and $(\sigma_y, -\tau_{xy})$ are connected and their intersection $C$ with the $\sigma$ axis determines the center $C$ of the circle with radius $R = CA = CB$. Line CA defines the $x$ axis and line $CB$ defines the $y$ axis. Since one stress is positive (tensile) and the other is negative (compressive), the positive stress is the maximum stress $\sigma_1$, and the negative stress must be the minimum stress corresponding to $\sigma_3$. In this case $\sigma_2 = 0$ consistent with the convention $\sigma_1 > \sigma_2 > \sigma_3$. From the geometry of Fig. 5.26 and noting that $\sigma_y = 0$, $\sigma_1$ is given by

$$\sigma_1 = R + \sigma_x / 2$$

where $R$ is the radius of the circle, and $\sigma_3$ is given by

$$\sigma_3 = R - \sigma_x / 2$$

![Figure 5.26](image_url)

**Figure 5.26** Graphical representation of Prob. 1 in the form of Mohr’s circle.
As can be seen from Fig. 5.26, the radius of the circle $R$ is given by

$$R = \left[\left(\frac{\sigma_x}{2}\right)^2 + \tau_{xy}^2\right]^{1/2}$$

Substituting the values of $\sigma_x$ and $\tau_{xy}$ in the above equation ($\sigma_x = 50$ MPa and $\tau_{xy} = 30$ MPa) results in

$$R = \left[(25)^2 + (30)^2\right]^{1/2} = 39 \text{ MPa}$$

and therefore

$$\sigma_1 = 25 + 39 = 64 \text{ MPa}$$
$$\sigma_3 = 25 - 39 = -14$$

Note that $\sigma_1 + \sigma_3 = 64 - 15 = 50 = \sigma_x + \sigma_y$, consistent with Eq 5.12. By definition, the maximum shear stress $\tau_{\text{max}} = (\sigma_1 - \sigma_3)/2 = (64 + 14)/2 = 39$ MPa, corresponding to the radius of the circle. From the geometry of Fig. 5.26, $\tan \alpha = \tau_{xy}/(\sigma_x - \sigma_y)/2 = 30/25$, and therefore $\alpha = 50^\circ$. It is recalled that the angles determined from Mohr’s circle are twice the real angles, and therefore the principal stress $\sigma_1$ is oriented at an angle of $50^\circ/2 = 25^\circ$ from the $x$ axis, and at an angle of $(180 - 50)/2 = 65^\circ$ from the $y$ axis. It is then evident that $\sigma_3$ is oriented at an angle of $65^\circ$ from the $x$ axis and $25^\circ$ from the $y$ axis.

To solve the above problem analytically, the cubic stress Eq. (5.9) is used. Since $\sigma_y = \sigma_z = \tau_{xz} = \tau_{yz} = 0$, the equation is reduced to

$$\sigma^3 - \sigma^2\sigma_x + \sigma(-\tau_{xy}^2) = 0$$

It is evident that $\sigma = 0$ satisfies the above equation, and therefore it is one of its roots. The other two roots are obtained from the following quadratic equation resulting from dividing the above equation by $\sigma$:

$$\sigma^2 - \sigma\sigma_x - \tau_{xy}^2 = 0$$

The above equation has two roots given by

$$\sigma' = \left(\sigma_x + \left(\sigma_x^2 + 4\tau_{xy}^2\right)^{1/2}\right)/2$$
$$\sigma'' = \left(\sigma_x - \left(\sigma_x^2 + 4\tau_{xy}^2\right)^{1/2}\right)/2$$

Substituting the values of $\sigma_x = 50$ and $\tau_{xy} = 30$ in the above equations gives $\sigma' = 64$ MPa, $\sigma'' = -14$ MPa. Since by convention $\sigma_1 > \sigma_2 > \sigma_3$, it becomes evident that $\sigma_1 = \sigma' = 64$ MPa, $\sigma_2 = 0$, and $\sigma_3 = -14$ MPa, and the maximum shear stress $\tau_{\text{max}} = (\sigma_1 - \sigma_3)/2 = (64 - (-14))/2 = 78/2 = 39$ MPa. To obtain the orientations of the principal planes, the above values of $\sigma_1$ and $\sigma_3$ are
substituted into Eq. (5.23) resulting in

\[
\begin{align*}
& l(\sigma_1 - \sigma_x) - m\tau_{xy} = 0 \\
& -l\tau_{xy} + m\sigma_1 = 0 \\
& n\sigma_1 = 0
\end{align*}
\]

Since \(\sigma_1\) does not equal 0, \(n = 0\), and therefore, the relationship \(l^2 + m^2 + n^2 = 1\) is reduced to \(l^2 + m^2 = 1\). Substituting the values of \(\sigma_1 = 64\) MPa, \(\sigma_x = 50\) MPa, and \(\tau_{xy} = 30\) MPa into the above equations results in

\[
\begin{align*}
& l(64 - 50) - m(30) = 0 \\
& -l(30) + m(64) = 0
\end{align*}
\]

Substituting \(l^2 + m^2 = 1\) into the above equations results in

\[
196l^2 = 900m^2 = 900(1 - l^2)
\]

that is \(1096l^2 = 900\), and therefore \(l = 0.91\). Using the same terminology as in the graphical solution, \(0.91 = \cos^{-1} \alpha/2\), i.e., \(\alpha/2 = 25^\circ\); i.e., \(\sigma_1\) is oriented at \(25^\circ\) from the \(x\) axis. Since \(l = 0.91, m = 0.41\) (from \(l^2 + m^2 = 1\)); therefore, \(\sigma_1\) is oriented at \(65^\circ\) from the \(y\) axis. From the values of \(\sigma_1\) and \(\sigma_3\), the maximum shear stress \(\tau_{\text{max}} = \frac{64 - (-14)}{2} = 39\) MPa. As can be seen, identical results are obtained by the graphical method.

**Problem 2**

Using the results of the preceding problem, determine the yield strength of a ductile material to be used in manufacturing the part for a static load application.

**Solution:** From the results of Prob. 1, the maximum normal stress \(\sigma_1 = 64\) MPa. Since a ductile material is to be used in static load application, the recommended factor of safety is 2 (see Table 5.1). According to Eq. (5.56), the design or allowable stress \(\sigma_d\) is given by

\[
\sigma_d = \sigma_1/N = \sigma_1/2
\]

Since the allowable stress cannot exceed the maximum stress developed in the part, i.e.,

\[
\sigma_d \leq \sigma_1
\]

the material must have a yield strength not less than \(2\sigma_1\), i.e., \(2 \times 64 = 128\) MPa.

**Problem 3**

A solid cylinder of a 1 m radius is acted upon by a torque of \(5 \times 10^8\) N\cdot m. Calculate the maximum principal stress and maximum shear stress developed in the cylinder.
Solution: To determine the principal stresses, a suitable coordinate system is selected as shown in Fig. 5.27, where the $x$ axis is taken parallel to the axis of the cylinder. An elemental cube is extracted from the cylinder to illustrate the state of stress at any point with reference to the selected coordinate system.

Among the three stress components completely defining the biaxial state of stress at any point, the only nonzero component is $\tau_{xy}$ as indicated by the loading condition. In this case, $\sigma_x = \sigma_y = 0$, and therefore, the cubic stress equation becomes

$$\sigma^3 + \sigma(-\tau_{xy})^2 = 0$$

Since $\sigma = 0$ satisfies the above equation, it is one of its roots, and the other two roots are obtained from solving the following quadratic equation resulting from dividing the above equation by $\sigma$:

$$\sigma^2 - \tau_{xy}^2 = 0$$

It is evident that the above equation has two solutions:

$$\sigma' = \tau_{xy}$$
$$\sigma'' = -\tau_{xy}$$

Using the convention $\sigma_1 > \sigma_2 > \sigma_3$, it is clear that $\sigma_1 = \sigma'$, $\sigma_2 = 0$, and $\sigma_3 = \sigma''$. It is recalled from Eq. (5.47), that the shear stress $\tau_{xy}$ developed in a solid cylinder of diameter $d$ as a result of torque $T$ is given by

$$\tau_{xy} = \frac{16T}{\pi d^3} = \frac{16(5 \times 10^8)}{\pi(8)} = 3.2 \times 10^8 \text{ N/m}^2 = 320 \text{ MPa}$$
and therefore:
\[ \sigma_1 = 320 \text{ MPa} \]
\[ \sigma_3 = -320 \text{ MPa} \]
and it is evident that the maximum shear \( \tau_{\text{max}} = (\sigma_1 - \sigma_3)/2 = 320 \text{ MPa} = \tau_{xy} \).

**Problem 4**

It is determined that the state of stress at a point within a part is \( \sigma_x = 600 \text{ MPa}, \sigma_y = 500 \text{ MPa}, \) and \( \tau_{xy} = 100 \text{ MPa} \). Calculate the principal stresses and maximum shear stress. If a brittle material is used to manufacture that part, what is its ultimate tensile strength?

**Solution:** Recalling that the state of stress at any point is defined in terms of \( \sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{xz}, \) and \( \tau_{yz} \), it is evident that \( \sigma_z = \tau_{xz} = \tau_{yz} = 0 \), and therefore the cubic stress equation is reduced to
\[ \sigma^3 - \sigma^2(\sigma_x + \sigma_y) + \sigma(-\tau_{xy}^2) = 0 \]

One of the three roots of the above equation is \( \sigma = 0 \), and the other two roots are obtained from solving the following quadratic equation resulting from dividing the above equation by \( \sigma \):
\[ \sigma^2 - \sigma(\sigma_x + \sigma_y) - \tau_{xy}^2 = 0. \]

The above equation has the following two roots:
\[ \sigma' = \frac{1}{2}(\sigma_x + \sigma_y) + \left\{\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right\}^{1/2} \]
\[ \sigma'' = \frac{1}{2}(\sigma_x + \sigma_y) - \left\{\frac{1}{2}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2\right\}^{1/2} \]

Substituting the values of \( \sigma_x, \sigma_y, \) and \( \tau_{xy} \) in the above equations results in
\[ \sigma' = 662 \text{ MPa} \]
\[ \sigma'' = 438 \text{ MPa} \]
It is then evident that \( \sigma_1 = \sigma' = 662 \text{ MPa}, \sigma_2 = \sigma'' = 438 \text{ MPa}, \) and \( \sigma_3 = 0 \). In this case, the maximum shear stress \( \tau_{\text{max}} = (\sigma_1 - \sigma_3)/2 = 662/2 = 331 \text{ MPa} \).

If a brittle material is used, Table 5.1 indicates that a factor of safety \( N = 6 \) on the basis of ultimate tensile strength should be used. Since \( \sigma_1 = 662 \text{ MPa} \), the ultimate strength of the material must not be less than \( 662 \times 6 = 3972 \text{ MPa} \).

**Problem 5**

At a given point within a part, the principal stresses are \( \sigma_1 = 70 \text{ MPa}, \sigma_2 = 30 \text{ MPa}, \) and \( \sigma_3 = 0 \). It is required to determine the normal stress \( \sigma \) and shear
stress $\tau$ acting on a plane oriented at $40^\circ$ from the direction of $\sigma_1$ using both the analytical and graphical methods.

**Solution:** According to Eq. (5.26), the normal stress $\sigma$ and shear stress $\tau$ acting on a plane whose direction cosines relative to the principal stress are

\[
\sigma = l^2 \sigma_1 + m^2 \sigma_2 \\
\tau^2 = l^2 \sigma_1^2 + m^2 \sigma_2^2
\]

Given:

$\sigma_1 = 70$ MPa, $\sigma_2 = 30$ MPa

$l = \cos 40^\circ$; $l^2 = 0.59$

$m = \cos 50^\circ$; $m^2 = 0.41$

Substituting these values into the above equation results in

$\sigma = (0.59)(70) + (0.41)(30) = 53.6$ MPa

and substituting the value of $\sigma$ in the shear stress equation gives

$\tau^2 = (0.59)(4900) + (0.41)(900) - 2873 = 387$ and

$\tau = 19.7$ MPa

To solve the problem graphically, Mohr’s circle is plotted as shown in Fig. 5.28 with its radius $R = (\sigma_1 - \sigma_2)/2 = (70 - 30)/2 = 20$ MPa. It is recalled that an angle in Mohr’s circle is represented by twice its true value.

![Figure 5.28](image-url)
Therefore, to determine the state of stress on the plane oriented at 40° from $\sigma_1$, a line is drawn from the center C such that it makes an angle of 40° with the normal stress axis. Its intersection A with the circle defines the respective state of stress by the coordinates $\sigma_x$ and $\tau_{xy}$. From the geometry of Fig. 5.28,

$$
\sigma_x = R \cos 80 + R + \sigma_3 = 20(1.17) + 30 = 53.5 \text{ MPa}
$$

$$
\tau_{xy} = R \sin 80 = 19.7 \text{ MPa}
$$

which are the same as those obtained by the analytical method.

Problem 6

A rectangular part of the dimensions shown in the Fig. 5.29 is subjected to a tensile force of 1000 N and a compressive force of 700 N, as illustrated. It is required to analytically determine the normal and shear stresses developed on the diagonal plane. Assume that the stresses are uniformly distributed.

Solution: As can be seen from Fig. 5.29, the tensile force acts on an area of $1 \times 0.1 = 0.1 \text{ m}^2$ producing a tensile stress $\sigma_t$ given by

$$
\sigma_t = +1000/0.1 = 10,000 \text{ N/m}^2
$$

An area of $0.6 \times 1 = 0.6 \text{ m}^2$ is acted upon by a compressive force producing a compressive stress $\sigma_c$ given by

$$
\sigma_c = -7000/0.6 = -11,667 \text{ N/m}^2
$$

It is evident that $\sigma_1 = \sigma_t$, $\sigma_2 = 0$, and $\sigma_3 = \sigma_c$. As shown in Fig. 5.29, the normal to the diagonal plane is oriented at an angle $\alpha$ from $\sigma_1$, where $\tan \alpha = 0.6$, and

![Figure 5.29](image_url)

**Figure 5.29**  Tensile and compressive stresses acting on a rectangular part (Prob. 6).
therefore \( \alpha = 31^\circ \). By definition, the direction cosine \( l = \cos 31^\circ \), and \( m = \cos (90 - 31) = \cos 59^\circ \), i.e., \( l^2 = 0.73 \) and \( m^2 = 0.265 \); therefore,

\[
\sigma = l^2 \sigma_1 + m^2 \sigma_2 = (0.73)(10,000) + (0.265)(-11,667) = 4208 \text{ N/m}^2
\]

\[
\tau^2 = (0.73)(108) + (0.265)(1.4 \times 108) - 0.2 \times 108 = 85 \times 106, \quad \text{and}
\]

\[
\tau = 9 \times 10^3 \text{ N/m}^2
\]

### Problem 7

At any arbitrary point within a part, the state of stress is defined by \( \sigma_x = 5000 \text{ MPa}, \ \sigma_y = -2000 \text{ MPa}, \ \text{and} \ \tau_{xy} = -3000 \text{ MPa} \). Using Mohr’s circle construction, determine (i) the principal stresses, (ii) the maximum shear stress and the normal stress acting on the plane of maximum shear stress, and (iii) the orientations of the planes of maximum shear stress relative to the \( x \) axis.

**Solution:** It is recalled that according to convention, a positive shear stress tends to produce a clockwise rotation, a negative shear stress produces an anticlockwise rotation, and \( \tau_{xy} = -\tau_{yx} \). In this case, \( \tau_{xy} = -3000 \text{ MPa} \), and therefore \( \tau_{yx} = +3000 \text{ MPa} \).

Mohr’s circle is then plotted as shown in Fig. 5.30. By definition, \( CA \) corresponds to the \( x \) axis and \( CB \) corresponds to the \( y \) axis. From the geometry of the Fig. 5.30, the radius \( R \) of the circle is given by

\[
R = \left( \frac{1}{2} \left( \sigma_x - \sigma_y \right) \right)^2 + \tau_{xy}^2 \right)^{1/2}
\]

Substituting the values of \( \sigma_x = 5000 \text{ MPa}, \ \sigma_y = -2000 \text{ MPa}, \ \text{and} \ \tau_{xy} = -3000 \text{ MPa} \) into the above equation results in \( R = 4610 \text{ MPa} \), and the principal stresses \( \sigma' \) and \( \sigma'' \) are given by.

\[
\sigma' = \frac{1}{2} (\sigma_x + \sigma_y) + R = 1500 + 4610 = 6110 \text{ MPa}
\]

\[
\sigma'' = \frac{1}{2} (\sigma_x + \sigma_y) - R = 1500 - 4610 = -3110 \text{ MPa}
\]

Since \( \sigma' \) is positive and \( \sigma'' \) is negative, \( \sigma_1 = \sigma' \), \( \sigma_2 = 0 \), and \( \sigma_3 = \sigma'' \). In this case, the maximum shear stress \( \tau = 4610 \text{ MPa} = (\sigma_1 - \sigma_3)/2 \). As shown in Fig. 5.30, the points corresponding to the maximum shear stress are \( D \) and \( D' \), i.e., \( \tau_{\max} = R = \pm 4610 \), and the normal shear stress acting on the planes of maximum shear stress corresponds to \( OC = \frac{1}{2} (\sigma_1 + \sigma_3) = \frac{1}{2} (\sigma' + \sigma'') = 3000/2 = 1500 \text{ MPa} \). From the geometry of Fig. 5.30, the angle between the direction of maximum shear stress and the \( x \) axis is \( 90^\circ + \alpha \), where \( \alpha = \tan^{-1} 3000/(5000 - 1500) = 40.6^\circ \). Therefore, the true angle between the normal to the plane of maximum shear stress and the \( x \) axis is \( (90 + 40.6)/2 = 65.3^\circ \).
Problem 8

A shaft (solid cylinder) 0.6 m in diameter is subjected to a bending moment of $5 \times 10^6$ J. Using Mohr’s circle construction, determine the torque, which can act simultaneously with the bending moment such that the maximum normal and shear stresses do not exceed 300 and 250 MPa, respectively.

Solution: It is recalled from Eq. (5.45) that when a part is subjected to a bending moment $M$, it produces a maximum normal stress $\sigma_{\text{max}}$ (tensile at the top and compressive at the bottom), given by

$$\sigma_{\text{max}} = \frac{Mh}{I}$$

where $h$ is the distance from the neutral axis to the top or bottom (in the case of a cylinder of diameter $d$, $h = d/2$) and $I$ is the moment of inertia of the part cross section. From Appendix B, $I$ for a circular cross section is $\pi d^4/64$. Therefore, the maximum tensile stress developed at the surface of the shaft is given by

$$\sigma_{\text{max}} = \frac{(64)(5 \times 10^6)(0.3)}{\pi (0.6)^4} = 200\,\text{MPa}$$

Figure 5.30 Graphical illustration of Prob. 7 by Mohr’s circle.
When the additional torque is applied, the normal stress resulting from the bending moment becomes combined with a shear stress. Such a state of stress is represented as shown in Fig. 5.31 by points $A$ and $B$ with coordinates of $(\sigma_x, \tau_{xy})$, and $(0, -\tau_{xy})$, respectively, where $\sigma_x = 200$ MPa and $\tau_{xy}$ is unknown. Since the maximum shear stress produced by the torque cannot exceed 250 MPa, the radius of the Mohr’s circle must be 250 MPa. Given this information, it is possible to construct a Mohr’s circle with center at point $C$ as shown in Fig. 5.31. It is clear from the geometry of Fig. 5.31 that $OC = 100$ MPa, and therefore

$$\tau_{xy} = (250^2 - 100^2)^{1/2} = 229 \text{ MPa} = 229 \times 10^6 \text{N/m}^2$$

Substituting the above value of $\tau_{xy}$ into Eq. (5.47):

$$\tau_{xy} = 16T/\pi d^3$$

![Figure 5.31](image) Construction of Mohr’s circle for Prob. 8.
results in:

\[ T = \tau_{xy} \frac{md^3}{16} = 9.7 \times 10^6 \text{N} \cdot \text{m} \]

**Problem 9**

Given a solid shaft subjected to torque \( T \), identify the normal and shear stress trajectories on its surface using the analytical method. If the shaft is made of a brittle material, draw sketches illustrating the fracture path for ductile and brittle materials.

**Solution:** To draw the stress trajectories, it is first required to identify the orientations of the principal stresses and maximum shear stresses on the surface of the shaft. Consider the shaft shown in Fig. 5.32, where its axis is selected to be the \( x \) axis. It is recalled from Prob. 3, that the only nonzero component of stress produced by the torque \( T \) is \( \tau_{xy} \), and therefore the cubic stress Eq. (5.9) is reduced to

\[
\sigma^3 + \sigma(-\tau_{xy}^2) = 0 \quad \text{or} \quad \sigma(\sigma^2 - \tau_{xy}^2) = 0
\]
It is evident that the solutions of the above equation are

\[ \sigma_1 = \tau_{xy} \text{ (tensile stress)} \]
\[ \sigma_2 = 0 \]
\[ \sigma_3 = -\tau_{xy} \text{ (compressive stress)} \]

From Eqs. (5.23) and noting that \( \sigma_x = \sigma_y = 0 \), the orientations of the principal stresses are given by

\[ l\sigma_1 - m\tau_{xy} = \tau_{xy}(l - m) = 0 \]
\[ -l\tau_{xy} + m\sigma_1 = 0 = \tau_{xy}(m - l) = 0 \]
\[ n\sigma_1 = n\tau_{xy} = 0 \]

Since \( \tau_{xy} \) does not equal 0, \( l = m, \) and \( n = 0, \) and from \( l^2 + m^2 = 1, \)
\( l = \pm \left(\frac{1}{2}\right)^{1/2}, \) and \( m = \pm \left(\frac{1}{2}\right)^{1/2}. \) Therefore, \( \sigma_1 \) and \( \sigma_3 \) are oriented at 45° from the principal stress, they must lie along the x and y axes. Figure 5.32 illustrates the tensile stress trajectories. Assuming that the material of the cylinder is brittle, it can be expected to fail on the planes normal to the maximum normal stress \( \sigma_1; \) i.e., the fracture path follows the trajectories of \( \sigma_3. \)

**Problem 10**

A thin-walled spherical pressure vessel of 1 m internal diameter is designed to carry a gas at a pressure of 10 MPa. If the material used to construct the vessel has a yield strength of 600 MPa, it is required to determine the wall thickness using a factor of safety \( N = 4. \)

**Solution:** According to Eq. (5.65), the maximum normal stress \( \sigma_1 \) developed in a thin-walled spherical vessel is given by

\[ \sigma_1 = \frac{pd}{4t} \]

where \( p \) is the internal pressure, \( d \) is the internal diameter, and \( t \) is the wall thickness. Given a yield strength of 600 MPa and a factor of safety \( N = 4, \) the maximum allowable stress or design stress \( \sigma_d \) is 600/4 = 150 MPa, i.e., \( \sigma_1 = 150 \) MPa. Substituting this value into the above equation (\( p = 10 \) MPa, \( d = 1 \) m) results in

\[ 150 = 10(l)/4t \quad \text{and} \quad t = 10/600 \text{ m} = 17 \text{ mm} \]

**Problem 11**

A thin-walled cylinder has an internal diameter of 1 m and a wall thickness of 3 mm. If the cylinder is designed to withstand an internal pressure of 1 MPa, it
is required to determine the yield strength of the cylinder material using 4 as a factor of safety.

Solution: It is recalled that the criterion for a thin-walled cylinder is that \( d/t > 10\) (\( d/t \) in this problem = 3000). It is recalled from Sec. 5.13.2 that the circumferential stress \( \sigma_1 \) and longitudinal stress \( \sigma_2 \) developed in a thin-walled cylinder as a result of an internal pressure \( p \) are given by

\[
\sigma_1 = \frac{pd}{2t} = \left(l/l_0\right)/2(0.003) = 166 \text{ MPa}
\]

\[
\sigma_2 = \frac{\sigma l}{2} = \frac{pd}{4t} = 83 \text{ MPa}
\]

Since \( \sigma_1 > \sigma_2 \), the yield strength of the material = \( \sigma_1 \times 4 = 332 \text{ MPa} \).
6

Macrosopic Aspects of Fracture and Fracture Mechanics

6.1 Definition of Fracture

Fracture is defined as the separation of a part into two or more pieces. It can result from a purely mechanical effect such as an applied or a residual stress, or from a combination of mechanical and environmental effects. Problems related to fracture have a wide scope, ranging from catastrophic failure of structures and machine parts to processes occurring on the submicroscopic scale and leading to separation of atomic planes within the crystals. In this chapter, emphasis is placed upon fracture related to purely mechanical effects. Environmentally assisted fracture is discussed in Chap. 9.

It is recognized that any fracture process involves two distinct stages: (i) crack initiation, or nucleation, and (ii) crack propagation as schematically illustrated in Fig. 6.1. Frequently, the terms crack and fracture are intermixed, which can lead to a great deal of confusion about the mechanism or exact process leading to fracture, particularly in failure analysis investigations. Cracks or discontinuities in a part can be inevitably introduced into a product during the primary and secondary fabrication techniques, as well as during joining processes, as pointed out in Chap. 2, e.g., voids in casting products, flaws in welded structures, and cracks produced by forming operations.

An initially existing crack can propagate through the part by various mechanisms depending upon a number of parameters, particularly: (i) the loading...
conditions, (ii) service temperature, (iii) geometry of the part, and (iv) nature of the environment. Regardless of the mechanism by which an initially existing crack propagates through the part, once it reaches a critical size $c_c$ and the part becomes overloaded because of the corresponding reduction in cross-sectional area supporting the load. (c) Final separation due to an overloading condition.

Figure 6.1  Schematics illustrating the process of fracture. (a) An initial crack of size $c_0$. (b) Crack propagation under the influence of a tensile load $P$; the crack reaches a critical size $c_c$ and the part becomes overloaded because of the corresponding reduction in cross-sectional area supporting the load. (c) Final separation due to an overloading condition.

6.2 Objective of Fracture Mechanics

Given an existing crack of a certain size in a part, the primary objective of fracture mechanics is to develop means by which the crack propagation rate can be
expressed in terms of some measurable parameters qualified to be characteristic materials properties such as yield strength and ultimate tensile strength. It is evident that an existing crack such as that shown in Fig. 6.1 reduces the part strength. Obviously, a direct correlation exists between the strength in the presence of a crack or residual strength and crack size. Such a correlation can be quantitatively developed by means of fracture mechanics. Also, it is possible to determine the maximum crack size, which can be accommodated by the part without the danger of catastrophic failure. This is called the allowable crack size by analogy to the allowable stress described in Chap. 5. Consequently, it becomes possible to design products with existing cracks of allowable size. Furthermore, the scope of fracture mechanics covers the incubation period of crack growth and the service life in the presence of existing cracks.

Application of fracture mechanics to failure analysis investigations can provide a great deal of information, which elucidates the cause of failure. For example, by means of fracture mechanics it is possible to determine the stress at the time of failure. If this stress is found to be considerably greater than the design stress defined in Chap. 5, it becomes evident that the machine part or structure has been overloaded. Conversely, if the failure stress is below or approaches the design stress, it can be concluded that the allowable crack size is greater than can be tolerated.

Use of proper terminology when discussing various aspects of fracture and fracture mechanics particularly in failure analysis investigations is of extreme importance. It is recalled that the results of failure analysis investigations can be used in legal proceedings to determine responsibility. As pointed out earlier, the final act in any fracture is the same regardless of the mechanism by which the cracks are propagated, yet terms such as brittle fracture and ductile fracture, or brittle failure and ductile failure, are very commonly used despite earlier attempts to avoid the use of such terms. Other misleading terms include slow fracture, fast fracture, creep fracture, fatigue fracture, and intermixing the definitions of such terms as type of failure and mode of failure. Therefore, it is extremely important to clarify the proper terminology, which must be used in relation to fracture.

6.3 Use of the Terms Brittle and Ductile in Fracture

Frequently the term brittle fracture is used to imply that the final fracture or separation is preceded by a small macroscopic plastic deformation. Conversely, a large amount of macroscopic plastic deformation manifested as obvious distortion of the part is taken to imply a ductile fracture. According to this classification, if a material is behaving in an ideally brittle manner during a tensile
test, fracture must be preceded by only elastic deformation, as shown in Fig. 6.2a. In this case the two pieces resulting from the fracture can be fitted together to restore the original geometry of the specimen. An ideal ductile behavior during a tensile test requires plastic deformation to continue in the necked cross section until it is reduced to a point as shown in Fig. 6.2b, where the fracture in this case is referred to as rupture. However, such an ideal behavior is rarely if ever observed in practice as explained below.

Some of the engineering alloys are susceptible to a phenomenon called cleavage, which leads to fracture as explained in Chap. 10. It suffices to mention here that cleavage can occur under a hydrostatic state of stress where no shear stress is generated. Since plastic deformation can occur by only a shear stress, it becomes evident that cleavage is a brittle phenomenon. However, fracture by a cleavage mechanism is almost always accompanied by localized plastic deformation on a microscopic scale. Furthermore, some materials are known to fracture by cleavage after a relatively large amount of macroscopic plastic deformation.

Although ideal ductile behavior requires the necked region of a tensile specimen to be reduced to a point, separation of a tensile specimen usually occurs before the necked cross section is greatly reduced. Another possibility is that a material classified as ductile may fracture due to localized deformation without appreciable macroscopic deformation.

It is evident from the above discussion that the terms brittle and ductile must be reserved to describe only the mechanism by which a crack propagates through the material, i.e., a crack may propagate by either a brittle or ductile mechanism.

**Figure 6.2** Ideal tensile fracture behavior of brittle and ductile materials. (a) Ideal brittle behavior. (b) Ideal ductile behavior.
6.4 Crack Loading Modes and Macroscopic Morphology of Fracture Surfaces

In terms of loading conditions of an existing crack, three basic modes can be distinguished, as schematically illustrated in Fig. 6.3: (i) the direct opening or mode I, (ii) the forward shear or edge sliding (mode II), and (iii) the sideways shear or tearing mode (mode III).

Mode I cracking is characterized by tensile stress acting normal to the crack surface (Fig. 6.3a). In this case, fracture results from crack propagation in a direction normal to its surface. For mode II (Fig. 6.3b), the crack surface is acted upon by a shear stress along a direction normal to the crack edge. As a result, the crack surfaces are sheared relative to each other along a direction normal to the crack edge. Also, the crack surface of mode III is acted upon by a shear stress; however, in this case, the direction of the shear stress is parallel to the crack edge, as shown in Fig. 6.3c. A generalized crack is simultaneously subjected to all three types of loading.

In terms of the crack loading modes described above, it is possible to classify the resulting fracture surfaces on the basis of their macroscopic appearance. Terms frequently used in connection with this classification include square and slant, representing two extreme cases as described below.

A fracture surface of a square morphology results when the cracks are propagated along a surface normal to an applied tensile stress corresponding to mode I cracking (Fig. 6.3a). In this case, the resulting fracture surface is flat, as shown in Fig. 6.4a. Such a flat surface can be produced when the cracks propagate by a brittle mechanism, e.g., cleavage.

Figure 6.3 Basic modes of fracture. (a) Mode I. (b) Mode II. (c) Mode III. The arrows indicate the direction of displacement of crack surfaces.
When the fracture surface is inclined at $45^\circ$ to the tensile stress axis, it is said to have a slant morphology, as shown in Fig. 6.4b. Cracks producing a slant fracture propagate along the surfaces of maximum shear stress. It is recalled from Chap. 5 that the planes of maximum shear stress are oriented at $45^\circ$ from the tensile stress axis. If the material consists of one crystal, i.e., single crystal, the cracks propagate across one surface resulting in the slant morphology shown in Fig. 6.4b. For an aggregate of crystals, i.e., polycrystalline, the cracks propagate across two surfaces, producing the slant morphology shown in Fig. 6.4c. Generally, a slant morphology results when plastic deformation can readily occur at the crack tip, as further explained later in this section. In terms of the crack loading mode, a slant morphology is produced from a combination of mode I and mode III loading, as schematically illustrated in Fig. 6.5.

Figure 6.4d illustrates the macroscopic appearance of a more generalized fracture surface. It consists of a mixture of square and slant morphology. As can be seen, the fracture surface consists of a central flat area (square morphology) and inclined areas usually referred to as shear lips corresponding to the slant morphology. On a macroscopic scale, the morphology of fracture surface is described in terms of the percentages of square morphology and slant morphology. A further insight into the development of the above morphologies is described below.

To develop the means by which the propagation of an existing crack can be described on a macroscopic scale, it is essential to characterize the nature of the stress field at the edge of the crack, i.e., the crack tip, corresponding to the...
three basic modes described above. Basically, this mechanical state is found to be entirely dependent upon the crack surface. Among the three cracking modes, the direct opening or mode I is considered to be the most critical for crack propagation. Therefore, it is of particular importance to consider propagation of a crack subjected to a tensile stress normal to its surface. By analogy to fracture surfaces of a square morphology, such a type of crack is said to be square.

It is generally recognized that when a crack propagates through a material, the crack tip or leading edge of the crack is always preceded by a plastically deforming zone; however, this plastic zone is influenced by the volume of metal ahead of the crack tip, i.e., thickness of the part. Depending upon the part thickness, three distinct stress fields at the crack tip are identified: (i) a state of plane strain, (ii) a state of plane stress, and (iii) a mixed state of plane strain and plane stress as described below.

### 6.5 Crack Propagation Under a Plane Strain Condition

Consider a thick plate containing a square crack where a tensile stress is acting normal to the crack surface as shown in Fig. 6.6. At the crack tip, the generalized

![Figure 6.5](image-url)
state of stress developed by the applied stress $\sigma$ is defined in terms of the three normal stress components $\sigma_x$, $\sigma_y$, and $\sigma_z$, and the three shear stress components $\tau_{xy}$, $\tau_{xz}$, and $\tau_{yz}$, as described in Chap. 5. However, if the plate is sufficiently thick, the rather large depth of metal ahead of the crack imposes a restriction on plastic deformation parallel to the tip of the crack as it advances. As a result, the applied tensile stress $\sigma$ develops a triaxial state of stress at the crack tip. Corresponding to this state of stress, the strain component $\varepsilon_z$ parallel to the crack tip (Fig. 6.6) is relaxed, i.e., $\varepsilon_z = 0$, reducing the mechanical state into a plane strain. Since plastic deformation occurs on the planes of maximum shear stress, which are oriented at 45° from the tensile stress axis the state of plane strain limits the deformation to those planes oriented at 45° from the surface of the crack, as shown in Fig. 6.6. As a result, the crack propagates under the influence of a tensile stress resulting in a fracture surface of a square morphology, which corresponds to mode I loading (Fig. 6.3a). Since plastic deformation is restricted at the crack tip, it can be stated the crack propagates by a brittle mechanism.

To summarize the above discussion, if a part of a thick section contains a square crack, the applied tensile stress develops a stress field at the crack tip corresponding to a state of plane strain (triaxial state of stress). Under this condition, the crack propagates with restricted plastic deformation at the crack tip, producing a fracture surface of a square morphology (mode I loading). A plain strain condition can develop in large structures such as ships, storage tanks, and bridges.
which fracture without general yielding. As pointed out earlier, fracture resulting from propagation of mode I cracking frequently results from cleavage.

### 6.6 Crack Propagation Under a Plane Stress Condition

If the plate shown in Fig. 6.6 is made sufficiently thin, the relatively small volume of metal ahead of the crack tip does not impose severe restriction on plastic deformation at the crack tip. Since plastic deformation occurs by shear stresses reaching maximum values on planes oriented at 45° from the tensile stress axis, plastic deformation can be expected to occur on planes inclined at 45° to the plane of the plate, as shown in Fig. 6.7. Consequently, as the crack propagates, plastic deformation occurs at its tip, relaxing the normal stress component parallel to the crack tip, i.e., $\sigma_x = 0$. As a result, the state of stress at the crack tip is reduced to a biaxial or a plane state of stress defined by the stress components $\sigma_x$, $\sigma_y$, and $\tau_{xy}$.

To summarize the above discussion, for parts of sufficiently thin sections, a tensile stress applied normal to the surface of an existing square crack develops a stress field at the crack tip corresponding to a plane state of stress (biaxial state of stress). Under this condition, fracture tends to occur by a shear mode on planes inclined at 45° to the tensile stress axis. In terms of the macroscopic morphology of fracture surfaces, the above sequence of events illustrates the development of

**Figure 6.7** Plane state of stress ahead of a crack propagating in a thin plate.
slant morphology from an initially square morphology. Typical examples include the tearing of an aircraft skin and the shearing of a pipeline.

It can be concluded from the discussion above that a gradual transition from a square to slant morphology of fracture surface occurs as a result of the corresponding change in the stress field at the crack tip from plane strain (tensile separation) into plane stress (shear separation) with decreasing thickness of the part. Evidently, at some intermediate thickness, both types of stress field can develop at the crack tip corresponding to a mixed state of plane strain and plane stress, as described below.

6.7 Crack Propagation Under a Mixed State of Plane Strain and Stress

For parts of intermediate thickness, an existing square crack can propagate under a mixed condition of plane strain and plane stress. In this case, the macroscopic morphology of the fracture surface is of the generalized type, i.e., a mixture of square and slant morphologies, as shown in Fig. 6.4d. Describing the fracture surface in terms of percentages of square and slant morphologies reflects the contribution of each type of stress field to crack propagation. A greater central flat area (square morphology) indicates that crack propagation is predominantly governed by a plane strain condition. Conversely, a smaller central flat area reflects the greater contribution of plane stress condition to crack propagation.

6.8 Sequence of Events Leading to Fracture

Once the inevitable presence of cracks in structural materials is accepted, every effort is made to minimize their size by proper selection of material, manufacturing processes as well as nondestructive testing. As pointed out earlier, the emphasis in design is to control the crack propagation rate. However, this requires an understanding of how cracks are initially nucleated. In this chapter, crack nucleation and propagation are discussed from a mechanical point of view. Related microscopic aspects are described in Chap. 10. When considering crack nucleation and propagation, it is important to distinguish between brittle materials and ductile materials.

For a perfectly brittle material, cracks are nucleated and propagated in the absence of plastic deformation. Conversely, plastic deformation is involved in the case of ductile materials. Although it is well recognized that some plastic deformation on a microscopic scale is involved in crack propagation, even in materials classified as brittle, it is very useful to analyze the case of perfectly brittle materials and then modify the results to accommodate the effect of localized plastic deformation.

In the preceding discussions, the inevitable existence of cracks in a part resulting from fabrication processes was considered, and the possible mechanical
conditions under which these cracks propagate to produce fracture were described. It is equally important to understand how cracks are nucleated and define the condition corresponding to the onset of crack propagation and its rate. These aspects are treated in the following section.

6.8.1 Crack Nucleation in Perfectly Brittle Materials

On a microscopic scale, nucleation of any crack involves some type of plastic deformation, as described in Chap. 10. From a macroscopic point of view, however, it is possible to describe crack nucleation in terms of the strain energy stored in a material as a result of fabrication processes.

It is recalled that within the elastic range of a given material, the strain energy stored per unit volume $U$ is given by

$$U = \frac{\sigma^2}{2E}$$

where $\sigma$ is the stress, $e$ is the strain, and $E$ is the elastic modulus. Evidently, the material is shifted into a higher energy state by the stored strain energy. Localized stress concentration resulting from the stored strain energy can be relieved by (i) plastic deformation if the material is sufficiently ductile or (ii) forming a crack if the material is brittle. Even if the applied stress is well below the yield strength, the localized stress may exceed the yield strength.

First, the case of a perfectly brittle material is considered where no plastic deformation occurs, which is known as the Griffith theory of fracture in brittle materials. From an energetic point of view, a strained brittle material can release some or all of its strain energy and in turn lower its energy by forming a crack. To illustrate the energetic of this process, consider that some of the strain energy stored in a thin plate of unit thickness acted upon by a stress $\sigma$ is released by forming a crack. As shown in Fig. 6.8, it is assumed that an edge crack (Fig. 6.8a) or a central crack (Fig. 6.8b) has an elliptical shape such that the length of its long axis is $2c$. Once the crack is formed, the internal stresses are relieved near the crack faces and fall to zero; however, at locations very far from the crack, the internal stresses remain unchanged. It can then be assumed that a circular region of radius $c$ around the crack is relieved from its elastic energy. Taking into account the two surfaces of the crack, the total area relieved from internal stresses is $2\pi c^2$. Therefore, the fraction of strain energy of the material is reduced by an amount $U_e$ given by

$$U_e = \text{(total strain energy)}(\text{crack width})(\text{total area relieved from strain energy})$$

$$= (\sigma^2/2E)(0)(2\pi c^2)$$

or

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

(6.2)
However, formation of the crack exposes a new surface, increasing the surface energy by an amount $U_s$. Since the crack has two faces of unit width and length $2c$, its surface area is $2 \times 2c = 4c$, and if $\gamma$ is the surface energy per unit area, $U_s$ is given by

$$U_s = 4\gamma c$$

Since the material is perfectly brittle, no energy is consumed in plastic deformation, and therefore the crack is formed when the released strain energy is just sufficient to create the crack surface, i.e.,

$$dU_e/dc = dU_s/dc$$

$$d/dc \left( \pi c^2 \sigma^2 / E \right) = d/dc \left( 4\gamma c \right)$$

$$2\pi c \sigma^2 / E = 4\gamma$$

$$\sigma = \left( 2\gamma E / \pi c \right)^{1/2} \quad (6.3)$$

It is noted that Eq. (6.3) is derived under an equilibrium condition where the strain energy released by the material is just sufficient to create the surfaces of an elliptical crack whose length is $2c$. Provided this condition is maintained, the crack is arrested and does not propagate.
If the material is not perfectly brittle, some of the elastic strain energy is consumed in plastic deformation at the crack tip, and thus becomes unavailable to provide surface energy. To allow for this localized deformation, Eq. (6.3) is modified to

$$\sigma = \left[ \frac{2E(g + \gamma_p)}{\pi c} \right]^{1/2}$$

where $\gamma_p$ is the portion of stored strain energy consumed in plastic deformation.

It is evident from the simple argument above that a crack of a given length is nucleated in a perfectly brittle material if the strain energy released by forming the crack is just sufficient to balance the energy to create the surface of the crack (surface energy). Also, it is clear that comparatively higher strain energy is required to nucleate a crack in a ductile material.

### 6.8.2 Crack Propagation Rate

An initially present crack in a given material may propagate at either a slow rate or at a fast rate. It is evident that the latter case can lead to catastrophic failure. When the crack propagation rate can be controlled by adjusting the applied stress such that its growth ceases when the stress is removed, it is said to be of a slow growth rate. Fast crack propagation occurs if the part is stressed in a manner permitting the crack to release more strain energy than the surface energy created by crack growth; i.e., according to Eq. (6.3), the crack propagates in a perfectly brittle material if:

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

and if the material is not perfectly brittle, i.e., localized plastic deformation occurs at the tip of the crack, the crack propagates if

$$\sigma = \sigma_f > \left[ \frac{2E(g + \gamma_p)}{\pi c} \right]^{1/2}$$

where $\sigma_f$ is called the fracture stress.

It can be seen from the above equations that whether plastic deformation occurs at the crack tip or not, the stress required to propagate the crack is inversely proportional to the critical crack length. As the critical crack length increases, the stress decreases, accelerating the propagation rate. Therefore, the fracture stress is determined by the largest preexisting crack length.

If plastic deformation occurs at the crack tip, ductile propagation of a crack may either stop after relieving some of the high local stresses, or it may continue to propagate if its size is sufficient and the stress is high enough to satisfy Eq. (6.6). Although this equation predicts that crack propagation may stop as a result of large local deformation, increasing the stress required to spread the crack, intense localized deformation may also lead to crack nucleation as occurs in...
the necked section of a tensile specimen. From an energy balance consideration, it is possible to determine the speed by which a crack propagates, as described below.

A crack can rapidly propagate by a brittle mechanism if the elastic strain energy released by forming a crack is more than that consumed in creating the crack surface. Excess energy is converted into kinetic energy (KE) causing the crack to gain speed and increase in size. From an energy balance, KE is expressed as

\[ KE = \frac{\text{strain energy released}}{C_0} - \frac{\text{surface energy of the crack}}{E} \]

and for the case described in the previous section,

\[ KE = \frac{\pi c^2 \sigma^2}{E} - 4 \gamma_c \]

By solving the above equation, the speed of the crack \( v_c \) is given by

\[ v_c = k v_l (1 - c_0/c) \]

(6.7)

where \( k \) is a dimensionless constant, \( v_l \) is the longitudinal speed of sound in the material, \( c_0 \) is the initial half crack length, and \( c \) is the half crack length at any instant. It can be concluded from the above equation that as long as the crack maintains a critical length \( c_0 \), i.e., \( c_0 = c \), propagation is halted. However, once the crack length exceeds the critical value \( c_0 \), it gains speed and starts to propagate. When the crack length becomes much larger than the critical length, i.e., \( c \gg c_0 \), \( c_0/c \approx 0 \), the crack speed reaches a maximum value of \( k v_l \), which is about \( \frac{1}{4} v_l \). Physically this means that the crack speed is limited by the speed of sound in the material consistent with the fact that atomic displacements in the vicinity of the crack can only be transmitted at the speed of sound. If plastic deformation occurs at the tip of the crack, the kinetic energy available for crack growth is reduced by the amount of energy consumed in plastic deformation. As a result, if plastic deformation can readily occur at the crack tip, it can be expected to propagate at a much slower rate in comparison with a crack propagating by a brittle mechanism.

It is recalled from earlier discussions that in terms of loading conditions of an existing crack, three cracking modes can be identified. Similarly, it is possible to classify the mode of crack propagation according to the loading conditions of the part, as explained below.

### 6.9 Classification of Crack Propagation Modes According to Loading Conditions

Generally, the mechanical loads applied to a part during service are classified into two extreme types: (i) static loads and (ii) dynamic loads. By definition, a static
load resulting from a single application of a force such as a dead weight remains
countant with time; i.e., a static load induces a constant stress in the part. Converse-
ly, the magnitude of a dynamic load and in turn the stress induced in the part
very with time i.e., a dynamic load induces cyclic stress. In most applications,
however, a part is subjected to a combination of static and dynamic load. Since
any part has to support its own weight, which is by definition a static
load, it is evident that a pure dynamic loading condition does not exist. Also,
if a part is subjected to a static load, cyclic stresses corresponding to a dynamic
loading condition can be generated as a result of thermal expansion and con traction
resulting from variations in temperature.

Even though a part may be subjected to a combination of static and
dynamic loading conditions, the designer can determine the principal load and
its nature by some preliminary calculations. A material is then selected for the
application on the basis of the principal load while taking into consideration
the effects of secondary loads. Before describing the various modes of crack
propagation, it is instructive to consider some aspects of static and dynamic loading
conditions as they relate to design.

6.9.1 Static Loading

Most of the static loads encountered in practice are of the tensile type; i.e., they
 induce tensile stresses in the part. When the static load is applied at sufficiently
low temperature with respect to the melting point of the material ($<0.5T_m$, where
$T_m$ is the melting point in degree Kelvin), the designer is primarily concerned
with short-term tensile properties determined from standard tensile tests, i.e.,
yield strength, ultimate tensile strength, and ductility (elongation or reduction
in cross-sectional area).

If the part is designed for service at higher temperatures $>0.5T_m$, the
designer becomes more concerned with a phenomenon known as creep. Even
if the applied static load induces a stress well below the respective yield strength
of the material derived from a short-term tensile test at the same temperature,
plastic deformation can still occur as a function of time. By definition, creep is
the accumulation of plastic deformation in the part as a function of time at elev ated
 temperatures under the influence of a low stress level.

Creep properties for engineering applications are determined from standard
creep tests where a specimen of specified dimensions is acted upon by a constant
load (constant engineering stress) at a given temperature, and the plastic strain is
recorded as a function of time, resulting in a creep curve such as that shown in
Fig. 4.14. As can be seen, the creep curve can be divided into three distinct
regions defining various creep stages: (i) primary creep, (ii) secondary or
steady-state creep, and (iii) tertiary creep. Initially, the creep strain rate is charac teristically high during the primary creep stage, causing the material to strain
harden. As the material strain hardens, the creep rate is progressively reduced toward the onset of steady-state creep. During this stage, strain hardening, reducing the strain rate, is balanced by softening processes such as recovery and recrystallization resulting in a constant strain rate. Eventually, voids form within the material and coalesce to form cracks which grow with time reducing the cross section of the specimen. As a result, the stress induced by the applied constant load increases, which accelerates the strain rate during the tertiary creep stage until the specimen is fractured.

Creep curves and in turn creep properties are sensitive functions of two important parameters: (i) temperature and (ii) stress. Increasing the temperature at a given stress has the effect of shifting the creep curve into higher strain levels and reducing the duration of steady-state creep until eventually the strain rate is rapidly accelerated from the beginning of the test until fracture occurs, as shown in Fig. 4.14. A similar effect is produced by increasing the stress at a constant temperature. Design parameters related to creep include the amount of a specified plastic strain, e.g., 0.5 or 1%, accumulated in the part at a given temperature and stress, and the corresponding time to fracture.

6.9.2 Dynamic Loading

Dynamic loads are classified into (i) impact loads and (ii) fatigue loads. An impact load may resemble a static load in that it consists of a single application of a force. However, unlike the case of a static load where strain is induced in the material rather slowly, the strain rate associated with an impact load is very high. It is recalled that flaws or cracks are inevitably present in any engineering alloy. Such defects represent localized areas of stress concentration. As pointed out earlier, localized stress concentration can be relieved by either plastic deformation or forming a crack. If the applied load is static, the low strain rate permits localized plastic deformation to take place to relieve the stress concentration and in turn homogenize the stress distribution. Conversely, in the case of an impact load, the associated very high strain rate severely restricts the occurrence of plastic deformation. This causes a sudden change in the stress level induced in the part, which can lead to rapid propagation of cracks resulting in fracture.

By definition, a fatigue load consists of a repeated or cyclic application of a force in contrast with a static or a dynamic load. It is considered to be the most common type of dynamic load. Typically, a fatigue load induces a stress usually below yield strength of the material. Although the maximum stress developed in each fatigue cycle is evidently safe if applied only once, its repeated application can result in fracture. Depending upon the magnitude of the stress induced in the part as a result of fatigue loading, two types of fatigue are distinguished: (i) low-stress, high-cycle fatigue and (ii) high-stress, low-cycle fatigue. Low-cycle fatigue differs from high-cycle fatigue in that every cycle involves
macroscopic strain of the part. Each type of fatigue is distinguished by the num-
ber of cycles preceding fracture. In the case of high-cycle fatigue, the number of
cycles to fracture is $10^6$ or higher. Fracture by low-cycle fatigue, however, can
occur after $10^5$ cycles or less. High-cycle fatigue properties are determined
from tests devised to determine the functional dependence of maximum stress
developed during the stress cycle (stress amplitude) on the number of cycles to
fracture. For low-cycle fatigue, the strain amplitude is determined as a function
of number of cycles to fracture. Fatigue life is expressed as number of cycles of
stress amplitude or strain amplitude that the part can withstand. Fracture by a fati-
gue mechanism is very sensitive to stress concentration sites resulting from flaws
or discontinuities particularly at the surface of the part.

Having defined the possible loading conditions of a part, it is possible to
classify the modes of crack propagation into two main types: (i) time independent
and (ii) time dependent. An account of each type is given below.

6.9.3 Time-Independent Crack Propagation

Evidently, a time-independent crack propagation results in an instantaneous
fracture by rapid propagation of one or more cracks. It occurs on first application
of either (i) a static load inducing a stress exceeding the ultimate tensile strength
as occurs in a tensile test or (ii) a sudden overload condition resulting on first
application of a high impact load.

Frequently, the instantaneous fracture resulting from time-independent
crack propagation is referred to as overload fracture. However, it is important
to realize that any fracture, regardless of how cracks are propagated, must natu-
rally result from an overloading condition. An equivalent statement is that frac-
ture does not occur unless the part is overloaded. In practice, an overloading
condition leading to fracture can be reached by various mechanisms, and it is
essential to identify those mechanisms before future fractures can be prevented.
Various mechanisms which can lead to an overloading condition are described
below.

6.9.4 Time-Dependent Crack Propagation

Most fractures encountered in practice result when one or more cracks existing in
the part are propagated slowly as a function of time, resulting in a progressive
reduction in cross-sectional area of the part. Associated with the reduction in
cross-sectional area of the part is a continued increase in the applied stress. Eventu-
ally, the cross-sectional area of the part is reduced to an extent where the
applied stress exceeds the ultimate strength of the material. Naturally, this corre-
sponds to an overloading condition leading to rapid propagation of cracks
resulting in instantaneous fracture. To summarize, most fractures result from
(i) time-dependent crack propagation leading to an overloading condition followed by (ii) time-independent crack propagation resulting in an instantaneous fracture. Evidently, the use of the phrase *overload fracture* can be applied to any fracture and therefore its use must be avoided.

Cracks can propagate as a function of time by various mechanisms. Depending upon many variables such as the loading conditions, service temperature, geometry of the part, state of stress, nature of the environment, and the material used in the application, the mechanisms of time-dependent crack propagation can be classified into (i) creep, (ii) fatigue, (iii) stress-corrosion, and (iv) hydrogen-assisted crack propagation. In all these cases, one or more of the above variables can permit the slow propagation of one or more cracks as a function of time, i.e., time-dependent crack propagation, until an overloading condition is reached, leading to instantaneous fracture. Both creep and fatigue loading have been described earlier. Cracks propagating by the combined effect of loading conditions and the environment are discussed in Chap. 9.

It can be concluded from the above brief discussion that similar to the phrase *overload fracture*, the use of such terms as *creep fracture* and *fatigue fracture*, etc. is rather misleading and must be avoided. A better description is that fracture has resulted from propagation of cracks by a creep or a fatigue mechanism or shortly fracture occurs by a creep or a fatigue mechanism. In short, it is highly recommended to use the word *fracture* without any adjective.

### 6.10 Variables Affecting Fracture Behavior

As demonstrated earlier, the stress required to propagate a crack in a brittle manner $\sigma_f$ is given by:

$$\sigma_f = \left(\frac{2E\gamma}{\pi c}\right)^{1/2}$$  \hspace{1cm} (6.8)

and when plastic deformation occurs at the tip of the crack, $\sigma_f$ is modified to

$$\sigma_f = \left[\frac{2E(\gamma + \gamma_p)}{\pi c}\right]^{1/2}$$  \hspace{1cm} (6.9)

Since both the elastic modulus $E$ and surface energy $\gamma$ are not significantly influenced by size and geometry of the part, state of stress, temperature, strain rate, and microstructure, the brittle fracture stress $\sigma_f$ [Eq. (6.8)] is independent of those variables. In contrast, plastic deformation, and in turn the term $\gamma_p$ in Eq. (6.9), is a sensitive function of the above variables, and the stress $\sigma_f$ required to propagate a crack in a ductile manner is very sensitive to all changes affecting plastic deformation. An account for the role of each variable in affecting fracture behavior is given below.
6.10.1 Effect of Part Size

It is recalled from Sec. 6.5 that parts having thin sections promote a state of plane strain associated with plastic deformation by shear stresses at the tip of the crack. By definition, such a crack propagates by a ductile mechanism. Parts of thicker sections, however, develop a state of plane stress suppressing plastic deformation at the tip of the crack, leading to crack propagation by a brittle mechanism. It is also possible to demonstrate the effect of part size by examining its influence on the fracture stress $s_f$, as illustrated below.

Consider a crack present in a stressed part of length $L$ and thickness $t$ as shown in Fig. 6.9. If the crack advances a small distance $Dx$ under the influence of the applied stress $\sigma$, two new surfaces are created (an upper surface and a lower surface), each of which has an area of $Dxt$, thus the total new surface area created is given by

\[
\text{Surface area} = 2 \Delta xt
\]

and if $\gamma$ and $\gamma_p$ are the surface energy and energy consumed in plastic deformation at the tip of the crack per unit area, respectively, the total increase in surface energy is given by

\[
\text{Increase in surface energy} = 2 \Delta x(\gamma + \gamma_p)
\]

Since the elastic strain energy per unit volume within the elastic range is $\sigma^2/2E$, as explained earlier, the decrease in strain energy associated with advancing the crack a distance $\Delta x$ is given by

\[
\text{Decrease in strain energy} = \sigma^2/2E(Lt \Delta x)
\]

![Figure 6.9](image_url) Illustration of the effect of part size on fracture behavior.
where $Lt \Delta x$ is the volume of material relieved from internal stresses. Therefore, the change in energy $\Delta U$ is

$$\Delta U = \text{increase in surface energy} - \text{decrease in strain energy} = 2 \Delta x (\gamma + \gamma_p) - (\sigma^2/2E)(Lt \Delta x)$$

As pointed out earlier, the crack propagates if the strain energy released is either equal to or larger than the surface energy absorbed, i.e., when the condition

$$\Delta U/\Delta X \geq 0$$

is satisfied, or

$$2(\gamma + \gamma_p) = \sigma_f^2 L/2E$$

Therefore, the fracture stress $\sigma_f$ is given by

$$\sigma_f = 2[E(\gamma + \gamma_p)/L]^{1/2} \tag{6.10}$$

It is evident from Eq. (6.10) that the fracture stress $\sigma_f$ decreases as the size of the part $L$ increases. A pipe containing a longitudinal crack can be created from the part shown in Fig. 6.9 by rolling it up and joining together the two fixed edges. In this case, the length $L$ becomes the circumference of the pipe cross section, and the stress $\sigma$ becomes the circumferential or hoop stress resulting from the internal pressure (see Sec. 5.13.2). If the stress $\sigma > \sigma_f$, the longitudinal crack can propagate at a speed of about one-third the speed of sound in the pipe material. However, the internal pressure release associated with forming the crack propagates at a much lower speed corresponding to the sonic speed in the fluid present in the pipe. Consequently, if a pipe containing a longitudinal crack has a large diameter $L$ and corresponding small fracture stress $\sigma_f$, instantaneous splitting can be observed over thousands of meters in length during proof testing at stresses below the yield strength of the pipe material. However, for a pipe of the same material and initial crack but of a smaller diameter, such catastrophic crack propagation is not observed.

### 6.10.2 Effect of State of Stress

It is recalled from Sec. 6.4 that from a macroscopic point of view, a square or brittle fracture is favored by high tensile stresses (mode I loading). A slant or ductile fracture, however, requires shear stresses (mode III loading). When a material can readily deform plastically by motion of dislocations requiring shear stresses, the effect of an applied or a residual stress tends to be relieved. Under this condition, it is not possible to maintain the high level of tensile stresses required for brittle fracture. Instead, an existing crack propagates by a ductile mechanism. In contrast, if the material becomes incapable of plastic deformation to an extent sufficient to
accommodate the effect of stress, the buildup of tensile stresses promotes brittle fracture. Evidently, then, the state of stress, and in particular the ratio of maximum tensile stress to maximum shear stress ($\sigma_{\text{max}} / \tau_{\text{max}}$), has a significant effect on fracture behavior of a given material as demonstrated by the following examples.

Consider a state of uniaxial stress as exemplified by a standard tensile test. It is recalled that the applied tensile stress $\sigma_z$ develops shear stresses, achieving a maximum values on planes inclined at $45^\circ$ to the tensile axis as shown in Fig. 6.10a. If another tensile stress $\sigma_y$ perpendicular to $\sigma_z$ is applied to the part, it develops maximum shear stresses in opposite directions to those developed by $\sigma_z$, as shown in Fig. 6.10b. In the case of the uniaxial state of stress, the maximum shear stress $\tau_z$ is given by

$$\tau_z = \frac{\sigma_z}{2}$$

However, when the state of stress becomes biaxial as shown in Fig. 6.10b, and assuming that $\sigma_z > \sigma_y$, the maximum shear stress $\tau_{\text{max}}$ is reduced to

$$\tau_{\text{max}} = \tau_z - \tau_y = (\sigma_z - \sigma_y)/2$$

It is then clear that as the state of stress is changed from uniaxial to biaxial, the ratio of maximum tensile stress to maximum shear stress is decreased. This ratio can be further reduced if a third stress $\sigma_x$ is applied along the $x$ axis, establishing a state of triaxial stress. Corresponding to the case of hydrostatic stress where $\sigma_z = \sigma_y = \sigma_x$, no shear stresses are developed, i.e., $\tau = 0$.

![Figure 6.10](image)

**Figure 6.10** Effect of state of stress on the maximum shear stress. (a) Uniaxial state of stress. (b) Biaxial state of stress.
To summarize, the ratio of maximum tensile stress to maximum shear stress is progressively increased, and the tendency for plastic deformation is suppressed as the state of stress is changed from uniaxial to biaxial and then to triaxial favoring brittle fracture over ductile fracture.

It can be concluded from the above discussion that even an inherently ductile material can fail by a brittle mechanism if it is subjected to a state of stress, suppressing the development of shear stresses and in turn plastic deformation. Therefore, as pointed out earlier, a distinction between a brittle fracture and ductile fracture on the basis of total macroscopic strain acquired during the entire deformation process leading to fracture can be misleading.

6.10.3 Effect of Part Geometry

It is possible to understand the effect of part geometry on fracture behavior in terms of its influence on the state of stress. Changes in part geometry such as cross-sectional area can have a significant effect on the local state of stress, as illustrated by the following example.

Consider a part containing a V notch and subjected to a force \( P \) as shown in Fig. 6.11a. As discussed earlier, the tensile force develops a uniaxial state of stress as shown in Fig. 6.11b. As can be seen, the V notch represents a localized reduction in cross-sectional area. Due to the corresponding higher stress level, plastic deformation tends to be initiated at the notched section earlier than the remainder of the specimen. Although this section tends to elongate in the direction of the applied stress (\( z \) axis), and contract along the transverse directions (\( x \) and \( y \) axes), it is constrained by the volume of the material above and below the notch. As a result, a state of triaxial stress is established at the notched section (Fig. 6.11c) where the principal stresses correspond to the applied stress \( \sigma_z \), and the stresses \( \sigma_x \) and \( \sigma_y \) are induced by the above constraint. It follows from the discussion of the previous section that the development of shear stresses is suppressed in the notched section, leading to a higher level of tensile stresses in comparison with the rest of the specimen; i.e., \( \tau_{\text{max}}/\sigma_{\text{max}} \) at the notched section is smaller than that at the rest of the part. Eventually, the specimen is fractured at the notched section.

6.10.4 Effect of Temperature and Strain Rate

Since plastic deformation is a thermally activated process, it is significantly affected by both temperature and time. Dislocations, which are immobile at lower temperature by various types of obstacles, can be freed and become mobile at higher temperatures. Decreasing the temperature has the effect of shifting the tensile stress-strain diagram to higher strength and lower ductility levels, increasing the tendency for brittle fracture. Because plastic deformation is also a time-dependent process, rapid application of a stress resulting in rapid strain affects
fracture behavior in a manner closely related to temperature. Increasing the strain rate produces an effect similar to that caused by reducing the temperature, i.e., shifting the stress-strain diagram to higher strength and lower ductility levels, and in turn increasing the tendency toward brittle fracture. This is also related to the effect of strain rate on dislocation mobility. Freeing dislocations from obstacles is progressively delayed as the strain rate increases, and therefore the yield strength increases with strain rate.

Although brittle fracture tends to occur as the temperature is reduced or the strain rate is increased, the extent of this behavior can vary considerably from one material to another according to the functional dependence of yield strength on temperature. Some metallic materials, particularly those having body-centered cubic structures, e.g., low-carbon steels, exhibit a strong dependency of yield strength on temperature, and therefore the effects of temperature and strain rate on fracture behavior are particularly pronounced in these materials. This is related to the lower mobility of dislocations in body-centered cubic structures in comparison with other structures such as the face-centered cubic, and also the role of interstitial elements, particularly carbon in locking dislocations in...
position. It is possible to determine the effect of temperature alone on fracture behavior from low-strain-rate tensile tests carried out at various temperatures. Typically, the transition from ductile to brittle fracture occurs over a temperature range known as the transition temperature range. Since high strain rates are usually achieved in impact tests (see Chap. 4), these tests are customarily used to determine the combined effect of temperature and strain rate. For a given material exhibiting a tendency to change its fracture behavior from ductile to brittle, the transition temperature range is significantly affected by a number of variables, as discussed below.

All variables which influence dislocation mobility and propagation of brittle cracks, e.g., chemical composition, microstructure including grain size, and strain rate, have significant effects on the transition temperature. For example, increasing the carbon content of a low-carbon steel is known to raise the transition temperature range. This can be related to the tendency of carbon to segregate at dislocation lines, immobilizing them and in turn promoting the tendency for brittle fracture. Also, reducing the grain size can be expected to produce a similar effect. In a fine-grained material, dislocations can only move relatively shorter distances before they can be intercepted by grain boundaries acting as obstacles to their motion; i.e., a finer grain size promotes the tendency for brittle fracture. Changing the microstructure of a given steel by heat treatment can also have a significant effect. In general, spherodization of cementite by quenching and tempering lowers the transition temperature range. Since increasing the strain rate produces a similar effect to that caused by lowering the temperature, i.e., reducing the dislocation mobility and in turn promoting the tendency for brittle fracture, increasing the strain rate tends to raise the transition temperature range.

### 6.10.5 Effect of Internal Structure of the Material

One of the most important structural parameters which influence fracture behavior is grain size. It is well known that the tensile yield strength of a given polycrystalline material varies with grain size according to the relationship

$$\sigma_y = \sigma_0 + Kd^{-1/2}$$  \hspace{1cm} (6.11)

where $\sigma_y$ is the tensile yield strength, $\sigma_0$ is a material constant called the friction stress, $K$ is an empirical constant, and $d$ is the grain diameter; i.e., the yield strength increases as the grain size decreases. An understanding of how grain size influences yield strength and fracture behavior can be developed from the microscopic aspects of plastic deformation discussed in Chap. 8. It suffices to mention here that on a microscopic scale, plastic deformation occurs by motion of a crystal defect called dislocation. Barriers to dislocation motion increase the external force required to cause plastic deformation. Among those barriers are
grain boundaries. Increasing the grain boundary area per unit volume by reducing the grain size can thus be expected to increase the yield strength consistent with Eq. (6.11).

When dislocations are held up against a grain boundary, a state of localized stress concentration is developed. In comparison with a coarse-grained material (smaller density of grain boundaries), dislocations in a fine-grained material can move smaller distances before they are held up against grain boundaries, as further explained in Chap. 8. However, as shown later, a coarse-grained material develops a higher level of stress concentration at a grain boundary. In general, stress concentration is relieved by either plastic deformation in a neighboring grain or opening a crack at the grain boundary. Because plastic deformation can occur more readily in a coarse-grained material, cracks are more likely to propagate by a ductile mechanism. In contrast, for a fine-grained material where deformation is suppressed, cracks tend to propagate by a brittle mechanism.

As explained in Chap. 8, phase transformations occurring in a material can also have a significant effect on its fracture behavior. For example, when an alloy is disordered, fracture occurs by a ductile mechanism. However, if it becomes ordered, fracture may occur by a brittle mechanism. Precipitation of secondary phases can also change the fracture behavior from ductile to brittle, e.g., precipitation of $\sigma$ phase.

6.11 Basic Principles of Fracture Mechanics

It is recalled from Chap. 4 that toughness of a material, a property combining both strength and ductility, is a measure of its resistance to fracture. Since toughness corresponds to the area under the uniaxial stress-strain diagram, its unit is energy per unit volume equivalent to force per unit length.

Basically, fracture mechanics aims at developing means by which crack propagation can be described on a macroscopic scale in terms of some measurable quantities. From the definition of toughness given in Chap. 4, it becomes evident that the resistance to crack propagation must be expressed in terms of a parameter having the unit of force per unit length, which is called the crack extension force $G$. Since force per unit length is equivalent to energy per unit area, they both have the same physical meaning. According to the Griffith theory of crack nucleation, surface energy per unit area or surface tension equivalent to force per unit area is essentially the energy required to form a crack in a perfectly brittle material. In contrast, $G$ is considered to be the force per unit length required to propagate a macroscopic crack.

It follows from the above discussion that the theoretical basis for fracture mechanics is reduced to deriving an expression for $G$ in terms of some
measurable parameters. For engineering design applications, however, it is more convenient to use another parameter $K$ called the stress intensity factor, as shown later. Although $K$ is closely related to $G$, it has the advantage of expressing the stress concentration at the tip of a crack in a convenient form, as described below.

### 6.12 Linear Elastic Fracture Mechanics (LEFM)

Successful application of fracture mechanics in engineering design to predict and prevent fracture during service is based upon developing a model representative of the events taking place at the tip of an existing crack, i.e., the stress field at the crack tip. For practical purposes, such a model must be as simple as possible, but it must include such variables as (i) crack length, (ii) state of stress, and (iii) fracture toughness.

Among the various types of deformation and possible states of stress and strain, it is possible to visualize the simplest useful model for the stress at the tip of a crack in terms of (i) a material behaving in a linear elastic manner and (ii) two-dimensional analysis of stress and strain. An approach based upon these two assumptions is referred to as linear elastic fracture mechanics (LEFM).

As pointed out earlier, experiment shows that crack propagation in engineering alloys is almost always preceded by a localized plastic zone at the crack tip. Evidently, this seems to violate the first assumption of linear elastic fracture mechanics. However, in many cases, the size of the plastic zone is rather small in comparison with the size of the crack, a condition referred to as small scale yielding. Under this condition, the plastic zone is visualized to be confined within a linear elastic field surrounding the crack tip. From an engineering design point of view, such an approach has been proved to be quite useful in providing adequate results, particularly if a small correction parameter is used to account for plastic deformation at the tip of the crack.

It is important to realize that if the local conditions at the crack tip permit large-scale plastic deformation, the linear elastic fracture mechanics approach becomes invalid. In this case, it becomes necessary to use the elastic-plastic approach. Although such an approach has not yet been advanced to the stage permitting its wide scope usage in engineering design, some of its results are described.

An important result of the approach of linear elastic fracture mechanics is the identification of what is known as the stress intensity factor $K$. It expresses the intensity of the normal stress field at the tip of an existing crack under elastic loading condition and can generally be expressed as

$$K = C\sigma(\pi c)^{1/2}$$
where $C$ is a constant depending upon the type of loading and geometry of the part away from the crack, $\sigma$ is the applied stress, and $c$ is the length of an edge crack or half the length of an embedded crack. As can be seen from Eq. (6.12), the unit of $K$ is (stress) (length)\(^{1/2}\).

A crack of the opening modes (mode I) described in Sec. 6.4 is the most critical to crack propagation. It is recalled that there are two extreme cases of stress field at the crack tip depending upon the size of the part: (i) plane stress condition and (ii) plane strain condition. For each condition, the crack extension force can be derived as described below.

### 6.12.1 Plane Stress Condition

As explained in Sec. 6.6, a plane stress condition (biaxial state of stress) is developed ahead of the crack by the applied stress if the part is rather thin. In this case, localized plastic deformation at the crack tip relaxes the normal stress component parallel to the crack edge or front, i.e., $\sigma_z = 0$ (Fig. 6.7), reducing the state of stress into biaxial. As a result of localized shear, the crack propagates by shearing of its surfaces relative to each other. In practice, a plane stress condition ahead of a crack is developed in such thin parts as an aircraft skin and a pipeline. Generally, a plane stress condition is defined by the condition that the size of the plastic zone ahead of the propagating crack is much larger than the thickness of the part.

To illustrate the fundamental basis for developing an expression for the crack extension force $G$ under plane stress condition, the linear-elastic fracture mechanics approach is considered where the applied stress is assumed to be well below the yield strength of the material; i.e., it is assumed that no plastic deformation occurs.

Consider a plate of a unit thickness subjected to an applied tensile stress $\sigma$ containing an edge crack of length $c$ such that $2c < 0.1w$ and $2c < 0.5t$, where $w$ and $t$ are the width and thickness of the plate, respectively, as shown in Fig. 6.12a. In order to evaluate the resistance of the material to crack propagation, the stress required to close the crack, i.e., to move the crack backward rather than forward, is considered. As shown in Fig. 6.12a, a stress $\sigma_y$ applied in a direction opposite to that of the stress $\sigma$ is assumed to be required to close the crack. After application of the stress $\sigma_y$, the crack tip moves from position 1 to position 2 through a distance $\Delta x$ as shown in Fig. 6.12a. It is evident that a mechanical work must be done on the plate to close the crack, which can be evaluated as described below.

An elemental portion of the volume of crack closed by moving the tip from position 1 to position 2 is considered as shown in Fig. 6.12b. If the height of this element is $2L$, it is evident that both the upper and lower surfaces must move a distance $L$ to close the crack, and if the width of this element is $dx$, elemental
The volume of crack closed by the stress is given by

\[
\sigma_y = L \Delta x t = L dx
\]

since the plate is assumed to have a unit thickness.

Assuming that the stress \( \sigma_y \) varies linearly with the displacement of the crack surfaces, the work done per unit volume \( dW \) to close the elemental crack is given by

\[
dW = \sigma_y L dx
\]
and the total work done $\Delta W$ to move the crack tip from position 1 to position 2, i.e., to close the crack by a distance $\Delta x$, is then given by

$$\Delta W = \int_{0}^{\Delta x} \sigma_y L \, dx$$

Under plane stress condition, it can readily be shown from the theory of elasticity that the stress $\sigma_y$ required to close the crack is given by

$$\sigma_y = \left( \sigma \sqrt{\frac{c}{2}} \right) / \sqrt{2x} \tag{6.13}$$

and the distance $L$ in Fig. 6.12b is given by

$$L = \frac{2\sigma \sqrt{c}}{E \sqrt{2(\Delta x - x)}}$$

where $c$ is length of an edge crack or half the length of a central crack, $x$ is the distance measured from the elemental volume to position 2 as shown in Fig. 6.12b, and $E$ is the elastic modulus. Therefore, $\Delta W$ is given by

$$\Delta W = \left( \sigma^2 c \pi \Delta x \right) / E$$

By definition, the derivative of work with respect to distance is a force; therefore $\Delta W / \Delta X$ in the above equation defines a force $G$ given by

$$G = \left( \sigma^2 c \pi \right) / E \tag{6.14}$$

From the above analysis, the force $G$ per unit length can be defined as the force required to extend or move the edge of the crack through the material. When the applied stress $\sigma$ reaches the level of the fracture stress $\sigma_f$, a critical extension force $G_c$ is defined such that

$$G_c = \left( \sigma_f^2 c \pi \right) / E \tag{6.15}$$

As can be seen from Eqs. (6.14) and (6.15), the crack extension force is expressed in terms of stress, half the crack length, and the elastic modulus. By comparing the above equation with the Griffith theory expressed by Eq. (6.3), it can be concluded that

$$G_c = 2\gamma \tag{6.16}$$

Physically, Eq. (6.16) means that the force required to extend a crack in a perfectly brittle material is equivalent to twice the specific surface energy. It is recalled that the Griffith theory states that the energy consumed to advance the crack is converted into surface energy of the two crack surfaces.

In Eq. (6.14), the square root of the parameter $\sigma^2 c$, i.e., $\sigma (c)^{1/2}$ has an important physical meaning. It determines the level of normal stress at any
point in front of the crack. Evidently, the larger the crack length or the greater the applied stress, the intensity of stress at any point ahead of the crack becomes more severe. Therefore, \( \sigma^2 c \pi \) in Eq. (6.14) is replaced by \( K^2 \), and therefore

\[
G = K^2 / E
\]  
(6.17)

By comparing Eqs. (6.14) and (6.16), the parameter \( K \) known as the stress intensity factor is given by

\[
K = \sigma (c \pi)^{1/2}
\]  
(6.18)

Combining Eqs. (6.13) and (6.18) results in

\[
\sigma_y = K/(2\pi c)^{1/2}
\]  
(6.19)

where \( \sigma_y \) is the normal stress at a distance \( x \) from the crack tip on the crack plane, and the stress intensity factor \( K \) represents the normal stress in the immediate vicinity of the crack expressed as a function combining the relationship between the applied stress and crack length.

Corresponding to the critical extension force given by Eq. (6.16), when the applied stress reaches the fracture stress, a critical stress intensity factor \( K_c \) is defined such that

\[
K_c = \sigma_f (c \pi)^{1/2}
\]  
(6.20)

where the critical stress intensity factor \( K_c \) is called the plane stress fracture toughness. It can be concluded from Eq. (6.20) that \( K_c \) is related to the critical crack extension force \( G_c \) by

\[
G_c = K_c^2 / E
\]  
(6.21)

From Eqs. (6.20) and (6.21), the fracture stress \( \sigma_f \) can be expressed as

\[
\sigma_f = (EG_c / \pi c)^{1/2}
\]  
(6.22)

and by substituting Eq. (6.16) into Eq. (6.22), \( \sigma_f \) can be expressed as

\[
\sigma_f = (2\pi y / \pi c)^{1/2}
\]  
(6.23)

which is the same as the Griffith equation (6.3).

It is important to realize that the above equations are derived assuming that no plastic deformation occurs at the tip of the crack and therefore, they can be applied to brittle materials. However, from a consideration of the more complex elastic-plastic approach, it is shown that the stress intensity factor \( K \) still expresses the relationship between applied stress and crack length. Furthermore, it is shown that both the size of the plastic zone and the strain at the crack tip are functions of \( K \) leading to a limited constant value of \( K = K_c \) at fracture.
6.12.2 Plane Strain Condition

When an existing crack propagates under a plane strain condition (a part of a sufficiently thick section), the critical stress intensity factor is labeled $K_{ic}$ and can be shown to be related to the crack extension force $G_{ic}$ by

$$K_{ic}^2 = EG_{ic}/(1 - v^2) \quad (6.24)$$

where $K_{ic}$ is called the plane strain fracture toughness and $v$ is Poisson’s ratio of the material. Corresponding to plane strain condition, the fracture stress $\sigma_f$ is given by

$$\sigma_f = (EG_{ic}/(1 - v^2)c_0)^{1/2} \quad (6.25)$$

where in this case $c_0$ is the critical half length of the crack permitting continuous propagation under plane strain conditions. It is recalled that under plane strain conditions, plastic deformation at the crack tip is severely restricted. Therefore, the resulting fracture is frequently referred to as resulting propagation of cracks by a brittle mechanism. A more precise definition, however, is that the size of the plastic zone at the crack tip is considerably smaller than the thickness of the part.

Most applications of fracture mechanics in engineering design and failure analysis investigations are limited to crack propagation under a plane strain condition for two reasons. First, a plane strain condition is considered to be more critical for crack growth because thick structures are more susceptible to crack propagation by a brittle mechanism. Second, as pointed out earlier, parts which can yield under plane stress condition are considered to be relatively safe. However, it is recalled that if the material has a limited ductility in the transverse-direction, crack propagation by a shear mode can lead to fracture of thin sections.

6.12.3 Effect of Plastic Deformation at the Crack Tip

To summarize the results of the above analysis, it is shown that in the application of linear elasticity principles to an infinite body containing either an edge crack of length $c$ or a buried crack of length $2c$, the stress intensity factor $K$ can be expressed as

$$K = \sigma(\pi c)^{1/2} \quad (6.26)$$

where $\sigma$ is the applied stress. For any body of finite dimensions, the above equation is modified to

$$K = C\sigma(\pi c)^{1/2} \quad (6.27)$$

where $C$ is a constant determined by the loading conditions and the geometry of the part away from the crack region. To account for the fact that plastic
deformation occurs at the crack tip, a simple correction parameter is applied to Eq. (6.27). It is possible to estimate the extent of yielding at the crack tip by considering a state of plane stress and replacing the normal stress component \( \sigma_y \) in Eq. (6.19), by the yield strength of the material \( \sigma_{yp} \). If the plastic zone at the crack tip is assumed to be a circular region of a radius \( r_p \), \( x \) in Eq. (6.19) can be replaced with \( r \) to result in

\[
\sigma_{yp} = K / (2 \pi r_p)
\]  
(6.28)

and therefore the radius \( r_p \) of the plastic zone under plane stress condition is given by

\[
r_p = 1/2 \pi (K / \sigma_{yp})^2
\]  
(6.29)

Under a plane strain condition, it can readily be shown that the yield strength \( \sigma_{yp} \) is increased by a factor of \( (3)^{1/2} \). Replacing \( \sigma_{yp} \) in Eq. (6.28) with \( (3)^{1/2} \sigma_{yp} \), the radius of the plastic zone under plane strain condition can be expressed as

\[
r_p = 1/6 \pi (K / \sigma_{yp})^2
\]  
(6.30)

Since fracture occurs when \( K = K_c \) for a plane stress condition, and \( K = K_{lc} \) for a plane strain condition, the above equations at fracture become

<table>
<thead>
<tr>
<th>Condition</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane stress</td>
<td>( r_p = 1/2 \pi (K_c / \sigma_{yp})^2 )</td>
</tr>
<tr>
<td>Plane strain</td>
<td>( r_p = 1/6 \pi (K_{lc} / \sigma_{yp})^2 )</td>
</tr>
</tbody>
</table>

When plastic deformation occurs at the crack tip, its geometry changes from a sharp edge into a blunted edge, increasing the effective crack length from \( c \) to \( c + r_p \), and therefore Eq. (6.27) becomes

\[
K = C \sigma \left( \pi (c + r_p) \right)^{1/2}
\]  
(6.33)

An important result of the above analysis is the identification of the parameters \( (K_c / \sigma_{yp}) \) and \( (K_{lc} / \sigma_{yp}) \), which define the relative toughness under plane stress and plane strain conditions, respectively. Also as shown by the above equations, these parameters determine the size of the plastic zone \( r_p \) at the crack tip measured in units of length on the crack propagation surface.

### 6.12.4 Criteria for Plane Stress and Plane Strain

It follows from the above discussion that the size of the plastic zone \( x_p \) relative to the thickness of the part determines whether a crack propagates under plane stress or plane strain condition. In practice, it is considered that plain strain conditions are valid when \( r_p < t/4 \), where \( t \) is the thickness of the part. When \( r_p = t/4 \), the stress field at the crack tip is significantly shifted from a plane strain condition.
into a plane stress condition corresponding to
\[
(K_{ic}/\sigma_{yp})^2(1/t) = 1.3
\] (6.34)

Standard tests are developed for measurements of plane strain and plane stress fracture toughness. For test purposes, a plane strain condition is considered valid when the radius of the plastic zone \( r_p < t/10 \), corresponding to \((K_{ic}/\sigma_{yp})^2(1/t) < 0.4 \), or
\[
t \geq 2.5(K_{ic}/\sigma_{yp})^2
\] (6.35)

Plane stress conditions are well developed when \( r_p = 2t \), corresponding to
\[
(K_{c}/\sigma_y)^2(1/t) = 5.2
\] (6.36)

A mixture of plane strain and plane stress conditions is developed at the crack tip when the size of the plastic zone \( x_p \) is smaller than \( 2t \) but larger than \( t/4 \), i.e., \( t/4 < r_p < 2t \). Within this range, a smooth transition occurs from crack propagation under plane strain \( (r_p < t/4) \) into plane stress \( (r_p = 2t) \).

As further explained in Chap. 7, some ferrous alloys exhibit a phenomenon called the ductile-to-brittle transition as the temperature is lowered. In terms of fracture mechanics, this means that above a critical temperature, cracks propagate by a ductile mechanism under a plane stress condition. Below the critical temperature, the cracks propagate by a brittle mechanism under a plane strain condition.

### 6.13 Use of Fracture Mechanics in Design

Fracture mechanics lead to the recognition that the plane stress fracture toughness \( K_c \) and plane strain fracture toughness \( K_{ic} \) are characteristic materials properties. Just as the yield strength of a given material is a measure of its load carrying ability, \( K_c \) and \( K_{ic} \) of a given material are considered measures of its resistance to fracture by cracks propagating under plane stress and plane strain conditions, respectively.

Similar to yield strength, both \( K_c \) and \( K_{ic} \) are independent of thickness of a given part. However, since toughness is a property combining both strength and ductility, both parameters can be expected to be dependent upon temperature and interdependent upon elastic modulus, yield strength, and strain hardening coefficient.

As pointed out earlier, fracture occurs when the stress intensity factor \( K \) reaches the critical value \( K_c \) or \( K_{ic} \); i.e., fracture occurs when
\[
\begin{align*}
\text{Plane stress condition:} & \quad K = K_c \\
\text{Plane strain condition:} & \quad K = K_{ic}
\end{align*}
\] (6.37)
Depending upon the particular stress field at the crack tip, prediction of fracture requires knowledge of $K$, and $K_{Ic}$ or $K_c$. Since designers are particularly concerned about fractures resulting from brittle propagation of cracks, more attention is given to fracture under plane strain condition.

It is important to realize that the stress intensity factor $K$ differs from the critical values $K_{Ic}$ and $K_c$ in that it is a variable dependent upon the geometry of the part as well as the crack. As pointed out earlier, the general expression for $K$ is

$$K = C\sigma (\pi c)^{1/2}$$

where $C$ is a constant, $\sigma$ is the applied stress, and $c$ is the crack length. Various expressions based upon the above general equation are developed for specific geometries, as demonstrated by the following examples.

For a single edge crack of length $c$ in a plate of width $W$ (Fig. 6.13a), $K$ is given by

$$C = 1.12 - 0.231(c/W) + 10.55(c/W)^2 - 21.72(c/W)^3 + 30.95(c/W)^4 \quad (6.38)$$

If the crack is rather small with respect to the width of the plate, it is possible to neglect the terms containing $c/W$, and in this case $K$ is given by

$$K = 1.12\sigma (\pi c)^{1/2} \quad (6.39)$$

Equation (6.39) is also valid if the plate contains two small edge cracks (Fig. 6.13b). If the length $c$ of each crack, however, is relatively large with respect to the width of the plate $w$, $C$ is given by

$$C = [1.122 - 0.561(c/w) - 0.205(c/w)^2 + 0.471(c/w)^3 - 0.190(c/w)^4]/(1 - c/w)^{1/2} \quad (6.40)$$

For a plate containing a central crack of length $2c$ (Fig. 6.13c), $C$ is given by

$$C = 1 + 0.256(c/w) - 1.152(c/w)^2 + 12.200(c/w)^3 \quad (6.41)$$

If the part contains a circular buried defect of radius $c$, as shown in Fig. 6.13d, the value of $K$ is given by

$$K = 0.64\sigma (\pi c)^{1/2} \quad (6.42)$$

For a semicircular surface-breaking defect of such as that shown in Fig. 6.13e, $K$ is given by

$$K = 0.71\sigma (\pi c)^{1/2} \quad (6.43)$$
where \( c \) is the depth of the defect. In the case of a semielliptical surface-breaking defect (Fig. 6.13f), the above equation becomes

\[
K = 1.12/(Q)^{1/2} \sigma (\pi c)
\]

(6.44)

where \( Q \) is a surface flaw shape parameter dependent upon (i) the ratio of the crack depth to length and (ii) the ratio of the applied stress to yield strength. Values of \( Q \) are available in tabular or graphical forms in the literature.

### 6.14 Concept of Allowable Crack Size

Since engineering materials must inevitably contain some type of cracks introduced during fabrication processes, it is important to determine the maximum or allowable crack size which can safely be tolerated during service. Fracture mechanics define the conditions under which an existing crack of a given size can propagate to result in fracture as follows.
It is recalled that if a part contains a crack of size \( c \) and is acted upon by a tensile stress \( \sigma \), the stress intensity factor \( K \) is given by

\[
K = C \sigma (\pi c)^{1/2}
\]

Under plane strain condition, fracture occurs when \( K = K_{lc} \), corresponding to \( \sigma = \sigma_f \) where \( \sigma_f \) is the fracture stress. Therefore, at fracture, the above equation can be rewritten as

\[
K_{lc} = \sigma_f (\pi c)^{1/2}
\]

or

\[
\sigma_f^2 c = K_{lc}^2 / \pi = \text{constant}
\]

Experiment shows that the above relationship holds for many materials provided a proper correction is made to account for plastic deformation at the crack tip. Provided fracture is expected to occur under plane strain condition, the above equation can be used to determine the maximum allowable crack size which can safely be tolerated from knowledge of \( K_{lc} \) and the applied stress.

Since in practice fracture usually occurs under a mixed condition of plane strain and plane stress, the constant of Eq. (6.45) must be determined experimentally for the particular thickness of the part and service temperature rather than from the values of \( K_{lc} \).

For welded components, the maximum allowable crack size \( c_{\text{max}} \) can be determined from

\[
c_{\text{max}} = A(K_{lc}/\sigma_{yp})^2
\]

where \( A \) is a constant ranging from 0.1 to 0.5, and \( \sigma_{yp} \) is the yield strength of the material. If fracture is expected to occur under plane strain condition, \( K_{lc} \) in the above equation is replaced by \( K_c \). Evidently, application of the above approach requires the availability of considerable data for many varieties of engineering materials.

As discussed in Chap. 4, various nondestructive tests can be used to measure the size of a flaw, discontinuity, or a crack in a part. By introducing the concept of allowable crack size, it is possible to design safe products with existing cracks. To illustrate the principle involved, consider a part of a rectangular cross section (width = 0.5 m and thickness = 0.1 m), as shown in Fig. 6.14. If the part is made of a steel grade whose yield strength is 1700 MPa and plain strain fracture toughness \( K_{lc} = 110 \text{MPa(m)}^{1/2} \), and nondestructive testing shows that the part contains an edge crack of 0.004 m through the thickness, it is required to determine the maximum tensile load \( P \) which can be supported by the part using 2 as a factor of safety. In order to calculate the maximum safe load, it is necessary to consider two possible causes of failure: (i) yielding
and (ii) brittle propagation of the existing crack, as explained below. By definition, the load $P$ is given by

$$P = \text{design stress } \sigma_d \times \text{cross-sectional area } A$$

and $\sigma_d$ is given by

$$\sigma_d = \sigma_y / N$$

where $\sigma_y$ is the yield strength and $N$ is the factor of safety. Substituting for $\sigma_y$ and $N$ into the above equation results in

$$\sigma_d = 1700 / 2 = 850 \text{ MPa}$$

Since the existing crack reduces the cross-sectional area of the part, it must be taken into account to calculate the actual cross-sectional area supporting the load. From the geometry of figure, the actual cross-sectional area $A$ is given by

$$A = (0.1 - 0.004)(0.5) = 0.048 \text{ m}^2$$

and therefore

$$P = (850 \times 106 \times 1.01972 \times 10^{-1} \text{ kg/m}^2)(0.048 \text{ m}^2) = 1.8 \times 10^8 \text{ kg}$$

To guard against the possibility of brittle propagation of the existing crack, the maximum load can be calculated using the principles of linear elastic fracture
mechanics described earlier. First, it is important to determine whether the existing crack propagates under plane strain or plane stress condition. Using the criterion defined by Eq. (6.35), plane strain condition is satisfied if

\[ t > 2.5\left(\frac{K_{lc}}{\sigma_y}\right)^2 = 2.5\left(\frac{110}{1700}\right)^2 = 0.01 \text{ m} \]

Since the thickness of the part is greater than 0.01 m, plane strain conditions are satisfied at the crack tip. It is then possible to use Eq. (6.27) to calculate the applied stress \( \sigma \), i.e.,

\[ K_{lc} = C\sigma\left(\pi c\right)^{1/2} \quad (6.47) \]

where \( C \) is a constant and \( c \) is the size of existing crack. With the help of available data in the literature, it is possible to estimate the value of the constant \( C \) for a plate with an edge crack of length \( c \). In this case the ratio of the crack length \( c \) to width of the cross section \( w \) is taken as \( c/w = 0.004/0.5 = 0.008 \), corresponding to

\[ C[1 - (c/w)^{3/2}] = 1.12 \]

or

\[ C(1 - 0.0007) = 1.12 \]

and therefore

\[ C = 1.12 \]

Substituting the values of \( K_{lc} \), \( C \), and \( c \) into Eq. (6.47) results in

\[ 110 = 1.12\sigma(3.14 \times 0.004)^{1/2} \]

giving

\[ \sigma = 877 \text{ MPa} \]

By applying 2 as a factor of safety, the design stress \( \sigma_d \) is given by

\[ \sigma_d = \frac{877}{2} = 439 \text{ MPa} \]

It is to be noted that the effect of the existing crack is already taken into account in calculating the stress \( \sigma_d \), and therefore the nominal area \( A = 0.5 \times 0.1 = 0.05 \text{ m}^2 \) is used to calculate the load \( P \), i.e.,

\[ P = \sigma_d \times A = 439 \times 106 \times 0.05 = 22 \times 10^6 \text{ kg} \]

By comparing the above values of \( P \), it can be seen that in order to guard against brittle crack propagation the maximum safe load is reduced by about \( \frac{1}{2} \). To demonstrate the effect of crack tip plasticity on the above calculation, it is recalled
from Eq. (6.32) that under plane strain condition, the radius of the plastic zone \( r_p \) is given by

\[ r_p = \frac{1}{6} \pi \left( \frac{K_{Ic}}{\sigma_y} \right)^2 \]

Substituting the given values of \( K_{Ic} \) and \( \sigma_y \) into the above equation results in

\[ r_p = 0.0002 \text{ m} \]

and therefore the effective crack size \( c_{\text{eff}} \) becomes

\[ c_{\text{eff}} = c + r_p = 0.004 + 0.0002 = 0.0042 \text{ m} \]

By substituting the above value into Eq. (6.33), the applied stress \( \sigma \) becomes:

\[ \sigma = 855 \text{ MPa} \]

and the design stress \( \sigma_d \) becomes

\[ \sigma_d = \frac{855}{2} = 428 \text{ MPa} \]

giving a load \( P \) of

\[ P = 428 \times 106(0.05) = 21 \times 10^6 \text{ kg} \]  \( (6.48) \)

It is evident from the above results that the correction due to crack tip plasticity is rather small. To further illustrate the use of the concept of allowable crack size, consider that during regular maintenance a surface-breaking crack is detected in the above part, as shown in Fig. 6.14. Nondestructive evaluation shows that the length of the crack is 0.005 m and its depth is 0.002 m, and it is required to determine the course of action. In this case, the maximum safe load must be evaluated on the basis of the crack detected using the proper form of Eq. (6.27). It is recalled that for a semielliptical surface-breaking crack such as that shown in Fig. 6.13f, \( K \) is expressed as

\[ K = 1.12/(Q)^{1/2} \sigma (\pi c)^{1/2} \]  \( (6.49) \)

To estimate the value of the shape parameter \( Q \), it is first required to determine (i) the ratio of the crack depth \( a \) to length \( c \) and (ii) the ratio of the applied stress \( \sigma \) to the yield strength \( \sigma_{\text{yp}} \). From available data it can be shown that

\[ \frac{a}{2c} = 0.002/0.005 = 0.4 \]
\[ \frac{\sigma}{\sigma_y} = (22/0.05)/1700 = 0.26 \]

and also \( Q = 1.9 \). Noting that \( c \) in Eq. (6.49) refers to the depth of the crack, substituting the values of \( K_{Ic} \) and \( Q \), results in

\[ \sigma = 1719 \text{ MPa} \quad \text{and} \quad \sigma_d = 1719/2 = 860 \text{ MPa} \]
giving a maximum safe load \( P \) of
\[
P = 860 \times 10^6 \times 0.05 = 43 \times 10^6 \text{ kg} \quad (6.50)
\]

By comparing Eqs. (6.48) and (6.50), it becomes evident that the part can still support a larger load than that calculated on the basis of the originally existing crack. Therefore, the part can safely be returned back to service. However, if the load is less than that calculated on the basis of the originally existing crack, either the operating load must be reduced or the part is taken out of service.

### 6.15 Use of Fracture Mechanics in Failure Analysis

Most fractures encountered in practice result from propagation of initially existing cracks or discontinuities by either (i) a brittle mechanism or (ii) a fatigue mechanism. Fractures resulting from brittle propagation of cracks can be investigated using the principles of linear elastic fracture mechanics described earlier. Likewise, fracture mechanics is a particularly useful investigative tool in failure analysis. It is the objective of this section to present the procedures of applying fracture mechanics in failure analysis investigations with particular emphasis on the above two cases of crack propagation.

As demonstrated in this section, regardless of the mechanism responsible for crack propagation leading to fracture, the use of fracture mechanics aids the analyst in distinguishing between two cases: (i) The operating stress at the time of fracture is greater than the design stress, and (ii) the operating stress at the time of fracture is smaller or closely approaches the design stress. If the operating stress is found to be greater than the design stress, it becomes evident that the initially existing crack is not responsible for the fracture, e.g., overheating. Subsequently, the analyst must investigate the possible causes of exceeding the design stress. Conversely, if the operating stress is found to be smaller or closely approaching the design stress, it becomes evident that the fracture has resulted from an initially existing defect in the product. In this case, the analyst focuses on the possible sources of this defect, e.g., design practice, fabrication, and service conditions.

In some cases, it is possible to identify the mechanism responsible for the fracture from a visual or a macroscopic inspection of the fracture surface. Such inspection can be carried out with the naked eye, a hand lens, or a low-power microscope at magnifications ranging from 1 to about 50. From the orientation of the fracture surface, it is possible to identify the type of stress as well as its path and direction.

It is recalled from earlier discussions that a flat or square fracture surface is produced by a tensile stress. In this case, the fracture path is normal to the
direction of maximum principal tensile stress. Characteristically, this corresponds to a state of plane strain as explained earlier. For parts of relatively thin cross sections, the entire fracture surface can be of the slant type since localized stress concentration through the thickness can be relaxed by plastic deformation (state of plane stress). Since the direction of maximum shear stress is at 45° from the direction of maximum principal tensile stress, a slant fracture surface lies at 45° from the direction of maximum principal tensile stress.

A mixed fracture surface corresponding to a mixed state of plane stress and plain strain contains a central flat area enclosed by a shear lip. Usually, formation of the flat fracture surface precedes formation of the shear lip. This is because as a crack propagates by a brittle mechanism under the influence of a tensile stress producing a flat surface, the localized stress concentration is relaxed and, as a result, continued propagation resulting in complete separation of the part occurs by shear producing a slant fracture surface.

Frequently, fracture surfaces contain characteristic markings, which are particularly useful in identifying the site from which the crack has propagated. Commonly, fractures resulting from brittle propagation of cracks expose surfaces containing “chevron” or “radial” marks pointing back at the fracture origin; i.e., the markings are emanated from the fracture origin. Both features run parallel to the crack growth direction. Usually, the fracture origin has a flat morphology, i.e., a square crack, but it may also contain a shear lip. Identifying the location and morphology of the fracture origin aid the analyst in modeling the initial crack, which is required to apply fracture mechanics to the problem.

Fracture surfaces resulting from crack propagation by fatigue usually contain “clam shell” or “beach” marks, which run normal to the crack growth. Typically, fractures produced by a fatigue mechanism originate from surface flaws. In this case, the fracture origin usually lacks the square or slant morphology. Having identified the origin of fracture, the principles of fracture mechanics can be applied as described below.

6.15.1 Crack Propagation by a Brittle Mechanism

If it is determined from macroscopic examination that fracture has resulted from crack propagation by a brittle mechanism, it is plausible to assume that a state of plane strain is developed at the crack tip. In this case, the relevant parameter is the plane strain fracture toughness $K_{IC}$. To ensure that a plane strain condition is applicable to the problem, the thickness of the part $t$ must satisfy the criterion defined by Eq. (6.35), i.e., $t \geq 2.5(K_{IC}/\sigma_{yp})^2$, where $\sigma_{yp}$ is the tensile yield strength of the material. The values of plane strain fracture toughness $K_{IC}$ and yield strength $\sigma_{yp}$ for a number of engineering alloys is available in the literature. If $K_{IC}$ is not known for the material, it can either be measured using the test
\[ K_{Ic} = \left( \frac{(0.05E\sigma_y n e_f)}{3} \right)^{1/2} \]  

(6.51)

where \( E \) is the elastic modulus, \( \sigma_y \) is the yield strength, \( n \) is the strain hardening exponent, and \( e_f \) is the true strain at fracture as determined from a standard tensile test. Both the elastic modulus \( E \) and tensile properties are tabulated for most of the engineering alloys. Typically, the tensile strain at fracture is tabulated as an engineering strain \( e \). As explained in Chap. 4, the true strain \( e \) is given by \( e = \ln(1 + \varepsilon) \), where \( \varepsilon \) is the engineering or nominal strain. Values of \( n \) vary from about 0.2 for carbon and low-alloy steels and 0.5 for austenitic, Ni-, and Cu-base alloys. Typically, the estimated values of \( K_{Ic} \) fit the experimental data reasonably well within \( \pm 30\% \) and can be used with some confidence when measured values of \( K_{Ic} \) are not available.

By means of macroscopic examination of the fracture surface, the fracture origin is determined and can be modeled, i.e., as an elliptical crack, semielliptical crack, etc. After modeling the initial crack, the proper expression for the stress intensity factor \( K \) is selected as described earlier. Since fracture occurs when \( K = K_{Ic} \), it is possible to estimate the fracture stress \( \sigma_f \), which determines the next stage of the investigation. As explained earlier, the results indicate whether \( \sigma_f > \sigma_d \) or \( \sigma_f \leq \sigma_d \), where \( \sigma_d \) is the design stress. If \( \sigma_f > \sigma_d \), it can be concluded that the initial crack is not responsible for the fracture. However, if \( \sigma_f \leq \sigma_d \), it becomes evident that the fracture is promoted by the initial crack.

### 6.15.2 Crack Propagation by a Fatigue Mechanism

Conventionally, the approach used to design parts against low stress high-cycle fatigue is based upon data obtained from \( S-N \) curves derived from fatigue tests. Usually, the data between \( 10^3 \) and \( 10^6 \) cycles is scattered on a straight line corresponding to a \( \log S - \log N \) plot. For many steels, the endurance limit of a standard test specimen \( (S_e) \) coinciding with about \( 10^6 \) cycles is about \( 0.5S_u \) where \( S_u \) is the ultimate tensile strength, and the \( 10^3 \) cycle strength is about \( 0.9S_u \). By plotting the above two points on a \( \log S \) vs. \( \log N \) diagram, a straight line is obtained whose equation is solved for \( S_N \) (strength corresponding to \( N \) cycles) or \( N \):

\[
S_N = S_e \left( \frac{10^6}{N} \right)^{1/3\log(0.9S_u/S_e)} \quad \text{for} \quad 10^3 < N < 10^6
\]

(6.52)

and

\[
N = 106(S_e/S_N)^{3/\log(0.9S_u/S_e)}
\]

(6.53)

Many other materials, however, do not exhibit a definite endurance strength, and in this case, each material is rated at a certain life time. For example, \( S_L \) is taken as \( 0.25-0.50S_u \) for Ni-base alloys at \( 10^4 \) cycles. Actual parts, however, can have a much smaller endurance limit \( S_L \) in comparison with that of a standard specimen.
$S'_e$, therefore the endurance limit is modified by a factor $K < 1$, i.e.,

$$S_e = K S'_e$$

(6.54)

where the factor $K$ is the product of a number of parameters including: a surface factor, size factor, reliability factor, temperature factor, shape factor, and a miscellaneous factor taking into account the effects of corrosive environments, brittle surface layer, and residual stresses. All these parameters can either be obtained from available charts or calculated. For example, the shape factor is a function of geometry, which is accounted for by the theoretical stress concentration factor $K_t$, defined as

$$K_t = \frac{\text{maximum stress at section}}{\text{nominal stress at section}}$$

(6.55)

Stress concentration factors can be calculated analytically for simple shapes and by experimental stress analysis methods for complex parts. Tables and charts are available in the literature giving $K_t$ for many types of geometry. Usually, in practice, the part is subjected to a fluctuating stress, as shown in Fig. 6.15. In this case, five stresses are defined as shown: the maximum stress $\sigma_{\text{max}}$, the minimum stress $\sigma_{\text{min}}$, peak-to-peak range of stress $\sigma_r$, the mean stress $\sigma_m$, and the alternating stress or stress amplitude $\sigma_a$. As can be seen from Fig. 6.15

$$\sigma_m = (\sigma_{\text{max}} + \sigma_{\text{min}})/2$$

$$\sigma_a = (\sigma_{\text{max}} - \sigma_{\text{min}})/2$$

(6.56)

With the aid of the mean stress $\sigma_m$, a fatigue failure diagram can be constructed providing a basis for safe fatigue design. A fatigue failure diagram is

**Figure 6.15** Fluctuating stresses. (a), (b), and (c) represent the functional dependence of the stress on time.
shown in Fig. 6.16a. Along the mean stress axis, the yield strength $\sigma_y$ and ultimate tensile strength $\sigma_u$ are plotted. Endurance strength $\sigma_e$ or $\sigma_N$ (strength corresponding to $N$ cycles), and also $\sigma_y$ are plotted on the stress amplitude axis. By connecting those points with straight lines as shown, the failure diagram is obtained indicating the safe region. Including a factor of safety $n$, i.e., dividing the respective strength by $n$, results in the design diagram shown in Fig. 6.16b. It is to be noted that the validity of the failure diagram must be verified by experiment.

It is to be noted that fatigue data obtained from standard tests are valid for a bending load. For the case of combined loading, however, the failure criterion becomes more complicated. In this case, a widely accepted method is the use

---

**Figure 6.16** Prediction of fatigue failure. (a) Fatigue failure diagram. (b) Design diagram.
of effective stress criterion $\sigma_{\text{eff}}$, defined as

$$
\sigma_{\text{eff}} = (\sigma_1^2 + \sigma_2^2 - \sigma_1 \sigma_2)^{1/2}
$$

(6.57)

where $\sigma_1$ and $\sigma_2$ are the principal stresses.

Fracture mechanics can be applied to determine the number of stress cycles preceding fracture resulting from crack propagation by a fatigue mechanism. When a part is subjected to a cyclic loading during service, the stress fluctuates between a maximum value $\sigma_{\text{max}}$ and a minimum value $\sigma_{\text{min}}$ during each cycle. An initially existing crack grows in size during each cycle (time-dependent crack propagation) until eventually reaching a critical size that causes time-independent crack propagation, which results in fracture. From this sequence of events, it becomes evident that three parameters must be known before the number of stress cycles preceding fracture can be determined: (i) initial crack size $c_0$, (ii) the crack growth rate per stress cycle $d\sigma/dN$, and (iii) the critical crack size $c_c$ required for time-independent crack propagation resulting in fracture.

According to the principles of linear elastic fracture mechanics, if the part contains an initial edge crack of length $c_0$ or an embedded crack of length $2c_0$ such that $c_0$ is sufficiently small with respect to the thickness of the part ($2c_0 < 0.5t$), where $t$ is the thickness of the part, the maximum stress intensity factor $K_{\text{max}}$ encountered during the cycle is given by

$$
K_{\text{max}} = C \sigma_{\text{max}} (\pi c_0)^{1/2}
$$

(6.58)

where $\sigma_{\text{max}}$ is the maximum applied stress during the cycle. Corresponding to the minimum stress $\sigma_{\text{min}}$, a minimum stress intensity factor $K_{\text{min}}$ is defined such that

$$
K_{\text{min}} = C \sigma_{\text{min}} (\pi c_0)^{1/2}
$$

(6.59)

Evidently, during each stress cycle, the part becomes subject to a range of stress intensity $\Delta K$ given by

$$
\Delta K = K_{\text{max}} - K_{\text{min}} = C \Delta \sigma (\pi c)^{1/2}
$$

(6.60)

where $\Delta \sigma = \sigma_{\text{max}} - \sigma_{\text{min}}$ and $c$ is the crack size after certain number of stress cycles. Experiments show that for many engineering alloys, the rate of crack growth per stress cycle $(d\sigma/dN)$ can be expressed by a relationship of the type

$$
d\sigma/dN = C_0 (\Delta K)^n
$$

(6.61)

where $C_0$ is an empirical constant for each material, and $n$ is another constant called the growth rate exponent. Integrating Eq. (6.63) results in

$$
c_N = c_0 + \int_0^N C_0 (\Delta K)^n \, dN
$$

(6.62)
where \( c_n \) is the crack length after \( N \) cycles. By taking the logarithms of the two sides of Eq. (6.63), it can be rewritten as

\[
\log(\frac{dc}{dN}) = \log C_0 + n \log(\Delta K) \tag{6.63}
\]

As schematically illustrated in Fig. 6.17a, a plot of \( \log(\frac{dc}{dN}) \) vs. \( \log(\Delta K) \) results in a straight line whose intercept with the \( dc/dN \) axis gives the value of \( C_0 \), and its slope is \( n \). In practice, however, the relationship between \( \log(\frac{dc}{dN}) \) derived from experimental data and \( \log(\Delta K) \) is not entirely linear, as shown in Fig. 6.17b. As can be seen, a generalized plot of \( \log(\frac{dc}{dN}) \) vs. \( \log(\Delta K) \) can be divided into three distinct regions. Region I corresponds to either the nucleation of a crack or the slow growth of an existing crack. During region II, the crack growth rate is characterized by a power law corresponding to Eq. (6.61); i.e., the crack growth rate reaches a steady state. Eventually, the crack size reaches a critical value \( c_c \), accelerating the growth rate during region III, which is terminated by fracture. It is important to realize that Eq. (6.61) is only applicable during region II of steady-state crack growth.

Although Eq. (6.61) is widely used to express the rate of fatigue crack growth, it does not take into account the existence of a threshold cyclic stress intensity range \( \Delta K_{th} \) at the boundary between regions I and II, as shown in Fig. 6.17b. If the cyclic stress intensity range \( \Delta K < \Delta K_{th} \), an existing crack does not propagate under cyclic loading, consistent with the concept of allowable crack size described earlier. Experiment shows that \( \Delta K_{th} \) is a sensitive function of the ratio \( \sigma_{min}/\sigma_{max} \) or \( K_{min}/K_{max} \). For a given value of \( \Delta K = K_{max} - K_{min} \), \( \Delta K_{th} \) decreases as the ratio \( \sigma_{min}/\sigma_{max} \) increases. Having identified the parameters required to estimate the number of stress cycles \( N \) preceding fracture, i.e., fatigue life, the procedure can be summarized as follows provided that \( dc/dN \) for the material is known within region II of Fig. 6.17b.

1. Given an initial crack size \( c_0 \) as evaluated from nondestructive testing, it is first essential to determine whether the stress field at the crack tip corresponds to a plane strain condition. If the plane strain fracture toughness and tensile yield strength of the material are \( K_{ic} \) and \( \sigma_{yp} \), respectively, the minimum thickness of the part \( t_{min} \) satisfying plane strain condition is

\[
t_{min} = 2.5\left(K_{ic}/\sigma_{yp}\right) \tag{6.64}
\]

If \( K_{ic} \) of the material is not known, it can be estimated using Eq. (6.51).

2. Frequently, the initial crack size \( c_0 \) must be corrected for plastic deformation at the crack tip. It is recalled from Eq. (6.32) that under plane strain condition, the radius of the plastic zone \( r_p \) at the crack tip is given by

\[
r_p = 1/6\pi \left(K_{ic}/\sigma_{yp}\right)^2
\]
therefore, the effective initial crack length $c_{\text{eff}}$ becomes

$$c_{\text{eff}} = c_0 + \frac{1}{6} \pi (K_{\text{lc}}/\sigma_{\text{yp}})^2$$  \hspace{1cm} (6.65)

3. To determine the critical crack size $c_c$, $K$ is replaced by $K_{\text{lc}}$ in the stress intensity equation, i.e.,

$$K = K_{\text{lc}} = C(\pi c_c)^{1/2}$$ \hspace{1cm} (6.66)

Figure 6.17  Shape of curves representing crack growth rate by a fatigue mechanism.  
(a) Theoretical curve.  (b) Experimental curve.
Since the constant $C$ in the above equation varies only slightly with crack size, it may be considered to be independent of crack size for all practical purposes. Therefore, the initial crack size $c_0$ can be used to estimate $C$ from the particular geometry of the crack and part, as described earlier.

4. Typically, the experimental value of $dc/dN$ for the material is expressed as $A(\Delta K)^n$, where $A$ and $n$ are known, i.e.,

$$dc/dN = A(\Delta K)^n \quad (6.67)$$

It is recalled from Eq. (6.60) that $\Delta K$ is given by

$$\Delta K = C \Delta \sigma (\pi c)^{1/2}$$

Combining Eq. (6.67) with the above equation results in

$$dc/dN = A[C \Delta \sigma (\pi c)^{1/2}]^n \quad (6.68)$$

It is possible to rewrite the above equation in the form

$$dc/c^{n/2} = A[C \Delta \sigma (\pi c)]^{n/2} dN \quad (6.69)$$

5. Finally, the number of stress cycles $N_f$ preceding fracture can be determined by integrating both sides of Eq. (6.69) between the limits $c = c_{\text{eff}} = c_0 + r_p$ corresponding to $N = 0$, and $c = c_c$ corresponding to $N = N_f$, as follows:

$$\int_{c_{\text{eff}}}^{c_c} (dc/c^{n/2}) = \int_0^{N_f} A[C(\Delta \sigma)(\pi c)^{1/2}]^{n/2} dN \quad (6.70)$$

When data is not available for the material, a reasonable approximation for steels is $A = 7.5 \times 10^{-12} \text{ m/cycle}$ and $n = 3$ for steels.

To further illustrate the above procedure, consider a part whose thickness is 0.03 m and width is 0.4 m. Nondestructive evaluation shows that the part contains an edge crack of a length 0.002 m. If the part is made of a material whose yield strength is 600 MPa and plane strain fracture toughness $K_{fc}$ of $50 \text{ MPa} \cdot \text{m}^{1/2}$, and fatigue crack growth rate $dc/dN$ of $7.5 \times 10^{-12} \Delta K^{3/2} \text{ m/cycle}$, it is required to estimate its fatigue life $N_f$ under a cyclic tensile load varying from 0 to 50,000 kg. The value of constant $C$ corresponding to the geometry of the part is given as 1.1.

First, it is important to determine whether the stress field at the initial crack tip corresponds to a plane strain condition from:

$$t_{\text{min}} = 2.5(K_{fc}/\sigma_{yt})^2 = 2.5(50/600)^2 = 0.017 \text{ m}$$

Since the thickness of the part 0.03 m $> t_{\text{min}}$, a plane strain condition is developed at the initial crack tip. Using Eq. (6.32), the radius of the plastic zone at
the crack tip $r_p$ is given by

$$r_p = (1/6)\pi (K_{ic}/\sigma_{yp})^2 = 0.0004 \text{ m}$$

and therefore the effective length of the initial crack $c_{eff}$ is given by

$$c_{eff} = c_0 + r_p = 0.002 + 0.0004 = 0.0024 \text{ m}$$

Given a maximum tensile load $P$ of 50,000 kg, $\sigma_{max}$ is given by

$$\sigma_{max} = P / \text{cross-sectional area} = 50,000 / (0.03)(0.4)$$

$$= 4.2 \times 16 \text{ kg/m}^2 = 41 \text{ MPa}$$

Since the load varies from 0 to 50,000 kg, $\sigma_{min} = 0$ and $\sigma_{max} = 41 \text{ MPa}$, i.e.,

$$\Delta \sigma = 41 \text{ MPa}$$

Using the equation

$$K = K_{ic} = C \sigma_{max} (\pi c_c)^{1/2}$$

the critical crack size $c_c$ is given by

$$c_c = K_{ic}^2 / C^2 \sigma_{max}^2 \pi = 0.39 \text{ m}$$

From Eq. (6.70),

$$\int_{c_{eff}}^{c_c} dc / \varepsilon^{3/2} = \int_0^{N_f} 7.5 \times 10^{-12} [1.1(41)^{1/2}]^3 dN$$

$$\int_{0.0024}^{0.39} dc / \varepsilon^{3/2} = (3.75 \times 10^{-8})N_f$$

or

$$\int_{0.0024}^{0.39} -2 / \varepsilon^{1/2} = (3.75 \times 10^{-8})N_f$$

$$-2/(0.39)^{1/2} + 2/(0.0024)^{1/2} = 3.75 \times 10^{-8}N_f$$

$$39 = 3.75 \times 10^{-8}N_f$$

and therefore, $N_f$ is given by

$$N_f = 39/3.75 \times 10^{-8} = 10^9 \text{ cycles}$$

To summarize, in the presence of an initial edge crack of length 0.002 m, the remaining life of the part under the specified maximum and minimum stresses during each cycle is about $10^9$ cycles.
6.16 Selection of Materials Resistant to Fracture

Structural components are usually designed for a certain load-carrying capacity as their principal functional requirement. It is recalled that mechanical loads can be broadly classified into two main types: (i) static loads and (ii) dynamic loads. In practice, parts are subjected to a combination of static and dynamic loads; however, the designer can determine both the magnitude and nature of the principal load encountered during service taking into account the effect of secondary loads as pointed out earlier.

It is a common practice in engineering design to consider the tensile yield strength of a given material as a measure of its load-carrying ability. In static applications where the principal load remains constant throughout the service life of the component or structure, it is required that the component or structure is capable of supporting the load without fracturing. From an engineering design point of view, however, the component or structure should not plastically deform during service, and elastic deformation is kept at a minimum predetermined value. To satisfy these requirements, materials are selected on the basis of their yield strength, elastic limit, modulus of elasticity, ultimate tensile strength, and in some cases ductility.

Dynamic applications differ from static applications in that the load on the component or structure may be applied only a few times during service life but in the form of a sudden shock load, or it may oscillate regularly for up to several million cycles between some definable limiting values. For these applications, the most important materials requirements are impact strength, impact toughness, fatigue strength (endurance limit), as well as general toughness and ductility.

Most structural applications involve complex combinations of static and dynamic loads. For example, pure dynamic loads cannot exist in practice because any component or structure must support its own weight, which is static in nature. However, by carrying out certain calculations, the design engineer determines the principal load and its nature in a given application and a structural material is selected on that basis while considering the effects of secondary loads in addition to other requirements. Therefore, classification of mechanically applied loads into two main categories remains to be a convenient means for selecting materials resistant to various types of fracture.

6.16.1 Materials for Static Load Applications

It is possible to classify the applications involving a static load into (i) large structures, (ii) fabricated structures, (iii) small components, and (iv) equipment housings. A brief account for each type of application is given below.

Examples of large structures include frames, housings, and machine beds. Because each of these structures represents a considerable weight, materials used in these applications must have a rather high load-carrying capacity. To satisfy
this requirement, various grades of cast iron as well as carbon steels are usually
used in these applications. In addition to their adequate mechanical strength, they
are characterized by low price and ease of fabricability. In some applications of
this nature, however, where weight saving is an overriding factor, nonferrous
light metals such as aluminum alloys are used.

Large structures are in general manufactured by two methods: (i) casting
using various techniques and (ii) fabricating (welding, riveting, bolting, etc.).
When the structure is rather large, it is extremely difficult to improve its mech-
anical strength by heat treatment. Therefore, it is essential that the desired proper-
ties are maintained in the as-manufactured condition requiring selection of
suitable materials as well as manufacturing methods.

Frequently, it is more economical as well as practical to build large struc-
tures from standard fabricated components such as bars, plates, rods, sheets, etc.
and assemble them by various joining processes. This practice has a number of
advantages; e.g., it permits the use of wrought metal products such as hot-rolled
plates and cold-rolled sheets which have better mechanical properties than
castings. Also, this practice permits the use of a wide range of fabrication
processes, and it speeds up production in comparison with castings.

Most of the fabricated components are made of hot-rolled low-carbon
steels; however, high-strength low-alloy steel can also be used because of its
superior strength and toughness. Another advantage of high-strength low-alloy
steels is their superior resistance to atmospheric corrosion. Even though AISI
alloy steels have still higher strength, they are generally undesirable for these
applications because of their inferior formability and fabricability. Although
the use of nonferrous alloys is rather restricted in these applications, light metals
such as aluminum, magnesium, and titanium alloys are used in a few applications
where weight saving is of prime importance.

Materials used in manufacturing smaller components, such as machines
and housings and casings for small instruments and appliances, can vary widely
depending upon the load and stress level during service, production quantity,
fabrication methods, and general appearance. Also, ferrous alloys are used in
many of these applications. When casting processes are suitable for production
of small components, various grades of cast iron are used. For example, malle-
able cast iron is particularly suited for applications requiring high strength and
ductility; however, for very high strength requirements, high-carbon steels or
even tool steels are preferred. Cast magnesium alloys are used in some appli-
cations where the load requirements are not too severe and light weight is an
important aspect. Because of their excellent castability, good mechanical proper-
ties, and corrosion resistance, cast aluminum alloys are also used in small
component applications. Furthermore, some of the cast copper alloys are suitable
for these applications. Although cast alloys have many small-part applications,
smaller parts and components can be best made from sheet metal by many
cold-working processes. Typical applications range from small brackets to complex base and frame configurations. Ferrous alloys are most commonly used in these applications, including low-carbon steel grades and the AISI grades; however, stainless steels are also used when excellent surface appearance and maximum corrosion resistance are required. Other materials used in these applications to a lesser extent than the ferrous alloys include aluminum alloys and copper alloys. Small components such as various fasteners, studs, and small flanges are usually manufactured from rod stock by means of what is known as screw machine. Carbon steels are used in the majority of these applications, but both aluminum and copper alloys can find some applications.

Equipment housings are used to protect operating units from mechanical damage and to some extent from environmental corrosion. Also, they guard operating personnel and others from possible injury by moving mechanical elements, electric shock, contact with hot area, and excessive noise. Ferrous sheet metals, particularly those of AISI grades, are useful in these applications. Some aluminum and copper alloys are used to a much less extent in certain applications.

6.16.2 Materials for Dynamic Load Applications

Design criteria for static load applications are rather simple and more straightforward in comparison with dynamic load applications. Dynamic loads include (i) impact loads which cause very rapid strain rate of the material and which can lead to fracture, and (ii) fatigue loads which are more commonly encountered and involve repeated application of a load smaller than the respective static breaking load. Although creep is sometimes treated as a dynamic loading condition, it is not as clearly dynamic as impact and fatigue loads, and therefore it is treated separately in the next section.

Impact loading conditions are encountered in many airborne parts subjected to severe shock and vibrations. However, fatigue loading conditions are developed in many engineering structures including rotating machine parts such as shafts. It is possible that both fatigue and impact loads are combined to produce very high strain rates.

It is recalled from Chap. 4 that the impact toughness test measures the capability of a given material for instantaneous and considerable plastic deformation under high localized stress, which is referred to as notch toughness. It is important to realize that microscopic notches acting as stress concentration sites can arise from microstructural inhomogeneities such as shrinkage cavities and gas holes in cast metals. Other microscopic notches, particularly in ferrous alloys, include sharp-edged graphite flakes in gray cast iron and the martensite needles in a soft retained austenite matrix. Therefore, a basic requirement for high-impact toughness in any ferrous alloy is a homogeneous microstructure. Among the steel grades suitable for impact toughness applications are those with maximum

Copyright © 2004 by Marcel Dekker, Inc.
ductility. Although high ductility must be combined with an adequate static strength, ductility remains to be the decisive factor. For example, a steel having a high level of tensile strength and lower ductility is inferior to a steel of lower strength and higher ductility. Performance capability is also determined by the ductile to brittle transition temperature range. Low-carbon steels (≤0.2% C) with a uniform fine pearlitic structure can satisfy those requirements. Extra-low nonmetallic content, particularly sulfur, phosphorus, nitrogen, and oxygen, further improve impact toughness, and therefore a thoroughly deoxidized steel grade is preferable. A fine-grained pearlitic structure promoting impact toughness can be obtained by quenching and tempering at a specified temperature usually around 650°C to produce a dispersion of fine spheroidal carbides in a ferrite matrix. However, since the hardenability of low-carbon steels with low carbon content is not adequate for such a treatment, low-carbon grades of AISI structural alloy steels are selected for applications requiring a combination of high static strength and impact toughness. Because of their low-impact toughness, high-carbon grades of both carbon and alloy steels are not recommended for such applications. However, some of the stainless steel grades with sufficient impact toughness can be used. Wrought nonferrous alloys can also be used in applications requiring high-impact toughness such as some of the high-strength precipitation hardened aluminum and titanium alloys.

Wrought steels are by far the most frequently selected structural materials for fatigue applications. One of the most important properties governing the selection of a steel for fatigue applications is its ultimate tensile strength. Usually, the endurance limit (see Sec. 4.3.6) is a fraction of the ultimate tensile strength in the order of 0.4. For example, if the ultimate strength of a steel grade is 690 MPa, its endurance limit is about 275 MPa. However, this value serves only as a guide because of the wide scatter usually observed in determining the endurance limit. To achieve high-strength levels suitable for fatigue applications, the steel must be quenched and tempered to produce tempered martensite, which requires high hardenability. Microstructure is also important for optimum fatigue strength. Spheroidal pearlite is preferred over lamellar pearlite. Since fatigue cracks originate at the surface, surface hardening treatments such as case hardening or carburizing significantly increases the fatigue strength. Mechanical treatments of the surface such as shot peening is also useful in increasing fatigue strength. Nonferrous alloys for fatigue applications include aluminum, titanium, and nickel alloys.

6.16.3 Materials for Creep Applications

Many materials are available for application requiring resistance to creep deformation and fracture at elevated temperatures. Among these materials are (i) ferritic steels, (ii) stainless steels, (iii) iron-base superalloys, (iv) nickel-base superalloys, (v) cobalt-base superalloys, and (vi) dispersion-hardened alloys.
Although carbon steels are used in many applications at ambient temperature, they lose their strength rapidly above about 430°C. However, addition of up to 0.5% Mo extends the temperature capability to about 480°C. Above this temperature, graphitization occurs (precipitation of carbon in the form of a graphite phase), considerably weakening the steel. Graphitization is suppressed and oxidation resistance is improved by addition of 1.5% Cr. Chromium-molybdenum steels with alloy contents in the range 1–10% Cr, 0.5–1.5% Mo are the most creep-resistant class of ferritic steels with temperatures capabilities extending to about 650°C. Commonly, these steels are used in either hot-worked or annealed condition.

Stainless steels containing more than 11% Cr are particularly useful for creep applications including the 200, 300, and 400 series of the AISI. Both the 200 and 300 series are austenitic; i.e., they have face-centered cubic structures. Because diffusion rates in fcc structures are lower than that in the body-centered cubic structure of ferritic steels, stainless steels have higher creep resistance than ferritic steels. Furthermore, the higher chromium content of the stainless steels offers a better resistance to high-temperature corrosion. Steels of the 400 series are ferritic; i.e., they have a bcc structure, and therefore they are inferior to those of the 200 and 300 series.

Essentially, the Fe-base superalloys are modifications of the 300 series of austenitic stainless steels. Such alloys are developed for applications at temperatures above 650°C. Other alloying elements present in Fe-base superalloys may include Co, Mo, W, V, Nb, Ti, Al, Zr, B, and Cu. Strengthening of these alloys is achieved by solid solution and precipitation of secondary phases, particularly carbides, nitrides, and borides. Precipitation of intermetallic compounds may also contribute to strengthening. However, the temperature capabilities of Fe-base superalloys range from about 760 to 900°C.

Nickel-base superalloys are the most widely used in high-temperature applications requiring creep resistance such as various components of gas turbine engines. Many of the Ni-base superalloys are primarily strengthened by an intermetallic compound based upon the Ni3Al composition and known as the γ’ phase. Because these alloys lack sufficient ductility, they are produced by various casting techniques. However, a few of the γ’-strengthened alloys are fabricated as wrought products, i.e., sheet, plate, and bar. Other wrought Ni-base superalloys derive their strength from solid solution. Generally, Ni-base superalloys are used in applications requiring creep resistance at temperatures higher than 900°C. Cobalt-base superalloys are primarily strengthened by solid solution and precipitation of carbide phases, and their temperature capability is the same or slightly better than the Ni-base superalloys. However, they are not as frequently used as the Ni-base superalloys.

Dispersion-hardened alloys are essentially Ni- or Fe-base superalloys strengthened by a fine dispersion of oxide phase such as yttria (Y2O3). Although they have better temperature capabilities than the conventional superalloys, their use is rather limited because of fabrication difficulties and higher cost.
7

Structure of Engineering Alloys

7.1 Introduction

Modern engineering technology vividly demonstrates the intimate relationship between matter and energy, as well as the vital importance of the structure of matter in design considerations. Of the three states of matter; solid, liquid and gaseous, matter in its solid state dominates the structural and mechanical applications in engineering design. As pointed out in Chap. 1, the properties of engineering materials are those qualities which determine their usefulness in certain applications. To effectively select and use materials for engineering applications, it is important to have a basic knowledge of the origin of these properties and how they can be controlled. Developing such a knowledge lies within the spectrum of materials science.

Most of the technologically important properties of engineering alloys, including mechanical, chemical, and physical properties, are determined by their internal structure. Therefore, an understanding of the principles governing the internal structural features is essential in the selection of the best material or the optimum combination of material and processing for specific engineering applications. Due to the intimate relationship between matter and energy, it is first instructive to review the principles of thermodynamics.

7.2 Principles of Thermodynamics

Thermodynamics is the branch of physical science dealing with energy, its conservation and its transformation from one form to another. Just as mechanics
introduces mass, length, and time as basic concepts, thermodynamics introduces temperature as a basic concept; it is a measure of thermal or heat energy. By means of some initial statements known as the laws of thermodynamics, it is possible to derive very useful relationships between measurable macroscopic properties of a substance such as temperature, solubility, and equilibrium constant. Such a substance which can either be a solid, liquid, gas, or any combination of these is called a system.

A phase of the system is a substance of a given composition, structure, and properties and is separated from any other substance by a definite bounding surface. If the system consists of a single phase, it is said to be homogeneous. An inhomogeneous system consists of two or more phases. In terms of thermodynamics, most of the engineering alloys are heterogeneous systems; i.e., they consist of two or more phases. Another character of a system is its components defined as the number of chemical species required to define the chemical composition of the system. For example, an alloy of Fe and C is a two-component system.

It is recalled from Chap. 5 that the state of stress at any point within a body is completely defined in terms of a minimum number of stress components. Similarly, the equilibrium state of a system is defined in terms of a number of variables, including temperature, pressure, composition, and volume. However, some of these variables are related requiring only the independent variables to be specified in order to completely define the state of the system. Since processes related to engineering alloys are carried out at constant pressure, it is excluded as a variable. If the system consists of \( p \) phases and \( c \) components, it can readily be shown that the number of independent variables or degrees of freedom \( f \) required to define the state of system at constant pressure is given by

\[
f = c - p + l
\]  

which is known as the phase rule.

A system in a state of thermodynamic equilibrium has a certain energy called the internal energy \( E \). It includes the kinetic energy of its atoms and molecules, as well as the energy of the system due to its position in space, i.e., the potential energy. Thermodynamics is only concerned with changes in energy, which are of practical significance than absolute energy values. When a system undergoes a change from equilibrium state to another, it exchanges energy with its surroundings. Such an exchange is governed by the first law of thermodynamics, which is a statement of the law of energy conservation. If a system absorbs an amount of heat \( \Delta Q \) from the surroundings and exerts a mechanical work \( \Delta W \) on the surroundings, changing its state from \( A \) to \( B \), the corresponding change in internal energy \( \Delta E \) is given by

\[
\Delta E = E_B - E_A = \Delta Q - \Delta W
\]
Since energy can neither be created nor destroyed, it is evident that the change in internal energy $\Delta E$ is independent of the path followed during the change from state $A$ to state $B$. Physically, this means that the internal energy $E$ is a function of only the thermodynamic state, or in short a state function such as temperature, pressure, and volume. Mathematically, a state function can be expressed as an exact differential and therefore, the above equation can be expressed as

$$dE = \Delta Q - \Delta W$$  \hspace{1cm} (7.2)

By means of the first law of thermodynamics, another state function called enthalpy or heat content can be defined as described below. If a process is carried out under constant pressure $P$, such as with many of the processes involving engineering alloys, the work done by the system is expressed as

$$\Delta W = P \Delta V$$

where $\Delta V = V_B - V_A$ is the change in volume of the system corresponding to the change from state $A$ into state $B$. Under constant pressure, Eq. (7.2) can be rewritten as

$$dE = E_B - E_A = \Delta Q - P(V_B - V_A) \quad \text{or} \quad (E_B + PV_B) - (E_A + PV_A) = \Delta Q$$

Since $E$, $P$, and $V$ are all state functions, $E + PV$ must be a state function, which is defined as the enthalpy $H$, or heat content, as

$$H = E + PV$$  \hspace{1cm} (7.3)

and

$$\Delta H = \Delta Q$$  \hspace{1cm} (7.4)

To illustrate the physical meaning of enthalpy, consider a chemical reaction whereby a substance $x_1$ combines with another substance $x_2$ to form a new substance $x_3$, i.e.,

$$x_1 + x_2 \rightarrow x_3$$

Differences in enthalpy $\Delta H$ between the reactants $x_1$ and $x_2$ and product $x_3$ is a measure of how tightly the atoms in substance $x_3$ are bound together, which is reflected on such properties as melting point and hardness. A higher $\Delta H$ means tighter bonding between the atoms, which in turn leads to higher hardness and higher melting point. A similar argument is applied to physical changes where a phase of a given substance is transformed into another phase.

According to the second law of thermodynamics, any system has a characteristic property called entropy $S$. If the state of the system is changed from $A$ to
B, the corresponding change in entropy \( dS \) is given by

\[
dS = \frac{\Delta Q}{T}
\]

(7.5)

where \( \Delta Q \) is the heat absorbed by the system and \( T \) is the temperature. Similar to internal energy \( E \) and enthalpy \( H \), the entropy \( S \) is a state function. Physically, entropy is linked with the degree of disorder or randomness of arranging objects together. A higher degree of disorder means a higher entropy and vice versa.

Substituting Eq. (7.5) into (7.2), and noting that at constant temperature \( T dS = d(ST) \), results in

\[
dE = T dS - \Delta W
\]

(7.6)

or

\[
d(E - TS) = -\Delta W
\]

Since \( \Delta W \) is the work done by the system, \( E - TS \) is called the work function \( A \), i.e.,

\[
A = E - TS
\]

(7.7)

and therefore for a process carried out at constant temperature,

\[
dA = -\Delta W
\]

(7.8)

Since \( E, T, \) and \( S \) are all state functions, \( A \) is also a state function called the Helmholtz free energy.

For a process carried out at constant temperature and pressure, the change in enthalpy can be obtained by differentiating Eq. (7.3), i.e.,

\[
dH = dE + P dV + V dP
\]

and since the pressure is constant, \( dP = 0 \), and therefore

\[
dH = dE + P dV
\]

Substituting Eq. (7.6) into the above equation results in

\[
dH = T dS - \Delta W + P dV
\]

and at constant temperature and pressure, the above equation can be rewritten as

\[
d(H - TS) = P dV - \Delta W
\]

Similar to the case of the work function \( A \), \( H - TS \) is another state function called the Gibbs free energy, i.e.,

\[
G = H - TS
\]

(7.9)
If the work done by the system is limited to that corresponding to expansion, 
\[ \Delta W = P \, dV, \]
and therefore
\[ dG = 0 \] (7.10)

According to the above equation, at constant temperature and pressure, a system is in equilibrium when its Gibbs free energy reaches a minimum value. Since many processes occur at constant temperature and pressure, the Gibbs free energy is used as a criterion to predict the occurrence of a given process. For example, consider a physical or chemical change whereby

\[ \alpha \rightarrow \beta \]

Three distinct conditions can be identified depending on the change in free energy \( \Delta G = G_B - G_A \), as follows:

1. At equilibrium, \( \Delta G = 0 \), i.e., the two phases coexist in equilibrium.
2. If \( \Delta G \) is negative, i.e., \( G_B < G_A \), the transformation from \( \alpha \) to \( \beta \) can occur.
3. If \( \Delta G \) is positive, i.e., \( G_B > G_A \), the transformation from \( \alpha \) to \( \beta \) cannot occur.

It is evident from Eq. (7.9) that a minimum in free energy corresponds to a maximum in entropy. Therefore, an equivalent statement for the condition of equilibrium is that the entropy of the system reaches a maximum value. Also, it is clear from the Eq. (7.9) that the role of entropy in determining the free energy of the system increases with temperature and diminishes with reducing temperature. At a temperature of absolute zero, i.e., \( T = 0 \), the third law of thermodynamics states that the entropy of the system \( S = 0 \). In practice, however, it is not possible to achieve such a condition. As a general rule, any system always seeks the condition of equilibrium corresponding to minimum free energy and maximum entropy.

To summarize the preceding discussion, any system always seeks a more stable state corresponding to the minimum possible free energy or, equivalently, the maximum possible entropy. For any transition to a more stable state, the driving force is the difference in free energy between the final and initial states. An important limitation of thermodynamics, however, is that it is unable to predict the rate of approach to equilibrium as described below.

Time required to reach an equilibrium state is called the relaxation time. For gases and most liquids, the relaxation time is so short that they are almost always in a well-defined state of complete equilibrium. In contrast, the relaxation time for most significant atomic processes in solids is so long that complete equilibrium is seldom, if ever, achieved. As a result, engineering alloys among other solid substances quite often exist for extended periods of time in metastable states not corresponding to the lowest possible free energy. Therefore, it is important to realize that thermodynamics is inherently limited in that it cannot predict the rate
of approach to equilibrium. An equivalent statement is that thermodynamics can predict whether a given process can take place, but it cannot predict its rate. As demonstrated later, many of the important technological applications of engineering alloys are dependent upon their existence in metastable states. Determination of the rate of approach to equilibrium lies within the scope of kinetic theory, as elucidated later in this chapter.

### 7.3 Elements of Internal Structure

A remarkable feature of engineering alloys among other materials is that their constituent atoms are arranged in certain geometrical patterns called crystals, as pointed out in Chap. 1. Only in a few applications, such as turbine blades of advanced gas turbine engines, are engineering alloys used in the form of a single crystal. In most applications, engineering alloys are used in the polycrystalline form where the alloy consists of an aggregate of crystals. It is recalled that each crystal in an aggregate is called a grain, and the surface separating each grain from another is referred to as a grain boundary. In a given alloy, all crystals of identical chemical composition and arrangement of atoms are referred to as a phase. An alloy may consist of one or more phases.

Generally, it is possible to classify the overall structural features of materials into two main types: (i) macrostructural features and (ii) microstructural features. By definition, macrostructure refers to structural characteristics, which can be revealed by the naked eye or by a low-power hand lens. Examples of such features include voids, porosity, and fracture characteristics. Macroscopic examination of failed products can reveal a great deal of information about the cause of failure, as explained in Chap. 6. Structural features which can only be revealed by means of a high-power microscope are collectively grouped as microstructure.

An important characteristic of the microstructure of engineering alloys is that it is not a fixed character. Rather, it is dependent upon thermomechanical history. By thermomechanical is meant all the mechanical and heat treatment processes carried out on the alloy. Furthermore, microstructural changes can occur during service of an engineering product, e.g., during exposure to elevated temperatures. As demonstrated in this chapter, microstructural changes can significantly alter the properties of engineering alloys. Some of these changes can have beneficial effects but others can be detrimental. Consequently, the failure analyst must be familiar with the relationship between microstructure and properties and the techniques used in microstructural characterization, which are the subjects of Chaps. 7 and 8.

In its broader sense, the microstructure of a given material can be classified into two main elements: (i) atomic structure and (ii) grain structure. Atomic structure consists of: (i) structure of individual atoms, (ii) binding forces, which hold
a large number of atoms together, and (iii) the manner in which atoms are
arranged relative to each other in crystals, i.e., the crystal structure. Grain struc-
ture includes such features as the morphology (size and distribution) and chemi-
cal composition of all phases present in the material. Evidently, a logical starting
point to study the structure of engineering alloys is the structure of individual
atoms. A simplified model of the atom as described below is useful in elucidating
the functional behavior of engineering alloys in engineering applications.

7.4 Structure of the Atom

To understand the functional behavior of materials in engineering applications, it
suffices to consider an atom as consisting of a central positively charged nucleus
surrounded by negatively charged electrons (Fig. 7.1a). Electrical neutrality of
the atom requires that the positive charge of the nucleus is exactly balanced by
the negative charge of the electrons. Each atom is distinguished by an atomic
number \( Z \) corresponding to the number of constituent electrons.

Experiment shows that the energy of electrons in an atom is “quantized.” By quantized is meant
that the electrons can only have certain allowed energies
and no others. Therefore, electrons in an atom must reside at their allowed energy
levels. Closely spaced levels are grouped into shells known as the K, L, M, N, . . .
in order of increasing energy. However, exceptions to this rule do exist as
explained later in this section. Each shell is identified by its characteristic number
\( n \) of energy levels where \( n = 1 \) for the K shell, \( n = 2 \) for the L shell, \( n = 3 \) for the
M shell, . . . . Also, an energy level is identified by an integer \( l \) reflecting its
energy relative to other levels such that \( l = 0, 1, 2, 3, 4, . . . \) in order of increasing
energy. A more common notation used to describe energy levels is to assign a

![Figure 7.1](image-url) Atomic structure. (a) A simplified model for the atom. (b) Energy levels of
electrons in an atom.
letter for each value of \( l \) such that the energy level corresponding to \( l = 0 \) is called \( s \), \( p \) corresponds to \( l = 1 \), \( d \) corresponds to \( l = 2 \), \( f \) corresponds to \( l = 3 \), and so on as shown in Fig. 7.1b.

It follows from the above discussion that the K shell (\( n = 1 \)) having the lowest energy consists of one \( s \) energy level identified as \( 1s \). For the next higher energy L shell for which \( n = 2 \), there are two levels identified as \( s \) and \( p \), i.e., \( 2s \) and \( 2p \). An M shell (\( n = 3 \)) has three levels \( s \), \( p \), and \( d \) identified as \( 3s \), \( 3p \), and \( 3d \); an N shell (\( n = 4 \)) has four levels \( s \), \( p \), \( d \), and \( f \) identified as \( 4s \), \( 4p \), \( 4d \), and \( 4f \); and so on.

Within any shell, the \( s \) level has the lowest energy followed by the \( p \), \( d \), \( f \), . . . levels in order of increasing energy, as schematically illustrated in Fig. 7.1b. As can be seen, the energy difference between levels within the same shell is considerably smaller than that between two shells.

Distribution of electrons in an atom is governed by the following rules:

1. According to the rules of thermodynamics described in Sec. 7.2, electrons always tend to reside at the lowest possible allowed energy levels. Once the lowest possible energy level is fully occupied, any remaining electrons seek the next higher allowed energy level, and so on. An electron has the lowest possible energy corresponding to the most stable state, when it resides in a completely filled energy level.

2. A given energy level can accommodate a maximum number of electrons \( N \) and no more, where \( N \) is given by

\[
N = 2(2l + 1)
\]

such that \( l = 0, 1, 2, 3, \ldots \). Therefore, an \( s \) level for which \( l = 0 \) can accommodate a maximum of two electrons; a \( p \) level can accommodate 6 electrons; a \( d \) level can accommodate 10 electrons; an \( f \) level can accommodate 14 electrons; etc. However, it is possible that an energy level contains fewer electrons than the maximum allowed number.

3. It follows from the preceding rule and the constituent energy levels of each shell described earlier that the K shell (\( n = 1 \)) can accommodate a maximum of 2 electrons, the L shell (\( n = 2 \)) can accommodate 8 electrons, the M shell can accommodate 18 electrons, etc. or in general a given shell can accommodate a maximum of \( 2n^2 \) electrons.

It can be concluded from the discussion above that the order of filling the energy levels with electrons follow the sequence: \( 1s \ 2s2p \ 3s3p3d \ 4s4p4d4f \ldots \) noting that the \( s \) level can accommodate 2 electrons, the \( p \) level can accommodate 6 electrons, the \( d \) level can accommodate 10 electrons, etc. However, this sequence is not strictly followed particularly in elements whose
atomic number $Z = 21$ or higher. In this case, the energy of the $4s$ level is lower than that of the $3d$ level. Likewise, the energy of the $5s$ level is lower than that of the $4d$ level, and so on.

Filling the energy levels of individual atoms with electrons defines the sequence of arranging the chemical elements in the periodic table shown in Fig. 7.2. As can be seen, the table is divided into 18 columns or groups, and 7 basic periods or rows with 2 supplementary periods. To illustrate how the periodic table is constructed based on principles governing the distribution of electrons in the atom, consider the simplest hydrogen atom whose atomic number $Z = 1$; i.e., the hydrogen atom contains only 1 electron.

It is evident that the single electron of the hydrogen atom must reside at the lowest energy level which corresponds to the $s$ level of the K shell (1s). For the helium atom where $Z = 2$, the 2 electrons reside at the $s$ level of the K shell, which then becomes completely filled (1$s^2$). Reference to Fig. 7.2 indicates the first row of the periodic table begins with hydrogen and is completed with helium, i.e., the first row defines the sequence of filling the K shell. Similarly, the sequence of arranging the elements within the second row follows the sequence of filling the L shell, as described below.

In the case of lithium where $Z = 3$, it becomes evident that 2 electrons fill the K shell and the third electron must reside at the $s$ level of the L shell. It is possible to envision that the Li atom consists of a core corresponding to the helium atom plus 1 electron residing in the 2s level.

For beryllium ($Z = 4$), 2 electrons fill the K shell (core of He atom) and the remaining 2 electrons reside in the $s$ level of the L shell. Ultimately, the second row is completed with neon ($Z = 10$) whose electronic structure can be written as

![Periodic Table](image-url)
1s\(^2\)2s\(^2\)2p\(^6\). It is noted that the L shell of neon is completely filled with electrons. Then, the third row is started with sodium (Z = 11) following the sequence of filling the M shell. By analogy with He and Li, the Na atom can be considered to consist of a core corresponding to the neon atom plus 1 electron residing in the 3s level. Similarly, the Mg atom (Z = 12) consists of a core of neon atom plus 2 electrons in the 3s level. Eventually, the third row is completed with argon (Z = 18) whose electronic structure can be written as 1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^6\). It is to be noted that the 3d level of argon is completely empty.

A fourth row begins with potassium (Z = 13). If the normal sequence of filling the levels is followed, the additional electron of potassium above that of argon, may be expected to occupy a 3d level. Instead, it resides in the 4s level, which has a lower energy. Therefore, the electronic structure of potassium can be written as 1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^6\)3d\(^0\)4s\(^1\). Similarly, for calcium whose Z = 14, the additional 2 electrons above those of argon reside in the 4s level; i.e., the electronic structure of Ca can be written as 1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^6\)3d\(^0\)4s\(^2\). Evidently, there are two common features of both potassium and calcium: (i) The energy of the 4s level is lower than that of the 3d level, and (ii) although the 4s level contains electrons, the 3d level is empty. A different behavior begins with strontium (Z = 15), as described below.

It is possible to visualize that the strontium atom is constructed by adding 1 electron to the calcium atom where the 3d level is empty and the 4s level is fully occupied. Evidently, the additional electron of the strontium atom either resides in a 4p level or 3d level depending upon which level has a lower energy. It turns out that the 3d level has a lower energy than the 4p level. Consequently, the electronic structure of strontium can be written as 1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^6\)3d\(^1\)4s\(^2\). As can be seen, strontium contains a partially filled d level, which is by definition a transition element. With increasing atomic number, however, irregularities occur in the progressive filling of the 3d shell as Z increases from 23 (vanadium) to 24 (chromium). In the case of vanadium, the electronic structure is 1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^6\)3d\(^3\)4s\(^2\); however, the electronic structure of Cr is 1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^6\)4d\(^5\)5s\(^1\). It is noted that 5 electrons instead of 4 electrons reside in the 3d level, leaving 1 electron in the 4s level. This behavior is related to a lower energy associated with 5 electrons in the 3d level. Progressive filling of the 3d level resumes from Z = 25 (manganese) to nickel (Z = 28), which contains 8 electrons in the 3d level. All elements from manganese to nickel contain 2 electrons in the 4s level.

Elements from strontium to nickel within the fourth row define the first series of transition elements with partially filled 3d level. Beginning with copper (Z = 29), the 3d level becomes completely filled; i.e., it contains 10 electrons. Ultimately, the fourth row is completed with krypton (an inert gas) with electronic structure of 1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^6\)3d\(^10\)4s\(^2\)4p\(^6\). Other series of transition elements are found in the fifth and sixth rows of the periodic table. Within
the sixth row, a group of transition elements from $Z = 57$ to $Z = 71$ is characterized by a filled $4f$ level. Such elements are known as the rare earths. Another group of rare earths is encountered in the seventh row.

7.5 Significance of the Electronic Structure of Atom

Electronic structure of individual atoms is the basis for understanding the nature of materials, as well as their behavior. It is evident from the previous discussion that an individual atom of a chemical element has a certain electronic structure characteristic of that element. More specifically, each atom is distinguished by characteristic electron energy levels different from those of any other atom. For example, the electron energy levels of a hydrogen atom are different from those of a sodium atom and so on. An equivalent statement is that a one-to-one correspondence exists between atomic number and electron energy levels. Once the electron energy levels of an atom are determined, its atomic number is fixed. Therefore, each chemical element can be identified by the electron energy levels of its atoms. This is the basis for several methods of chemical analysis.

Any material consists of a large number of individual atoms, which must be held together by some kinds of binding forces. Such interatomic binding forces differ in their nature depending upon the electronic structure of individual atoms. A direct correlation exists between the nature of these forces and materials properties dependent upon the transport of electrons such as electrical, thermal, and magnetic properties. The nature of interatomic binding forces also dictates the manner in which individual atoms are arranged relative to each other in a material, which is the single most important factor determining the mechanical properties of materials, as demonstrated later.

Another importance of the electronic structure of individual atoms is its relationship to chemical properties of materials. Chemical properties are those qualities which determine the tendency of materials to combine with other substances to form compounds. Any chemical reaction takes place between individual atoms to lower their free energy. It is possible to understand the origin of these properties in terms of the electronic structure of individual atoms, as follows.

7.6 Electronic Structure and Chemical Properties: Classes of Elements

A common feature of the chemical elements constituting each of the 18 columns or groups of the periodic table (Fig. 7.2) is that their atoms contain the same
number of electrons in the outermost shells. It is recalled from the previous dis-
cussion that each period or row of the periodic table begins with filling an elec-
tron shell and ends with completely filling that shell. As a consequence, the
outermost shells of all elements of the last group are completely filled with elec-
trons, and elements of all other groups have partially filled outermost shells.
Since an electron has its lowest energy when it resides in a completely filled
shell, it becomes evident that the electronic structure of elements of the last
group corresponds to the most stable state:

He: \(1s^2\)
Ne: \(1s^2 2s^2 2p^6\)
Ar: \(1s^2 2s^2 2p^6 3s^2 3p^6\)
Kr: \(1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 4p^6\)

In practice, it is found that the above elements are chemically inactive; i.e.,
they are reluctant to undergo any type of chemical reactions. This behavior is cor-
related with their outermost filled shells, which corresponds to a stable state. With
the exception of He, all atoms of the inert gases contain filled outermost \(s\) and \(p\)
levels; i.e., they contain 8 electrons in the outermost shell, 2 in the \(s\) level, and 6 in
the \(p\) level. Chemical activity can then be correlated with the number of electrons
in the outermost shells and their ability to either donate or accept electrons. Since
atoms of the inert gases have already completely filled shells, they cannot donate
or accept electrons, and therefore they are chemically inactive.

Experiment shows that all atoms of the chemical elements, with the excep-
tion of H, always seek the state of lowest free energy corresponding to filled out-
nermost \(s\) and \(p\) levels; i.e. they naturally tend to produce an outermost shell
containing 8 electrons. Depending upon the number of electrons already present
in the outermost shell, four cases are distinguished, leading with the well-known
four classes of the elements: (i) metals, (ii) nonmetals, (iii) metalloids, and (iv)
inert gases. When an atom loses or gains electrons, its electrical neutrality
becomes imbalanced; i.e., the atom becomes electrically charged. An electrically
charged atom is said to be ionized. If it loses electrons, it contains an excess of
positive charge, i.e., becomes positively charged, and it is said to be a positive
ion. Conversely, if the atom gains electrons, it becomes negatively charged
and it is said to be a negative ion.

From a chemical point of view, metals are those chemical elements whose
atoms contain a few electrons in their outermost shells (less than 4). They occupy
the left and middle sections of the periodic table (Fig. 7.2), constituting about
two-thirds of the chemical elements. Because of the relatively few electrons in
the outermost shell of a metallic atom, e.g., Li, Na, K, Cu, Fe, Ni, etc., it is
more energetically favorable to lose electrons and form a positive ion. To remove
electrons from the outermost shells, energy must be supplied to atom, e.g., in the
form of heat. This energy is called the ionization energy. Metallic elements are
distinguished by relatively low ionization energies. Conversely, nonmetallic elements which occupy the upper-right-hand section of the periodic table contain a larger number of electrons (5–7) in their outermost shells; e.g., Cl with 7 electrons in its outermost shells requires high ionization energy to remove those electrons. Instead, it is energetically more favorable for a nonmetallic element to gain electrons and become a negative ion. Therefore, metals are distinguished by low ionization energy and low electron affinity and are said to be electropositive elements. In contrast, nonmetals are distinguished by high ionization energy and high electron affinity and are said to be electronegative elements. Elements with intermediate number of electrons in their outermost shells (usually 4) such as Si and Ge (group IV) have equal tendency to either lose or gain electrons; i.e., in some reactions they behave as metals and in others they behave as nonmetals—hence the name *metalloids*. As pointed out earlier, inert gases with completely filled shells have no tendency to either lose or gain electrons.

It can be concluded from the above discussion that the number of electrons in the outermost shell of an individual atom is the governing factor in determining the chemical properties of the elements. For example, consider the electronic structure of Na (a metal) and Cl (a nonmetal):

\[
\begin{align*}
\text{Na (Z = 11):} & \quad 1s^2 2s^2 2p^6 3s^1 \\
\text{Cl (Z = 17):} & \quad 1s^2 2s^2 2p^6 3s^2 3p^5
\end{align*}
\]

Sodium can readily lose its 3s electron to form a positive ion with 8 electrons in its outermost shell, i.e.,

\[
\text{Na} = \text{Na}^+ (1s^2 2s^2 2p^6) + e
\]

and Cl can readily gain one electron to form a negative ion with 8 electrons in its outermost shell, i.e.,

\[
\text{Cl} + e = \text{Cl}^- (1s^2 2s^2 2p^6 3s^2 3p^5)
\]

Therefore, there is a natural tendency for Na to react with Cl to form sea salt by the reaction

\[
\text{Na}^+ + \text{Cl}^- = \text{NaCl}
\]

resulting in a net decrease in free energy.

It follows from the above example that a direct correlation exists between the number of electrons in the outermost shell of an atom and its chemical properties. As a result, elements contained within the same group of the periodic table (same number of electrons in the outermost shell) have similar chemical properties. By definition, the number of electrons either lost or gained to complete the outermost shell of an atom is called the chemical valency of that element. According to this definition, both Na and Cl have a valency of 1. Transition
elements characterized by partially filled inner shells present a special case of valency. For example, Fe, whose electronic structure is $1s^2 2s^2 2p^6 3s^2 3p^6 3d^6 4s^2$, can have more than one valency. It can either lose the 2 electrons in the 4s level to form $Fe^{2+}$, or it can lose those two electrons in addition to one electron from the $3d$ level to produce $Fe^{3+}$. Since atoms of the inert gases already have completely filled outermost shells, they are assigned a valency of 0.

Since the hydrogen atom contains only 1 electron, it presents a special case of group I elements. To achieve the most stable state, the hydrogen atom can gain 1 electron to have a filled K shell. This can be achieved by 2 hydrogen atoms sharing their electrons to form a hydrogen molecule ($H_2$), giving rise to what is known as the hydrogen bond. Also, it is possible that hydrogen loses its electron to another atom such as Cl to form HCl.

An argument similar to that described above to explain the origin of chemical properties can also be applied to explain the origin of interatomic binding forces holding atoms together in a material. Interatomic binding forces originate from the fact that individual atoms of elements other than the inert gases are unstable because of their partially filled outermost electron shells. As a result, when atoms are brought into close contact, natural forces are developed to hold them together. By natural is meant that these forces result in a net decrease in free energy when individual atoms are grouped together as described below.

### 7.7 Origin of Interatomic Binding Forces

It is possible to understand the origin of interatomic binding forces by considering the reaction between Na and Cl to form NaCl, described in the previous section. Naturally, Na tends to become a positive ion by losing its 3s electron, and Cl tends to become a negative ion by gaining an electron. When Na and Cl atoms are brought into close contact, a bond is established by the attractive force between the positive and negative ions. In the presence of this bond, both the $Na^+$ and $Cl^-$ ions have completely filled outermost electron shells, as explained in the previous section, corresponding to the most stable state. Although there are various types of interatomic binding forces as explained in the next section, a common feature of all these forces is that they involve attractive interaction between positive and negative electric charges when atoms are brought into close contact.

It can be concluded from the above discussion that the origin of interatomic binding forces lies in the natural tendency of individual atoms to seek the most stable state corresponding to outermost shells completely filled with electrons. This is the same principle governing the origin of chemical properties, explained earlier. As a result, chemical elements naturally exist in the form of compounds such as oxides, sulfides, and chlorides. Evidently, the reader may wonder why metals extracted from their ores by various processes do exist. It is recalled from Chap. 2 that although thermodynamics predicts that any system should
seek the most stable state corresponding to the minimum possible free energy, systems frequently exist in a metastable state. Metals extracted from their ores present a typical example. A technologically important consequence of their metastability is that metals have a natural tendency to react with other species in the environment to form oxides, sulfides, chlorides, etc. We call this natural tendency corrosion, which is an important cause of failure of many products. Basically, corrosion is a chemical reaction between a material and other species in its environment to achieve a more stable state. Depending upon the mechanism by which individual atoms are held together in a material, various types of interatomic binding forces are distinguished as described below.

7.8 Types of Interatomic Binding Forces

Interatomic binding forces which hold atoms together in a material into two main types: (i) primary, or large, forces and (ii) secondary, or small, forces. Atoms of most materials are predominantly held by one type of force; however, it is possible to find more than one type of force in a given material. Regardless of the type of interatomic binding force, it can be visualized that forces of attractive nature are developed between positive and negative charges when atoms are brought into close contact. One type of such an attractive force between sodium and chlorine ions has been elucidated in the previous section (ionic bond). Primary interatomic binding forces result in strong bonds, which include (i) the ionic bond, (ii) the metallic bond, and (iii) the covalent bond. Secondary bonds include the hydrogen and Van der Waals bonds.

7.8.1 Ionic Bond

An ionic bond is developed between strongly electropositive elements (metals) and strongly electronegative elements (nonmetals), as typified by the case of Na and Cl described earlier. To illustrate how an interatomic binding force develops between Na and Cl, consider the change in energy of a thermodynamic system consisting of a single Na atom and a single Cl atom as the two atoms are brought into close contact. Assume that the Na atom is placed at the origin of an x-y coordinate system such that the x axis represents the position of the Cl atom, and the y axis represents energy as shown in Fig. 7.3a. In this case, the energy of the system corresponds to potential energy.

It is recalled that the Na atom (1s² 2s² 2p⁶ 3s¹) has a natural tendency to lose the electron in the 3s level and become a positive ion containing 8 electrons in its outermost shell. Conversely, the Cl atom (1s² 2s² 2p⁶ 3s² 3p⁵) has a natural tendency to gain 1 electron and become a negative ion containing 8 electrons in its outermost shell. When the Cl atom is moved along the x axis to become close enough to the Na atom, as shown in Fig. 7.3a, it captures the 3s electron of the Na
Figure 7.3  Effect of interatomic distance on the potential energy of an ionic bond developed between one chlorine atom and one sodium atom. (a) Variation of repulsive and attractive energies with distance. (b) Variation of net potential energy with distance.
atom into its $3p^5$ level. As a result, the Na atom becomes a positive ion, and the Cl atom becomes a negative ion. This develops an electrostatic attractive force between the two oppositely charged ions. Corresponding to this force is an attractive energy, which is considered negative (Fig. 7.3a). With continued decrease in distance between the two ions, the attractive energy becomes more negative. However, when the two ions become too close to each other, their completely filled shells begin to overlap. This develops a repulsive force and corresponding repulsive energy, which is considered to be positive (Fig. 7.3a). Evidently, this repulsive force tends to push the two ions apart.

It follows from the discussion above that the total change in potential energy of the system as a function of distance between the two ions is the sum of attractive and repulsive energy. Since the repulsive energy increases more rapidly with decreasing distance than the attractive energy, the change in potential energy at a small distance is dominated by the increase in repulsive energy. Equilibrium is established at a critical distance $r_0$ between the two ions where the total potential energy (attractive + repulsive) reaches a minimum value, as shown in Fig. 7.3b. By definition, the distance $r_0$ is called the inter-atomic distance. It is possible to correlate the net energy $U_0$ in Fig. 7.3b with the enthalpy of formation of sodium chloride (NaCl). Evidently, the strength of the bond is directly correlated with the value of $U_0$. Because the repulsive energy increases more rapidly with distance than the attractive energy, the shape of the potential energy diagram around the equilibrium interatomic distance becomes asymmetric.

7.8.2 Metallic Bond

In a metallic bond, the attractive forces holding a large number of atoms together are developed between a negative electron cloud and positive ions. It is recalled from earlier discussion that atoms of metallic elements are characterized by a small number of electrons in their outermost shells. When metallic atoms are brought into close contact, they can readily lose those electrons to become positive ions and produce completely filled outermost shells. Electrons lost by a large number of atoms can be visualized to form a cloud, as illustrated in Fig. 7.4. Attractive forces are developed between the negatively charged electron cloud and the positively charged positive ions; i.e., the electron cloud acts as a glue holding the positive ions together.

A common feature of the ionic and metallic bonds is that the attractive forces holding the atoms together are nondirectional; i.e., the magnitude of the force is independent of direction or spherically symmetric. This is reflected on the manner in which the atoms are arranged relative to each other in ionic metallic materials, as described later. However, the two bonds significantly differ in
the nature of the negative charge, which is reflected by the properties of the metallic and ionic materials described below.

In the case of an ionic bond, the negative charge can be represented by rigid spheres resulting from ionization of nonmetallic atoms, and therefore the negative and positive ions are directly bonded to each other. Conversely, the negative charge in a metallic bond is represented by a cloud of valency electrons filling the space between the positive ions. Consequently, while the positive ions are directly bonded to the electron cloud, they are indirectly bonded to each other. Therefore, positive ions of a given metal can readily be replaced by those of another metal without disrupting the bond. In practice, this characteristic of the metallic bond permits the mixing of two or more metals to produce a metallic alloy. Furthermore, the positive ions in a metallic bond can readily be displaced relative to each other while maintaining the electron cloud in between, which explains the characteristic ductility of metallic materials. Also, the relative ease by which electrons in the cloud can move explains the relatively high electrical and thermal conductivities of metallic materials.

7.8.3 Covalent Bond

A covalent bond differs from both the ionic and metallic bonds in that the forces holding the atoms together are directional. Typically, this type of bond is established among atoms, which have equal tendency to lose or gain electrons in order to achieve the most stable state particularly elements of group IV of the periodic table. Also, a covalent-type bond is established between two hydrogen atoms to form a hydrogen molecule (H₂). In this case, when the two hydrogen atoms, each with 1 electron in the 1s level, are brought into close contact, the two electrons filling a common K shell become shared between the two atoms. Another
example is that of C, which has 4 electrons in its outermost shell ($1s^2 2s^2 2p^6 3s^2 3p^2$). Individual C atoms share their outermost electrons with neighboring atoms. Other elements of group IV such as Si, Ge, Ga, and Se form partly covalent and partly metallic bond. Transition elements such as Fe, Cr, Mn, Mo, W, etc., which have partially filled inner shells can also have some fraction of a covalent bond.

Similar to a metallic bond, the valence electrons of atoms held together by a covalent bond form a cloud. However, in contrast with a metallic bond, the electron cloud of a covalent bond is localized, giving rise to directional interatomic binding forces.

As in an ionic bond, atoms of different chemical elements can be held together by a covalent bond forming a compound. A typical example is that of methane ($\text{CH}_4$) where one carbon atom ($1s^2 2s^2 2p^6 3s^2 3p^2$) is bonded to four hydrogen atoms ($1s^1$), as shown in Fig. 7.5.

### 7.8.4 Secondary Bonds

It is recalled that the inert gases have outermost shells completely filled with electrons. Evidently, when an inert gas is made to solidify at sufficiently low temperatures, the atoms lack valency electrons to participate in forming a bond. A similar situation arises in molecules such as $\text{H}_2$, $\text{CH}_4$, CO, CO$_2$, etc., where the requirement of a completely filled outermost shell is satisfied. However, such gases can also be made to solidify at sufficiently low temperatures. Evidently, some type of bond can still be established even when valency electrons are unavailable as described below.

Any atom whose outermost shell is completely filled with electrons consists of a positively charged nucleus electrically balanced by the negative charge of an electron cloud. On an average basis, such an atom is electrically neutral.

![Figure 7.5](image)

**Figure 7.5** A schematic illustration of the covalent bond developed between one carbon atom (C) and four hydrogen atoms (H) in methane (CH$_4$).

Copyright © 2004 by Marcel Dekker, Inc.
Although electrons in the cloud are in constant motion, the center of negative charge may not necessarily coincide with that of the positive charge. This creates what is known as a dipole, producing an electrostatic field capable of exerting a force on other electric charges, as illustrated in Fig. 7.6a. When atoms of an inert gas are brought sufficiently close to each other, the instantaneous dipole created in one atom can induce a dipole in another neighboring atom. As shown in Fig. 7.6b, an attractive force is developed between the positive and negative ends of the two dipoles. Such a bond is called Van der Waals bond and it is evidently much weaker than a primary bond.

Another type of a secondary bond is established with the aid of a hydrogen atom, and therefore it is called the hydrogen bond. It can either be established between a hydrogen atom and another single atom of an electronegative element, or between hydrogen and more than one type of atoms. In the former case, the bond is a directional covalent bond such as that formed between hydrogen and chlorine in HCl. However, in the latter case, which is commonly found in organic solids, the hydrogen atom acts as bridge holding together strongly electronegative atoms.

### 7.9 Bond Strength and Properties of Materials

It can be concluded from the discussion above that a direct correlation exists between the bond strength and the decrease in free energy, resulting from aggregating a group of atoms together. As the decrease in free energy becomes more negative, the bond becomes stronger. An equivalent statement is that the amount of heat required to break the bond and vaporize the material, i.e., dissociate it into its constituent atoms, is a measure of the bond strength. By definition, this heat corresponds to the enthalpy of formation of the material.

![Figure 7.6](image.png)

**Figure 7.6** Development of a van der Waals bond. (a) A schematic illustrating an electrical dipole. (b) Interaction between dipoles in two atoms of an inert gas.
Having developed some basic understanding of the structure of atom and interatomic binding forces, it becomes possible to understand the origin of some material properties. It is recalled that the binding energy holding atoms together is the decrease in potential energy resulting from placing the atoms at positions corresponding to the equilibrium interatomic distance, as explained earlier for the ionic bond. This argument can be applied to any type of bond. Consider the potential energy vs. displacement diagram shown in Fig. 7.7, where one atom A is placed at the origin of coordinates and another atom B is placed at the equilibrium interatomic distance \( r_0 \), and the bond between the two atoms is represented by a spring.

To illustrate the origin of elastic properties of materials, consider that a tensile force \( F \) is applied to the material. As a result, atom B is displaced a distance \( \Delta r \), increasing the equilibrium interatomic distance from \( r_0 \) to \( r_0 + \Delta r \) and stretching the bond by the same distance. Without going into the mathematical details, it suffices to mention here that when \( \Delta r \) is small as it is in elastic deformation, the variation of potential energy \( U \) with displacement \( r \) can be written as

\[
U(r) = U_0 + (dU/dr)r_0 \Delta r + \frac{1}{2}(d^2U/dr^2)r_0 \Delta r^2 \quad (7.11)
\]

Noting that \( dU/dr = 0 \) when \( r = r_0 \), the potential energy corresponding to a displacement \( \Delta r \) is given by

\[
U(\Delta r) = U_0 + \frac{1}{2}(d^2U/dr^2)r_0 \Delta r^2 \quad (7.12)
\]

Figure 7.7 Schematic illustrating the origin of elastic properties.
Since by definition energy $= \text{force} \times \text{displacement}$, the force $F$ is given by

$$F = -\frac{dU}{d(\Delta r)} = -(d^2U/dr^2)r_0 \Delta r$$  \hspace{1cm} (7.13)$$

where the negative sign indicates a negative potential energy. As can be seen from the above equation, when the displacement is very small, the force is linearly proportional to displacement. By definition, the constant of proportionality $d^2U/dr^2$ at $r = r_0$ is the curvature of the potential energy curve at $r = r_0$. If force and displacement are replaced by stress and strain, the result is Hooke’s law and the proportionality constant is the elastic modulus, as described in Chap. 2. As shown in Fig. 7.7, as the potential energy diagram at $r = r_0$ becomes steeper with increasing $U_0$, the proportionality constant and in turn the elastic modulus increases; i.e., a direct correlation exists between the binding energy and elastic modulus—the higher the binding energy, the higher is the elastic modulus.

It is clear from the discussion above that the potential energy of the atoms is increased when they are displaced from the equilibrium interatomic distance. When the applied force is removed, thermodynamics predicts that atom B must return to its equilibrium position lowering the energy of the system, which corresponds to the elastic behavior of materials described in Chap. 2.

Origin of thermal properties of materials can also be understood in terms of the potential energy diagram of Fig. 7.7. At any temperature above 0 K, atoms in a solid vibrate about their equilibrium position. With increasing temperature, the atoms acquire more thermal energy, increasing the amplitude of vibration, i.e., the extent of displacement from the equilibrium interatomic distance. Since the shape of the potential energy diagram is asymmetric, the atoms tend to move apart with increasing temperature, leading to the phenomenon of thermal expansion. It can be expected from Fig. 7.7 that the extent of this displacement is dependent upon the shape of the potential energy diagram. If the binding energy is small, i.e., a weaker bond, the diagram becomes shallow and the extent of displacement by thermal energy is relatively large. Conversely, if the binding energy is large, the diagram becomes steeper, limiting the extent of displacement to a smaller value. Table 7.1 summarizes some properties of selected materials illustrating their dependence on bond strength.

<table>
<thead>
<tr>
<th>Table 7.1</th>
<th>Properties of Selected Materials Illustrating the Effect of Bond Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>Type of bond</td>
</tr>
<tr>
<td>Diamond</td>
<td>Covalent</td>
</tr>
<tr>
<td>NaCl</td>
<td>Ionic</td>
</tr>
<tr>
<td>Fe</td>
<td>Metallic</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
7.10 Arrangement of Atoms in Perfect Crystals

A direct correlation exists between the nature of interatomic binding forces and the manner in which a large number of atoms are arranged relative to each other. It is recalled from earlier discussions that a common feature of the ionic and metallic bonds is that the attractive forces between the negative and positive charges are nondirectional. Since in an ionic material, the negative charge is in the form of rigid spherical ions, electrical neutrality must be preserved in arranging positive and negative ions; i.e., positive and negative ions must alternate in a symmetrical fashion. Conversely, in a metallic material where the negative charge is in the form of an electron cloud filling the space between positive ions, electrical neutrality is always preserved. As a result, the only rule governing the arrangement of positive ions in a metallic material is only geometric. If the attractive force is predominantly non directional, the positive ions are arranged in a symmetrical fashion such that they occupy the least possible volume, i.e., as closely packed as possible. However, if the attractive forces contain a non directional component as in the case of transition elements, the arrangement becomes not as closely packed.

Since the atoms of a metallic material are arranged in a symmetric fashion, they are said to be crystalline. It is evident that the crystalline nature of metallic materials is a direct consequence of the nature of metallic bond. By definition, a crystal is a three-dimensional network of atoms, groups of atoms, or molecules arranged in an orderly pattern, which periodically repeats itself.

Having recognized that metallic materials are crystalline, all possible arrangements of atoms in crystal can be determined in terms of a useful concept known as the space lattice. By definition, the space lattice is a three-dimensional construction consisting of three sets of nonparallel lines, as shown in Fig. 7.8a. Each set consists of equally spaced lines parallel to one the coordinate axes \( x \), \( y \), and \( z \). Intersections of the three sets of lines define points of a space lattice. It can readily be shown from Fig. 7.8a that each lattice point has identical surroundings; i.e., each point has the same number of nearest neighboring points located at the same distances. To completely describe the geometry of a space lattice, it is sufficient to refer to the smallest lattice unit, called unit cell, which maintains the geometry of the entire lattice (Fig. 7.8b). If the axial lengths of the unit cell along the three-coordinate axes \( x \), \( y \), and \( z \) are \( a \), \( b \), and \( c \), respectively, it is evident that the entire lattice can be generated from the unit cell by translating it equivalent distances along the three axes.

An identification of the geometry of unit cell such as that shown in Fig. 7.8b requires knowledge of its axial lengths \( a \), \( b \), and \( c \), as well as its axial angles \( \alpha \), \( \beta \), and \( \gamma \). For example, if \( a = b = c \) and \( \alpha = \beta = \gamma = 90^\circ \), the lattice is cubic. Collectively, these axial parameters are called lattice parameters or lattice constants. By considering all possible variations of axial
relationships, only seven geometries of a space lattice are recognized, defining the seven crystal systems listed in Table 7.2.

For the space lattice shown in Fig. 7.8, lattice points are located only at the corners of the unit cell. A unit cell of this type is called simple or primitive.

Table 7.2 Seven Crystal Systems

<table>
<thead>
<tr>
<th>Crystal system</th>
<th>Axial relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic</td>
<td>$a = b = c$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>$a = b \neq c$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>$a \neq b \neq c$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>$a = b \neq c$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 120^\circ, \beta = \gamma = 90^\circ$</td>
</tr>
<tr>
<td>Rhombohedral</td>
<td>$a = b = c$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = \beta = \gamma \neq 90^\circ$</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>$a \neq b \neq c$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = \gamma = 90^\circ \neq \beta$</td>
</tr>
<tr>
<td>Triclinic</td>
<td>$a \neq b \neq c$</td>
</tr>
<tr>
<td></td>
<td>$\alpha \neq \beta \neq \gamma \neq 90^\circ$</td>
</tr>
</tbody>
</table>
However, it is also possible to place additional lattice points and still maintain the requirement of identical surroundings described earlier. Locations of such additional lattice points include (i) body center of the unit cell, (ii) centers of two opposite faces, or (iii) center of each face defining body-centered, base-centered, and face-centered unit cells, respectively. Therefore, there are four types of unit cells, as shown in Fig. 7.9. By considering all possible arrangements of lattice points of identical surroundings with each of the seven crystal systems listed in Table 7.2, only 14 space lattices are identified, as summarized in Table 7.3.

A perfect crystal can be envisioned to be created from a space lattice by placing atoms, groups of atoms, or molecules at all lattice points. At these positions, the atoms are in their lowest energy states, and the distance between atoms is the equilibrium interatomic distance. In practice, perfect crystals never exist. However, the concept of a perfect crystal can be used to illustrate important geometrical features of crystalline materials, as explained below. Also, it can be used to understand the origin of the mechanical strength of real crystals.

**Figure 7.9** The four types of unit cell. (a) Simple (primitive). (b) Body centered. (c) Base centered. (d) Face centered.
As pointed out earlier, the nature of the metallic bond dictates that atoms of a metallic material are arranged as closely packed as possible. Experiments show that crystals of most metallic materials assume either the face-centered cubic, body-centered cubic, or hexagonal close-packed structures illustrated in Fig. 7.10. Geometrically, the face-centered cubic (fcc) and hexagonal close-packed (hcp) crystals are equivalent in terms of the degree of close packing. Both structures are built from close-packed layers of atoms such as that shown in Fig. 7.10. However, they differ in the stacking sequence of the close-packed layer, as shown in Fig. 7.11. An hcp crystal is characterized by a stacking sequence of the type ABABA... i.e., the third layer lies exactly above the first layer. In contrast, the stacking sequence of an fcc crystal is ABCABCA..., i.e., the fourth layer lies exactly above the first layer. Such a difference in stacking sequence has a significant effect on the mechanical properties of hcp and fcc crystals, as shown later. Examples of fcc metals are Na, Cu, Ni, Ag, and Au. Some of the hcp metals include Zn and Mg.

Body-centered cubic (bcc) crystals are not as close packed as the hcp and fcc crystals. Sometimes bcc crystals are said to have an open structure. Arrangement of atoms in a bcc crystal is a direct consequence of the presence of a directional or covalent component of the interatomic binding force. Consequently, the bcc structure is found in transition metals such as Cr, Mo, and W.

Interpretation of some experimental data related to the crystalline nature of materials, such as those derived from X-ray diffraction experiments as well as understanding certain physical phenomena responsible for many of the useful

<table>
<thead>
<tr>
<th>Crystal system</th>
<th>Type of unit cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic</td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>Body centered</td>
</tr>
<tr>
<td></td>
<td>Face centered</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>Body centered</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>Body centered</td>
</tr>
<tr>
<td></td>
<td>Base centered</td>
</tr>
<tr>
<td></td>
<td>Face centered</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>Simple</td>
</tr>
<tr>
<td>Rhombohedral</td>
<td>Simple</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>Base centered</td>
</tr>
<tr>
<td>Triclinic</td>
<td>Simple</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
properties of these materials, requires a knowledge of the orientations of planes and directions within the crystal. A convenient means for describing those orientations is the Miller index system described below.

Figure 7.12 illustrates the basis for the Miller index system. Consider the hatched face of the unit cell shown in Fig. 7.12a. To determine its indices, its
intercepts along the three coordinate axes (x, y, and z) are determined in multiples or fractions of the respective axial lengths (a, b, and c). Along the x axis, the intercept of the plane can be seen to be \(1a\), but it is parallel to both the y and z axes, and therefore its respective intercepts are \(b\) and \(c\). Dividing the intercepts by the respective axial lengths results in \(1\), \(\infty\), and \(\infty\). To avoid the appearance of \(\infty\) in the final indices, the reciprocals of the intercepts are taken to become \((hkl)\), where \(h = 1\), \(k = 0\), and \(l = 0\). Finally, the reciprocals of the intercepts are

\[
\begin{align*}
\text{Figure 7.11} & \quad \text{Stacking sequence of close-packed layers in the hcp structure (ABAB . . .) and fcc structure (ABCABC . . .).}
\end{align*}
\]

\[
\begin{align*}
\text{Figure 7.12} & \quad \text{An illustration of the Miller index system used to describe the orientation of planes and directions in a crystal. (a) Orientation of planes. (b) Orientation of directions.}
\end{align*}
\]
reduced to the smallest integers which remain to be 100 in this example and then enclosed in parentheses to read (100).

Since the choice of origin of coordinates to determine indices of crystal planes is arbitrary, it is evident that all the planes parallel to that shown in Fig. 7.12a are equally spaced and have the same indices. Therefore, the indices (100) describe a set of equally spaced parallel planes of the same atoms arrangement. It is evident that the indices of the other faces of the unit cell are (010) and (001), as shown in Fig. 7.12a. Since the arrangements of atoms on all three faces are the same, they are collectively referred to as a family of planes. Indices of one member of the family can be used to describe the entire family. To distinguish a family of planes from a set, the family is expressed as \( \{100\}, \{010\}, \text{or} \{001\} \). A negative intercept is indicated by a bar above the respective number.

Indices of a direction are determined from the coordinates of a point along a line drawn from the origin parallel to that direction. Similar to the case of planes, the coordinates of a point are determined in multiples or fractions of axial lengths and then directly reduced to the smallest integers, resulting in the indices of the direction \([uvw]\). Examples are illustrated in Fig. 7.12b. A family of directions is expressed as \( \langle uvw \rangle \). To maintain similar indices for planes of identical atoms arrangements in hexagonal crystals, a four-rather than three-index system is used. If in the three-index system described above, the indices of a plane are \((hkl)\), they become \((hkil)\) in the four-index system such that \(h + k + i = 0\) or \(i = -(h + k)\). Also, if the three indices of a direction are \([UVW]\), they become \([uvw]\) in the four-index system, where

\[
\begin{align*}
u &= (2U - V)/3 \\
v &= (2V - U)/3 \\
t &= -(U + V)/3 \\
w &= W
\end{align*}
\]  

It is recalled from earlier discussion that elastic properties of materials can be understood in terms of the interatomic binding forces which hold atoms together. Atoms arrangement in crystals, however, can be used to elucidate some important features of the plastic properties of crystals described below.

### 7.11 Understanding the Microscopic Plasticity of Perfect Crystals

It is recalled that plastic deformation is a permanent change in dimensions or shape of a material resulting from the application of a stress exceeding the yield strength of a given material. When the surface of a plastically deformed material as in a tensile test is examined by means of a microscope, a series of
markings or steps are observed to run continuously around the surface of the specimen, as schematically illustrated in Fig. 7.13. Such steps can only result if layers of the material are slipped or sheared relative to each other under the influence of the applied stress. Since plastic deformation of crystalline materials preserves the arrangement of atoms in a crystal, it must involve displacement of atoms from one equilibrium position into another under the influence of a shear stress. Using the concept of a perfect crystal, this stress can be calculated as follows.

Consider two layers of atoms as shown in Fig. 7.14 such that the interatomic distance is \( b \) and the interplanar spacing is \( d \). Under the influence of a shear stress \( \tau \), atoms in a row of the top layer are consecutively displaced from their equilibrium positions. A displacement \( x \) of an atom such as \( A \) from its equilibrium position causes an increase in the free energy of the system requiring the shear stress to increase with displacement. However, when the displacement becomes equivalent to \( b/2 \), atom \( A \) becomes in an unstable position. It spontaneously moves into another equilibrium position \( B \). At this position, the shear stress falls to 0. Evidently, the shear stress reaches a maximum value between \( x = 0 \) and \( x = b/2 \). This variation can reasonably be approximated by a sinusoidal function of period \( b \), i.e.,

\[
\tau = K \sin \frac{2\pi x}{b}
\]  

(7.15)

where \( \tau \) is the shear stress and \( K \) is a constant. For a very small displacement, the above equation can be written as

\[
\tau \approx 2\pi Kx/b
\]  

(7.16)

Figure 7.13 A schematic illustration of surface steps resulting from plastic deformation.
It is recalled that for a very small strain corresponding to a very small displacement, the shear stress can be expressed as

\[ \tau = \frac{Gx}{d} \]  

where \( G \) is the shear modulus. Combining the above two equations results in

\[ K = \frac{Gb}{2\pi d} \]

and therefore

\[ \tau = \frac{Gb}{2\pi d} \sin 2\pi x/d \]

Figure 7.14  A schematic illustrating the basis for estimating the theoretical shear strength of a crystal.

It can be concluded from the above equation that the shear stress required to cause plastic deformation is directly proportional to \( b \) (interatomic distance) and inversely proportional to \( d \) (interplanar spacing); i.e., it requires less shear stress to cause plastic deformation along directions where \( b \) has the smallest value (close-packed direction) on planes of larger \( d \) spacing. Such two requirements are satisfied by crystal planes of closest packing of atoms. Therefore, plastic
deformation is energetically more favorable if it occurs by shearing of close-packed planes along close-packed directions.

In practice, plastic deformation of crystalline materials is indeed found to occur more favorably by shearing of close-packed planes along close-packed directions. For a face-centered cubic crystal, there are four sets of close-packed planes with indices of (111), (1̅11), (1̅1̅1), and (111). Using the terminology described earlier, the family of close-packed planes is expressed as {111}. It can readily be shown from Fig. 7.11 that there are six close-packed directions in a face-centered cubic crystal with indices of [110], [1̅10], [101], [1̅01], [011], and [0̅1̅1]; i.e., the close-packed directions are of the type ⟨110⟩. Each of the four close-packed {111} planes contains three ⟨110⟩ close-packed directions. A combination of a close-packed plane and a close-packed direction is called a slip system. Therefore, an fcc crystal has 12 slip systems. For a hexagonal close-packed crystal, there is only one set of close-packed planes, namely the ⟨0001⟩, and three close-packed directions of the type ⟨2̅1̅00⟩; i.e., an hcp crystal has only 3 slip systems. In the case of a body-centered cubic crystal (Fig. 2.5b), there are six close-packed planes of the type ⟨110⟩ and four close-packed directions of the type ⟨111⟩. Each of the six close-packed planes contains two close-packed directions. However, those planes are not as close packed as those of an fcc or hcp crystal.

Because of the close-packed nature of fcc crystals and their relatively large number of close-packed slip systems, engineering alloys having fcc structure are more ductile than hcp and bcc materials. Materials having hcp structures have very limited ductility because of the rather small number of slip systems. Ductility intermediate between that of fcc and hcp crystals is found in bcc crystals.

Although some of the important features of plastic deformation of crystalline materials can be elucidated from a consideration of a perfect crystal, the stress required to cause plastic deformation of real crystals is found to be smaller by a factor of about 10^5 than that calculated from equation. Such a large discrepancy can be accounted for by the fact that real crystals are far from being perfect. As shown below, real crystals contain various types of imperfections. One of those imperfections is responsible for the observed large discrepancy between the theoretical and observed shear strengths.

### 7.12 Crystal Imperfections

It is recalled that a perfect crystal is constructed from a space lattice by placing atoms, groups of atoms, or molecules at all lattice points such that each atom, groups of atoms, or molecules have identical surroundings. Real crystals, however, are far from being perfect. They contain various types of imperfections in the arrangement of atoms, which can significantly influence their properties.
A real crystal can be visualized to consist of a number of imperfections superimposed on a perfect crystal, i.e.,

Real crystal = perfect crystal + imperfections

In terms of their geometrical characteristics, crystal imperfections can be classified into (i) point imperfections (dimensionless), (ii) line imperfections (one-dimensional), (iii) surface or planar imperfections (two-dimensional), and (iv) volume or bulk imperfections (three-dimensional). An account of each type of imperfections and its influence on the properties of crystals is given below.

### 7.12.1 Point Imperfections

As their name implies, point imperfections are associated with lattice points. Various types of point imperfections include (i) vacant lattice sites or shortly vacancies, (ii) substitutional imperfections, and (iii) interstitial imperfections.

Vacancies are natural consequence of the fact that crystals exist and are used at temperatures above 0 K. It is recalled from the Gibbs free energy equation

\[ G = H - TS \]

that the free energy \( G \) of a system is dominated by the entropy \( S \) as the temperature increases. At any temperature above 0 K, the atoms are in a constant state of thermal vibration. Since entropy is a measure of the randomness of the system, as the temperature increases, the instantaneous positions of atoms in a crystal become increasingly disorganized.

By definition, a vacancy is a lattice point not occupied by an atom as a result of thermal energy. It is possible to visualize how a vacancy is created, as schematically illustrated in Fig. 7.15. As shown in the next section, atoms at the external surface of a crystal have higher energy in comparison with those in the interior of the crystal. By means of thermal energy, an atom such as A at the surface of the crystal can jump into a position such as that shown in Fig. 7.15, leaving behind a vacant lattice site. It is possible to fill the surface

![Figure 7.15](image-url)
As a result a vacancy is created, which wanders at random throughout the crystal. Since vacancies increase the entropy of the crystal, a certain number of vacancies dependent upon temperature can exist in equilibrium within the crystal. If the total number of lattice sites is $N$, it can readily be shown that the number of vacancies $N_v$ existing in equilibrium is given by

$$N_v = N \exp \left( -\frac{\Delta G}{RT} \right)$$

(7.19)

corresponding to minimum free energy, where $\Delta G$ is the free energy of formation of a mole of vacancies, $R$ is the universal gas constant, and $T$ is the temperature in degrees Kelvin. Evidently, the concentration of vacancies is directly proportional to temperature.

A large number of vacancies can cluster to form other defects. For example, if a crystal is cooled rapidly from high temperature as in some heat treatment processes encountered in practice, there is insufficient time for vacancies to reach the surface and be annihilated. In this case, a large number of vacancies can cluster to form a void (bulk imperfection). Also, vacancies can cluster to form line and planar imperfections of the types described later in this section.

Substitutional imperfections are foreign atoms present in the crystal either as an inevitable impurity or intentionally introduced (alloying elements) to promote certain properties of engineering alloys. It is recalled that the metallic bond is rather flexible in the sense that one type of atoms can readily be replaced by another leading to the formation of a solid solution. By definition, substitutional atoms of element B occupy normal sites in the lattice of element A, as shown in Fig. 7.16a. Because the size of the substitutional atom is different from that of the host atoms, the lattice becomes strained giving rise to the concept of

---

**Figure 7.16** Schematics illustrating substitutional point defects. (a) Random solid solution. (b) Ordered solid solution.
strengthening by solid solution. In engineering alloys, if the substitutional atoms are randomly distributed within the lattice as shown in Fig. 7.16a, the solid solution is said to be random or disordered. Conversely, if there is a tendency for substitutional atoms B to have atoms A as their nearest neighbors, as shown in Fig. 7.16b, the solid solution is said to be ordered.

Interstitial imperfections are of two types, as shown in Fig. 7.17. An atom occupying a normal site can be forced to occupy an interstitial site (see Fig. 7.17a) by severe plastic deformation. Also, smaller atoms present as impurities or alloying elements can occupy interstitial sites as shown in Fig. 7.17b. A classical example is that of carbon in steels.

### 7.12.2 Line Imperfections: Dislocations

Only one type of line imperfection exists in crystals, namely, dislocations. It is of extreme importance to understand the mechanical properties of real crystals, as well as other phenomena. As explained earlier, the shear stress required to initiate plastic deformation in a perfect crystal is found to be greater by a factor of about $10^4$ in comparison with real crystals. It is recalled that this shear stress is estimated on the assumption that plastic deformation occurs by shearing or displacement of entire atomic planes relative to each other. Real crystals are found to contain dislocations, which can be visualized as shear nuclei on the slip planes. Dislocations can directly be observed by means of transmission electron microscopy, as shown in Chap. 8. During plastic deformation, slip occurs by motion of dislocation across the slip plane rather than by shearing an entire

![Interstitial Atom](image)

**Figure 7.17** Schematics illustrating interstitial point defects. (a) Self interstitial atom. (b) Foreign interstitial atom.

Copyright © 2004 by Marcel Dekker, Inc.
atomic plane, as illustrated later. First, some geometrical aspects of dislocations are described.

Atomic planes in real crystals are partially slipped or sheared relative to each other on the close-packed planes and along close-packed directions as a result of the presence of dislocations. By definition, a dislocation is the line on the slip plane separating the slipped and unslipped areas, and the shear displacement associated with a dislocation is called its Burgers vector \( b \). Dislocation density in a crystal is expressed in terms of the length of dislocation per unit volume, i.e., \( \text{cm/cm}^3 \) or \( \text{cm}^{-2} \). In well-annealed metallic crystals, the density of dislocation is on the order of \( 10^6 \text{ cm/cm}^3 \).

If the Burgers vector of a dislocation connects atoms of identical surroundings, the dislocation is said to be total dislocation. Its motion through the crystal maintains the same atoms arrangement. For fcc crystals, the Burgers vector of a total dislocation is of the type \( 1/2(110) \), and it is of the type \( 1/2(111) \) for bcc crystals. In case of hcp crystals, the Burgers vector of a total dislocation is of the type \( 1/3(1120) \). If the Burgers vector does not connect atoms of identical surroundings, the dislocation is said to be partial. Partial dislocations are associated with planar imperfections. In this section, the discussion is limited to total dislocations. Partial dislocations related to planar imperfections are described in the next section.

Depending upon the direction of the Burgers vector relative to the dislocation line, two extreme orientations of dislocation are distinguished: (i) edge orientation, where the Burgers vector is perpendicular to the dislocation line, and (ii) screw orientation, where the Burgers vector is parallel to the dislocation line as illustrated in Fig. 7.18. Since the dislocation line is the boundary between slipped and unslipped areas, a dislocation line cannot end abruptly in a crystal. As a boundary of a surface, the dislocation line must either form a closed loop or end at another surface, e.g., the external surface of the crystal or an internal surface such as a grain boundary. A dislocation loop such as that shown in Fig. 7.19 is of a mixed character, i.e., it is partly edge and partly screw.

Figure 7.20 illustrates the atoms arrangement in the vicinity of an edge dislocation. It is possible to visualize that an edge dislocation is introduced into the crystal by making a cut normal to the slip plane and then inserting an extra half plane of atoms, as shown in Fig. 7.20. By definition, the dislocation line corresponds to the edge of the extra half plane of atoms. An edge dislocation is represented by the symbol \( \perp \) or \( \top \). If the extra half plane of atoms lies above the slip plane, as shown in Fig. 7.20, the dislocation is represented by \( \perp \) and is said to be positive. Conversely, if the extra half plane of atoms lies below the slip plane, the dislocation is represented by \( \top \) and is said to be negative. In the vicinity of dislocation, the atoms are displaced from their normal lattice sites and this displacement decreases with distance from the dislocation. As a result there is an elastic strain energy associated with a dislocation, which can readily be shown to be proportional to \( b^2 \), where \( b \) is its Burgers vector. It can
be appreciated from Fig. 7.20 that the atomic planes above a positive edge dislocation are compressed and those below the dislocation are stretched.

In a screw dislocation, the atomic planes in the vicinity of the dislocation line form a spiral arrangement as in helical ramp with the dislocation line at its center. If the helical ramp advances one atomic plane when a clockwise circuit

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure}
\caption{A schematic illustrating the mixed character of a dislocation loop of a Burgers vector $b$.}
\end{figure}
is made around the dislocation, it is said to be right-handed or positive, and if the same observation is made for an anticlockwise circuit, the dislocation is said to be left-handed or negative.

Because dislocations increase the free energy of the crystal, they are classified as nonequilibrium imperfections unlike vacancies and impurity atoms. Therefore, real crystals containing dislocations exist in a state of metastable equilibrium. Consequently, there is always a natural tendency for dislocations to assume configurations of lower energy by various processes as explained later in section.

To illustrate the consequence of the presence of dislocations in real crystals on the process of plastic deformation, consider a perfect crystal and a real crystal of the same material containing an edge dislocation of Burgers vector \( \mathbf{b} \). Under the influence of a shear stress, it is required to displace the upper part of each crystal relative to the lower part by an amount \( \mathbf{b} \) on the slip plane. In the case of the perfect crystal, the entire atomic planes of the upper part must be displaced by a distance \( \mathbf{b} \), as shown in Fig. 7.21a. However, the same shear displacement can be produced in the real crystal as a result of moving the dislocation across the slip plane by only slight adjustments of atoms corresponding to a fraction of the atomic distance \( \mathbf{b} \), as shown in Fig. 7.21b. When the dislocation sweeps the entire area of the slip plane, it is annihilated at the surface producing a surface step corresponding to a slip line. Because of the smaller atoms displacement involved in moving a dislocation, it becomes clear why the shear stress required to plastically deform real crystals is much less in comparison with that of a perfect crystal.

Figure 7.20  Arrangement of atoms and nature of strain in the vicinity of a positive edge dislocation represented by the \( \perp \) symbol.
Dislocation motion is generally classified into (i) conservative and (ii) non-conservative. By conservative is meant that atoms need not to be transported into or removed from the vicinity of a dislocation as it moves. An example of a conservative motion is shown in Fig. 7.21b. Conservative motion of a single dislocation is termed glide. It is evident that a measurable macroscopic strain requires that a large number of dislocations glide across the slip plane. Such a process is referred to as a slip. In contrast with conservative motion, nonconservative motion requires transport of atoms into or removal of atoms from the vicinity of a dislocation as it moves. For example, motion of a positive edge dislocation by one interatomic distance below the slip plane requires transport of interstitial atoms into the dislocation, as shown in Fig. 7.22a. Conversely, if vacancies are transported into the dislocation, it moves one interatomic distance above

Figure 7.21  Effect of dislocations on plastic deformation. (a) Deformation of a perfect crystal free of dislocations. (b) Deformation of a real crystal containing a dislocation ($\tau_2 \ll \tau_1$).
the slip plane (Fig. 7.22b). Since in both cases the dislocation moves out of its slip plane, the process shown in Fig. 7.22 is called climb. Evidently, climb is a thermally activated process and therefore it becomes an important mode of dislocation motion during high-temperature deformation processes such as creep.

If a unit vector $\mathbf{e}$ is assigned to a dislocation line, the normal to the slip plane $n$ is given by the cross product $n = \mathbf{e} \times b$. For an edge dislocation where $b$ is normal to $\mathbf{e}$, the glide plane is uniquely defined; i.e., an edge dislocation is confined to move by glide on its original slip plane. Conversely, since $b$ is parallel to $\mathbf{e}$ for a screw dislocation, it can move by glide on any slip plane containing its Burgers vector; i.e., a screw dislocation moving on a given slip plane by glide can continue its motion on another slip plane provided it contains its Burgers vector. Such a process is called cross slip.

It is evident from the discussion above that motion of an edge dislocation by glide across a slip plane causes a translational displacement of the atomic
planes above the slip plane relative to those below the slip plane. In contrast, motion of a screw dislocation causes a rotational displacement of the slip planes. Therefore, when a general dislocation of a mixed character moves, it causes translational and rotational displacement of the slip planes.

There are various processes by which dislocations are nucleated in real crystals. One of these processes is encountered during solidification where dendrites are nucleated in the molten metal, as described earlier. When dendrites of different atoms orientation are adjoined, the mismatch is accommodated by forming dislocations. Another process is condensation of vacancies during cooling from high temperatures as encountered in heat treatment processes. Furthermore, during cooling from high temperatures, the local stresses generated because of the differences in thermal expansion characteristics of inclusions or secondary phase particles and the surrounding matrix can be accommodated by forming dislocations at the particle-matrix interface.

In an annealed metallic material, dislocations exist in the form of a three-dimensional network intersecting the slip planes. On a given slip plane, there exists a certain number of dislocations capable of moving and producing plastic deformation under the influence of a shear stress. When these dislocations move, they are encountered by other dislocations intersecting the slip plane as well as by other imperfections such as impurity atoms, twin boundaries, grain boundaries, and particles of secondary phases with which they interact. Dislocation mobility is strongly influenced by this interaction, which is of extreme importance in understanding many processes related to plastic deformation, strain hardening, and fracture. Some of the important results of these interactions include (i) immobilization of moving dislocations, (ii) increasing the density of dislocations, and (iii) nucleating cracks.

Any obstacle to moving dislocations, including other dislocations as well as various types of imperfections, increases the stress required to continue plastic deformation. In effect, two important mechanical properties are controlled by dislocations in an opposite manner: (i) ductility requiring an increased dislocation mobility and (ii) strength requiring a decreased dislocation mobility. It is possible to understand many of the important results of interaction between a moving dislocation and other imperfections in terms of the atoms arrangement in the vicinity of an edge dislocation, as described below.

Interaction between a moving dislocation and another dislocation is governed by three parameters: (i) stress fields of the two dislocations, (ii) separation of the two dislocations, and (iii) the natural tendency to lower the energy of dislocations. In considering this interaction, two cases are distinguished: (i) parallel dislocations and (ii) nonparallel dislocations. When two parallel edge dislocation of opposite sign approach each other, their stress fields tend to cancel each other, lowering the energy of the crystal, and therefore an attractive force is developed between them. If the two dislocations lie on the same slip plane,
they annihilate each other, leaving behind a perfect region of the crystal. However, if they lie on adjacent parallel slip planes, they still annihilate each other but leave behind a row of vacancies or interstitial atoms. Conversely, if the two dislocations have the same sign, the type of force developed between them is dependent upon their separation. If they lie on the same or nearby slip planes, a repulsive force is developed leading to the formation of what is known as a dipole. A dipole consisting of two parallel dislocation lines of the same sign is highly immobile and therefore, it contributes to the phenomenon of strain hardening, as further explained later.

It can readily be shown that if parallel edge dislocations of the same sign lie on widely separated slip planes, an attractive or repulsive force is developed depending upon whether the angle between the slip direction and the line joining the dislocations is greater or less than $45^\circ$, respectively. At an angle of $45^\circ$, a pair of dislocations of opposite sign form a stable arrangement known as a dipole. When the angle becomes $90^\circ$, edge dislocations of the same sign are aligned above each other to form another stable arrangement, called a low-angle tilt boundary. Another type of a low-angle boundary is formed by screw dislocations. Since there are three vectors of the type $\frac{1}{2}(110)$ on a given $\{111\}$ plane in a fcc crystal, three total dislocations in screw orientation can meet at a point called a node. At this point the sum of the respective Burgers vectors vanishes, i.e., $b_1 + b_2 + b_3 = 0$. Therefore, screw dislocations on $\{111\}$ planes of fcc crystals tend to form a two-dimensional hexagonal network.

Low-angle boundaries divide individual grains of the material into subgrains. Since the degree of misorientation between subgrains is smaller than that between the grains, subboundaries are termed low angle, and grain boundaries are referred to as high angle. Usually, the degree of misorientation across low-angle boundaries is less than $10–15^\circ$. High-angle grain boundaries are characterized by a degree of misorientation greater than about $10–15^\circ$. As explained later, low-angle boundaries can form during high-temperature deformation processes, as well as during heat treatment processes following cold-work. Interactions between nonparallel dislocations is of special importance in understanding the relative ease with which slip can take place.

As a general rule, when two nonparallel dislocations intersect each other, each dislocation line acquires a step or a segment of a size equivalent to the Burgers vector of the other dislocation. If the step lies on the slip plane, it is called a kink, and if it lies normal to the slip plane, it is called a jog. A kink acquired by the edge dislocation has a screw character, while that acquired by the screw dislocation has an edge character. Such kinks can easily be eliminated by slip. A jog in an edge dislocation remains to have an edge character. In this case, the dislocation can still move by glide but on a steeped slip plane. Also, a jog in a screw dislocation has an edge character. However, in this case, the jog limits the mobility of dislocation by cross slip, particularly at low temperatures.
recalled that a screw dislocation can move on any slip plane containing its Burgers vector. In contrast, an edge dislocation can only move out of its glide plane by climb.

Because of the characteristic atoms arrangement of an edge dislocation and the associated stress field, there is a natural tendency for the dislocation to attract impurity atoms to lower its strain energy. As shown in Fig. 7.20, there is a tensile stress field associated with the extra half plane of atoms, and a compressive stress field associated with the missing half plane. Smaller substitutional atoms tend to replace larger matrix atoms in the tensile side, while small interstitial atoms tend to segregate to the compressive side. As a result of this interaction, the dislocation is locked in position, which could explain what is known as the yield point phenomena explained later.

Planar imperfections described in the next section such as twin and grain boundaries act as barriers to dislocation motion. It suffices to mention here that because the slip planes are misoriented across a boundary, moving dislocations are held in position by the boundary forming a pile-up. This interaction contributes to the phenomena of strain hardening and can be used to explain the effect of grain size on the yield strength as shown later. Moving dislocations also interact with secondary phase particles; however, the exact mechanism of interaction is dependent upon the nature and strength of these particles, as further explained later.

To summarize the discussion above, plastic deformation in real crystals occurs by motion of dislocations. When dislocations move, they multiply and increase in density, which increases the stress required to continue plastic deformation giving rise to the phenomenon of strain hardening. Any obstacle to dislocation motion increases the stress required to initiate plastic deformation. In conclusion, any method used to strengthen real crystals must involve the introduction of some obstacle to dislocation motion. However, it is important to point out here that dislocations play important roles in other phenomena such as precipitation of secondary phases and phase changes, as explained later.

### 7.12.3 Surface or Planar Imperfections

An obvious imperfection in a real crystal is its external or free surface. In a perfect crystal, each atom has identical surroundings; i.e., each atom has the same number of nearest neighboring atoms located at the same distances. Evidently, atoms at the external surface of a crystal cannot have identical surroundings. As a result, each atom at the external surface is in a higher energy state in comparison with atoms in the interior of the crystal giving rise to surface energy. To lower their energy, surface atoms have a natural tendency to attract other atoms or molecules from the environment. It is for this reason that the external surface of most metallic crystals is covered with an oxide layer resulting from continuing
chemical reactions between surface atoms and the atmosphere, as illustrated in Fig. 7.23. Such a surface oxide layer can have marked effects on many properties.

Another surface defect is a grain boundary, which is the surface separating two adjacent crystals in a polycrystalline material. It is recalled from Chap. 1 (see Fig. 1.5) that a grain boundary is developed because atoms in a given grain are oriented differently from those in its neighbor. Usually, an atom at the boundary cannot simultaneously occupy a lattice site in both crystals. Therefore, atoms at a grain boundary have fewer nearest neighboring atoms lying at equilibrium lattice positions in comparison with an atom in the interior of the crystal. Similar to the case of an external surface, an atom at a grain boundary is in a higher energy state in comparison with an atom within a grain giving rise to a grain boundary energy. However, the energy associated with a grain boundary energy is lower than that of external surface. Energy associated with grain boundaries is a function of relative orientation of the two neighboring grains. Grain boundaries act as catalyst for many processes occurring in materials, such as precipitation of secondary phases and elemental segregation, as shown later. In certain cases, precipitation of secondary phases at grain boundaries can lead to failure.

![Figure 7.23](image.png)  
*Figure 7.23*  Schematic illustrating the tendency of metallic materials to be covered by an oxide layer to lower the energy of surface atoms.
As can be concluded from the above discussion, a grain boundary arises from the misorientation between two neighboring grains. Certain orientation relationships, however, can exist between two neighboring grains where the boundary atoms can simultaneously occupy lattice sites in both crystals. Each boundary atom has the same number and spacing of nearest neighboring atoms as in the interior of the crystal, however, mistakes exist in the arrangement of second nearest neighbors. Such an orientation relationship defines a coherent twin boundary, as shown in Fig. 7.24 where the two crystals separated by the boundary are mirror images of each other. Energy of a twin boundary is about 1 order of magnitude less than the energy of a typical grain boundary. Coherent twin boundaries in face-centered cubic crystals lie along the \{111\} close-packed planes. In this case, the stacking sequence across the boundary is locally changed from

\[
\text{ABCABCABCABCAB}
\]

to

\[
\text{ABCABCABCABCAB}
\]

Figure 7.24  A face-centered cubic crystal viewed along a (110) direction to illustrate a twin on a (111) plane. The two crystals labeled \(T\) have the same orientation but each crystal is in a twin orientation with the crystal labeled \(T_1\).
where the twin boundary lies along the bolded layer (C). It is evident that the stacking of close-packed layers on one side of the boundary is a mirror image to that on the other side (see Fig. 7.24). Also, it is noted that the characteristic stacking sequence of a face-centered cubic crystal is maintained across the boundary. Observation of the morphology of twin boundaries can provide a great deal of information about the mechanical state of the material. For example, observation of a large density of straight twin boundaries as shown later indicates that the material is relatively free of cold work, i.e., fully recrystallized. However, if the boundaries are bent, this indicates that the material is deformed. Twin boundaries in fcc crystals can readily be shown to be created by motion of partial dislocations of Burgers vector $\frac{1}{6}[112]$ on the close-packed $\{111\}$ planes, which are called the twinning planes. In hcp crystals, the most common twinning planes are of the type $\{10\bar{1}\}$; however, other possible twinning planes include $\{10\bar{1}2\}, \{10\bar{1}3\}, \{11\bar{2}1\}$, and $\{11\bar{2}2\}$. Twins are not as frequently observed in bcc crystals as in fcc and hcp crystals. Common twinning planes in bcc crystals are of the type $\{112\}$. Another surface defect closely related to a twin boundary is known as a stacking fault, as shown below.

Essentially, stacking faults are formed as a consequence of the natural tendency of a total dislocation to lower its energy. It is recalled that the Burgers vector of a total dislocation in a fcc crystal is of the type $\frac{1}{2}[110]$. Such a total dislocation can lower its energy by splitting into two partial dislocations having Burgers vectors of the type $\frac{1}{6}[112]$ and separated by a ribbon of a stacking fault such that the partial dislocations form the boundaries of the fault, as shown in Fig. 7.25. To illustrate the physical meaning of a stacking fault, consider the characteristic stacking sequence of a face-centered cubic crystals, i.e.,

$$\text{ABCABCABCABCAB}$$

If the bolded C layer is removed, the stacking sequence is locally changed to become

$$\text{ABCABCABCA}$$

Such a change can be affected by motion of $\frac{1}{6}[112]$ partial dislocation on $\{111\}$ planes. As indicated by the bolded layers (BAB), the stacking sequence is locally changed from that characteristic of a face-centered cubic crystal, i.e., ABCA... to that characteristic of a hexagonal-close-packed crystal, i.e., BAB, in contrast with a twin boundary. Similarly, a stacking fault in a close-packed hexagonal crystal represents a thin layer of a face-centered cubic crystal. Such a surface defect is called an intrinsic stacking fault. Its energy is about twice that of a coherent twin boundary. A similar fault results if instead of removing a close-packed layer, an additional layer is inserted. In this case, the defect is called an extrinsic stacking fault and it has about the same energy as an intrinsic, and the characteristic displacement vector of the fault is of the type $\frac{1}{2}[111]$. By definition, a stacking fault is a very thin layer of a twin, and therefore a correlation exists between the frequency of observing twins...
in the microstructure and stacking fault energy. A high density of twins is usually associated with low stacking fault energy.

Other surface defects include (i) antiphase boundaries and (ii) phase interfaces. Antiphase boundaries are surface imperfections which exist in ordered solid solutions. In an ordered crystal, there is a preference for an atom A to have unlike atoms B as its nearest neighbors. This preference, however, is distributed along an antiphase boundary which contains like atoms A-A and B-B. A phase interface is the surface separating two phases in the material. Since secondary phases in a material can be coherent, partially coherent, or noncoherent, three types of phase interfaces are distinguished.

7.12.4 Volume Imperfections

Volume imperfections are three-dimensional in character. They include secondary phase particles inevitably introduced into the material during processing such
as oxide and sulfide particles. Other volume imperfections include voids and cracks. Voids can be created by condensation of a large number of vacancies. Cracks are initiated by separation of surfaces, particularly at the interfaces separating secondary phase particles from the matrix.

7.13 Understanding the Microscopic Plasticity of Real Crystals

Having recognized that plastic deformation of crystals occurs by motion of dislocations, it becomes possible to explain the microscopic aspects of plasticity and relate these aspects to macroscopic behavior. In discussing the microscopic aspects of plasticity, two cases are distinguished: (i) low-temperature deformation, where diffusion of atoms is insignificant, and (ii) high-temperature deformation, where diffusion of atoms is significant.

7.13.1 Low-Temperature Deformation

Low-temperature deformation behavior can be understood by considering the sequence of events which takes place during a standard tensile test. To illustrate these events, consider the grain structure of a tensile specimen as schematically illustrated in Fig. 7.26. Also shown in Fig. 7.26 are mobile dislocations laying on

![Figure 7.26](image-url)  

**Figure 7.26** Schematics illustrating the procedure for determining resolved shear stress resulting from an applied tensile load.
their respective slip planes. Before those dislocations can move and contribute to plastic strain, the shear stress component of the applied tensile stress must reach the level sufficient to overcome the inherent resistance of the crystal to dislocation motion. This is called the critical resolved shear stress. If the cross-sectional area of the specimen is \( A_0 \), and \( \theta \) is the angle between the direction of the applied tensile load \( P \) and the normal to the slip plane \( n \), the area of the slip plane \( A \) is given by

\[
A = A_0 / \cos \theta
\]

and if the angle between the tensile load \( P \) and the slip direction is \( \phi \), the component of shear force \( P_\tau \) acting along the slip direction is given by

\[
P_\tau = P \cos \phi
\]

Therefore, the resolved shear stress \( \tau \) can be expressed as

\[
\tau = P_\tau / A = P \cos \phi / (A_0 \cos \theta) = P / A_0 \cos \theta \cos \phi
\]

By definition, the applied tensile stress \( \sigma = P / A_0 \), and thus resolved shear stress \( \tau \) can be expressed in terms of the applied tensile stress \( s \) as

\[
\tau = \sigma \cos \theta \cos \phi \quad (7.20)
\]

Dislocations begin to move on their slip plane when \( \tau \) reaches the critical value. Because of the misorientation between neighboring grains, the orientation factor in the above equation (\( \cos \theta \cos \phi \)), which is called the Schmid factor, varies with the orientation of slip planes relative to the applied tensile stress. As a result, dislocation motion is initiated in those most favorably oriented grains where the Schmid factor allows \( \tau \) to reach the critical value. It is important to distinguish between the macroscopic yield strength determined from a tensile test and the critical shear stress component required to initiate dislocation motion. Experiment shows that dislocations begin to move at stresses well below the observed yield strength, which is more precisely defined later in this section.

When dislocation motion is locally initiated as described above, slip occurs in the most favorably oriented grain. It is important to realize that no measurable macroscopic strain is realized at this stage. Since grain boundaries act as strong obstacles to dislocation motion particularly at low temperatures, the leading dislocation in the row shown in Fig. 7.26 is repelled by the grain boundary, causing it to be immobilized. Likewise, the leading dislocation exerts a repulsive force on the next dislocation, and so on, causing the row of dislocations in Fig. 7.26 to be immobilized and forming a pileup at the grain boundary. It can readily be shown that if the resolved shear stress is \( \tau \), the force \( F \) exerted on a dislocation per unit length is given by

\[
F = \tau b \quad (7.21)
\]
where \( b \) is the Burgers vector of the dislocation. If the number of dislocations in the pileup is \( n \), the stress exerted by the leading dislocation on the grain boundary is \( n\tau \). As a result, a localized high level of stress concentration is developed at the grain boundary. It then becomes evident that deformation of the most favorably oriented grain must be accommodated by the neighboring grain. To accommodate this deformation and relieve the high level of stress concentration at the grain boundary, either slip is initiated in the neighboring grain or cracks are developed at the boundary, causing grain separation depending upon the structure of the material and in turn the number of available slip systems.

It can readily be shown that when two differently oriented crystals are adjoined, the misorientation can be accommodated by five independent ways: three ways in which each crystal is tilted or twisted relative to the other crystal and two ways in which the boundary is aligned between the two crystals. As a result, the high level of stress concentration at a grain boundary can be relieved by slip in the neighboring grain if at least five independent slip systems are available. It is recalled that for the closest-packed crystals, there are 12 independent slip systems in fcc crystals, while hcp crystals have only 3 slip systems. Although in the less close-packed bcc crystals, the slip planes are not as well defined as in the close-packed crystals, slip still occurs along the \((111)\) close-packed direction.

Based upon the discussion above, it can be concluded that slip can readily propagate from one grain to another in fcc crystals because of the availability of a large number of slip systems. Stress concentration developed at the grain boundary can activate dislocation motion in a neighboring grain. In contrast, hcp crystals are likely to develop cracks at grain boundaries before large macroscopic strain is realized, and therefore they have much less ductility in comparison with fcc crystals. Body-centered cubic crystals exhibit a behavior intermediate between that of fcc and hcp crystals.

Slip initiated locally in the most favorably oriented grains propagates from one grain to other as described above until the entire material is deformed and its surface is covered with slip lines. This corresponds to a macroscopic strain in the order of 1% extension or less. Yielding marks the onset of the growth of slip lines into visible slip bands; i.e., the yield strength can be defined as the stress required to grow the initially formed slip lines into bands. Growth of a slip band involves three different processes, which occur simultaneously: (i) plastic strain multiplication, (ii) multiplication of mobile dislocations, and (iii) multiplication of immobile dislocations. Sometimes, these processes are intermixed, leading to some confusion regarding the microscopic aspects of plastic deformation and strain hardening.

It is important to realize that the amount of plastic shear strain is determined by the number of dislocations that have traversed the slip planes and annihilated at the external surface leaving behind surface steps in the form of slip lines. Since the shear displacement \( b \) of a single dislocation is very small,
a large number of those dislocations must be involved in producing a measurable macroscopic strain. Therefore, it is possible to define strain multiplication as the shearing by more than one Burgers vector associated with motion of dislocations and their annihilation at the external surface. Evidently, those dislocations can never be observed in a deformed material. Since strain multiplication must involve the continued generation of mobile dislocations, the earliest mechanism proposed to explain strain multiplication is known as the Frank-Reed source. Although other more plausible mechanisms are later identified, they are all considered modifications of the Frank-Reed source described below.

Consider a segment of a dislocation line on a slip plane pinned at two points $A$ and $B$ by, e.g., impurity atoms, as shown in Fig. 7.27a. Under the influence of the shear component of an applied stress $\tau$, the segment is bowed between points $A$ and $B$, causing slip in the area (Fig. 7.27b). As the dislocation continues to bow, two segments in screw orientation of opposite sign $x$ and $x'$ come into contact and annihilate each other as shown in Fig. 7.27c, d, leaving behind a dislocation loop and the original segment $AB$. With continued application of stress, the above process is repeated, causing the generation of additional loops. Another means of creating a Frank-Reed source from a single loop of moving dislocation is cross slip of screw segments.

Plastic strain can also be multiplied by means of what is known as the crossover mechanism. A moving dislocation can develop an irregular shape as a result of encountering a random distribution of obstacles such as impurity atoms. When dislocation segments of opposite sense approach each other, they

Figure 7.27  A schematic illustration of the Frank-Reed source for dislocation multiplication. (a) A dislocation line on a slip plane pinned at two points $A$ and $B$. (b) Bowing of the dislocation line between $A$ and $B$ under the influence of a shear stress $\tau$. (c) Configuration of the dislocation as it continues to bow around $A$ and $B$, creating two screw segments $x$ and $x'$ of opposite sign. (d) Creation of a dislocation loop when the segments $x$ and $x'$ are annihilated and return of segment $AB$ to its original shape.
can either (i) annihilate each other, (ii) form an immobile close pair, i.e., a dipole, or (iii) pass each other depending upon the separation between the glide planes. If the two segments lie on the same glide plane or are near screw orientation, annihilation occurs, leaving behind an elongated prismatic loop. By prismatic is meant that the Burgers vector of the dislocation is normal to the glide plane, and therefore this loop can only move by climb. A dipole is formed if the separation between the two glide planes is below a critical value, and crossover occurs if the separation is above the critical value. Other possible mechanisms of strain multiplication include the annihilation mechanism and boundary multiplication.

To summarize the discussion above, strain multiplication and increasing the length of mobile dislocations are closely related processes. It may be noted that all strain multiplication mechanisms involving an increase in the length of mobile dislocation require dislocation segments of opposite sense to annihilate each other. Also, it is recognized that the alternative to annihilation is the formation of an immobile close pair of dislocations or a dipole. Therefore, another related process is the increase in length of immobile dislocations contributing to strain hardening.

Dislocations observed in deformed crystals by means of transmission electron microscopy (see Chap. 8) reflect the damage left behind as mobile dislocations traverse the slip planes. It is possible to correlate the maximum shear displacement and the strain hardening rate with the amount of damage in the form of close pairs and prismatic loops as well as their distribution. Increasing the amount of damage and its random distribution within the grains can be expected to increase the strain hardening rate and reduce the ductility. However, when the damage can be collected into subgrain boundaries with relative ease, the material is softened and more ductility can be realized.

An important factor controlling the amount of damage and its distribution particularly in fcc crystals is the stacking fault energy. This governs the relative ease by which dislocations in screw orientation can cross slip. In fcc crystals with relatively low stacking fault energy, total dislocations are split into partials. Before those partials can cross slip, they must be constricted into a total dislocation by eliminating the fault separating them. Evidently, work must be done by the applied stress to eliminate the fault and the lower the stacking fault energy the higher is the stress required to constrict the partials. As a result, fcc crystals with lower stacking fault energy strain harden at a greater rate and exhibit less tendency to collect the damage into subgrain boundaries in comparison with crystals of lower stacking fault energy.

Having recognized that the macroscopic yield strength marks the onset of growth of slip bands and that strain multiplication requires the propagation of slip from one grain to another, it is possible to understand the effect of grain size on yield strength. To illustrate this effect, consider a given material of coarse and fine grain sizes as shown in Fig. 7.28a, b. Evidently, both the length and width
of the slip band increase with the grain size and in turn the number of dislocation pileups as well as the number of dislocations in each pileup as can be seen by comparing both parts of Fig. 7.28. Consequently, increasing the grain size raises the level of stress concentration exerted by dislocation pileups on the grain boundary. As a result, a lower stress level is required to propagate slip from one grain to another as the grain size is increased. Experiment shows that the yield strength is related to grain size by

$$\sigma_y = \sigma_0 + kd^{-1/2}$$  (7.22)

where $\sigma_y$ is the tensile yield strength, $\sigma_0$ is a material constant called the friction stress, $K$ is an empirical constant, and $d$ is the grain diameter; i.e., the yield strength increases as the grain size decreases. A similar relationship exists between yield strength and subgrain size.

Another phenomenon that can be explained in terms of a dislocation mechanism is that related to the upper and lower yield points, exhibited by some ferrous alloys as shown in Fig. 7.29. Existence of such a yield point phenomenon is of particular importance to manufacturers of such parts as thin sheets of automobile bodies. As shown in Fig. 7.29, the existence of an upper and lower yield points indicate that once plastic deformation is initiated during forming operations, the material is effectively softened causing relatively large plastic strain. Since a state of stress concentration exists at the boundary, separating the deformed from undeformed regions, plastic deformation is spread into the region which has yielded at the upper yield point. In a tensile test, the yield point phenomenon is manifested by what is known as the Luders bands,
illustrated in Fig. 7.29. Due to the high level of stress concentration in the region of specimen adjacent to the grips, yielding is initiated with relative ease within the Luders bands (upper yield point) and then is spread into the gage length at nearly a constant stress (lower yield point). It is emphasized that the Luders bands associated with the yield point phenomenon differ from the slip bands. One possible explanation of yield point phenomenon is the locking or immobilization of dislocations by interstitial atoms, particularly C. Before a slip band can grow, dislocations must break away from the interstitial atoms, requiring the applied stress to increase up to the yield point. Once the dislocations are freed, the slip band can grow at a lower stress giving rise to the lower yield point.

7.13.2 High-Temperature Deformation

It is recalled that plastic deformation is a thermally activated process; i.e., the stress required to induce a certain plastic strain in a given material decreases as the temperature increases. Evidently, thermal energy can substitute some of the mechanical work required to move dislocations. Some of the important effects of temperature on plastic deformation relate to dislocation mobility and the role of grain boundaries. At sufficiently high temperatures where diffusion of atoms becomes significant, dislocations can move by climb. Although grain boundaries act as a source of strengthening at low temperatures, as demonstrated in the previous section, they become a source of weakness at elevated temperatures.

In general, there are a number of strain-producing mechanisms which can only operate at high temperatures, contributing to creep strain, such as
(i) dislocation climb, (ii) thermally activated glide, (iii) thermally activated cross slip, (iv) grain boundary shear or sliding, and (v) direct mass transfer by diffusion. However, the relative importance of these mechanisms is dependent upon both temperature and the stress level. Diffusion-controlled dislocation climb is believed to be a major contributor to creep strain at temperatures and stresses of engineering significance. Direct mass transport by diffusion becomes important at very high temperatures and low stress levels where dislocation-related mechanisms are rather suppressed. Typically, however, more than one mechanism operating simultaneously can contribute to creep strain, particularly in the case of complex multicomponent engineering alloys.

To illustrate the microscopic aspects of creep deformation, consider the typical creep curve shown in Fig. 7.30. It is recalled that such a curve is generated by subjecting a specimen to a constant load. As the specimen is elongated, its cross-sectional area is reduced, as shown in Fig. 7.30.

During the stage of primary creep, the strain rate is initially high and then decreases toward the end of this stage. An initial high strain rate reflects the relative ease by which dislocations initially present in the material can move by various mechanisms, particularly climb, and contribute to creep strain. As explained earlier in the case of low-temperature deformation, as dislocations traverse their slip planes they leave behind damage in the form of immobile dislocations, which can contribute to strain hardening and in turn reduce the creep strain rate toward the end of the primary stage. Because the initial dislocation structure is dependent on the prior history of the material, the primary creep behavior is quite history dependent. Other structural features can significantly influence the primary creep behavior.
Unlike the stage of primary creep, the creep strain rate during the secondary stage is dependent only on temperature and applied stress. In effect, the primary stage acts to erase the dependence of creep rate on history of the material. Since the creep strain rate becomes constant during the secondary stage, it becomes evident that the rate of strain hardening is balanced by the rate of recovery processes. It is also possible that recrystallization occurs during creep. As described earlier, strain hardening occurs by damage left behind in the form immobile dislocations, e.g., dislocations held in position by obstacles and generation of immobile dislocations. Recovery and associated softening of the material occurs by grouping the above damage into low-angle boundaries, which is facilitated by climb. Evidently, when the rates of the above two opposing processes are balanced, the creep strain rate reaches a steady state.

Because steady-state creep is characterized by an unchanging internal structure, it can be expected to continue until deformation becomes localized, giving rise to the phenomenon of necking during the tertiary creep stage. Under this condition, the creep strain rate becomes greater than the rate of strain hardening, causing the cross section to be eventually reduced to zero. Although this behavior is observed in practice for a limited number of materials particularly at very slow creep rates, the tertiary creep stage for most materials marks the onset of developing internal cracks or voids particularly at grain boundaries or precipitate particles prior to fracture. Such cracks or voids localize the creep strain at the beginning of the tertiary stage and then continue to grow and link up to cause fracture.

In terms of the dislocation mechanisms of creep described above, it is possible to understand the practical significance of some parameters influencing the creep rate, particularly grain size, stacking fault energy, alloying elements, and second phase particles. A brief account for the role of each parameter is given below.

Experiment shows that the creep rate at a given temperature and stress level is reduced by increasing the grain size. Grain boundaries can play two important roles during creep deformation. First, grain boundaries act as effective sources of vacancies required for dislocation climb. However, it is important to realize that the effectiveness of a grain boundary as a source of vacancies is dependent upon the degree of misorientation and in turn on the grain structure. High-angle grain boundaries are more effective as sources of vacancies in comparison with low-angle boundaries. It is recalled that vacancies can be created rather easily at the external surface. In a polycrystalline material, however, the area of grain boundaries per unit volume is far greater than the external surface area. An atom excessively agitated by thermal energy can diffuse into void or hole in a boundary, leaving behind a lattice vacancy. Presence of such voids or holes is critically dependent upon the degree of misorientation. As the degree of misorientation increases, the density of such voids or holes increases. Furthermore, because
of the high mobility of atoms at a high-angle boundary, it can readily accommodate the effect of losing vacancies. High-angle grain boundaries also act as preferred sites for nucleation of voids or cracks. To summarize, the creep rate increases as the high-angle grain boundary area per unit volume increases or grain size decreases, and therefore a material of a large grain size (less high-angle grain boundary area per unit volume) is preferred for creep applications. This is in opposition to low-temperature deformation, where a fine-grained structure of improved strength is preferred.

Recognizing the preceding effects of high-angle grain boundaries has led to developing materials of preferred grain structures for high-temperature creep applications as exemplified by the case of gas turbine blades. All these materials have fcc structure. One preferred grain structure is developed by directional solidification. In this case, the grains are aligned parallel to the applied tensile stress axis, which coincides with the \(100\) direction. First, the degree of misorientation across the grain boundaries is rather small, reducing their effectiveness as sources of vacancies. Second, the shear stress component parallel to the plane of the boundaries is nearly zero, which eliminates or minimizes the occurrence of grain boundary sliding. Another development is the use of single-crystal materials free of grain boundaries.

A material of lower stacking fault energy is generally preferred for creep applications. Lowering the stacking fault energy reduces the dislocation mobility by thermally activated cross slip, which strengthens the material. Slow-diffusing refractory transition elements such as Mo and W, whose atoms are relative large, are among the most preferred alloying elements for creep applications. Such elements are most effective in filling holes or voids at high-angle grain boundaries, reducing their effectiveness as sources of vacancies. Because of their low mobility, these elements deter the boundaries from readjusting their structure to continue to act as sources of vacancies. Second-phase particles improve the creep strength by acting as effective barriers to dislocation motion.

### 7.14 Alloy Phases and Phase Change

From a thermodynamic point of view, engineering alloys are considered to be multicomponent systems. Typically, an engineering alloy consists of more than one phase; i.e., it is a heterogeneous system. Properties of engineering alloys are closely related to the nature and distribution of their constituent phases. Detailed information about the number, nature, and ratio of phases in a particular alloy is of critical importance not only to the materials scientist and process engineer but also to the design engineer who selects and specifies materials for newly developed components. Such information can be obtained from the phase or equilibrium diagram of the alloy system involved.

Copyright © 2004 by Marcel Dekker, Inc.
7.14.1 Alloy Phases and Metastability of Engineering Alloys

It is recalled from earlier discussions that a phase in a thermodynamic system is defined as a substance of a given composition and properties and is separated from other substance by a definite bounding surface. Considering a system of an engineering alloy, a phase of an engineering alloy is defined as a substance having a certain crystal structure, composition, and properties, and is separated from other phases by a phase boundary.

Most engineering alloys consist of (i) a host or matrix phase consisting of the base metal containing various elements (substitutional and interstitial) in solid solution and (ii) secondary phases. Secondary phases can include carbides (compounds of carbon and metallic elements), borides (compounds of boron and metallic elements), nitrides (compounds of nitrogen and metallic elements), and intermetallic compounds (compounds of metallic elements).

Typically, engineering alloys are supplied by the manufacturers in what is known as the mill-annealed condition. In this condition, the alloy has a given set of properties corresponding to a certain microstructure. Quite often the microstructure corresponding to the mill-annealed condition exists in a metastable condition. As a result, there is always a natural tendency for phases in engineering alloys to seek a more stable state, giving rise to phase changes as described below.

7.14.2 Phase Changes

It is recalled that the constituent atoms of a solid material at any temperature above 0 K are always in a state of thermal vibration. In effect, the microscopic state of a system in equilibrium is not static. Rather, the system perpetually undergoes rapid fluctuations. By means of these fluctuations, the system explores other states of lower free energy in accordance with the second law of thermodynamics described earlier.

By definition, a phase change is any rearrangement of atoms resulting in the formation of a more stable phase(s). Solidification of molten metal as a result of cooling is an example of a phase change. Phase changes occurring in the solid state are of extreme importance in controlling the properties of engineering alloys, particularly ferrous alloys, as explained later in this section. In general, a phase change can result in a change in crystal structure and/or composition. One phase may decompose into two phases. Also, two phases may combine to form a new phase. Depending upon the nature and distribution of the phase(s) resulting from a phase change, it can either have beneficial or detrimental effects on the properties of engineering alloys, as demonstrated later.

There are two main aspects for any type of phase change: its thermodynamics and kinetics. Thermodynamically, the driving force for any phase change
is the free-energy difference between the initial and final states. A phase change can proceed only if it results in a decrease in free energy. However, from a kinetic point of view, there exists an energy barrier for any phase change. To illustrate the principle involved, consider a system consisting of an object such as a box placed on a table such that its height from the floor is $h$, as shown in Fig. 7.31a. In this case, the free energy of the system can be considered to be its potential energy. If the plane of the table is considered to be the reference for measuring energy, the potential energy of the box at the table is 0. With respect to the floor, however, its potential energy is $-mgh$, where $m$ is the mass of the box and $g$ is the acceleration of gravity. Evidently, the box has a lower free energy if it lies at the level of the floor. Therefore, thermodynamics predicts that the box must fall from the table to the floor to lower its free energy; however, this process cannot occur spontaneously because of the existence of an energy barrier, as described below.

![Diagram](image)

**Figure 7.31** Schematics illustrating (a) change in the position of the box from $A$ to $C$ and (b) change in free energy of a system from $A$ to $C$. 

Copyright © 2004 by Marcel Dekker, Inc.
It is clear that the box can lie on the table indefinitely unless it is provided with sufficient energy to bring it to the position shown in Fig. 7.31a. At this position, the box spontaneously falls to the floor under the influence of gravity and assumes a lower free energy. On a relative basis, when the box lies at the table, it is said to be in a metastable state. Energy spent to bring the box to the position shown in Fig. 7.31a is called activation energy. At this position, where the box falls spontaneously, it is said to be in an unstable state.

In terms of free energy, the sequence of events taking place as the box falls from the table to the floor as described above can be represented as shown in Fig. 7.31b. Generally, Fig. 7.31b represents the change in free energy of any system when it undergoes a transition to a more stable state. As can be seen, the total change in free energy can be divided into two components $\Delta G$ and $\Delta G^*$, which have different physical meanings. Also, the values of $\Delta G$ and $\Delta G^*$ are independent of each other. By definition, $\Delta G = G_B - G_A$ is the driving force of the transition ($B$ is the final state and $A$ is the initial state), which is a thermodynamic parameter. In contrast, $\Delta G^*$ is the activation energy of the transition, which is a kinetic parameter. Since many energy diagrams can be drawn between points $B$ and $A$ such that $\Delta G$ is fixed, it becomes evident that the value of $\Delta G^*$ is independent of the value of $\Delta G$. It clear that the rate of the transition is dependent on the value of $\Delta G^*$. Those processes for which $\Delta G^*$ is rather high are more sluggish or occur at a slower rate in comparison with those processes for which $\Delta G^*$ is smaller. Within this context, a catalyst is defined as an agent whose function is to lower the activation energy for a given transition and in turn increases its rate. Various agents which can act as catalysts for phase transitions in materials include various imperfections such as grain boundaries, twin boundaries, stacking faults, and dislocations. Frequently, it is more energetically favorable if the change from the initial into the final state occurs in steps, giving rise to intermediate transient phases.

Most of the phase changes in engineering alloys occur as a result of changing temperature. Some phase changes require massive rearrangement of atoms, requiring transport or diffusion of atoms over large distances. In this regard, vacancies play an important role in phase changes. Atomic diffusion is a thermally activated process; i.e., its rate increases with temperature and decreases with decreasing temperature. Also, diffusion is a time-dependent process. Therefore, the time required to affect a given extent of phase change decreases with increasing temperature. Some of the phase changes, however, are diffusionless. Such phase changes are only dependent on temperature. Diffusionless phase changes can also be stress induced, as shown later.

A temperature-dependent phase change can occur either at a fixed temperature or over a range of temperature. Solidification of a pure metal is an example of a phase change occurring at a fixed temperature. In contrast, an alloy solidifies over a range of temperature.
To illustrate the various thermodynamic and kinetic aspects of a phase change, consider that a phase $\alpha$ tends to transform into another phase $\beta$ during cooling from a high temperature. Assume that the transformation begins at a temperature $T_0$ and is completed at a temperature $T$. It is possible to represent the time required to initiate the transformation on a time-temperature-transformation (TTT) diagram such as that shown in Fig. 7.32. Because of the characteristic C shape of the diagram, it is sometimes called the C curve. Its characteristic shape can be explained in terms of the opposing effects of thermodynamics and kinetics as follows.

At the temperature $T_0$ in Fig. 7.32, the two phases can coexist in equilibrium where the difference in free energy between the two phases $\Delta G = 0$. This temperature is called the equilibrium transition temperature. Evidently, at the equilibrium transition temperature, the driving force for the phase change $\Delta G = 0$, and therefore no phase change can take place.

If the temperature is lowered to just below $T_0$, e.g., $T_1$ in Fig. 7.32, a finite driving force $\Delta G$ exists. Although the transformation is kinetically favorable at the relatively high temperature $T_1$ because of higher diffusion rate, the driving force for the transformation is rather small. As a result, the time required to initiate the transformation at $T_1$ is rather long, as shown in Fig. 7.32. If the temperature is lowered to $T_2$, the driving force becomes larger; however, the diffusion rate becomes relatively slower in comparison with that at $T_1$. However,

**Figure 7.32** Time-temperature-transformation (TTT) diagram showing the transformation of $\alpha$ to $\beta$ during cooling from a high temperature.
this is outweighed by the large driving force shortening the time required to initiate the transformation. At some intermediate temperature \( T_c \), the best compromise is reached between the opposing effects of thermodynamics and kinetics, and the transformation is initiated at the shortest possible time \( t_c \). Below the temperature \( T_c \), the driving force continues to be large; however, the sluggishness of the diffusion rate at lower temperatures slows down the rate of transformation, and therefore the time required to initiate the transformation increases again.

A similar argument to that described above can be applied to the time required to complete the transformation. A series of other C curves can be constructed to represent the completion of certain percentages of the transformation, as shown in Fig. 7.33, e.g. 10%, 20%, 50%, etc. It is customarily to call point C in Fig. 7.33 the nose of the C curve.

Lines drawn on the TTT diagram shown in Fig. 7.34 are by definition cooling curves; they represent the cooling rate from a high temperature where the \( \alpha \) phase is stable. It is clear that the cooling rate increases as the cooling curve is shifted to the left. A critical cooling rate is defined such that the corresponding curve touches the nose of the diagram. If a material consisting of phase \( \alpha \) is cooled rapidly enough such that the cooling rate exceeds the critical value, it is possible to skip the transformation into the \( \beta \) phase. In this case the material is said to exist in a metastable state. By means of rapid cooling, the atoms of phase \( \alpha \) do not have sufficient time to rearrange themselves into the structure

![Figure 7.33](image) A series of C curves that represent the completion of certain percentages of transformation.

Copyright © 2004 by Marcel Dekker, Inc.
of β phase. An equivalent statement is that the material is supercooled. Evidently, if the material is reheated to the temperature range where the β phase exists, the transformation can take place.

In practice, processing of engineering alloys can quite often lead to a metastable structure by a sequence of events similar to those described above. Various types of phase change can take place during exposure to elevated temperatures either during manufacturing processes or during service. Since these phase changes can have significant effects on properties, better use and selection of engineering alloys requires knowledge of their thermal stability. By thermal stability is meant the structural changes which occur during exposure to elevated temperatures and the corresponding effects on properties.

Experiment shows that the kinetics of many processes can be expressed in terms of an Arrhenius equation of the type

$$K = K_0 \exp(\Delta G^*/RT)$$

(7.23)

Customarily, the activation energy $\Delta G^*$ is represented by $Q$, and the above equation is usually written as

$$K = K_0 \exp(Q/RT)$$

(7.24)

Figure 7.34  Cooling curves on the TTT diagram represent different cooling rates of α phase. The cooling rate increases as the cooling curve is shifted to the left.
where $K$ is the temperature-dependent reaction rate constant, $K_0$ is a constant for a given process, $R$ is the universal gas constant, and $T$ is the temperature expressed in degrees Kelvin (K) and $e = 2.71828$ (natural logarithm). A plot of log $K$ vs. $1/T$ results in a straight line. From the slope of the line, the activation energy $Q$ can be calculated.

Another important experimental observation related to the kinetics of phase changes is the coexistence of the reactants and products. For example, if a phase change involves the transformation of phase $\alpha$ into phase $\beta$, at an intermediate stage during the transformation two phases coexist. It is evident from this observation that all atoms in a system do not undergo transformation from one phase into another at one and at the same time. Instead, at any instant, a certain fraction of the total atoms is in the process of transformation and the remainder is either completely transformed or untransformed. Therefore, the extent to which a given phase change has progressed after a given time $t$ can be expressed by the fractional transformation $\alpha(t)$, defined as the ratio of the number of atoms per unit volume in the final state to the total number of atoms available for the transformation at $t = 0$. At any time $t$, the rate of the transformation is simply $dy/dt$, which is in general a function of time.

It follows from the discussion above that the rate $K$ at which a given phase change occurs is inversely proportional to the time $t$ required for the transformation. Therefore, the equation can be rewritten as

$$K = K_0 \exp\left(-\frac{Q}{RT}\right)$$

$$\frac{1}{t} = C \exp\left(-\frac{Q}{RT}\right)$$

where $C$ is a constant. Taking the logarithms of both sides of the above equation results in

$$\log \frac{1}{t} = \log C - \frac{Q}{R} \log e 1/T$$

where $e$ is the natural logarithm ($e = 2.71828$). Substituting for $\log e = 0.43$ and $R = 8.3143$ J/degree, the above equation becomes

$$\log \frac{1}{t} = \log C - 0.05 Q 1/T$$

A plot of log $1/t$ vs. $1/T$ results in a straight line whose slope is $-0.05Q$ from which $Q$ can be determined in units of Joules/mol (1 kJ = 1000 J). To obtain $Q$ in cal/mol, the result in J/mol is multiplied by 4.1868 (1 kcal = 1000 cal).

To illustrate the use of the preceding equation in determining $Q$, consider that as a result of a phase change, the yield strength of a given material varies with temperature and time, as shown in Fig. 7.35a. A horizontal line is drawn at a given percentage increase in yield strength, e.g., 50%. Intersection of this line
with each curve gives the time required to cause a 50% increase in yield strength at each temperature. By plotting \( \log \frac{1}{t} \) vs. \( T \) where \( t \) and \( T \) are obtained from Fig. 7.35a, a straight line is obtained as shown in Fig. 7.35b. From the slope of the line, \( Q \) can be determined as described above.

**Figure 7.35** (a) Change of yield strength of a material with temperature and time. (b) Plot of \( \log \frac{1}{t} \) vs. \( T \) showing a straight line. \( Q \) can be determined from the slope of the line.
7.14.3 Types of Phase Changes and Their Effect on Properties

Although there many types of phase changes, they can broadly be classified into (i) diffusion-controlled phase changes, (ii) diffusionless phase changes. Diffusion-controlled phase changes include precipitation reactions and disorder-order transformations. A classical example of a diffusionless phase change in the martensitic transformation which occurs in steels.

Precipitation reactions results in the formation of secondary phases when the solubility limits of certain elements in an alloy are exceeded. Secondary phases have different chemical composition from that of the matrix phase, but they can have either the same or different crystal structure. Also, they alter the chemical composition of the matrix phase. Examples of secondary phases resulting from precipitation reactions include carbides, nitrides, borides, and intermetallic compounds. Various types of imperfections act as catalysts for various precipitation reactions. Among these imperfections are grain boundaries, dislocations, stacking faults, and twin boundaries. In terms of the relationship between the crystal structure of the precipitate and matrix, secondary precipitates can be classified into (i) coherent, (ii) partially coherent, and (iii) incoherent.

A precipitate particle is said to be coherent if its crystal structure is the same as that of the matrix phase. However, its crystal spacings can either be the same as that of the matrix or slightly different. Because of the close similarity between the structure of the precipitate and matrix, any mismatch at the interface is accommodated by elastic strain. Coherent precipitates are effective means of strengthening engineering alloys. Typically, precipitation of coherent particles occurs on a very fine scale within the matrix of the alloy and cannot be revealed by a light optical microscope.

Similar to coherent precipitate, a partially coherent precipitate has the same crystal structure as that of the matrix phase. However, because of the larger difference in crystal spacings, the mismatch cannot be accommodated only elastically. In this case, some of the misfit is accommodated by forming dislocations at the interface. Some of the partially coherent precipitates are coarse enough to be detected by a light optical microscope. Precipitation of partially coherent particles can occur within the matrix of the alloy as well as at grain boundaries. Also, they provide a source of strengthening the engineering alloys.

A noncoherent interface has a crystal structure and/or crystal spacings widely different from those of the matrix. In this case, the interface resembles a grain boundary. Usually, noncoherent precipitates are rather coarse and can be easily detected by a light optical microscope. Precipitation of noncoherent particles can occur within the matrix as well as at grain boundaries. However, they are not as effective strengtheners as coherent and partially coherent precipitates.
In terms of performance of engineering alloys, precipitation of secondary phases can either have beneficial or detrimental effects on properties. Some precipitation reactions are intentionally induced to promote the mechanical properties. A classical example is the strengthening of a series of Ni-base alloys used to manufacture gas turbine blades by what is known as the γ' phase. Basically, the γ' phase is an intermetallic compound based upon the Ni₃Al composition. Such alloys are commonly known as the superalloys. A similar phase is used to strengthen some of the precipitation hardened steels.

Some secondary phases, however, can have a detrimental effect on properties. A classical example of such phases is what is known as the sigma phase. It precipitates in some stainless steels and nonferrous alloys. In most of these alloys, the sigma phase is based upon the FeCr composition. Because it is extremely hard and brittle, it can cause severe embrittlement of the alloy.

It is important to realize that the same phase can have either beneficial or detrimental effects on properties depending upon its morphology. A typical example of such a phase is a carbide based upon the Cr₂₃C₆ composition. It precipitates in many stainless steels and nonferrous alloys. This carbide is of the partially coherent type. It can precipitate within the grains as well as at grain boundaries. Depending upon its morphology at grain boundaries, it can either be beneficial or detrimental, as described below.

Precipitation of a secondary phase at grain boundaries can occur either by a continuous or discontinuous mechanism. As its name implies, continuous precipitation occurs by continued diffusional transport of elements in excess of the solubility limit into the grain boundary. As a result, the precipitate assumes the morphology of discrete particles. Such discrete particles can strengthen the grain boundaries particularly during high-temperature service. Under certain conditions, however, precipitation of the same phase at grain boundaries can occur discontinuously. By discontinuously is meant that the precipitation takes place only when the grain boundary migrates leaving behind a cellular structure. As can be seen, the cellular structure consists of alternating lamellae of the carbide and a Cr-depleted matrix phase. Such a grain boundary lamellar structure is very detrimental to mechanical strength, particularly at elevated temperatures.

Disorder-order transformation occurs in some of the nonferrous alloys. It is recalled that a solid solution can either be ordered or disordered. Depending upon the composition of the alloy, the matrix phase is transformed from a disordered into an ordered phase below a critical temperature. Overall ordering of the alloy requires its composition to be at or very close to certain ratios of the constituent elements such as A₂B, A₃B, . . . . In this case, the transformation does not result in a change in chemical composition, however, it can result in a change in crystal structure. However, usually the composition of engineering alloys which undergo disorder-order transformation deviates slightly from the exact value. Although the alloy may remain to contain a disordered portion, the
major portion becomes ordered. Ordering can result in severe loss of ductility. In some cases, however, the ductility loss is not as severe and the ordering transformation can be beneficial.

Martensitic transformations are essentially diffusionless phase changes. In a few cases, however, the transformation can also be controlled by diffusion. A martensitic phase transformation results in a change in crystal structure but not in composition. Such a transformation is of particular importance in steels, however, it also occurs in some nonferrous alloys. In steels, the martensitic transformation involves a change in crystal structure from body-centered cubic into body-centered tetragonal. Another example of such a phase change occurs in some of the Co- and Ti-base alloys where the fcc structure is transformed into hcp structure. A diffusionless martensitic transformation occurs independent of time. It is only dependent on temperature. Another feature of this transformation is that it can be stress-induced. For example, the fcc into hcp transformation in nonferrous alloys can occur by motion of $\frac{1}{3} \langle 112 \rangle$ partial dislocations on $\langle 111 \rangle$ planes under the influence of the shear stress component of an applied stress. Because martensitic phases are extremely hard and brittle, they result in severe loss of ductility.

7.15 Equilibrium Phase Diagrams

By definition, a phase diagram is a map illustrating the effect of two thermodynamic variables, temperature and composition, on the phases existing in equilibrium within an alloy system. Phase diagrams are constructed at constant atmospheric pressure, eliminating both pressure and volume as state variables. Most of the phase diagrams are developed for binary alloys, i.e., two-component systems. Ternary diagrams for some alloys systems are also available; however, in this case, the diagram is constructed in sections, each of which corresponds to a given temperature.

To understand the basis for constructing a phase diagram, consider the solidification, i.e., change from liquid to solid state, of a pure metal A and an alloy of B in A as represented by the cooling curves of Fig. 7.36 such that the melting point of A is higher than that of B. As can be seen, a cooling curve is a plot of temperature vs. time. In discussing the significance of a cooling curve, it is important to distinguish between sensible heat and latent heat. By definition, sensible heat is the thermal energy added or removed from the system as a result of increasing or decreasing its temperature, respectively. Conversely, latent heat is the thermal energy associated with a phase change, e.g., latent heat of melting.

When a pure molten metal is cooled through the liquid state, the cooling curve changes smoothly as sensible heat is removed (Fig. 7.36). A sudden change in the slope of the cooling curve signifies the onset of a phase change, e.g., change from the liquid to solid state. As a pure metal solidifies, the composition of the remaining liquid remains unchanged and in turn its freezing point. Consequently,
the pure metal solidifies at a constant temperature as latent heat is extracted, giving rise to what is known as thermal arrest in the cooling curve. Once solidification is completed, the cooling curve changes smoothly again as sensible heat is extracted from the solid.

In contrast with the above case, a molten alloy solidifies over a temperature range. As the first solid is formed, the composition of remaining liquid becomes enriched in element B and its melting point decreases. Further solidification can only occur if the temperature is decreased. Over the temperature range where solidification occurs both latent heat and sensible heat are extracted, causing the slope of the curve to change (Fig. 7.36). Upon completion of solidification, further cooling occurs by extracting sensible heat from the solid phase. Various types of phase diagrams are distinguished depending upon the properties of the two components of the system, as described below.

7.15.1 Complete Solid Solubility

When two elements have similar atomic structures, the difference in their atomic size does not exceed about 15%, and their crystal structure is the same, the solid solubility of each element in the other is unlimited. Nickel and copper are
examples of such elements. In this case, the phase diagram is of the type shown in Fig. 7.37. Composition in weight or atomic percent is represented on the $x$ axis, and the $y$ axis represents temperature.

As can be seen from Fig. 7.37, the diagram consists of two curves, a liquidus and solidus, separating three phase fields. Above the liquidus curve, an alloy of A and B of any composition consists entirely of one liquid phase. Within the phase field bounded by the liquidus and solidus, any alloy exists as a mixture two phases, a solid phase and a liquid phase. Below the solidus curve an alloy of any composition exists in the form of one solid phase, a solid solution. To illustrate the phase changes occurring during cooling, consider an alloy of composition $x$ represented by a vertical line.

Above the liquidus curve, the alloy consists of one liquid phase. By applying the phase rule [Eq. (7.1)], $f = c - p + l$, where $c = 2$ (number of components), and $p = 1$ (number of phases), the number of degrees of freedom $f = 2$. Physically, this means that the two variables, temperature and composition, can be changed independently. It is possible to change composition while the temperature remains constant. Conversely, the temperature can be changed while the composition remains constant.

Figure 7.37  Binary phase diagram showing complete solubility in liquid and solid states.
Solidification during cooling begins at a point corresponding to a temperature $T_a$. At any point $b$ within the two phase fields, the composition of each phase is determined by the intersection of a horizontal line drawn from point $b$ to intersect the liquidus and solidus curves. Intersection of the line with the solidus $x_s$ gives the composition of the solid phase, and its intersection with the liquidus $x_l$ gives the composition of the liquid phase. Since the number of phases $p = 2$, the number of degrees of freedom $f = 1$; i.e., either temperature or composition cannot be changed independently. To determine the ratio of the liquid and solid phases within the phase field bounded by the liquidus and solidus, a simple mass balance is used leading to what is known as the lever rule principle, as described below.

Consider an alloy of mass $m$ whose composition is $x$ as shown in Fig. 7.37. It is evident that $x$ corresponds to the percentage of element B in the alloy. If the mass of the liquid phase is $m_l$ and the mass of the solid phase is $m_s$, then

$$m = m_l + m_s$$

and the mass of element B in the alloy must equal the sum of its masses in the two phases, i.e.,

$$x_m = x_1 m_l + x_2 m_s = x_1 (m - m_l) + x_2 m_s, \quad \text{or} \quad m_l (x_1 - x_2) = m (x_1 - x),$$

and therefore

$$m_l / m = (x - x_1)/(x_2 - x_1)$$

Similarly, the ratio of the mass of solid phase to that of the alloy is given by

$$m_s / m = (x_2 - x)/(x_2 - x_1)$$

As shown in Fig. 7.37, when the alloy begins to solidify at temperature $T_a$, the concentration of element B in the first solid phase to form is $c_1$. At a lower temperature $T_2$, the concentration of element B in the corresponding solid phase becomes $c_2$ where $c_2$ is less than $c_1$, as shown in Fig. 7.37. With further cooling the concentration of element B in the solid continues to decrease until it reaches the value corresponding to $x$ at the solidus curve at temperature $T_d$, where solidification is completed. Such a phenomenon is called coring and it results in a concentration gradient within the solid solution phase. Evidently, coring leads to inhomogeneous properties. To eliminate the effect of coring, the solid solution alloy must be homogenized by a heat treatment allowing the atoms to readily diffuse, which removes the concentration gradient. Just below the solidus curve, the alloy consists of one phase (a solid solution). Both a pure metal and a solid-solution phase have similar microstructures when viewed under a light optical microscope.
7.15.2 Partial Solid Solubility

Most metals have rather limited solubility in the solid state, which can lead to a binary phase diagram such as that shown in Fig. 7.38, which contains two terminal solid-solution phases \( \alpha \) (a solid solution of \( B \) in \( A \)) and \( \beta \) (a solid solution of \( A \) in \( B \)) separated by a phase field consisting of a mixture of \( \alpha \) and \( \beta \). It is to be noted that the crystal structure of the \( \alpha \) phase is the same as that of metal \( A \), and the crystal structure of the \( \beta \) phase is the same as that of metal \( B \).

In the phase diagram of Fig. 7.38, line \( aeb \) represents the liquidus, and line \( ad'b'b \) corresponds to the solidus. Solvus lines \( a'C \) and \( b'D \) represent the solid solubility limits of element \( B \) in the \( \alpha \) phase and element \( A \) in the \( \beta \) phase, respectively. A distinguishing feature of the phase diagram shown in Fig. 7.38 is point \( e \), which is known as the eutectic point. When an alloy of a composition corresponding to point \( e \) is cooled from the liquid state, it solidifies by a reaction called the eutectic reaction where the liquid phase is decomposed into two solid solutions \( \alpha \) and \( \beta \), i.e.,

\[
\text{Liquid (cooling) } \rightarrow \alpha_{(\text{solid})} + \beta_{(\text{solid})}
\]

![Figure 7.38](image-url)  
Phase diagram showing partial solid solubility of two metals.
At point $e$ three phases coexist: the liquid phase and two solid phases $\alpha$ and $\beta$, i.e., $p = 3$. Application of the phase rule results in $f = 0$, i.e., the eutectic point is uniquely defined. As can be seen from Fig. 7.38, the composition of the $\alpha$ phase at the eutectic temperature is given by point $a'$ and that of the $\beta$ phase is given by point $b'$. Due to the nature of the eutectic reaction, the liquid phase solidifies into alternating lamellae of the $\alpha$ and $\beta$ phases. To understand the sequence of events leading to the development of phase structure as a function of composition, consider the alloys marked 1, 2, 3, and 4 in Fig. 7.38.

When alloy 1 is cooled from the liquid state, it begins to solidify into the $\alpha$ phase at a temperature just below the respective liquidus temperature $T_1$. Solidification occurs over the temperature range $T_1 - T_2$, where $T_2$ is the solidus temperature. At any given temperature, the composition of the solid phase is represented by line $aa'$ and that of the liquid phase is represented by line $ae$. Just below $T_2$, the alloy consists of a single phase, $\alpha$ solid solution. No phase change occurs until the temperature is lowered to $T_3$ marking the limit of solubility of element B in the $\alpha$ phase. Below $T_3$, the alloy contains excess concentration of element B, which cannot be accommodated in the $\alpha$ solid solution. However, since element B can accommodate a certain concentration of element A in solid solution, the excess concentration of B precipitates within the $\alpha$ phase in the form of a $\beta$ phase. With further cooling, the amount of $\beta$ phase precipitated within the $\alpha$ phase continues to increase, and the ratio of the two phases at any given temperature can be obtained by applying the lever rule principle described earlier. It is evident from the phase diagram of Fig. 7.38 that the behavior of alloy 2 differs from that of alloy 1 as described below.

Solidification of alloy 2 occurs at a temperature $T_4$ just below the liquidus resulting in the formation of $\alpha$ phase within the liquid phase. With continued cooling, the compositions of the solid and liquid phases vary according to lines $aa'$ and $ae$, respectively. As can be seen from the phase diagram of Fig. 7.38, when the temperature reaches the eutectic temperature the composition of the liquid phase corresponds to the eutectic composition $e$ and the composition of the $\alpha$ phase is given by point $a'$. At this temperature, the remaining liquid phase solidifies by the eutectic reaction described earlier, resulting in a phase mixture of $\alpha$ and $\beta$. It is to be noted that the $\alpha$ phase formed by the eutectic reaction is the same as that formed from the liquid phase at higher temperature, which is commonly referred to as primary $\alpha$ phase; i.e., when the alloy is completely solidified, it consists of two phases: $\alpha$ and $\beta$. With further cooling, the compositions of the $\alpha$ and $\beta$ phases vary according to lines $a'C$ and $b'D$, respectively. Alloy 3 solidifies in a similar fashion to alloy 2 except the primary $\alpha$ phase is replaced by primary $\beta$ phase. Also, the solidification of alloy 4 is similar to that of alloy 1 with $\beta$ phase replacing the $\alpha$ phase.

Other features of phase diagrams include the formation of intermediate solid phases, intermetallic compounds, and solid-state reactions. Intermediate
phases have finite homogeneity range. An intermediate phase has a crystal structure different from those of its constituent metals. Intermetallic compounds resemble intermediate phases; however, they are characterized by a narrower homogeneity range, which can approach a straight line.

A solid-state reaction of particular importance to engineering alloy steels is known as the eutectoid reaction because it resembles the eutectic reaction described earlier. In this case, a solid phase $\gamma$ is decomposed into two solid phases $\alpha$ and $\beta$ upon cooling, i.e.,

$$\gamma_{(\text{solid})} \text{(cooling)} \rightarrow \alpha_{(\text{solid})} + \beta_{(\text{solid})}$$

Also, for an eutectoid reaction involving three phases, both the temperature and composition of the eutectoid point are uniquely defined by the condition $f = 0$. Other three-phase reactions which can occur in binary alloy systems are the peritectic and peritectoid reactions. In a peritectic reaction, a liquid phase combines with a solid phase during cooling to form another solid phase, i.e.,

$$\text{Liquid} + \text{solid (cooling)} \rightarrow \text{solid}$$

A peritectoid reaction is the reverse of eutectoid reaction; i.e. upon cooling, two solid phases combine to form another solid phase:

$$\alpha_{(\text{solid})} + \beta_{(\text{solid})} \text{ (cooling)} \rightarrow \gamma_{(\text{solid})}$$

7.15.3 Iron-Carbon Phase Diagram

Pure Fe is known to undergo a type of phase transformation that involves a change in crystal structure as the temperature is changed. Usually, this type of phase change is called polymorphic or allotropic transformation. When pure Fe solidifies from the liquid state, it has a body-centered cubic structure called the \(\delta\) phase. As the temperature is lowered, the \(\delta\) phase remains stable until a temperature of 1390°C, where the body-centered cubic \(\delta\) phase is transformed into a face-centered cubic structure called the \(\gamma\) phase, or austenite. With continued cooling the \(\gamma\) phase remains stable until a temperature of 910°C, where the \(\gamma\) phase is transformed into a body-centered cubic structure called the \(\alpha\) phase, or ferrite. Ferrite remains stable at lower temperatures. Among the allotropic forms of Fe, only austenite and ferrite are of particular significance to ferrous alloys.

For practical purposes, the Fe-C phase diagram is plotted up to a C concentration of 6.67%, as shown in Fig. 7.39. As described earlier for other types of phase diagrams, the development of the phase structure of ferrous alloys can be understood by considering the sequence of events taking place during cooling from the liquid state. It is observed from Fig. 7.39 that the addition of C to Fe results in three types of reactions, each of which involves three phases: Two
reactions contain a liquid phase (peritectic and eutectic), and one solid-state reaction (eutectoid), as described below.

1. Peritectic point: 1493°C, 0.16% C
   \[ \text{Liquid} + \delta \text{Fe(solid)(cooling)} \rightarrow \gamma \text{Fe(solid)} \]

2. Eutectic point: 1147°C, 4.3% C
   \[ \text{Liquid(cooling)} \rightarrow \gamma \text{Fe(solid)} + \text{cementite Fe}_3\text{C(solid)} \]

3. Eutectoid point: 723°C, 0.8% C
   \[ \gamma \text{Fe(solid)(cooling)} \rightarrow \alpha \text{Fe(solid)} + \text{cementite Fe}_3\text{C(solid)} \]

Among the preceding three reactions, the solid-state eutectoid reaction assumes a primary importance in engineering applications of steels. A mixture of ferrite and cementite produced by the eutectoid reaction is called pearlite; it consists of alternating lamellae of ferrite and cementite. To develop a basic understanding of the effect of C on the phase structure of steels and how it can be controlled to produce desirable properties, it suffices to describe the sequence of events occurring in the solid state.

Figure 7.39 The iron–iron carbide phase diagram.
When C is added to Fe, the austenite to ferrite transformation starts at a lower temperature in comparison with pure Fe, as indicated by line $AB$ in Fig. 7.39, and is completed over a temperature range determined by the C content. For alloys containing less than about 0.025% C, line $AD$ indicates the completion of the transformation. At 723°C, the solubility of C in ferrite is 0.025% as indicated by point $D$, and it decreases with temperature according to line $DF$ until it reaches 0.005% at room temperature. With continued cooling of an Fe-C alloy containing $<0.025\%$ C, excess C which cannot be accommodated in solid solution is precipitated at the grain boundaries of ferrite in the form of iron carbide (cementite). A completely different sequence of events take place in Fe-C alloys containing more than 0.025% C, as described below.

Consider, for example, an alloy containing 0.2% C. Above a temperature of about 850°C, the alloy consists of austenite. When the temperature is lowered, the transformation of austenite into ferrite begins at 850°C, and with continued cooling the amount of ferrite increases and that of austenite decreases. Within the area enclosed by lines $ABD$, the alloy consists of two phases: ferrite and austenite. Ferrite formed from austenite at a temperature above 723°C (eutectoid temperature) is called primary ferrite to distinguish it from ferrite formed by the eutectoid reaction. As the temperature reaches 723°C, the composition of ferrite is indicated by point $D$ (0.025% C), and that of austenite is given by point $B$ (0.8% C) corresponding to the eutectoid point. Therefore, the remaining austenite is decomposed into pearlite, which is nucleated at the grain boundaries of the austenite. With continued cooling to room temperature, the solubility of C in ferrite decreases according to line $DF$, and therefore excess C precipitates as cementite. Since ferrite is the major phase in an alloy containing 0.2% C, and pearlite is the minor phase, the phase structure consists of colonies of pearlite distributed in a matrix of primary ferrite.

Structural carbon steels are produced with carbon contents in the range of 0.06–0.95%. Steels containing less than 0.8% C are called hypoeutectoid, and those containing more than 0.8% C are called hypereutectoid. Most of these steels, however, are of the hypoeutectoid type and therefore, according to the Fe-C phase diagram, they consist of ferrite and pearlite. It is evident from the discussion above that the amount of pearlite increases and that of primary ferrite decreases as the C content is increased.

It is emphasized that the phase changes predicted from the Fe-C phase diagram as described above occur under equilibrium conditions. Other phase changes, however, can occur under nonequilibrium conditions. Such phase changes are used to control the mechanical strength of steels.

To summarize the discussion above, phase diagrams provide a very useful tool in determining the constituent phases of engineering alloys. Furthermore, they provide an important basis for manipulating the constituent phases by heat treatment particularly in the case of ferrous alloys.
7.16 Methods of Strengthening Engineering Alloys

Any method used to strengthen engineering alloys is based upon introducing into the structure an obstacle to dislocation motion. Various strengthening methods include (i) plastic deformation or cold work: strain hardening, (ii) solid solution, and (iii) phase change, (iv) dispersion hardening, and (v) refining the grain size.

It is recalled from earlier discussions that during plastic deformation, dislocations move and multiply. Interactions among individual dislocations results in certain configurations acting as barriers to further dislocation motion. This results in an increase in the yield strength of the material. In this condition, the material is said to be strain hardened.

Strengthening by solid solution (alloying with other elements) is based upon lattice distortion caused by alloying elements. When the lattice becomes distorted, its inherent resistance to dislocation motion is increased resulting in an increase in strength.

Phase change can be used to strengthen a material by various methods, including (i) precipitation of a secondary phase (precipitation hardening), (ii) a disorder-order transformation, and (iii) changing the crystal structure of the material as in martensitic transformations.

Dispersion hardening is a strengthening method based upon introducing a fine dispersion of insoluble particles usually oxides. This method is employed in engineering alloys processed by the powder method.

Increasing the strength by secondary phase particles can be explained as follows. When a moving dislocation approaches a secondary phase particles, it can either bypass the particles or cut through the particle depending upon the mechanical strength of the particle. Bypassing occurs if the particles are rather hard and cannot be sheared as in dispersion hardening. In this case, the shear stress \( \tau \) required to bypass the dislocation is approximately given by

\[
\tau = \frac{G b}{d}
\]

where \( G \) is the shear modulus of the matrix phase, \( b \) is the Burgers vector of the dislocation, and \( d \) is the interparticle spacing. Evidently, the stress required to move the dislocation increases with decreasing \( d \). As dislocations bypass the particle, they multiply, leaving behind dislocation loops enveloping the particle and decreasing the effective interparticle spacing \( d \) and in turn increasing the stress \( \tau \).

If the particles are relatively soft, a moving dislocation can shear the particle under the influence of a stress \( \tau \) given by

\[
\tau = \frac{\pi g r}{2 b d}
\]
where $g$ is the surface energy associated with the surface imperfection created by shearing the particle, $r$ is the particle radius, $b$ is the Burgers vector of the dislocation, and $d$ is the interparticle spacing.

In the case of a structural change as in a disorder-order transformation and martensitic transformation, strengthening can be produced by the limited number of slip systems in the resulting structures. Strengthening by refining the grain size for low-temperature applications is based upon the effectiveness of a grain boundary as a barrier to dislocation motion as described earlier. However, it is recalled that for high-temperature creep applications a large grain size is preferred.

Generally, any method used to increase the strength of a given material is associated with a corresponding decrease in ductility. Therefore, various strengthening methods must be controlled to obtain an optimum combination of strength and ductility. Usually, a given engineering alloy is strengthened by at least two mechanisms. Solid-solution strengthening is almost always applicable to all engineering alloys. Most of the effective strengthening methods, however, are based upon a phase change which can be affected by heat treatment processes, as described below.

### 7.16.1 Heat Treatment

By definition, a heat treatment process consists of heating and cooling cycles designed to affect specific structural changes to promote certain properties. Structural changes produced by a heat treatment process can be classified into two main types: (i) recovery and recrystallization, which are collectively called annealing processes, and (ii) phase changes.

Both recovery and recrystallization are processes by means of which the effects of plastic deformation (cold work) can be eliminated either partially (recovery) or completely (recrystallization). Recrystallization is also used to control the grain size.

Recovery is the process by which dislocations in a cold worked material are permitted to rearrange themselves into a configuration of lower energy. During this process, dislocations of opposite sense cancel each other, and dislocations of the same sign arrange themselves into walls or subgrain boundaries. A recovery heat treatment is carried out at a temperature below that required to form a new set of grains by the process of recrystallization described below.

As its name implies, recrystallization involves the nucleation and growth of a new set of grains within the material. It is carried out by heating to a temperature higher than that required for recrystallization. An inverse relationship exists between the minimum recrystallization temperature and the amount of cold work. Increasing the amount of cold work reduces the minimum recrystallization temperature. For a given material, the final grain size produced by recrystallization is
influenced by (i) the amount of cold work and (ii) the heat treatment temperature and the exposure time at that temperature. Increasing the amount of cold work reduces the grain size produced by heating to a given temperature for a given time. Conversely, for a given amount of cold work, the grain size increases with increasing the temperature and exposure time.

Heat treatment processes designed to precipitate secondary phase particles (precipitation hardening) consist of first heating the material to a sufficiently high temperature so that all or most elements go into solid solution, i.e., produce an essentially single-phase material. Subsequently, the material is rapidly cooled, e.g., by quenching in water or other media. Finally, the material is heated to an intermediate temperature and held at that temperature for a given time determined by the particular phase to be precipitated. Finally, the material is cooled to ambient temperature.

In steels, the heat treatment processes are designed on the basis of the respective TTT diagram of the steel. It is recalled from earlier discussion that the TTT diagram reflects the combined effects of thermodynamics and kinetics on a given phase change. To illustrate the principles involved in the heat treatment of steels, consider the TTT diagram of a steel containing 0.2% C, schematically shown in Fig. 7.40. Commonly, the temperature at which the austenite

![TTT Diagram](image)

**Figure 7.40** The TTT diagram of a steel containing 0.2% C.
begins to transform into ferrite is called the $A_3$ temperature, and the eutectoid temperature is labeled $A_1$.

First, the steel is heated to a temperature where it consists entirely of austenite. Subsequently, it is cooled at a rate dependent upon the level of hardness desired. It is recalled that equilibrium phase changes can be avoided by cooling rates faster than the critical cooling rate.

If the steel is cooled at a rate slower than the critical cooling rate, the austenite is transformed into ferrite and pearlite. However, the resulting grain size is dependent upon the exact cooling rate. Increasing the cooling rate reduces the grain size, and vice versa. Eventually, the transformation of austenite into ferrite and pearlite can be avoided when the cooling rate becomes fast enough to exceed the critical cooling rate. This behavior can be attributed to the insufficient time available for the atoms to diffuse and rearrange themselves into the structures of ferrite and pearlite. Avoiding the equilibrium transformation by rapid cooling places the steel in a state of metastable equilibrium. To lower its free energy, the steel transforms into other phases of lower free energy. Depending upon the exact cooling rate, the austenite transforms into either bainite or martensite.

If the cooling rate is relatively slow such that the transformation occurs at a higher temperature, the resultant phase is called bainite. It resembles pearlite in that it consists of a mixture of ferrite and cementite. However, the structure of bainite is considerably finer than that of pearlite. Instead of forming as alternating lamellae of cementite and ferrite, the carbide phase in bainite assumes the morphology of finely dispersed particles in a ferrite matrix consisting of elongated lathelike crystals because the large driving force causes the transformation to occur very rapidly in the absence of diffusion; i.e., the transformation is diffusionless. Bainite is stronger and harder than pearlite.

Provided the cooling rate is fast enough to avoid the region of bainite transformation, austenite is transformed into another phase called martensite. Similar to the bainite transformation, the martensite transformation is characterized by being diffusionless. It occurs by shear similar to plastic deformation processes, and therefore it proceeds very rapidly independent of time but dependent on temperature. During rapid cooling, the temperature at which the transformation starts is indicated by the $M_S$ temperature in the TTT diagram, and the temperature at which the transformation is completed is indicated by the temperature $M_F$.

Because of the nature of martensite transformation, all the C contained in austenite is trapped into the martensite, and therefore martensite is an extremely hard and brittle phase. Typically, martensite has a platelet type morphology similar to bainite. Since the transformation of austenite into either bainite or martensite occurs under nonequilibrium conditions, it is not indicated on the Fe-C phase diagram shown in Fig. 7.39.

Since the critical cooling rate determined from the TTT diagram determines the response of the steel to be hardened by rapid cooling, it is considered...
to be a measure of what is known as hardenability. Since martensite is an extremely hard and brittle phase, it is of no engineering significance. To improve the ductility of the steel, it is reheated to an intermediate temperature to release some of the carbon trapped in martensite. Such a heat treatment is called tempering. It results in the decomposition of martensite in various stages which ultimately leads to the formation of spheroidal particles of cementite in a ferrite matrix.

Hardenability is considered to be one of the most important characteristics of a steel. When a steel part is hardened by rapid cooling or quenching from the austenitic region, the surface is cooled more rapidly than the interior of the part; i.e. the surface is hardened first followed by the interior of the part. Whether the steel has a high or low hardenability is determined by measuring the hardness as a function of depth. A steel of a high hardenability exhibits a slight variation of hardness with depth. Conversely, a considerable variation of hardness with depth is observed if the steel has a low hardenability.

In practice, hardenability is measured by means of a standard test known as the Jominy test. In this test a standard cylindrical specimen is first heated to the austenitic region and then suspended in a fixture containing a nozzle, which sprays water at one end of the specimen for a certain time. After the specimen is completely cooled to room temperature, thin flat strips are ground on two diametrically opposite sides of the specimen and the hardness is measured at the quenched end and in regular intervals of one-sixteenth of an inch, and then the hardness is plotted as a function distance from the quenched end. Hardenability is influenced by a number of variables, particularly (i) grain size of austenite, (ii) carbon content, and (iii) alloying elements.

Since pearlite is nucleated at the grain boundaries of austenite, increasing the grain boundary area per unit volume of austenite; i.e., decreasing the grain size of austenite facilitates the transformation. Consequently, hardenability is increased as the grain size of austenite becomes coarser. It is possible to coarsen the grain size of austenite by heating to a higher temperature within the austenitic region. However, quenching from a very high temperature can cause cracks to develop in the steel, and therefore this method of increasing hardenability is not commonly practiced.

Increasing the C content of the steel has the effect of shifting the TTT diagram shown in Fig. 7.40 to the right, corresponding to an increase in hardenability. Therefore, steels of relatively higher C content are easier to harden by heat treatment in comparison with steels containing very small concentrations of C. Hardenable carbon steels usually contain from 0.35 to 0.95% C.

Addition of alloying elements to carbon steels can have significant effects on hardenability. Most of the structural alloy steels are of the low-alloy grades where the concentration of alloying elements is generally less than 5% and frequently much less. Alloying elements in these steels include Mn, Si, Ni, Cr, Mo, and V. A low-alloy steel contains one or more of these elements; however,
both Mn and Si are considered as alloying elements if present in concentrations greater than those in the corresponding carbon steel. Most of the low-alloy steels contain 0.35–0.60% C and therefore they are hardenable. Since all of the above elements suppress the decomposition of austenite into pearlite upon cooling, low-alloy steels have a better hardenability than the corresponding carbon steels. When a low-alloy steel is not hardened, its microstructure is similar to that of carbon steel, i.e. it consists of ferrite and pearlite; however, its grain structure is finer because the transformation of austenite occurs at lower temperatures.
8 Materials Characterization

8.1 Introduction

Materials characterization is almost always involved in failure analysis investigations. It requires knowledge of the overall composition of the material, as well as its microstructural features. Chemical composition refers to the elemental constituents of the material and their concentrations. By definition, microstructure refers to the internal structural features of a material which cannot be visually observed. They are to be distinguished from macrostructural features such as large cracks, porosity, and fracture characteristics which can be observed either visually or by using a low-power hand lens.

Complete microstructural characterization of a given material requires determination of (i) morphology (size, shape, and distribution of all phases present), (ii) crystal structure of all phases present, and (iii) elemental composition of all phases present, including localized segregation and depletion. Since some techniques for chemical analysis are associated with certain methods used to reveal the microstructure, techniques for microstructural characterization are described first. Various techniques for chemical analysis are described in Sec. 8.3. As described below, there are various techniques available for microstructural characterization. Selection of one technique or another depends upon the type of information required.

8.2 Techniques for Microstructural Characterization

Microstructural features of a material are revealed by means of a high-power microscope. Its primary function is to form an image of an object, as schematically.
Illustrated in Fig. 8.1. To observe the microstructure of a material, a representative specimen is illuminated by a suitable source of radiation. Radiation emitted by the specimen is then gathered by an objective lens and focused into an image. Other lenses can be used to magnify the image formed by the objective lens.

Various techniques for microstructural characterization can be conveniently classified according to the type of radiation used to illuminate the object. Two types of radiation are used, namely, (i) visible light (electromagnetic radiation) and (ii) electrons. Figure 8.2 summarizes the techniques used for microstructural characterization classified according to the type of radiation used to illuminate the object.

Each of the techniques listed in Fig. 8.2 is characterized by a *resolving power*, or *resolution*. It is defined as the smallest distance between two point objects which can be resolved as separate and distinct. Comparative useful resolving ranges of various techniques are illustrated in Fig. 8.3. In general, the resolution of electron microscopy is superior to that of light optical microscopy. However, it must be realized that various techniques for microstructural characterization complement rather than compete with each other. Selection of one technique or another for analysis is dependent upon the scale of information required. A brief account for each technique, including its capabilities, applications, and limitations, is given below.

### 8.2.1 Light Optical Microscopy

Light optical microscopy is an extremely useful technique for characterizing the overall or gross microstructural features of a material such as grain structure (grain size, shape, and orientation). A typical light optical microscope consists of (i) light source, (ii) specimen stage, (iii) a glass objective lens, (iv) glass magnifying lenses, (v) viewing screen, and (vi) an image recording system. In this technique, the microstructure is revealed by reflection of light radiation from the surface of a bulk specimen.
Specimens for examination by light optical microscopy are prepared in three successive stages: (i) grinding, (ii) polishing, and (iii) etching. First, the surface of the specimen is ground with successively finer grits of silicon nitride starting with 80 grit followed by 180, 320, and finally 600 grit. Second, the surface is intermediately polished using a 6 and 1 μm diamond paste. Final polishing is carried out using a suspended solution of 0.05 μm gamma-alumina. A proper etchant (chemical reagent) is then applied to the surface to outline the gross microstructural features.

Figure 8.2 Various techniques for microstructural characterization, classified according to the type of radiation used to illuminate the object.

Figure 8.3 Comparative useful resolving ranges of various techniques for microstructural characterization.
Etching is essentially a chemical reaction between the material of the specimen and the etchant resulting in dissolution of the material. However, depending upon the etchant used, this dissolution does not occur uniformly over the entire surface. Since atoms at grain boundaries are in a higher energy state than atoms within the grains, preferential etching or dissolution can occur at grain boundaries, making them appear as shallow steps at the surface of the specimen. Also, it is possible that the rate of dissolution varies with the orientation of the grains and their compositions.

To summarize the above discussion, light radiation reflected across the surface of the specimen varies in intensity according to changes in composition, structure, or by local changes in surface angles due to preferential etching. Such variations in intensity give rise to contrast in light optical images.

Specimens can be examined in the as-polished and etched conditions. Generally, it is not possible to observe the grain structure in the as-polished condition. However, images derived in the as-polished condition can be useful in observing certain microstructural features. Examples of such features include secondary phase particles considerably harder than the surrounding matrix, leading them to stand in relief (Fig. 8.4a). Also, voids and microcracks are clearly visible in the as-polished condition. As shown in Fig. 8.4b, the grain structure can be observed after etching due to preferential etching at the grain boundaries and around precipitate particles.

Reference to Fig. 8.3 indicates that the resolving power of light optical microscopy is limited to about 0.2 \( \mu m \); i.e., objects smaller than about 0.2 \( \mu m \) cannot be resolved in light optical images. Also, light optical microscopy is characterized by a relatively small depth of field, limiting the thickness of surface layer from which a focused image can be derived to about 0.1 \( \mu m \). Typically, useful light optical images can be obtained at magnifications ranging from 10 to 2000\( \times \). Therefore, if it is required to resolve finer microstructural details, the more powerful techniques of electron microscopy described below must be utilized. Despite the above limitations, light optical microscopy is very useful in obtaining low resolution view of microstructure illustrating the overall or gross microstructural features. One of the most important microstructural parameters influencing many properties particularly mechanical strength is grain size. Light optical microscopy is used to determine grain size. Following ASTM specifications, the number of grains within area of 1 in\(^2\) at \( \times \)100 magnification is counted to obtain a number reflecting the grain size, as summarized in Table 8.1. As can be seen, the grain size decreases as the ASTM grain size number increases.

Generally, it is always recommended that light optical microscopy precedes the higher resolution techniques so that the proper resolution level is selected to reveal specific microstructural features.
Figure 8.4  (a) Image derived from a 304 stainless steel sample showing secondary phase particles and voids. (b) Image showing relative orientations of grains in a Ni–22 Cr–9Mo alloy.
8.2.2 Scanning Electron Microscopy (SEM)

In a scanning electron microscope, the specimen is illuminated by scanning its surface with a fine electron beam accelerated through a potential difference of up to 50 kV. However, metallic materials are usually examined at an accelerating voltage of 20–30 kV. A typical scanning electron microscope consists of (i) electron source, (ii) electromagnetic condenser lenses, (iii) scanning coils, (iv) specimen chamber, (v) an electromagnetic objective lens, (vi) magnifying lenses, (vii) secondary electron detector, (viii) backscattered electron detector, (ix) image viewing system, and (x) image recording system. Among the optional attachments is an X ray spectrometer for microchemical analysis. Usually, a tungsten filament is used as an electron source; however, other filaments with better capabilities, such as a single crystal of LaB₆, are also available.

When the filament is heated by passing through it an electric current, it emits electrons. In this manner, the filament acts as a cathode. Electrons emitted by the filament are accelerated through a potential difference toward an anode. Subsequently, the electrons are focused into a fine beam on the specimen surface using the condenser lens. Since electrons are easily scattered by air, the components of the microscope are enclosed in an evacuated column.

Similar to light optical microscopy, a bulk specimen can be examined. In contrast, prior specimen preparation is minimal and as-received specimens can directly be examined provided that they have relatively high electrical conductivity, e.g., metallic materials. Nonconducting specimens and presence of poorly conducting species in an otherwise conducting specimen can lead to charging of the surface, making it difficult to focus the image. In this case, it becomes

<table>
<thead>
<tr>
<th>ASTM number</th>
<th>Number of grains per square inch at ×100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
</tr>
<tr>
<td>9</td>
<td>256</td>
</tr>
<tr>
<td>10</td>
<td>512</td>
</tr>
</tbody>
</table>
necessary to coat the surface with a thin layer of a conducting material, e.g., a carbon or a gold film. However, some of the recent versions of scanning electron microscopes are designed to permit nonconducting materials to be directly examined in the as-received condition without the need to coat the surface.

In comparison with about 0.2 μm resolution of light optical microscopy, the resolution of scanning electron microscopy is reduced to about 0.05 μm, permitting finer microstructural details to be resolved. Also, the depth of field of scanning electron microscopy dependent upon magnification is about 30 μm at ×1000 in comparison with 0.1 μm for light optical microscopy. Scanning images can be obtained at magnifications up to ×200,000–×250,000. Because of these advantages over light optical microscopy, scanning electron microscopy has a wider scale of applications in materials science, such as studying fracture surfaces, oxidation processes, and electrodeposition.

When the incident electrons interact with the specimen in a scanning electron microscope, they are scattered and spread into an onion-shaped volume of interaction about a few micrometers in size depending on the applied voltage and the type of material, as schematically shown in Fig. 8.5a. Interaction between the incident electron beam and specimen results in various types of signals which can be processed into useful microstructural information. Signals of particular importance include (i) secondary electrons, (ii) backscattered electron, and (iii) X rays.

Both secondary and backscattered electrons can be processed into images providing microstructural information. Each signal is displayed in synchronization on a cathode-ray tube (CRT) so that an image of the total scanned area is obtained with a magnification corresponding to the ratio of the CRT area to the scanned area of the specimen. Characteristic X rays emitted by the specimen provide information about the local chemical composition, as described later in this chapter. Figure 8.5b illustrates the energy distribution among the electrons emitted by the specimen. It is a plot of the number of electrons \( N(E) \) emitted with a given energy \( E \) vs. energy. Two peaks are observed in the energy distribution. One peak covers energies in the range of 0–50 eV, defining secondary electrons. Another peak exists at a higher energy slightly less than that of the incident beam \( E_0 \), which defines backscattered electrons.

Secondary electrons result from interaction between electrons in the incident beam and electrons in the specimen. Low-energy secondary electrons (<50 eV) are emitted from surface regions about 0.01 mm in thickness. Such electrons are those residing in outermost shells, therefore weakly bound to the nuclei of the specimen. Because secondary electrons are created before the electron beam is spread within the specimen, secondary electron images have a greater spatial resolution in comparison with backscattered images. A distinctive feature of secondary electron images is their sensitivity to surface topography. Brightness of secondary images varies according to changes in the angle \( \theta \).
between the incident beam and the local normal to the specimen. As the angle \( \theta \) increases, the number of secondary electrons increases, and vice versa. Accordingly, surface features whose local normals are inclined at larger angles to the incident beam appear brighter than those inclined at smaller angles. Essentially, the contrast in secondary images is the same as that in light optical images making them easy to interpret visually. Because of their sensitivity to surface topography and high spatial resolution, secondary electron images are frequently used to

---

**Figure 8.5**  (a) Schematic illustrating the interaction of electron beam with the specimen. (b) A plot of energy distribution of secondary electrons (SE) and backscattered electrons (BE) emitted from the sample.
characterize surface morphology such as fracture surfaces and corrosion products. However, they carry little information about chemical composition which can be found in backscattered images as described below.

In contrast to secondary electrons, backscattered electrons result from scattering of incident beam electrons by the nuclei in the specimen. Since the energy of backscattered electrons is considerably higher than that of secondary electrons, they are emitted from deeper regions in the specimen. As a result, backscattered electrons carry much less information about surface topography and have less spatial resolution in comparison with secondary electrons. However, the intensity of backscattered electrons is a sensitive function of atomic number $Z$, and thus they contain important information about local chemical composition. With increasing atomic number and in turn the positive charge of the nuclei, the density of backscattered electrons increases. The regions in the specimen of a higher average atomic number, e.g., second-phase particles, exhibit a brighter contrast in comparison with regions of lower atomic number. Backscattered electrons carrying chemical composition information can be separated from those carrying topographic information by means of suitable detectors, giving rise to backscattered electron composition images and backscattered electron topographic images. Backscattered electron composition images are thus useful in selecting regions for microchemical analysis.

In general, the clarity of backscattered electron topographic images is considerably reduced in comparison with that of secondary electron images. Typically, the brightness of a backscattered topographic image is uneven because the detector must be situated such that secondary electrons are excluded from contributing to the image. Therefore, very few or no backscattered electrons are collected from those regions which are tilted away from the detector.

To summarize, microstructure is revealed by three types of images in a scanning electron microscope, namely, (i) secondary electron image; (ii) backscattered electron composition image, and (iii) backscattered electron topographic image. However, because of their higher quality, secondary electron images are preferred for surface morphology studies.

8.2.3 Transmission Electron Microscopy (TEM)

Transmission electron microscopy is not used as frequently as scanning electron microscopy in failure analysis studies; however, it can assume a special importance in solving certain problems. Similar to a scanning electron microscope, a transmission electron microscope consists of an electron source; a set of electromagnetic lenses, including condenser; objective, and magnifying lenses; specimen chamber; viewing screen; and an image recording system all enclosed in an evacuated column.

Unlike the case of scanning electron microscopy where an electron image is formed by electrons emitted from the specimen surface, transmission electron
microscopy is based upon forming an image by electrons passing through the specimen and accelerated through a potential difference which can be as high as 1000 kV. Usually, however, metallic materials are examined at 100–200 kV. Due to the low penetrating power of electrons through solid objects, the applications of TEM are limited to specimens in the form of thin films or foils transparent to electrons. In practice, the thickness of a specimen transparent to electrons is a function of the accelerating voltage. For example, at 100 kV, the useful thickness of specimen is about 0.1 mm. As the accelerating voltage is increased to, e.g., 200 kV, the penetrating power of electrons is increased by about 1.6 times, reducing the useful thickness of specimen.

When a thin crystalline specimen is irradiated by an energetic electron beam, electrons emerging from the bottom surface are distributed between (i) transmitted electrons which pass through in the same direction as the incident beam and (ii) diffracted electrons, which are scattered in well defined directions depending upon the crystal structure of the specimen. In addition, the specimen emits secondary and backscattered electrons as well as characteristic X-rays similar to the case of SEM.

A distinguishing feature of TEM is its superior resolution over other techniques for microstructural characterization (see Fig. 8.3). At an accelerating voltage of 100 kV, the point-to-point resolution is about 0.35 nm, and is reduced to about 0.3 nm at 200 kV. One of the specialized techniques of TEM is high-resolution lattice imaging. In this case, an ultimate resolution of 0.1–0.2 nm can be realized. At 100 kV, the depth of field in TEM is about 350 nm considerably greater than the thickness of a specimen transparent to electrons (100 nm). Physically, this means that both the top and bottom surfaces of a TEM specimen can be focused simultaneously. Therefore, a TEM image is essentially a two-dimensional projection of the three-dimensional features of the specimen.

Typically, TEM images can be obtained at magnifications as high as 500,000×. To realize a resolution of 0.35 nm, the magnification must be about 300,000×. Formation of a transmission electron microscopy image is based upon the phenomena of electron diffraction by crystals, as described below.

When the incident electron beam passes through the specimen, a portion is scattered in different directions by the massive nuclei of the specimen and the remainder passes through undeviated. Along certain directions defined by the positions of atoms in the crystal, the scattered electrons combine and intensify to form a diffracted beam. No diffracted beams are formed along other directions. By means of an objective lens, both the transmitted and diffracted beams can be focused to form a diffraction pattern, reflecting both the geometry and size of the crystal. Directions along which diffraction occurs are defined by the well-known Bragg’s law:

$$\lambda = 2d \sin \theta$$
where $\lambda$ is the wavelength of the incident electron beam, $d$ is the spacing of atomic planes diffracting the electrons, and $\theta$ is the incident angle of the electron beam, as schematically illustrated in Fig. 8.6. From a consideration of atoms positions in the crystal it can readily be shown that for a face-centered cubic crystal, diffracted beams are formed only by sets of planes of unmixed indices; i.e., all indices are either even or odd numbers (0 is considered an even number). These are called allowed reflections such as (111), (200), (220), (311), etc. For a body-centered cubic crystal, the allowed reflections are those for which $h + k + l =$ even number, e.g., (110), (200), (211), (222), etc. In the case of a hexagonal close-packed crystal, the allowed reflections are those for which:

- $h + 2k = 3n$, $l$ is an odd number
- $h + 2k = 3n$, $l$ is an even number
- $h + 2k = 3n + 1$, $l$ is an odd number
- $h + 2k = 3n + 1$, $l$ is an even number

where $n = 0, 1, 2, 3, 4, \ldots$.

If only one grain or crystal contributes to diffraction, a spot pattern such as that shown in Fig. 8.7a is obtained. As can be seen, the pattern consists of a geometrical arrangement of spots containing a central intense spot corresponding to the transmitted beam and number of weaker spots corresponding to the diffracted beams. A diffraction pattern such as that shown in Fig. 8.7a is a magnified image

![Figure 8.6](image)

**Figure 8.6** Schematic illustration of Bragg’s law of diffraction.
of the pattern formed at the back focal plane of the objective lens. It can readily be shown that distances measured from the central spot to diffracted spots are related to spacings of crystal planes by the relationship

$$\lambda L = rd$$

where $\lambda$ is the wavelength of the incident electron beam, $L$ is a microscope parameter called the camera length corresponding to the distance between the specimen and the viewing screen, $r$ is the distance from the transmitted spot to a given diffracted spot, and $d$ is the spacing of the set of atomic planes giving rise to the diffracted spot. For a given diffraction pattern, $AL$ is a known constant fixed by the accelerating voltage known as the camera constant. It is evident from the above relationship that from measurements of the distance $r$, the spacing $d$ of atomic planes giving rise to the respective diffracted spot can be calculated. If the crystal structure of the specimen is known, it is possible to index the diffraction pattern by assigning the proper indices to the diffracted spots and calculate their spacings ($d$ values). Indices of the transmitted spot are always (000). For example, in the case of a cubic crystal, the spacing $d$ of a set of ($hkl$) planes is

Figure 8.7  (a) Spot diffraction pattern obtained from a single crystal orientated at [100]. (b) Ring diffraction pattern obtained from a multigrained material.
given by

\[ 1/d^2 = (h^2 + k^2 + l^2)/a^2 \]

where \( a \) is the lattice constant.

Since the pattern shown in Fig. 8.7a is derived from a face-centered cubic crystal, the allowed reflections are (111), (200), (220), etc. By combining this information with the above relationships, the indices are assigned. In practice, it is only required to index the two nearest spots to the origin, and the remainder can be obtained by simple additions. Complete indexing of a spot pattern requires knowledge of its orientation \([uvw]\), i.e., the direction normal to the pattern. This can simply be obtained from the indices of any two spots in the pattern such as \((h_1k_1l_1)\) and \((h_2k_2l_2)\), as follows:

\[
\begin{align*}
  u &= (k_1l_2 - l_1k_2) \\
  v &= -(h_1l_2 - l_1h_2) \\
  w &= (h_1k_2 - k_1h_2)
\end{align*}
\]

The orientation of the pattern shown in Fig. 8.7a is [100]. When an aggregate of fine crystals contribute to an electron diffraction pattern, spots located at equal distances from the transmitted spot merge to form a ring, as shown in Fig. 8.7b. Indexing of a ring pattern is the same as that of a spot pattern with the distance \( r \) corresponding to the radius of a ring.

If the crystal structure of the specimen is unknown, several selected-area spot patterns must be derived in different orientations and a set of \( d \) spacings is obtained and matched with those in the ASTM card file, similar to the case of X ray diffraction experiments. However, from the geometries of spot patterns derived in different orientations, it is possible to make an educated guess about the geometry of the crystal. This difficulty is to a great extent eliminated in modern microscopes equipped with advanced lens systems permitting diffraction patterns containing three-dimensional information to be derived. From a single convergent beam diffraction pattern, it is possible to determine the geometry and size of the crystal. A convergent beam pattern derived from a face-centered cubic crystal is shown in Fig. 8.8. As can be seen, the pattern consists of a central spot layer similar to a selected-area pattern surrounded by a number of rings, each of which contains sets of diffracted beams. Orientation of the pattern can be determined from indexing the central spot layer similar to the case of selected-area diffraction described above. In this example, the orientation of the pattern is [100]. To obtain three-dimensional information, the distances \( r \) from the transmitted spot to the rings are measured. It can readily be shown that the spacing \( H \) of crystal planes...
along the direction of the incident beam is given by

\[ H = 1/(rL/\lambda (\lambda/2)) \]

To confirm that the crystal is cubic, the lattice constant \( a \) determined from the central square array of spots must be the same as the value of \( H \). If the structure is unknown, selection of the proper orientation of the pattern where the spots of the central layer are most closely spaced, i.e., the lowest index pattern, can greatly facilitate the determination of the geometry and size of the crystal from a single pattern.

Microstructure can be revealed in the TEM by various types of images depending upon the electron beam(s) selected to form the image. Conventionally, the microstructure of a crystalline specimen is revealed by bright- and dark-field

**Figure 8.8** Convergent beam pattern with [100] orientation obtained from a face-centered cubic crystal.
images with the aid of the corresponding selected-area diffraction pattern, as described below.

If the transmitted and diffracted beams emerging from the bottom surface of the specimen are focused on the back focal plane of the objective lens, selected-area diffraction patterns such as those shown in Fig. 8.7a,b are observed on the viewing screen. By definition, a TEM image results from magnifying one or more of the spots or rings shown in Fig. 8.7a,b. Only one spot or ring is used to form a conventional TEM image with the aid of an objective aperture. Magnifying the transmitted spot results only in a bright-field image. It is also possible to magnify only one of the diffracted spots or rings to produce a complementary dark-field image of the same area of specimen. Contrast in such images is known as amplitude or diffraction contrast. In a bright-field image, any defect in the crystal resulting in a localized change in its periodicity appears dark on a bright background.

Contrast is reversed in a dark-field image; i.e., the defect appears bright on a dark background. For example, Fig. 8.9a illustrates a bright-field image of a second phase particle in a nickel-base alloy. As can be seen, the particle appears dark on a bright background. This image was formed by the transmitted spot in the selected-area diffraction pattern (Fig. 8.9b) derived from the particle. If a diffracted spot in Fig. 8.9b is used to form the image, the contrast is reversed; i.e., the particle appears bright on a dark background, as shown in Fig. 8.9c.

High-resolution images can be produced using both the transmitted spot and one or more of the diffracted spots in forming the image. Such images, known as lattice images, consist of sets of fringes (alternating parallel and bright lines) directly exhibiting the periodicity of the crystal. A one-dimensional lattice image is produced if the transmitted spot and one or more diffracted spots corresponding to one set of crystal planes are used to form the image, as shown in Fig. 8.10a. A lattice structural image is formed if two sets of diffracted spots in addition to the transmitted spot are used to form the image, as shown in Fig. 8.10b.

Because of its superior resolution, TEM is an extremely powerful technique for characterizing defects in crystals. Some of the typical materials science applications include studying plastic deformation processes in crystalline materials, phase transformation, and phase identification and characterization of fine precipitates as illustrated by the following examples.

Dislocations responsible for many of the useful properties of metallic materials appear as straight lines in transmission electron microscopy images, as shown in Fig. 8.11a. Thin planar defects such as stacking faults appear as alternating dark and bright fringes, as shown in Fig. 8.11b. Various types of interactions between secondary phase particles and dislocations can be identified, as illustrated in Fig. 8.11c–e. In many cases, metallic materials are strengthened
8.2.4 Scanning Transmission Electron Microscopy (STEM)

Scanning transmission electron microscopy combines the advantages of SEM and TEM. Similar to a TEM, a STEM utilizes a thin specimen transparent to electrons, and as in the case of SEM, the specimen is scanned by a fine electron beam. Both bright- and dark-field STEM images can be obtained as in the case of TEM. However, since the electron probe used in STEM is much finer, it is possible to derive diffraction patterns from much smaller areas in comparison with TEM. Contrast in STEM images is similar to that in TEM images. The advantage of STEM is demonstrated by the bright-field image shown in Fig. 8.12a, where fine-sized precipitates, which can only be revealed by transmission electron microscopy. An example is given in Fig. 8.11f.

Figure 8.9 (a) Bright field image of a second phase particle. (b) Corresponding diffraction pattern. (c) Dark field image of the same particle.
Figure 8.10  (a) A lattice image formed by using transmitted spot and one set of diffracted spot. (b) A lattice structural image formed by using transmitted spot and two sets of diffracted spots.
Figure 8.11  (a) Bright field TEM image of dislocations. (b) Bright field TEM image of a stacking fault. (c)–(e) Examples of interaction between second phase particles and dislocations. (f) Bright field TEM image of fine γ” phase in a Ni–22Cr–9Mo–3.5Nb alloy.
Figure 8.11 Continued.
precipitates are clearly visible. The diffraction pattern obtained from this region is shown in Fig. 8.12b, and the corresponding dark field image in Fig. 8.12c, which highlights the very fine precipitates present in the bulk alloy. These precipitates were imaged with the spot shown encircled and labeled 1 in the diffraction pattern. The elemental information obtained from the region is shown in Fig. 8.12d.

Formation of a STEM image is similar to that of an SEM image; i.e., the signal is displayed in synchronization on a cathode-ray tube (CRT) so that an image of the total scanned area is obtained with a magnification corresponding to the ratio of the total CRT area to the scanned area of the specimen. Generally, the quality of a TEM image is superior to that of a STEM image. In materials science applications STEM images do not provide additional information over that provided by TEM images. However, STEM images are useful for selecting fine microstructural features for analysis because of utilizing a finer electron probe.

Figure 8.12  (a) Bright field STEM image of Fe-Cr-Mo alloy exhibiting fine precipitation. (b) Selected area diffraction pattern from the area. (c) Corresponding dark field image. (d) EDS spectrum obtained from the region.
Some of the modern TEM instruments are fitted with STEM capabilities and in some cases with SEM, in addition to an X-ray spectrometer for microchemical analysis. An instrument of this type is known as an analytical electron microscope (AEM) and can be operated in the TEM, STEM, and SEM modes. Because of specimen size limitations, specimens about 3 mm in diameter and about 0.5 mm in thickness can only be examined in the SEM mode. However, since an accelerating voltage as high as 200 kV can be utilized, SEM analysis can be carried out at a much higher resolution level in comparison with conventional SEM described earlier.

### 8.3 Techniques for Chemical Analysis

Chemical analysis can broadly be classified into two main types: (i) elemental analysis and (ii) phase analysis. Elemental analysis refers to identifying the overall elemental constituents of a given material and their concentrations. As pointed out earlier, a material can either be homogeneous, i.e., consisting of a single phase, or heterogeneous, i.e., consisting of more than one phase. A single-phase material may undergo a phase change resulting in its separation into two phases, each of which has a given elemental composition and crystal structure, e.g., the transformation of austenite into ferrite and cementite. However, the overall elemental composition remains unchanged. Transformation of austenite into martensite is an example where the phase change is accompanied only by a change in crystal structure. It is then evident that elemental analysis does not provide any information about the state of presence of the elemental constituents of the material. This information is provided by phase analysis, i.e., identifying the phases present in the material.

Many failure analysis studies involve chemical analysis including elemental and phase analysis, e.g., material verification, identification of secondary phases, and analysis of corrosion deposits. However, the level of accuracy required can significantly vary from one case to another. Therefore, the analyst must be familiar with the various techniques available for chemical analysis and their limitations. According to the volume of material analyzed, it is possible to classify those techniques into two main types: (i) techniques for bulk chemical analysis, where the entire volume of a bulk specimen is analyzed, and (ii) techniques for microchemical analysis, where a selected micro feature of a specimen is analyzed. Techniques for microchemical analysis are typically combined with microscopy techniques, particularly scanning, transmission, and scanning transmission electron microscopy.

There are many techniques available for elemental analysis. However, all techniques are based upon the same principles, as described below.
8.3.1 Basic Principles of Elemental Analysis

All techniques of chemical analysis used to determine the elemental composition of a given material are based upon analyzing the electron energy levels within the atom by a process called excitation. It involves the transition of electrons between allowed energy levels, as described below.

Electrons in atoms can move from one allowed energy level to another, which is referred to as electron transition. Since the outermost shells of elements other than the inert gases contain empty or partially filled energy levels, an electron can make a transition from, e.g., an s level to the higher energy p level, if provided with sufficient energy. In this state, the atom is said to be in an excited or high-energy state. To lower its energy, the electron falls back into the lower energy level. Simultaneously, the energy difference between s and p levels is emitted by the atom in the form of radiant energy, as schematically illustrated in Fig. 8.13.

By definition, radiant energy is defined as energy transformed in the form of radiation having characteristic wavelengths, particularly electromagnetic radiation of which visible light, ultraviolet light, and X rays are parts. If instead, the atom is provided with a much higher energy, it is possible to eject an electron from, e.g., the K shell, leaving behind an electron vacancy as shown in Fig. 8.14. Again the atom becomes excited or attains a higher energy state. However, in this case, the atom can lower its energy if an electron from the L shell falls into the K shell to fill that vacancy (Fig. 8.14). As a result of this transition, the atom also emits radiant energy $K_{\alpha}$ equivalent to the energy difference between the K and L shells, which is much higher than that resulting from transitions in the outermost shells. In practice, differences in energy levels are used to

Figure 8.13 Schematic illustrating the electron transitions within an atom between s and p energy levels resulting in the emission of radiant energy.
Identify the chemical element, i.e., to conduct chemical analysis, as described below.

In general, radiant energy such as electromagnetic radiation is expressed in units, each of which is called a photon whose energy $E$ is given by

$$E = h\nu$$

where $h$ is Planck’s constant ($h = 6.626 \times 10^{-34} \text{ J.s}$) and $\nu$ is the frequency of the radiation. Since electromagnetic radiation travels with the speed of light $c$ ($c = 2.998 \times 10^8 \text{ m/s}$), the frequency of the radiation $\nu$ is given by

$$\nu = c/\lambda$$

where $\lambda$ is the wavelength of the radiation. Therefore, the energy $E$ can be expressed as

$$E = hc/\lambda$$

and substituting for the constants $h$ and $c$ results in

$$E = (6.626 \times 10^{-34} \text{ J.s})(2.998 \times 10^{18} \text{ m/s})/\lambda$$

Typically, the energy $E$ is expressed in units of electron volts (eV) or kilo electron volts (keV), where $1 \text{ keV} = 103 \text{ eV}$ (1 eV = $1.602 \times 10^{-19} \text{ J}$), and the wavelength is expressed in units of Å (1 Å = $10^{-10} \text{ m}$). In terms of these units, $E$ is

---

**Figure 8.14** Schematic illustrating the electron transitions within an atom as a result of interaction with primary electrons that leads to the generation of $K_\alpha$ radiation.
given by

\[ E = \frac{(6.626 \times 10^{-34}/1.602 \times 10^{-19})(2.998 \times 10^8 \times 10^{+10})}{\lambda} \]
\[ = (12.4 \times 10^3)/\lambda \]

where \( E \) is expressed in units of keV and \( \lambda \) is expressed in units of Å.

It is evident from the above equation that the energy \( E \) is inversely related to wavelength \( \lambda \); i.e., as the energy increases, the wavelength decreases. It can be concluded from the above discussion that the energy of electromagnetic radiation \( E \) emitted by an atom as a result of its excitation can be expressed as

\[ E = E_2 - E_1 = (12.4 \times 10^3)/\lambda \]

where \( E_2 \) can be the energy of a \( p \) level and \( E_1 \) is the energy of \( s \) level if electron transition occurs in the outermost valency shells, or \( E_2 \) corresponds to the energy of a K shell electron and \( E_1 \) is the energy of an L shell electron if transitions occurs between the inner shells. Each form of electromagnetic radiation such as visible light, ultraviolet radiation, and X rays is characterized by certain wavelengths, as described below.

When the radiant energy emitted by an atom is resolved into its constituent wavelengths or energies, the result is called a spectrum. A spectrum produced by an atom as a result of electron transitions in the outermost valency shells usually lies within the range of visible light and ultraviolet radiation. However, a spectrum produced by electron transitions between the inner electron shells lies within the range of X rays.

Resolution of visible white light into its components results in a series of radiations each of which is characterized by a given color and corresponding wavelength, as summarized below.

Ultraviolet radiation has wavelengths in the range of 50–4000 Å, i.e., the longest waves of ultraviolet light have wavelengths just shorter than violet light. In contrast, X rays have extremely shorter wavelengths in the range of 0.5–2.5 Å.

<table>
<thead>
<tr>
<th>Color</th>
<th>Wavelength (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>6470–7000</td>
</tr>
<tr>
<td>Orange</td>
<td>5850–6470</td>
</tr>
<tr>
<td>Yellow</td>
<td>5750–5850</td>
</tr>
<tr>
<td>Green</td>
<td>4912–5750</td>
</tr>
<tr>
<td>Blue</td>
<td>4240–4912</td>
</tr>
<tr>
<td>Violet</td>
<td>4000–4240</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
As indicated earlier, electron transitions between the innermost electron shells produce X rays, and since their wavelengths are characteristic of the emitting atom, they are termed characteristic x rays. Characteristic X rays are identified by the electron shells from which electrons are ejected and those contributing electrons to fill the respective vacancies. For example, radiation resulting from filling an electron vacancy in the K shell is termed K radiation. If this vacancy is filled by an electron from the L shell, the radiation is termed Kα radiation. Radiation resulting from filling a K shell vacancy by an electron from the M shell is known as Kβ. It is evident that the energy of Kβ radiation is higher than that of Kα radiation; however, the intensity of Kα radiation is greater than that of the Kβ radiation. Typically, the intensity of Kβ radiation is about 0.10–0.25 that of the respective Kα radiation, which can be useful in differentiating between different elements. Likewise, the radiation resulting from filling a vacancy in the L, M, etc. shells are identified as Lα, Lβ, etc. and Mα, Mβ, etc., respectively, and so on.

An important property of the energy $E$ of characteristic X rays is its dependence on atomic number $z$ according to the following relationship:

$$E = c_1(z - c_2)^2$$

where $c_1$ and $c_2$ are constants for a given characteristic peak. As can be seen from the equation, the energy of a given characteristic radiation, e.g., Kα or Kβ, increases with the atomic number. Therefore, characteristic X rays act as fingerprints of the chemical elements leading to their identification.

Just as a specimen emits electromagnetic radiations having certain wavelengths and corresponding energies characteristic of the elements present, it can also absorb characteristic radiations according to the same principles described above. In practice, any technique for chemical analysis must provide means for

1. Excitation of atoms in the specimen to be analyzed using a suitable source of energy so that the specimen can emit electromagnetic radiations characteristic of its elements
2. Resolving the emitted spectrum into its constituent wavelengths or energies
3. Recording the emitted spectrum

Examination of the chemical composition by the above means is referred to as spectroscopy. An instrument used to resolve the emitted spectrum into its constituent wavelengths or energies is called a spectrometer. Spectrophotometry is a term reserved to describe the process by which the constituent wavelengths of light radiation are analyzed. A spectrum emitted by the specimen consists of a series of lines or peaks, each of which is called a spectral peak. Each spectral
peak has a characteristic wavelength and corresponding energy. By matching the observed wavelengths or energies with those of known elements, the elemental composition of the specimen is determined. This is known as *qualitative elemental analysis*. In principle, the intensities of spectral peaks of a given element are proportional to its concentration in the specimen. Depending upon the particular technique used for elemental analysis, various methods are available to convert the observed intensities in concentrations, which is known as *quantitative analysis*. Almost all of the above processes used to analyze the observed spectrum are performed with the aid of computer software that controls the operation of the system. Some of the techniques used for bulk elemental analysis are described below.

### 8.3.2 Techniques for Bulk Elemental Analysis

Several techniques are available to analyze the bulk chemical composition of metallic materials, i.e., the overall chemical composition of a representative specimen. Some techniques allow the sample to be directly analyzed in the solid-state. Other techniques, however, require the sample to be first converted in the liquid state by dissolution in a suitable acid.

Selected techniques used for bulk chemical analysis include

1. Absorption spectrophotometry
2. Optical emission spectroscopy
3. X ray fluorescence (XRF)
4. X ray diffraction (XRD)

Absorption spectrophotometry is one of the oldest methods used for chemical analysis and is still used by industrial laboratories. In this technique, a specimen of a predetermined mass is dissolved in an acid such as hydrochloric acid. Certain reagents are then added to the prepared solution of the specimen such that a given reagent produces substances of specific colors for each element present. Subsequently, the solution is allowed to absorb light radiation. As pointed out earlier, absorption of radiant energy is governed by the same principles as emission; i.e., only certain wavelengths are absorbed by a given element depending upon the energy levels of its atoms, particularly in the outermost or valency shells. Absorption of certain wavelengths is then measured using an ultraviolet/visible light (UV/VIS) spectrophotometer and the elements present in the specimen are identified one by one by matching the absorbed wavelengths with those of known elements. To quantify the results, the intensity of absorbed radiation is related to concentration by a relationship known as Beer’s law. Since in this technique, a solution of the sample is analyzed, this method among others is sometimes called *wet chemical analysis*.

Optical emission spectroscopy is more commonly used for elemental analysis. It includes a number of techniques, each of which is distinguished by the
excitation source. Each of these techniques is based upon analyzing the spectral peaks emitted by atoms of the specimen as a result of electron transitions within the outermost valency shells. One of those techniques utilizes a spark discharge as an excitation source. Another technique uses a plasma (a stream of an ionized gas consisting of equal numbers of positively charge ions and negatively charged electrons) as a source of excitation. In this technique, known as inductively coupled plasma (ICP), an ionized argon gas and its associated electrons are maintained at a very high temperature. A solid sample to be analyzed is first converted into a liquid by dissolution in a suitable acid. Subsequently, the solution of the sample is atomized by a process called nebulization and is then introduced into the plasma for excitation and emission of characteristic spectral peaks. This technique is particularly useful in analyzing residual elements present in very small concentrations, e.g., <0.01 wt%. It is to be noted that some failure modes of structural alloys have been related to segregation of residual elements at grain boundaries. In all of these methods, the light radiation emitted by the specimen passes through a slit into the spectrometer where it is dispersed into its constituent wavelengths by means of an optical grating and then recorded.

As its name implies, X ray fluorescence uses an energetic beam of X rays as a source of excitation. It is a commonly used technique for bulk chemical analysis. Because X rays have a much higher energy in comparison with the sources used in the above techniques, atoms in the specimen emit their characteristic X ray spectral peaks resulting from electron transitions between the innermost electron shells. In this technique, a solid sample can directly be analyzed. Characteristic X rays emitted by the specimen can be analyzed in terms of either its energy (energy dispersive X ray spectroscopy) or wavelength (wavelength dispersive X ray spectroscopy).

In the case of energy dispersive spectroscopy, the emitted radiation is detected by a Si single crystal containing Li, commonly referred to as Si(Li) detector. After the energy spectrum emitted by the specimen is analyzed by the spectrometer, it is displayed on the monitor screen of the external computer controlling the operation of the system in a form such as that shown in Fig. 8.15. It is a plot of X ray energy vs. intensity. Spectral peaks are observed at certain energies dependent upon the elemental composition of the specimen superimposed on a background. This background corresponds to the X rays emitted by the specimen as a result of collision between the X rays used as an excitation source and the massive nuclei of the specimen. Since these x-rays have a continuous range of wavelengths and energies, they are called continuum radiation to distinguish them from the characteristic radiation. Qualitative analysis is based upon correlating the energies of the observed spectral peaks with atomic number leading to identification of the elements present in the specimen. It is always recommended to identify a given element by more than one peak. Sometimes, however, this is not possible because weaker peaks may become buried in the background particularly
The energy dispersive X-ray spectra of the Ni-Mo-Cr alloy showing peaks at energy positions unique to elemental constituents of the material.

Figure 8.15  Energy dispersive X-ray spectrum obtained from a Ni-Mo-Cr alloy showing peaks at energy positions unique to elemental constituents of the material.

if the element is present in small concentrations. Table 8.3 lists the excitation energies of characteristic X rays for selected elements.

One limitation of energy dispersive X-ray spectroscopy is its relatively low spectral resolution. Usually, the quality of the spectrometer is judged from the width of the spectral peak at one half of its maximum intensity, which is referred to as the spectral resolution. Typical spectral resolution is in the range of 135–155 eV. Spectral peaks from different elements whose energy difference lies within the spectral resolution tend to overlap, degrading the accuracy of the analysis as described below.

As an example, Table 8.3 indicates that the energy difference between the $L_\alpha$ peak of Mo (2.29 keV) and $K_\alpha$ peak of S (2.31 keV) is 0.02 keV = 20 eV, well below the spectral resolution leading to overlapping of the two peaks. Therefore, if the specimen contains Mo and S, both elements can contribute to the observed peak. Such a problem, however, can be resolved if the nominal composition of the specimen is known. Another example is the difficulty of identifying Mn in the presence of Cr. As shown in Table 8.3, the Mn-$K_\alpha$ and Cr-$K_\beta$ peaks are expected to overlap because their corresponding energy difference is 60 eV. However, this problem can be resolved if the specimen does not contain Fe, whose $K_\alpha$ peak and the $K_\beta$ peak of Mn are overlapped (see Table 8.3). In the absence of Fe, the presence of Mn can be identified on the basis of the relative intensities of the $K_\alpha$ and $K_\beta$ peaks of Cr. As pointed out earlier, the intensity of the $K_\beta$ peak is usually about 0.10–0.25 that of the $K_\alpha$ peak of a given element. Provided the intensity is accurately measured, an unusually high intensity of the $K_\beta$ peak of Cr is an indication that Mn is present.
Another limitation of energy dispersive analysis is its elemental detection limit dependent upon the type of spectrometer. In principle, the strongest characteristic peak of a given element is that corresponding to $K_a$ radiation which has the lowest energy in the spectrum. In practice, however, as the atomic number increases, $K$ radiation becomes less frequently observed and radiation corresponding to higher energy shells such as the L and M shells become more noticeable.

### Table 8.3  Excitation Energies of Characteristic X-rays for Selected Elements (keV)

<table>
<thead>
<tr>
<th>Element</th>
<th>Atomic number</th>
<th>$K_a$</th>
<th>$K_b$</th>
<th>$L_a$</th>
<th>$L_{b1}$</th>
<th>$L_{b2}$</th>
<th>$L_{y1}$</th>
<th>$L_{y3}$</th>
<th>$M_a$</th>
<th>$M_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>13</td>
<td>1.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca</td>
<td>20</td>
<td>3.69</td>
<td>4.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cl</td>
<td>17</td>
<td>2.62</td>
<td>2.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co</td>
<td>27</td>
<td>6.92</td>
<td>7.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cr</td>
<td>24</td>
<td>5.41</td>
<td>5.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>29</td>
<td>8.04</td>
<td>8.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>9</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe</td>
<td>26</td>
<td>6.40</td>
<td>7.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hf</td>
<td>72</td>
<td></td>
<td></td>
<td>7.90</td>
<td>9.02</td>
<td>9.35</td>
<td></td>
<td></td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>19</td>
<td>3.31</td>
<td>3.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La</td>
<td>57</td>
<td></td>
<td></td>
<td>4.66</td>
<td>5.04</td>
<td>5.14</td>
<td>5.79</td>
<td>6.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg</td>
<td>12</td>
<td>1.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td>25</td>
<td>5.89</td>
<td>6.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mo</td>
<td>42</td>
<td></td>
<td></td>
<td>2.29</td>
<td>2.39</td>
<td>2.47</td>
<td>2.63</td>
<td>2.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>7</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na</td>
<td>11</td>
<td>1.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nb</td>
<td>41</td>
<td></td>
<td></td>
<td>2.26</td>
<td>2.33</td>
<td>2.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>28</td>
<td>7.47</td>
<td>8.26</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>8</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>15</td>
<td>2.01</td>
<td>2.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Re</td>
<td>75</td>
<td></td>
<td></td>
<td>8.65</td>
<td>10.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>16</td>
<td>2.31</td>
<td>2.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.84</td>
</tr>
<tr>
<td>Si</td>
<td>14</td>
<td>1.74</td>
<td>1.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ta</td>
<td>74</td>
<td></td>
<td></td>
<td>8.15</td>
<td>9.34</td>
<td>9.665</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ti</td>
<td>22</td>
<td>4.15</td>
<td>4.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>23</td>
<td>4.95</td>
<td>4.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.74</td>
<td>2.04</td>
</tr>
<tr>
<td>Y</td>
<td>39</td>
<td></td>
<td></td>
<td>1.92</td>
<td>2.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zr</td>
<td>40</td>
<td></td>
<td></td>
<td>2.04</td>
<td>2.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aStrongest line.*
In addition to the above limitations, elements which can be detected are dependent upon the type of spectrometer used in the analysis. Typically, the detecting crystal of the spectrometer, i.e., Si(Li), is enclosed in a tube sealed with a window transparent to X rays. Conventionally, the window is made of beryllium. An important disadvantage of a Be window detector is that it absorbs X rays whose energies are less than about 2 keV. Reference to Table 8.3 indicates that this corresponds to energies of characteristic X rays emitted by elements lighter than sodium (Z = 11). Thus, if a Be window detector is used, elements lighter than sodium cannot be detected. However, more advanced ultrathin window detectors, as well as windowless detectors, are now available extending the detection limit down to B (Z = 5).

Wavelength dispersive X ray spectroscopy is based upon the fact that X rays can be diffracted by crystals, as described earlier, according to Bragg’s law:

\[ \lambda = 2d \sin \theta \]

Therefore, by using suitable diffracting crystals, the constituent wavelengths of the characteristic spectrum emitted by the specimen can be resolved as schematically illustrated in Fig. 8.16. Some of the diffracting crystals used for this purpose include lithium fluoride (LiF), germanium (Ge), ptaeythritol (PET),

![Figure 8.16](image)

Figure 8.16  Schematic illustration of elemental analysis set up by wavelength dispersive X-ray spectroscopy.
One of the advantages of this technique is its superior spectral resolution over energy dispersive spectroscopy. Although the spectral resolution of energy dispersive spectroscopy is in the range of 135–155 eV, the spectral resolution of wavelength dispersive spectroscopy is about 5 eV. Therefore, the problem of overlapped peaks of different elements is avoided to a great extent.

Both the results of energy dispersive and wavelength dispersive X-ray spectroscopy can be quantified by essentially the same methods. Although there are a number of methods for quantitative analysis, the empirical-coefficient method is the most commonly used for engineering alloys where five or more elements must be analyzed. If the intensity of a characteristic peak of element A in the alloy is \( I_A \), a relative intensity ratio \( R_A \) is defined such that

\[
R_A = \frac{I_A}{I_{0A}}
\]

where \( I_{0A} \) is the peak intensity of the same element in the pure state. To a first approximation, the weight fraction of element A in the alloy is given by

\[
w_A = R_A
\]

Since the presence of other elements in the alloy influences the characteristic peak intensity of a given element, the above equation is modified to

\[
w_A = R_A(a_{AA}w_A + a_{AB}w_B + a_{AC}w_C + \cdots)
\]

where \( a_{AB} \) is called the influence coefficient of element B, \( a_{AC} \) is the influence coefficient of element C, etc. Standard samples of known composition covering the range being analyzed are required to determine the values of the coefficients \( a_{ij} \). At least, the number of standard samples is the same as the number of elements to be analyzed. For each standard sample, the value of \( R_i \) is measured and in turn the respective coefficient \( a_{ij} \). By combining the equations expressing the values of \( R_i \) and \( a_{ij} \) with the auxiliary equation

\[
w_A + w_B + w_C + \cdots = 1
\]

the values of \( w_A, w_B, w_C, \ldots \) can be calculated. Such calculations are carried out by means of a software.

### 8.3.3 Basic Principles of Phase Analysis

As pointed out earlier, elemental analysis of a specimen reveals only its elemental constituents and their concentrations. It does not provide any information about the state of presence of these elements or the microconstituents. Almost always, engineering alloys consist of more than one phase. If, for example, the specimen contains a compound or a secondary phase whose composition is \( A_xB_y \), elemental analysis only reveals the presence of elements A and B. In many cases, it is required to determine the state of presence of various elements. For example,
during service of an engineering product, secondary phases having detrimental effects on properties can precipitate. Also, various corrosion products consist of compounds such as oxides, sulfides, and chlorides. Therefore, it becomes necessary to identify those phases.

In contrast with elemental analysis, phase analysis reveals the actual state of presence of various elements. If, for example, the specimen contains a compound such as $A_xB_y$, phase analysis identifies that compound as such rather than as elements A and B. Phase analysis is based upon the phenomenon of diffraction by crystals governed by Bragg’s law:

$$\lambda = 2d \sin \theta$$

where $\lambda$ is the wavelength of the radiation, $d$ is the interplanar spacing of the diffracting crystal, and $\theta$ is one-half the Bragg angle. As described below, an X ray diffraction experiment provides a set of $d$ spacings, which can be matched with those of known phases listed in ASTM or JCPDS card files for phase identification.

### 8.3.4 Technique of Bulk Phase Analysis

Bulk phase analysis is carried out by X ray diffraction. Either an X ray diffractometer or an X ray powder diffraction camera can be used. Frequently, however, the diffractometer is more convenient to use particularly in failure analysis investigations. Specimens examined include bulk metal, corrosion deposit on the surface of a metal, or corrosion deposit separated from the metal.

Phase analysis by X ray diffraction provides information about the average crystal structure of the specimen. However, to identify a certain phase, it must be present in an amount equivalent to at least about 10% of the specimen by volume. Furthermore, unique phase identification can sometimes be made difficult by the presence of more than two phases particularly if the phases, present are isomorphous. By isomorphous is meant that two phases of different chemical compositions can have the same crystal structure and about the same interplanar spacings. To eliminate some of these difficulties, it is recommended to combine the results of X ray diffraction with those derived from another technique such as microchemical analysis in a scanning electron microscope, and in special cases, it may be required to use the technique of electron diffraction described later. An example illustrating the usefulness of phase analysis by X ray diffraction is given below.

Elemental chemical analysis of a Ni-based alloy shows its overall composition to be 64Ni-26Mo-7.7Cr-1Fe-0.5Co-0.6Mn-0.2Si-0.1Al, all in weight %. An X ray diffraction pattern derived from that sample after oxidation in air for 24 h at 800°C is shown in Fig. 8.17. In many experiments, Cu-K$_\alpha$ radiation with a wavelength $\lambda = 1.54$ Å is used to irradiate the specimen. As can be
seen, the diffraction pattern is a plot of the Bragg angle $2\theta$ vs. X ray counts or relative intensity. When the Bragg angle satisfies the Bragg law for a certain set of crystal planes (a given $d$ spacing), an intensity peak is observed as shown in Fig. 8.17. Such a peak is called a Bragg reflection, or a reflection. From knowledge of the angle $\theta$ and the wavelength $\lambda$, the corresponding $d$ spacing can be calculated using Bragg’s law. To summarize, an X ray diffraction pattern provides a set of $d$ spacings. All the reflections shown in Fig. 8.17 are indexed in terms of the bulk alloy and various phases of oxides, i.e., NiO, Cr$_2$O$_3$, and NiCr$_2$O$_4$, formed at its surface.

8.3.5 Techniques for Microchemical Analysis

Microchemical analysis can be carried out in combination with microstructural characterization using various techniques. Energy dispersive X ray spectroscopy can be used in combination with SEM, TEM, and STEM, and therefore it has a wide range of applicability. Both bulk specimens and thin-foil specimens can be analyzed, depending upon the electron optical system used, to characterize the microstructure. Bulk specimens can be analyzed in the conventional SEM as well as in the SEM mode of an analytical electron microscope. Both the TEM and STEM can be used to analyze thin-foil specimens. Wavelength dispersive X ray spectroscopy is usually used to analyze bulk specimens in an electron probe microanalyzer. However, conventional SEM can also be fitted with a wavelength spectrometer. Microchemical analysis by electron energy loss
spectroscopy is limited to specimens transparent to electrons, i.e., thin foils or films, and therefore it can only be used in combination with TEM and STEM.

Various surface analysis technique are used to determine the chemical composition of thin surface layers about 2 nm in thickness. These techniques include (i) Auger electron spectroscopy (AES), (ii) electron spectroscopy for chemical analysis (ESCA), (iii) secondary ion mass spectroscopy (SIMS), and (iv) X ray for photoelectron spectroscopy (XPS). Both bulk and thin-foil specimens can be analyzed, and SEM images can be obtained in the case of scanning Auger microprobes. Table 8.4 summarizes some the main features of the above techniques including their capabilities and limitations.

As shown in Table 8.4, energy dispersive X ray spectroscopy is conventionally limited to detecting sodium and heavier elements only. However, more advanced detectors are capable of detecting lighter elements down to carbon and possibly boron, as further described later. An account for each of the techniques listed in Table 8.4 is given below.

**Energy Dispersive X-Ray Spectroscopy**

Microchemical analysis by energy dispersive X ray spectroscopy is based upon the same principles of X ray fluorescence (XRF) described earlier. However, in this case, while observing the microstructure in an SEM or TEM, a certain feature is selected for the analysis, and its elemental composition is determined. As a result, an energy dispersive X ray spectrum such as that shown in Fig. 8.15 is obtained.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Elements detected</th>
<th>Resolution</th>
<th>Detection limit</th>
<th>Relative accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy dispersive X-ray spectroscopy</td>
<td>$Z &gt; 4$ (boron and heavier elements)</td>
<td>Lateral: 10 nm in STEM and 1 μm in SEM; Depth: 10 nm in STEM and 1 μm in SEM</td>
<td>0.02 wt%</td>
<td>1–5%</td>
</tr>
<tr>
<td>Wavelength dispersive X-ray spectroscopy</td>
<td>$Z &gt; 4$ (boron and heavier elements)</td>
<td>Lateral: 1 μm; Depth: 1 μm</td>
<td>0.01 wt%</td>
<td>1–5%</td>
</tr>
<tr>
<td>Electron energy loss spectroscopy</td>
<td>$Z &gt; 2$ (lithium and heavier elements)</td>
<td>Lateral: 10 nm; Depth: 10 nm</td>
<td>0.05 wt%</td>
<td>10%</td>
</tr>
<tr>
<td>Auger electron spectroscopy</td>
<td>$Z &gt; 2$ (lithium and heavier elements)</td>
<td>Lateral: 0.1–1 μm; Depth: 1 nm</td>
<td>0.1 wt%</td>
<td>10%</td>
</tr>
</tbody>
</table>
An important experimental factor influencing the intensity of characteristic X rays emitted by a given atom in a specimen being examined in an electron microscope is the accelerating voltage. Increasing the accelerating voltage increases the X ray signal, and vice versa. At a given accelerating voltage, emission of characteristic X rays by the specimen being examined is influenced by three parameters, namely, (i) atomic number of the emitting atom and average atomic number of the specimen, (ii) absorption, and (iii) secondary fluorescence. Since these parameters can significantly influence the results of quantitative analysis, a brief account for each of them is given below.

Atomic number of the emitting atom influences the X ray signal by its effect on (i) ionization cross section and (ii) fluorescent yield. As pointed out earlier, characteristic X rays are produced by creation of electron vacancies in lower energy shells, and the subsequent filling of those vacancies by electrons from higher energy shells. Ionization cross section is defined as the probability that an electron vacancy is created when the specimen is irradiated by an electron beam. Fluorescent yield is defined as the probability of producing characteristic X rays once an electron vacancy is created. It is to be noted that not all of the energy of incident electron beam can be utilized to create electron vacancies. Some of this energy becomes unavailable to create electron vacancies by other scattering processes, which is a function of the average atomic number of the specimen.

Absorption refers to the possibility that characteristic X rays emitted by a given atom may be absorbed by another atom while traveling through the specimen. Physically, absorption involves ejection of loosely bound electrons from the atom by the characteristic X rays provided its energy just exceeds the binding energy of the electron, which is known as the absorption edge. It is then evident that the probability of absorption increases as the X ray energy approaches the binding energy. Also, absorption is dependent upon the thickness of the specimen. Since X rays must travel longer distances before leaving a thicker specimen, the probability of absorption increases with specimen thickness.

Fluorescence is related to absorption. As a result of ejecting an electron from a given atom by the characteristic X rays, the atom becomes in a higher energy or an excited state. To lower its energy electron, transition occurs within the atom as described earlier, emitting secondary characteristic X rays. Since fluorescence is directly related to absorption, the probability of its occurrence also increases with the specimen thickness.

After the emitted X rays modified by the effects of atomic number, absorption, and fluorescence emerge from the specimen, they enter into a spectrometer consisting of a collimator, window, and a detector. It is the primary function of the spectrometer to analyze and measure the energy distribution of the collected X ray signal. All energy dispersive spectrometers operate by means of several software control programs through an external computer.
Quantification of X-ray spectral data is based upon correlating the intensity of characteristic peaks with concentration. Prior to quantification, however, it is necessary to subtract from the observed spectrum both the background and what is known as escape peaks. As indicated earlier, the background corresponds to the continuum component of X rays emitted by the specimen. Escape peaks result from the interaction between the X rays emitted by the specimen and the Si single crystal of the detector. As a result of this interaction, each peak of a major element present in the specimen is associated with an escape peak at an energy corresponding to 1.74 keV below that corresponding to the respective element. Both the background and escape peaks can be easily subtracted using the software of the computer controlling the operation of the system.

As pointed out earlier, emission of characteristic X rays is complicated by three parameters resulting from atomic number $Z$, absorption $A$, and fluorescence $F$ requiring corrections for each parameter before the spectral data is quantified. Almost all quantification methods are based upon the ZAF method, where $Z$ accounts for the atomic number correction, $A$ for the absorption correction, and $F$ for the fluorescence correction. Without going through the mathematical details involved, the ZAF method is based upon assigning a $K$ ratio for each element such that

$$K = \frac{I}{I_0}$$

where $I$ is the integrated intensity of the spectral peak corresponding to a given element (area enclosed by the peak after subtracting the background), and $I_0$ is the corresponding intensity of the same element in a standard specimen of a known composition. With the aid of software, a set of $K$ values are calculated and normalized to 100%. To a first approximation, each initial $K$ value is assumed to be equivalent to the concentration of the respective element. From these preliminary estimates, the initial ZAF correction parameters are calculated for each element. Subsequently, those correction parameters are applied to the preliminary estimates of concentrations to obtain a second set of concentrations serving as the basis for recalculating the correction parameters. By repeating this iteration process, the changes in estimated concentrations as well as the correction parameters become progressively less. Eventually, the results converge to reflect the most accurate measurement of concentration. It is possible, however, to carry out standardless analysis by making certain reasonable assumptions in the mathematical treatment to avoid the need for standards and be able to calculate the ZAF correction parameters, and substitute the intensity of a given element in a standard specimen with a theoretical calculated intensity in order to calculate the $K$ ratio. Having determined the values of the correction parameters $Z$, $A$, and $F$, the concentration $C$ of a given element is given by

$$C = K(ZAF)$$
For example, Table 8.5 illustrates the results of analyzing the spectral data of an energy dispersive X ray spectrum derived in a scanning electron microscope from an alloy with a nominal chemical composition of Ni–23Cr–10Mo–5Fe (wt%, maximum). The spectrum was analyzed by the ZAF method using 2 iterations. It is evident that the observed composition corresponds well with the nominal composition of the alloy.

Quantification of energy dispersive spectral data derived from thin-foil specimens in the TEM and STEM is more straightforward in comparison with the case of bulk specimens described above, provided what is known as the thin film criterion is satisfied. Generally, a thin-foil specimen transparent to electrons is considered to satisfy the thin-film criterion. In this case, it is possible to neglect the correction parameters associated with the effects of absorption and fluorescence. This approach is known as the Cliff-Lorimer method. It represents a reasonable compromise between accuracy and simplicity. According to this method, the relationship between concentration C and peak intensity can be expressed as

$$C_A / C_B = K_{AB} (I_A / I_B)$$

where $C_A$ and $C_B$ are the concentrations of elements A and B in the specimen, respectively, $K_{AB}$ is a constant known as the Cliff-Lorimer $K$ factor, and $I_A$ and $I_B$ are the integrated intensities of the most intense peaks of elements A and B, respectively. To calculate the concentrations of elements present in the specimen, the acquired spectrum is first assigned a ratio peak corresponding to the intensity of the strongest peak in the entire spectrum and defining a reference element. Then the concentration of each element is calculated from the intensity ratio of its most intense peak and that of the reference element. As an example, Figure 8.18 illustrates an energy dispersive spectrum derived in a transmission electron microscope from a thin-foil specimen of an alloy. Table 8.6 shows the results of quantifying the spectral data using the Cliff-Lorimer method.

In addition to X ray spectra, energy dispersive X ray data can also be displayed in the form of X ray mapping images or dot maps exhibiting the distribution of various elements throughout the observed microstructure. For each element observed in the spectrum, a dot map can be derived. A direct correlation exists between the density of dots the map and concentration of the respective element.

<table>
<thead>
<tr>
<th>Element</th>
<th>$K$ ratio</th>
<th>$Z$</th>
<th>$A$</th>
<th>$F$</th>
<th>ZAF</th>
<th>Atomic %</th>
<th>Weight %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni–K</td>
<td>0.647</td>
<td>0.985</td>
<td>1.050</td>
<td>1.000</td>
<td>1.035</td>
<td>64.71</td>
<td>64.07</td>
</tr>
<tr>
<td>Cr–K</td>
<td>0.231</td>
<td>1.008</td>
<td>1.047</td>
<td>0.919</td>
<td>0.971</td>
<td>24.63</td>
<td>21.49</td>
</tr>
<tr>
<td>Mo–L</td>
<td>0.077</td>
<td>1.074</td>
<td>1.326</td>
<td>0.996</td>
<td>1.419</td>
<td>6.57</td>
<td>10.58</td>
</tr>
<tr>
<td>Fe–K</td>
<td>0.042</td>
<td>1.004</td>
<td>1.079</td>
<td>0.872</td>
<td>0.946</td>
<td>3.10</td>
<td>3.10</td>
</tr>
</tbody>
</table>
It is evident from the above examples that the accuracy of microchemical analysis by energy dispersive X-ray spectroscopy is in general sufficient to verify the material of a failed component, particularly in the SEM since specimen preparation is rather minimal. In many cases of failure analysis studies, it may also be required to analyze the composition of secondary phase particles present within the material as well as products of a corrosion reaction. An example of a secondary phase particles is shown in Fig. 8.19, where platelets of sigma phase have precipitated in a Ni-based alloy. Precipitation of sigma phase particularly in a platelet-type morphology such as that shown in Fig. 8.19 can cause considerable embrittlement of the material leading to its failure. It is known to precipitate in Fe-based alloys and some Ni-based alloys after long-term exposure at elevated temperatures. Energy dispersive X-ray spectroscopy can be used to identify such a phase.

**Table 8.6** Elemental Analysis of Alloy Using the Cliff-Lorimer Method

<table>
<thead>
<tr>
<th>Element line</th>
<th>Peak intensity (counts)</th>
<th>$K$ factor</th>
<th>CEL/CREF</th>
<th>Atomic %</th>
<th>Weight %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni–K</td>
<td>39,262</td>
<td>1.000</td>
<td>1.000</td>
<td>47.43</td>
<td>41.73</td>
</tr>
<tr>
<td>Mo–L</td>
<td>21,216</td>
<td>1.953</td>
<td>1.055</td>
<td>30.77</td>
<td>44.04</td>
</tr>
<tr>
<td>Si–K</td>
<td>3,589</td>
<td>0.846</td>
<td>0.077</td>
<td>7.74</td>
<td>3.23</td>
</tr>
<tr>
<td>Fe–K</td>
<td>1,461</td>
<td>0.931</td>
<td>0.035</td>
<td>1.73</td>
<td>1.45</td>
</tr>
<tr>
<td>Cr–K</td>
<td>10,325</td>
<td>0.871</td>
<td>0.229</td>
<td>12.33</td>
<td>9.56</td>
</tr>
</tbody>
</table>
An example demonstrating the application of the technique in analyzing corrosion products is given in Fig. 8.20 and Table 8.7. It illustrates the analysis of scale formed on a carbon steel as a result of atmospheric corrosion using an ultrathin window detector capable of detecting light elements. As can be seen from the results of the analysis, the scale is an iron oxide. However, it must be emphasized that quantitative analysis of light elements, e.g., oxygen, cannot be carried out with the same level of accuracy as elements heavier than sodium because of complications related to absorption and contamination. In this case, it is important to combine microchemical analysis with another technique such as X ray diffraction to positively identify the nature of the oxide phase. Using X ray diffraction, the oxide phase observed in Fig. 8.20 is identified to be Fe₃O₄.

Even though microchemical analysis by energy dispersive X ray spectroscopy can be very useful in failure analysis studies as demonstrated above, in certain cases it may be required to use more accurate techniques, particularly if relatively small concentrations of certain elements must be analyzed. One of those techniques is wavelength dispersive X ray spectroscopy described below.

**Wavelength Dispersive X-Ray Spectroscopy**

Similar to energy dispersive X ray spectroscopy, microchemical analysis by wavelength dispersive X ray spectroscopy is based upon analyzing the characteristic X rays emitted by a specimen. However, as its name implies, the wavelength spectrum is analyzed instead of the energy spectrum. As pointed out earlier, the energy \( E \) and wavelength \( \lambda \) of characteristic X rays are related by

\[
E = \frac{hc}{\lambda}
\]
where $h$ is Planck’s constant and $c$ is the speed of light. Because of the inverse relationship between energy and wavelength, low-energy spectral peaks in an energy spectrum ($K\alpha$ and $K\beta$) appear at large wavelengths in a wavelength spectrum.

Wavelength spectroscopy displays its strength in detecting light elements, and considerably smaller concentrations in comparison with energy spectroscopy permitting analysis to be carried out at a high accuracy level. With the availability of ultrathin window detectors in energy spectrometers, detection of light elements is no longer an advantage. However, in addition to its characteristic

<table>
<thead>
<tr>
<th>Table 8.7</th>
<th>Analysis of Fe$_3$O$_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Element</td>
<td>Atomic %</td>
</tr>
<tr>
<td>Fe</td>
<td>45.05</td>
</tr>
<tr>
<td>O</td>
<td>54.25</td>
</tr>
<tr>
<td>Mn</td>
<td>0.29</td>
</tr>
<tr>
<td>Si</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
mass detection limit, wavelength dispersive spectroscopy is also characterized by a considerably greater spectral resolution, permitting distinction to be made between elements whose characteristic X ray energy difference lies within the spectral resolution of energy spectroscopy, e.g., Mo and S.

Dispersion of characteristic X rays emitted by the specimen according to their wavelengths is based upon diffraction of X rays by crystalline materials according to Bragg’s law described earlier, i.e.,

$$\lambda = 2d \sin \theta$$

**Surface Analysis Techniques**

Generally, materials have two main types of surfaces: (i) external or free surface, which is exposed to the surrounding environment and (ii) internal surface such as grain boundaries and phase interfaces. Many of the technologically important properties of engineering alloys are influenced by the chemical compositions of both external and internal surfaces. Some of the properties influenced by the chemistry of external surfaces are (i) fatigue, (ii) hardness, (iii) surface reactivity, (iv) surface diffusion, and (v) catalysis. Internal surfaces influence such properties as plastic deformation processes and fracture. Development of various techniques designed to characterize surface chemistry has significantly contributed to understanding the relationship between properties and surface chemistry.

Auger electron spectroscopy (AES) and X ray photoelectron spectroscopy (XPS) are the most common techniques used to determine the chemical composition of surface layers of metallic materials ranging in thickness from 0.5 to 2 nm. In this section, the basic principles of each technique are described and some typical applications are illustrated. Also, other surface analysis techniques are briefly outlined.

It is possible to illustrate the basic principle of AES with reference to the electron energy level diagram of an atom such as that shown in Fig. 7.1. Auger spectroscopy is based upon exciting surface atoms of the material being examined by removal or ejection of electrons from an inner shell such as the K shell. Atom excitation is accomplished by using an energetic electron beam accelerated through a potential difference of up to 30 keV. In this state, the atom is said to be in a primary excited state. To lower the energy of the primary excited atom, an electron from the higher energy L shell falls into the K shell to fill the vacancy created by the initial excitation. However, in this state the atom remains to be in a higher energy state relative to its state prior to primary excitation, and therefore it is said to be in a secondary excited state. Evidently, in the secondary excited state, the atom gains an energy equal to

$$\text{Gain in energy} = E_K - E_L$$

where $E_K$ and $E_L$ are the energies of an electron in the K and L shells, respectively.
Deexcitation of the atom occurs if the above gain in energy is released by emitting an electron from the L shell, which is called Auger electron, leaving the atom in a doubly ionized state. If the kinetic energy of the emitted electron is $E_{KE}$, and $\Phi$ is the work function of the material (energy which must be supplied to enable an electron to escape from the material), then

$$E_K - E_L = E_{KE} + \Phi + E_L \quad \text{or} \quad E_{KE} = E_K - 2E_L - \Phi$$

Auger electron kinetic energy is characteristic of the emitting atom and is independent of the energy used for primary excitation. Therefore, it can be used as a fingerprint to identify the elemental composition of the surface.

It is the primary function of an Auger spectrometer to determine the kinetic energies of Auger electrons, which form characteristic peaks superimposed on a background. To eliminate the high background resulting from the large number of backscattered electrons, an Auger spectrum is constructed by plotting kinetic energy of Auger electrons vs. the first derivative of the electron intensity $dN/dE$, as illustrated in the spectrum of Fig. 8.21 obtained from a catalyst sample.

![Auger Spectrum](image)

**Figure 8.21** Auger spectrum obtained from a catalyst sample showing the presence of Al, S, C, and O.
X ray photoelectron spectroscopy (XPS) differs from AES in that a beam of X rays is used for primary excitation. A photon of the incident X ray beam ejects an electron from the inner K shell, which is called a photoelectron. In terms of energy, the photon energy $h\nu$ is transformed into binding energy of the photoelectron $E_{\text{BE}}$, its kinetic energy $E_{\text{KE}}$, and the work function of the material $\Phi$, i.e.,

$$h\nu = E_{\text{BE}} + E_{\text{KE}} + \Phi$$

By definition, the binding energy $E_{\text{BE}}$ of the photoelectron is energy difference between the Fermi level and the ground state of the atom where the Fermi level is the energy of an electron in the outermost shell, and the ground state corresponds to the lowest energy level ($1s$). Photoelectron binding energies are characteristic of the emitting atoms and therefore can be used to identify the elemental constituents of the surface. An XPS spectrum is a plot of the number $N$ of photoelectrons with energy $E$, i.e., $N(E)$ vs. the binding energy such as that shown in Fig. 8.22 from a catalyst sample. Therefore, XPS can be used to determine elemental constituents as well as their state of oxidation; i.e., chemical species can be distinguished.

**Figure 8.22** X-ray photoelectron spectrum obtained from a catalyst sample.
Both Auger and XPS spectral data are quantified using the following relationship:

\[ C_x = \frac{I_x}{S_x \left( \sum_{i} \frac{I_i}{S_i} \right)^2} \]

where \( C_x \) is the atomic concentration of element \( x \), \( I_x \) is the peak-to-peak height of element \( x \), and \( S_x \) is a relative elemental sensitivity factor measured from a pure elemental standard, and \( i \) is summed over all elements present.

Other techniques for surface analysis include electron spectroscopy for chemical analysis (ESCA) and secondary ion mass spectroscopy (SIMS). ESCA is essentially the same as XPS described above. However, it differs from XPS in that it can be used to study the surface chemistry of organic solids nondestructively. A SIMS experiment consists of bombarding the surface of a solid material with an ion beam. Secondary ions emitted by the surface are analyzed in terms of their mass using a mass spectrometer. Both positive and negative ions can be analyzed.

Chemical profiles normal to the surface are of particular importance in surface analysis when the surface composition is significantly different from that of the bulk. Depth profiles of various elements can be generated by sputtering of surface layers, i.e., physically removing surface layers of the material by ion bombardment using an inert gas such as Ar. Ions of the inert gas sputter atoms off the surface while surface analysis is being carried out. Also, in situ fracture of a specimen being examined in the Auger can be carried out. This is particularly useful in the case of brittle materials. Some of the typical applications of AES and XPS in the surface analysis of engineering alloys include (i) elemental segregation and depletion, (ii) corrosion, (iii) hydrogen charging, and (iv) fracture analysis.

### 8.4 Microstructure of Engineering Alloys

In this section, some general microstructural features of the most common structural alloys are described with illustrative examples. Correlation of microstructure with observed properties is briefly outlined so that desirable and undesirable microstructural features as they relate to certain applications are identified.

#### 8.4.1 Ferrous Alloys

Equilibrium microstructures of ferrous alloys can best be illustrated with reference to the binary Fe–C diagram shown in Fig. 7.39. An example is given in the optical micrograph of Fig. 8.23a to illustrate a typical microstructure of a carbon steel. As explained earlier, the microstructure consists of (dark) colonies of pearlite in a (bright) ferrite matrix. When viewed at higher magnification, the characteristic lamellar structure of pearlite is observed as shown in the SEM.
Figure 8.23  (a) Optical micrograph of a carbon steel showing dark pearlite within white ferrite matrix, (b) SEM image showing lamellar structure of pearlite.
image of Fig. 8.23b. It is recalled [Eq. (4.19)] that yield strength is related to grain size by

\[ \sigma_y = \sigma_0 + K_y d^{-1/2} \]

where \( \sigma_y \) is the yield strength, \( \sigma_0 \) and \( K_y \) are constants, and \( d \) is the grain size.

When a low-alloy steel is not hardened by heat treatment, its microstructure is similar to that of carbon steel, i.e., it consists of ferrite and pearlite; however, its grain structure is finer because the transformation of austenite occurs at lower temperatures. If the same steel is rapidly cooled so that austenite is transformed into martensite rather than ferrite and pearlite, its microstructure becomes as shown in Fig. 8.24. As can be seen, the microstructure contains fine platelets of martensite.

High-alloy wrought steel grades are classified as stainless steels. There are three types of stainless steels: (i) austenitic, (ii) ferritic, and (iii) duplex. As its name implies, an austenitic stainless steel consists entirely of austenite stabilized by the addition of a minimum of 8% Ni. Such steels have a face-centered cubic structure, and because of the lack of allotropic transformation, they are not heat treatable. Essentially, they consist of a single phase (solid-solution), as illustrated in Fig. 8.25. Conversely, ferritic stainless steels consist entirely of ferrite stabilized by the addition of large concentrations of Cr in the range of 12–25%. However, the microstructures of ferritic stainless steels are not as homogeneous as the austenitic grades. Since the solubility of C in ferrite is very small, excess C is precipitated as Cr-rich carbides. Also, the ferritic stainless steels are heat

![SEM image showing formation of needlelike martensite upon rapid cooling of austenite.](image)
treatable because of the very low critical cooling rates. Duplex stainless steels consist of a mixture of austenite and ferrite, and therefore they are characterized by high concentrations of Ni (austenite stabilizer) and Cr (ferrite stabilizer). A typical microstructure of a duplex stainless steel is shown in Fig. 8.26.

Other than the austenitic stainless steels, heat-resistant steels are also austenitic. These steels contain high concentrations of Ni and Cr, and derive their strength from carbide precipitates requiring the presence of 0.3–0.5% C. Because of their high C content, heat-resistant steels cannot be processed into wrought products; rather they are produced as castings having an inherently large grain size and low room temperature ductility. Alloing elements in these steels include Si, and some grades contain refractory elements such as Nb or W. Heat-resistant steel castings consist of an austenite matrix (face-centered cubic solid solution of Fe, Cr, and Ni) and Cr-rich carbide precipitates as illustrated in Fig. 8.27.

Figure 8.25  Microstructure of a single-phase 316 stainless steel.
8.4.2 Nonferrous Alloys

Superalloys are an important class of nonferrous alloys that are used for high-temperature applications. Superalloys are based on group VIIIA of the periodic table and exhibit high strength during long exposures at elevated temperatures as well as excellent surface stability. Superalloys are mainly Ni, Co, and Fe based with additions of Cr, W, Mo, Ta, Nb, Re, Hf, Ti, and Al. Important classes of superalloys are discussed below.

Gamma Prime (\(\gamma^0\)) Strengthened Ni-Based Superalloys

High strength in precipitation-hardened Ni-based superalloys is primarily imparted by the formation of a coherent stable intermetallic compound called gamma prime (\(\gamma^0\)) [(Ni, Co)\(_3\)(Al, Ti)]. The \(\gamma^0\) phase has an L1\(_2\) fcc structure and is homogeneously distributed throughout the matrix. Depending on matrix-lattice mismatch, the \(\gamma^0\) phase can precipitate in several morphologies, i.e., spherical, cubical, elongated, and fine, as shown in Fig. 8.28a–d. The volume fraction of \(\gamma^0\) in certain alloys can be as high as 70%. The strength of \(\gamma^0\) increases with temperature and due to its ductile nature the fracture toughness of the alloy is not adversely affected by the presence of \(\gamma^0\) precipitates. Compositionally, Co
substitutes for Ni, whereas Ti, Nb, Ta and Hf substitutes for Al in $\gamma'$. Mo, Cr and Fe can substitute for both Ni and Al positions within the ordered phase.

In addition to $\gamma'$, carbides, borides and topologically close-packed (TCP) phases are also present in Ni-based superalloys. The presence of carbon in the amount of 0.05–0.2% can react with Ti, Ta, and Hf to form their respective carbides. The most common forms of carbides are MC, $M_{23}C_6$, and $M_6C$. These carbides have fcc structure. $M_{23}C_6$ and $M_6C$ carbide phases usually precipitate at the grain boundaries as irregular blocky particles, as shown in Fig. 8.29, while MC is distributed in a heterogeneous fashion throughout the superalloy matrix. The presence of $M_{23}C_6$ at the grain boundaries can increase the alloy’s rupture strength significantly. However, these same precipitates can initiate rupture failure by fracture. $M_6C$ carbides are more stable at higher temperatures compared to $M_{23}C_6$ carbides.

Boron is an essential ingredient in superalloys as it serves to increase their creep rupture strength. It is present in low amounts (i.e., up to 500 ppm) and reacts to form $M_3B_2$ type borides at the grain boundaries.

Figure 8.27  Morphology of Cr-carbide precipitates in a heat resistant steel.
Improper heat treatment or inappropriate service conditions can form undesirable precipitates called TCP phases in superalloys, as shown in Fig. 8.30. Commonly found TCP phases in Ni alloys are \( \sigma \) and \( \mu \). Sigma phase results in brittle failure at low temperatures and a reduced rupture strength at elevated temperatures.

Directional solidification technique is also employed to produce cast Ni alloys with single grains. This results in a significant improvement in strength over multigrained alloys.

Ni-based superalloys are a preferred choice for use in hot sections of turbine engines. They can be used at incredible temperatures of 90% of their
melting points. They carry 50% of the total weight of aircraft engines being manufactured at present times.

**Solid-Solution Strengthened Ni-Based Superalloys**

A large number of alloying additions in the Ni-based alloys serve to increase their strength and creep resistance. The matrix is face-centered cubic austenite (γ) comprising Ni, Cr, and W with possible additions of Co, Nb, Al, Ti, Mo, Re, Ta, Zr, Hf, Bo, Fe, etc. Most alloying additions exhibit a partitioning tendency to either the matrix or the precipitated phase. Common solid solution elements in γ are Co, Fe, Cr, Mo, W, Ti, and Al. A microstructure of such an alloy is shown in Fig. 8.31. Solid solution strengthening is provided by virtue of the atomic size difference between Ni and that of the solute atoms resulting in lattice expansion. This difference can vary between 1 to 13% depending upon the type of solute atoms. The larger the difference in atomic diameters of the solute and the solvent, the greater the strengthening effect. In addition to the difference in atomic size, the extent of elastic modulus misfit and the development of short-range order contribute to overall strengthening. Al is considered to be a powerful solid solution strengtheners, followed by W, Mo, and Cr. Other elements such as
Fe, Ti, Co, and V impart solid solution strengthening to a lesser extent. Low diffusion rates of Mo and W make them more effective strengtheners at temperatures above $0.6T_M$ due to their low diffusion rates. Their presence, therefore, serves to increase high temperature creep resistance of superalloys.

**Figure 8.30** Dark field TEM image of a mu phase observed in a Ni-Cr-Fe-Mo alloy.

Fe, Ti, Co, and V impart solid solution strengthening to a lesser extent. Low diffusion rates of Mo and W make them more effective strengtheners at temperatures above $0.6T_M$ due to their low diffusion rates. Their presence, therefore, serves to increase high temperature creep resistance of superalloys.

**Figure 8.31** Bright field TEM image of a Ni-based solid-solution strengthened alloy.
Cobalt-Based Superalloys

Cobalt-based superalloys hold a secondary position to Ni-based alloys in the gas turbine industry due to a lack of hardening mechanism comparable to γ′ precipitation. The Co-based alloys can contain alloying additions of Ni, Cr, W, Mo, Ta, Nb, Zr, Hf, etc. and derive their strength through solid solution strengthening and carbide precipitation. Addition of 10–20 wt% Ni enhances the stability of austenitic fcc cobalt matrix at high temperatures. Chromium provides a measure of solid solution strengthening, oxidation and hot corrosion resistance. Aluminium additions have also been used to improve upon high-temperature oxidation. Tungsten and Molybdenum are the primary solid-solution strengtheners, while Ta, Nb, Zr and Hf serve to impart strength by carbide formation. Carbides formed within Co alloys are M₃C₂, M₇C₃, and M₂₃C₆, while the C content within the alloy can vary between 0.25 and 1.0 wt%. Co-based alloys are primarily produced by air or argon melting, thus reducing costs incurred by vacuum melting techniques. Co-based alloys have higher melting temperatures, better hot corrosion resistance due to higher Cr contents, and superior thermal fatigue resistance compared to Ni-based alloys.

Iron-Based Superalloys

The Fe content in Fe-based superalloys can range from 15 to 60%. The microstructure consists of an austenitic matrix that exhibits precipitation within the alloy as well as at the grain boundaries. Depending upon the alloy type, strengthening is achieved by precipitation of γ′, γ″, and/or formation of carbides, nitrides, and carbonitrides. Chromium and Molybdenum are the primary solid solution strengtheners. Fe-based alloys contain elements like Fe, Ni, Cr, Mo, Ti, Al, Nb, etc. Major alloy types are Inconel, Incoloy, HMN, and CRMD series. Higher Fe content in the alloy improves forgeability and lowers cost. The Fe alloys containing Ti and Al form ordered fcc γ′ phase while those containing Nb form bct γ″ phase that acts as the primary source of precipitation hardening. Most of the Fe alloys are available in both cast and wrought alloys.
9

Corrosion

9.1 Introduction

In general terms, corrosion is defined as the degradation of a structural material by chemical reactions with the environment. From a chemical point of view, any type of corrosion is basically an oxidation reaction where a metal atom \( M \) loses a number \( x \) of negatively charged electrons \( e \) to become a positively charged ion, i.e.,

\[
M \rightarrow M^{+x} + xe
\]

Metals in particular are the most susceptible to the above type of reaction because of their characteristic electronic structure. It is recalled from Chap. 2 that metals have a few valency electrons in the outermost electron shells, which are loosely bound to the nucleus. As a result, these electrons can be easily separated from the respective atoms, leaving behind positive ions as indicated by the above reaction. Naturally, metals exist in the form of oxides or salts corresponding to the most stable or lowest energy state. Therefore, a metal recovered from its naturally existing ore always exhibits a tendency to transform to its most stable state resulting in corrosion; i.e., corrosion is basically the result of metals seeking their lowest energy state. It is then evident that corrosion products consist of oxides or salts.

Structural materials always function in some kind of an environment, e.g., normal atmosphere, aqueous medium, gaseous medium, etc. During service, a structural material always tends to react with its environment. Depending upon
the temperature of the environment as well as its nature, two main classes of corrosion are identified: (i) low-temperature aqueous corrosion and (ii) high-temperature gaseous corrosion. Corrosion is considered to be one of the most important causes of failure of structural materials during service. As a result of metal wastage to a corrosion product by the above basic reaction, both the structural integrity of the material as well as its mechanical strength can be severely degraded. In this chapter, the two main classes of corrosion are discussed. Examples are given to illustrate the specific types of corrosion within each class and how they can be characterized using various investigative tools.

9.2 Low-Temperature Aqueous Corrosion

Low-temperature aqueous corrosion refers to corrosion attack of a structural material in the presence of a liquid medium. Corrosion proceeds when the corroding material has a significant solubility in the liquid. It is recalled that surface atoms are in a higher energy state than bulk atoms because of their fewer nearest neighboring atoms. As a result, surface atoms always exhibit a tendency to react with other species in the environment to lower their energy. Therefore, corrosion can only proceed if the associated chemical reaction results in a net decrease in free energy. By definition, free energy is a thermodynamic quantity representing the energy released or absorbed during a chemical reaction or any other physical change. To illustrate the principle involved, consider the reaction of magnesium (a reactive metal) and gold (a noble metal) with the normal atmosphere:

\[
\begin{align*}
\text{Mg} + \left[ \text{H}_2\text{O(water)} + \frac{1}{2} \text{O}_2(\text{oxygen}) \right] & \rightarrow \text{Mg(OH)}_2, \quad \Delta G = -142 \text{kcal/mol} \\
\text{Au} + \left[ \frac{3}{2} \text{H}_2\text{O(water)} + \frac{3}{4} \text{O}_2(\text{oxygen}) \right] & \rightarrow \text{Au(OH)}_3, \quad \Delta G = +15.7 \text{kcal/mol}
\end{align*}
\]

As can be seen, the reaction leading to the formation of Mg(OH)\(_2\) is accompanied by a decrease in free energy \(\Delta G\), as indicated by the minus sign, and therefore it is thermodynamically favorable. In contrast, formation of Au(OH)\(_3\) is accompanied by an increase in free energy, and therefore it is thermodynamically unfavorable. It is evident that the reverse reaction, i.e., decomposition of Au(OH)\(_3\) into Au, H\(_2\)O, and O\(_2\), is accompanied by a decrease in free energy, and therefore it is thermodynamically favorable.

Although thermodynamics is able to predict whether a given reaction can occur, it is inherently limited in that it does not provide any information about the reaction rate, which falls under the subject of kinetics. It is possible that a given reaction is accompanied by a decrease in free energy as determined from thermodynamic data. On the other hand, the reaction rate can be extremely sluggish to the extent that, for practical purposes, the reaction is considered not to occur. Thus, in principle, corrosion tendencies can be predicted from thermodynamic data. However, this approach cannot explain the “selectivity” of the environment, i.e.,
specific metal in specific environment. In contrast, the electrochemical approach can explain the selectivity of the environment as described below.

Most corrosion processes occur by an electrochemical mechanism involving the interconversion of electrical and chemical energy. Basically, corrosion proceeds by removal of electrons from an atom converting it into a positive ion. Electrons released by this oxidation reaction provide an electric current either between two electronically different regions of the same material or between two different materials. Differences in electronic structure are reflected by an electromotive force or a potential acting as the driving force of the corrosion reaction. An electrolyte must be present to complete the reaction path. Positively charged ions dissolve preferentially in the electrolyte and migrate to another region where they can regain the lost electrons balancing their electronic structure.

It is evident from the above discussion that certain requirements must be satisfied before electrochemical corrosion can occur. Such requirements define the components of an electrochemical cell, which are (i) a liquid environment acting as an electrolyte, (ii) both anodic and cathodic regions acting as a galvanic couple, and (iii) a metallic conductor between the anode and cathode. An electrolyte is a liquid solution of a compound, salt, or an acid capable of conducting electricity. By definition, an anode is a positive electrode, i.e., an electric conductor by means of which electric current enters the electrolyte. A cathode is a negative electrode by means of which electric current leaves the electrolyte.

Galvanic couples initiating electrochemical corrosion attack can be established by various processes described later. To illustrate the process of electrochemical corrosion, consider the example of Fig. 9.1, where a piece of

![Figure 9.1](image)

**Figure 9.1** Illustration of electrochemical corrosion where anodic and cathodic regions are formed at the surface of Fe in contact with an electrolyte.
iron is placed in contact with an electrolyte in which it can dissolve. It is assumed that a galvanic couple, i.e., anodic and cathodic regions, is established within the piece of iron. At the anodic region, the following oxidation reaction occurs:

$$\text{Fe} \rightarrow \text{Fe}^{2+} + 2e^-$$

At the cathodic region, another chemical reaction occurs resulting in either the reduction of oxygen or liberation of hydrogen dissolved in the electrolyte depending upon whether the electrolyte is aerated, i.e., containing dissolved oxygen such as alkaline solutions, or deaerated (acidic solution).

Aerated (alkaline): $\text{O}_2(\text{in solution}) + 4e^- + \text{H}_2\text{O} \rightarrow 4\text{OH}^- (\text{in solution})$

In an aerated solution, $\text{Fe}^{2+}$ is further oxidized by

$$\text{Fe}^{2+} \rightarrow \text{Fe}^{3+} + e^-$$

and then combines with $(\text{OH}^-)$ by

$$\text{Fe}^{3+} + 3\text{OH}^- \rightarrow \text{Fe}($$\text{OH})_3$$

resulting in the formation of rust deposited at the cathodic region.

In contrast, if the electrolyte is deaerated, i.e., an acidic solution, the following reaction occurs at the cathodic region:

$$2\text{H}^+(\text{in solution}) + 2e^- \rightarrow \text{H}_2(\text{gas})$$

To summarize, corrosion by an electrochemical mechanism consists of an oxidation reaction at the anode and a reduction reaction at the cathode. It is evident from the above discussion that actual corrosion occurs at the anode where the metal dissolves in the electrolyte by an oxidation reaction. Very little or no corrosion occurs at the cathode.

In view of the electrochemical mechanism of corrosion, the tendency of different materials to corrode can be expressed in terms of electromotive force (EMF) or electrode potential, and the rate of corrosion can be expressed in terms of an electric current as described below.

### 9.2.1 Electrode Potential: Corrosion Tendency

In the above example, a galvanic couple is assumed to be established within the same metal. It is also possible to establish a galvanic couple between two dissimilar metals. By definition, electrode potential determines which metal acts as an anode, and which acts as a cathode in a galvanic cell. To define electrode potential, consider a simple electrochemical cell such as that shown in Fig. 9.2. It consists of a Zn electrode immersed in a compartment containing a solution of zinc salt, and a Cu electrode immersed in another compartment containing a solution.
of Cu salt. In this cell, the Zn electrode acts as an anode, and the Cu electrode as a cathode. At the Zn anode, an oxidation reaction occurs such that a Zn atom loses two electrons and becomes a positive ion, i.e.,

$$\text{Zn} \rightarrow \text{Zn}^{2+} + 2e^-$$

and at the Cu cathode, a reduction reaction occurs whereby a positive ion of Cu (Cu$^{2+}$) in solution gains the two electrons lost by the Zn, and deposits as metallic Cu on the cathode, i.e.,

$$\text{Cu}^{2+} + 2e^- \rightarrow \text{Cu}$$

and the net cell reaction is obtained by adding the above two reactions:

$$\text{Zn} + \text{Cu}^{2+} \rightarrow \text{Zn}^{2+} + \text{Cu}$$

By definition, an electric current is a directed flow of electrons, and therefore, for the above reaction to proceed, the electrons released by the Zn atoms are made to flow into the cathode by means of an external electric circuit, as shown in Fig. 9.2. To complete the circuit, ions must be transferred from the cathode into the anode by means of the electrolyte, and this can achieved by using what is known as a salt bridge, e.g., an inverted U tube containing an inert electrolyte, which does not take part in the electrode reactions. If an electric charge $Q$ flows through the cell across a potential difference or electromotive force $E$, the maximum work $W$ done on the cell is given by

$$W = -EQ$$
By convention, work done on the cell is indicated by a minus sign, to distinguish it from work done by the cell, which is considered positive. For each mole of the cell reaction, $Q$ is given by

$$Q = N A e$$

where $N$ is the ionic charge or the number of electrons transferred in the cell per mole of reaction, $A$ is Avogadro’s number, and $e$ is the charge on the electron. Since the product $A e$ is by definition the Faraday $F$, the work $W$ can be expressed as

$$W = -NEF$$

Under equilibrium conditions, the chemical free energy change $\Delta G$ of the cell reaction must supply the energy required for the electrical work, and therefore

$$\Delta G = -NEF$$

However, under nonequilibrium conditions, the chemical free energy change of the reaction $\Delta G$ becomes

$$\Delta G = \Delta G_0 + RT \ln K$$

where $\Delta G_0$ is called the standard free energy change corresponding to a reaction occurring under standard conditions corresponding to unit activity, $R$ is the universal gas constant, $T$ is the temperature in degrees Kelvin, and $K$ is the equilibrium constant of the reaction ($K = a_p/a_r$ where $a_p$ and $a_r$ are the activities of the products and reactants; activity is proportional to concentration). Similar to $\Delta G$, the standard free energy change $\Delta G_0$ can be written as

$$\Delta G_0 = -NE_0 F$$

Combining the above equations results in

$$E = E_0 + RT/NEF \ln K$$

where $E$ is the electrode potential for a given concentration of the electrolyte and $E_0$ is called the standard electrode potential corresponding to unit activity of ions in the electrolyte. By using a potentiometer, the standard values of the electrode potential can be determined under equilibrium conditions when no net current is flowing in the electrochemical cell. It is not possible to measure the value of $E_0$ on an absolute scale; therefore, it is determined with reference to a standard electrode. By convention, a hydrogen electrode represented by the reaction

$$2H^+ + 2e^- \rightarrow H_2$$

is arbitrarily selected as a reference with $E = 0$. A hydrogen electrode consists of a platinum wire in a 1 normal hydrochloric acid solution (1 g equivalent of HCl per liter) in contact with hydrogen gas at 1 atmospheric pressure. Table 9.1 summarizes
the standard electrode potentials of selected metals, defining what is known as the standard electromotive (EMF) series. When two metals are coupled in a galvanic cell, the values of their electrode potentials determine which metal acts as cathode, and which is anodic. As a general rule, the metal having the most positive electrode potential acts as cathode, and the difference in electrode potential is a measure of the driving force of corrosion of the anodic metal. With increasing potential difference, corrosion of the anodic metal is accelerated, and vice versa.

It is a common practice to illustrate an electrode in the form \( M|M^{n+}Ne \). To illustrate the physical significance of the standard electromotive series shown in Table 9.1, consider the case of \( Fe|Fe^{2+} \) electrode with a standard electrode potential of \(-0.444 \) V. If the potential of Fe is maintained at \(-0.444\) V, the \( Fe^{2+} \) ions in the electrolyte at unit activity remain in equilibrium. However, if Fe is electrically connected to a \( Sn|Sn^{2+} \) electrode \((E_0 = -0.136)\), the potential of Fe tends to rise. In this case, the \( Fe^{2+} \) ions can be envisioned as moving down the potential gradient, out of the metal and into the electrolyte, which by definition is oxidation or corrosion. Conversely, if Fe is electrically connected to a \( Zn|Zn^{2+} \) electrode \((E_0 = -0.736)\), the \( Fe^{3+} \) can be envisioned to move up the potential difference, resulting in reduction of \( Fe^{2+} \) into Fe, and therefore the Fe is protected against corrosion.

### Table 9.1 Standard EMF Series of Metals

<table>
<thead>
<tr>
<th>Metal reaction</th>
<th>( E_0 ) (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Au = Au^{3+} + 3e^- )</td>
<td>+1.498 Cathodic</td>
</tr>
<tr>
<td>( Pt = Pt^{2+} + 2e^- )</td>
<td>+1.20</td>
</tr>
<tr>
<td>( Pd = Pd^{2+} + 2e^- )</td>
<td>+0.987</td>
</tr>
<tr>
<td>( Ag = Ag^{3+} + 3e^- )</td>
<td>+0.799</td>
</tr>
<tr>
<td>( Fe^{3+} + e = Fe^{2+} )</td>
<td>+0.771</td>
</tr>
<tr>
<td>( Cu = Cu^{2+} + 2e^- )</td>
<td>+0.337</td>
</tr>
<tr>
<td>( Sn^{4+} + 2e^- = Sn^{2+} )</td>
<td>+0.15</td>
</tr>
<tr>
<td>( 2H^+ + 2e^- = H_2 )</td>
<td>0 (arbitrary reference)</td>
</tr>
<tr>
<td>( Pb = Pb^{2+} + 2e^- )</td>
<td>-0.126</td>
</tr>
<tr>
<td>( Sn = Sn^{4+} + 4e^- )</td>
<td>-0.136</td>
</tr>
<tr>
<td>( Ni = Ni^{2+} + 2e^- )</td>
<td>-0.250</td>
</tr>
<tr>
<td>( Co = Co^{3+} + 2e^- )</td>
<td>-0.277</td>
</tr>
<tr>
<td>( Cd = Cd^{2+} + 2e^- )</td>
<td>-0.403</td>
</tr>
<tr>
<td>( Fe = Fe^{3+} + 2e^- )</td>
<td>-0.440</td>
</tr>
<tr>
<td>( Cr = Cr^{3+} + 3e^- )</td>
<td>-0.744</td>
</tr>
<tr>
<td>( Zn = Zn^{2+} + 2e^- )</td>
<td>-0.763</td>
</tr>
<tr>
<td>( Al = Al^{3+} + 3e^- )</td>
<td>-1.662</td>
</tr>
<tr>
<td>( Mg = Mg^{2+} + 2e^- )</td>
<td>-2.363 Anodic</td>
</tr>
</tbody>
</table>
From a practical point of view, the above example illustrates that Zn provides a better corrosion protection of iron or steel in comparison with Sn. Although Sn is commonly used as a protective layer, any discontinuity can lead to corrosion of Fe, which is anodic with respect to Sn. In contrast, for galvanized steel containing a surface where Zn is anodic with respect to Fe, Zn corrodes rather than Fe. However, this requires the Zn layer to be replaced periodically.

Although the standard electrode potentials listed in Table 9.1 can serve as useful guides in comparing corrosive tendencies of pure metals, considerable deviation may occur in practice for various reasons. First, the oxidation tendencies of metals are dependent upon the particular environment. Second, galvanic corrosion occurring in practice usually involves alloys rather than pure metals, and therefore it is highly unlikely that the corrosion reaction involves a pure metal in equilibrium with its ions, as listed in Table 9.1. It is therefore of more practical significance to compare the corrosive tendencies of various metals under actual service conditions. For example, Table 9.2 lists a modified galvanic series in sea water.

It is evident from the above discussion that in order to predict the corrosive tendencies of different materials, a series such as that shown in Table 9.2 is required for each specific environment. Obviously, this is impractical, because it requires an endless number of corrosion tests.

### 9.2.2 Corrosion Current: Corrosion Rate

In practice, knowledge of corrosion rates of different materials in specific environments is of utmost importance in (i) selection of proper materials for certain applications, (ii) ranking the corrosion resistance of different materials, and (iii) maximizing the efficiency of operation as well as planned shutdowns.

**Table 9.2** Galvanic Series of Commercial Metals and Alloys in Sea Water

<table>
<thead>
<tr>
<th>Gold</th>
<th>Nickel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphite</td>
<td>Tin</td>
</tr>
<tr>
<td>Titanium</td>
<td>Lead</td>
</tr>
<tr>
<td>Silver</td>
<td>18Cr-8Ni stainless steel</td>
</tr>
<tr>
<td>18Cr-8Ni stainless steel</td>
<td>Steel or iron</td>
</tr>
<tr>
<td>11–30% Cr stainless steel</td>
<td>Cast iron</td>
</tr>
<tr>
<td>Silver</td>
<td>Aluminium–4.5% Cu-1.5% Mg-0.6% Mn</td>
</tr>
<tr>
<td>Monel alloy (70% Ni, 30% Cu)</td>
<td>Cadmium</td>
</tr>
<tr>
<td>Bronzes (Cu-Sn)</td>
<td>Aluminium</td>
</tr>
<tr>
<td>Copper</td>
<td>Zinc</td>
</tr>
<tr>
<td>Brasses (Cu-Zn)</td>
<td>Magnesium</td>
</tr>
</tbody>
</table>
for repair and maintenance. As explained in the previous section, electrochemical corrosion involves the flow of an electric current from the anode into the cathode. Obviously, a direct correlation exists between the current density, i.e., the flow of current per unit area, and corrosion rate. Corrosion rates can be expressed by various methods such as milligrams per square decimeter per day (mdd), and mils per year (mpy) where 1 mil = 0.001 in. However, expressing corrosion rate in terms of mpy is generally preferred because it provides the rate of penetration or the thinning of the part as a result of corrosion attack.

To illustrate the correlation between corrosion rate and electric current, consider the mdd expression, where

\[
1 \text{ mdd} = 1 \text{ mg} / (100 \text{ cm}^2)(24 \text{ h}) = \left[ (10^{-3} \text{ g})(10^{-2}) \right] / (\text{cm}^2)(24 \text{ h})
\]

By definition, when a mass of a metal equivalent to its atomic weight in grams is dissolved in an electrolyte, the amount of electric charge \( Q_0 \) (expressed in units of coulomb) which flows in the electrochemical cell is given by

\[
Q_0 = NF \text{ coulomb}
\]

where \( N \) is the ionic charge and \( F \) is a constant called the Faraday (\( F = 96,485 \text{ coulomb} \)). Therefore, the charge \( Q \) associated with dissolution of 1 g is

\[
Q = NF / \text{atomic weight} \quad \text{coulomb/g}
\]

Electric current \( I \) expressed in units of amperes (A) is defined as the rate of flow of electric charge, i.e., electric current = \( Q \) in coulomb/time in seconds (c/s). Therefore, the electric current density equivalent to 1 mdd is given by

\[
I = [96,485N(10^{-5})]/[(\text{atomic weight})(24 \times 60 \times 60)]
\]

\[
= 1.17N(10^{-5}) / \text{atomic weight} \quad \text{A/cm}^2
\]

To summarize, 1 mdd is equivalent to \((1.17 N/\text{atomic weight}) \times 10^{-5}\). Experimental techniques have been developed which can directly convert the measured electric current associated with corrosion into corrosion rate. For most engineering purposes, however, the corrosion rate is determined from weight loss measurements. Test specimens are exposed to a given environment for a predetermined period of time, and the measured weight loss is converted into mpy by the following relationship:

\[
\text{mpy} = (534W)(DAT)
\]

where \( W \) is the weight loss in mg (1 mg = \( 10^{-3} \) gm), \( D \) is the density of the material in g/cm\(^3\), \( A \) is the surface area of the specimen in in\(^2\), and \( T \) is the
exposure time in hours. In terms of corrosion rate expressed as mpy, different materials are in general classified into three classes as follows:

- Less than 5 mpy: Good corrosion resistance
- 5 < mpy < 50: Acceptable
- More than 50 mpy: Unsatisfactory

One of the most important factors influencing the rate of actual corrosion processes is the deviation of the electrode potential from the equilibrium value $E_0$ leading to a phenomenon known as polarization. By polarization is meant that the rate of electrochemical reactions occurring at the surface is limited. If the potential is increased above the equilibrium value, the reaction is said to be anodically polarized. In this case, the material is said to be active and the corrosion current or corrosion rate increases with potential, as shown in Fig. 9.3. Conversely, if the potential is reduced below the equilibrium value, the reaction is said to be cathodically polarized where the corrosion current decreases with increasing potential, as shown in Fig. 9.3.

Electrochemical reactions can be polarized in two distinct manners: (i) activation polarization and (ii) concentration polarization. In the case of activation polarization, the corrosion rate or current is limited by the rate of some reaction
involved in a corrosion process such as the adsorption of a species to the surface. Concentration polarization is a process where the corrosion rate is limited by the diffusion of a certain species in the electrolyte. Evidently, in the absence of polarization, the corrosion rate increases. However, many metal-electrolyte systems undergo some type of polarization, limiting the corrosion rate. One of the most important types of activation polarization is passivity, as described below.

By definition, passivity is the loss of chemical reactivity between a metal and its environment. It results from the formation of a thin layer usually of an oxide phase on the surface of the metal. Such a passive layer acts as a barrier between the metal and its environment. Typically, when a metal inherently capable of developing a protective passive layer is first placed in contact with an electrolyte, the corrosion rate increases, exhibiting the behavior of an active material. Subsequently, the corrosion rate decreases while the passive layer is being developed. During this stage, the metal exhibits a cathodic polarization behavior. Once the passive layer is developed, no further corrosion takes place until the passive layer is disrupted by various means as described later. To summarize, the corrosion rate or current of a passive metal as a function of potential or the oxidizing power of the electrolyte can be represented by a diagram as shown in Fig. 9.4.

Regardless of the mechanism contributing to corrosion, the corrosion rate is influenced by a number of factors dependent upon the nature of the material and

![Figure 9.4](image-url)  

**Figure 9.4** Schematic illustration of passive behavior of a metal. The corrosion rate drops sharply within the passive region.

Copyright © 2004 by Marcel Dekker, Inc.
the environment. Materials-related factors include (i) electrode potential, (ii) surface homogeneity, and (iii) inherent formation of a surface protective film. When anodic and cathodic regions are established within the material, the corrosion rate is accelerated the farther apart are those regions in the galvanic series, and vice versa. Metallic and nonmetallic inclusions in the material, as well as surface areas of local stress concentration, surface irregularities such as machining marks, and microcracks, contribute to accelerating the corrosion rate. If the material is inherently capable of developing a continuous surface protective film, it is said to be passivated; i.e., it becomes resistant to corrosion. For example, stainless steels are passivated by a surface layer of a Cr-rich oxide. It is to be noted that this layer results from a corrosion reaction; i.e., it is a corrosion product. Provided the surface layer is thermodynamically stable, well adhered to the metal, and of a slow growth rate, it acts as an effective barrier between the environment and the underlying metal. However, the corrosion resistance is degraded when the surface layer becomes discontinuous, or is removed and the material becomes unable to reestablish that layer. Environmental-related factors include (i) presence of a suitable electrolyte, (ii) nature of corrosion product, and (iii) duration of exposure to the environment. Both an electrolyte and a proper supply of oxygen are necessary agents for galvanic corrosion to take place. Corrosion is minimized or inhibited by the absence or limited access of either agent. In contrast with surface protective layers (passive layers), spongy and poorly adhered corrosion products contribute to accelerating the corrosion rate. To some extent, the extent of damage caused by corrosion is influenced by the exposure time to a corrosive environment.

9.2.3 Types of Galvanic Couples

In practice, three types of galvanic couples can initiate electrochemical corrosion attack: (i) composition couples, (ii) stress couples, and (iii) concentration couples. Proper design can minimize the corrosive effects of these couples, as described below.

A composition couple consists of regions of different chemical composition or structure, such as two different metals in contact. Fasteners, e.g., a steel bolt holding together aluminum components, are a typical example for setting up galvanic composition cells. It is possible to judge the electrode potential generated in such a cell from the data of Table 9.2. As pointed out earlier, metals or alloys toward the bottom of the electromotive or galvanic series behave in an anodic manner, i.e., corrode, if placed in contact with another metal or alloy above them. Electrode potential increases, and in turn the corrosion rate is accelerated as the two members of a couple become farther apart in the galvanic series. Although aluminum is anodic with respect to iron, the relatively small difference in electrode potential limits the corrosion rate of aluminum when it is in contact with steel. Conversely,
because copper and zinc are more widely separated in the series, zinc corrodes at a higher rate when it is in contact with copper as typified by dezincification of brass (an alloy of copper and zinc). When zinc corrodes, it leaves behind holes in the alloy, rendering it porous, weak, and prone to complete breakdown. It is to be noted that in this case the two metals are present in the form of phases within the same heterogeneous alloy. In general, all multiphase alloys are prone to corrosion by composition couples, as further illustrated in the following example.

Nickel-molybdenum alloys containing 26–30 wt% Mo are known to be highly resistant to reducing media such as hydrochloric acid provided Mo remains in solid solution. However, precipitation of a Mo-rich phase, e.g., a carbide or an intermetallic, leads to establishing anodic and cathodic regions in the alloy. In this example, the Mo-rich phase becomes cathodic with respect to the surrounding Mo-depleted zones, which are anodic. When the material is placed in contact with hydrochloric acid, preferential corrosion attack occurs in the Mo-depleted zones (anodic regions). This behavior can be demonstrated by testing specimens of the Ni-27 Mo alloy in hydrochloric acid after given different types of heat treatments to produce different microstructures. In the heat-treated condition, the alloy consists essentially of a single phase, i.e., a solid solution of Ni and Mo, as shown in the light optical micrograph of Fig. 9.5a. Figure 9.5b illustrates the microstructure of the alloy after 24 h of exposure in boiling 20% hydrochloric acid solution. As can be seen, the microstructure remains similar to that in the heat-treated condition, indicating that corrosion occurs uniformly over the entire test specimen, i.e., by uniform dissolution of the metal. Figure 9.5c illustrates the alloy microstructure after 1000 h of exposure at 700°C. It is observed that a secondary phase of a lamellar structure has precipitated at the grain boundaries. Detailed analysis shows that this phase is Ni₄Mo, and its precipitation has resulted in Mo-depleted zones at the grain boundaries. If the same corrosion test (24 h of exposure in boiling 20% hydrochloric acid solution) is carried out in this condition, extensive corrosion attack at grain boundaries is observed, as shown in Fig. 9.5d. It is clear from the above example that the corrosion behavior of a given material in the same environment is a sensitive function of its microstructure.

Stress couples are developed when parts or regions of the same structure contain higher residual stresses in comparison with adjacent regions. Typical examples include rivet heads, any part where localized stress concentration occurs, and welded components. Residual stresses increase the energy of the atoms, and therefore the stressed regions become anodic and prone to corrosion attack. Just like grain boundaries provide a source of a composition couple as described above, they also act as a source of a stress couple. It is recalled that grain boundaries contain a larger concentration of lattice defects relative to the matrix. As a result, the boundaries become locally stressed leading to the development of a stress couple.
Concentration couples usually develop locally in partly shielded areas as a result of changes in environmental conditions from one surface region of the part to another. For example, variation in the oxygen content of an electrolyte such as water can develop anodic regions depleted in oxygen (corroding) and cathodic regions enriched in oxygen (protected). In practice, variation in oxygen content of an electrolyte can occur in such parts as gasket surfaces, lap weld joints, surface deposits, and crevices under bolt heads and rivet heads. An electrolyte such as water can develop anodic regions depleted in oxygen (corroding) and cathodic regions enriched in oxygen (protected). In practice, variation in oxygen content of an electrolyte can occur in such parts as gasket surfaces, lap weld joints, surface deposits, and crevices under bolt heads and rivet heads. An electrolyte such as

**Figure 9.5** Optical micrographs of Ni-27 Mo alloy after (a) heat treatment showing a single phase of Ni and Mo, (b) 24 h exposure in boiling 20% hydrochloric acid solution showing uniform corrosion, (c) 1000 h exposure at 700 °C showing the precipitation of lamellar Ni<sub>4</sub>Mo phase (marked A) that results in Mo-depleted zones at the grain boundaries, and (d) 24 h exposure in boiling 20% hydrochloric acid solution of depleted alloy resulting in extensive corrosion attack at grain boundaries.
as water can seep into the crevice where it is entrapped. Because of the low oxygen concentration of the entrapped water, the crevice region becomes anodic and preferentially corrodes. A concentration cell is also developed by a speck of dirt or a scale on an exposed metal surface, which becomes a preferential site for corrosion in the presence of atmospheric moisture, even a single drop of water. Grain boundaries intersecting the exposed surface of a part can also act as microscopic crevices conducive to the formation of concentration couples.

9.2.4 Nature of Corrosion Products

It is recalled that corrosion is basically the result of metals seeking their lowest energy state corresponding to their naturally existing ores, such as oxides and salts. Consequently, products resulting from corrosion reactions, which are deposited at the cathodic sites are oxides or salts of the corroding metal. Most of these compounds are porous and are built up as weak, spongy layers with poor adherence to the metal substrate. Usually, corrosion products can be easily separated from the metal surface by scraping, vibration, or friction.

Characterization of corrosion products is of particular importance in failure analysis studies. It provides a great deal of information about the environmental conditions leading to corrosion attack.

9.2.5 Types of Corrosion Attack

It is possible to classify the various types of corrosion into five main types: (i) uniform corrosion, (ii) localized corrosion, (iii) environmental stress cracking, (iv) erosion corrosion, and (v) selective leaching. An account for each of the above types of corrosion is given below.

Uniform or General Corrosion

Uniform or general corrosion proceeds uniformly over the exposed metal surface, resulting in uniform thinning of the part. Sometimes, this type of corrosion, which prevails in all materials, is referred to as chemical corrosion. Corrosion occurs when the material is in contact with a liquid solution in which it has a significant solubility. It is evaluated and quantified by corrosion rates (mpy and mdd) as described earlier.

As explained in the previous section, grain boundaries can lead to the development of composition, stress, and concentration couples. Consequently, grain boundaries can be expected to corrode more rapidly relative to the matrix. In fact, this is the basis for metallographically revealing the grain structure of a material by etching. This may lead to the conclusion that the uniform corrosion rate of a given material in a certain environment increases with decreasing its grain size or increasing the grain boundary area per unit volume. However, the exact rate at which a grain boundary corrodes relative to the matrix is dependent
upon its structure and composition. Because of the larger degree of misorientation and in turn higher energy associated with high-angle grain boundaries, they corrode more rapidly than low-angle boundaries. Furthermore, the boundaries can become extremely reactive (highly anodic) if they contain certain impurities. In practice, corrosion-resistant materials are heat-treated to optimize their resistance to corrosion. Within this context, grain size has little or no significant effect on the uniform corrosion rate, as demonstrated in Table 9.3 for a Ni-Mo alloy resistant to HCl. When the alloy is heat treated to produce a grain size corresponding to ASTM No. 5-6, the uniform corrosion rate in boiling 20% HCl solution is 203 mpy. At a grain size of ASTM No. 2-3, the corrosion rate becomes 229 mpy. In contrast, if the grain boundaries become highly anodic or sensitized by impurities or precipitates, the boundaries can preferentially corrode at a high rate relative to the matrix.

An isocorrosion diagram is a practical tool for presenting uniform corrosion rate data for a given material, indicating its resistance to a specific environment such as an acid. Isocorrosion diagrams resemble the time-temperature-transformation (TTT) diagram in that they contain three variables in a two-dimensional plot. Variables contained in an isocorrosion diagram include temperature, concentration of a certain corrosive species in an aqueous solution, and corrosion rate. As shown in Fig. 9.6, the x axis represents concentration and the y axis represents temperature. Each diagram indicates the effect of concentration on the boiling point of the aqueous solution (boiling point curve). Other curves in the diagram indicates the effects of temperature and concentration on given corrosion rates such as 5, 20, 50 mpy, etc.

To illustrate the usefulness of isocorrosion curves, it is recalled from Sec. 9.2.2 that from an engineering point of view, a material is considered to have a good corrosion resistance to a specific environment if the corrosion rate is less than 5 mpy. The plot of Fig. 9.6 represents the resistance of stainless steel type 316L to formic acid. It is evident that it has good resistance at all concentrations of formic acid provided the service temperature is below that represented by boiling point curve.

It is evident from the above discussion that uniform corrosion results in an overall reduction in cross-sectional area. If the part is stressed, uniform corrosion has the effect of increasing the applied stress. However, since the uniform

**Table 9.3** Effect of Grain Size on the Corrosion Rate of a Ni-Mo Alloy in boiling 20% HCl

<table>
<thead>
<tr>
<th>Grain size (ATM number)</th>
<th>Corrosion rate, μm/year (mpy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5–6</td>
<td>203</td>
</tr>
<tr>
<td>2–3</td>
<td>229</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
corrosion rate is usually slow, its weakening effect on the cross section is rather small. Furthermore, it can easily be detected in its early stages.

Localized Corrosion

When the anodic process becomes localized and limited to specific regions, corrosion becomes of the localized type. Examples include (i) intergranular corrosion of sensitized material (grain boundary attack), (ii) preferential corrosion attack at precipitate-matrix interface, (iii) heat-affected zone attack in welded components, (iv) pitting, which refers to corrosion attack at specific sites of the material, and (v) crevice, which results from oxygen depleted sites related to environmental conditions.

In comparison with uniform corrosion, all types of localized corrosion pose a more serious problem, which can lead to deep pits and even perforations severely weakening the cross section. For example, under fatigue loading conditions, the pits act as notches or localized regions of stress concentration accelerating fatigue failure. Another problem of localized corrosion is that it frequently occurs at spots hidden from direct view, and as a result the damage is not detected until it has progressed into an advanced stage.

Localized intergranular corrosion can occur by three mechanisms: (i) preferential attack of grain boundary precipitates whose composition renders them
anodic in a specific environment, (ii) preferential attack of grain boundaries resulting from segregation of certain impurities at grain boundaries, and (iii) preferential attack of alloy zones adjacent to grain boundary precipitates which deplete those zones in certain alloying elements. Under severe conditions, intergranular corrosion can lead to grain separation causing the material to lose its structural integrity.

Preferential attack of grain boundary precipitates can be exemplified by precipitation of the Mo-rich mu phase in a Ni-Mo-Cr alloy. In an oxidizing environment, the mu phase behaves in anodic manner, and therefore it is preferentially attacked. Since free machining steel grades contain larger concentrations of P and S in comparison with other steels, portions of these elements diffuse into the grain boundaries, rendering them conducive to corrosion. Preferential attack of alloy-depleted zones alongside grain boundaries can result from grain boundary precipitation of the Cr-rich \( \text{M}_{23}\text{C}_{6} \) carbide in some stainless steels and Ni-based alloys. Since the smaller C atoms diffuse to the grain boundaries more rapidly than the larger Cr atoms, precipitation of the Cr-rich phase leaves behind a zone depleted in Cr. Because Cr in solid solution provides resistance to corrosion in oxidizing environments, the Cr-depleted zones alongside grain boundaries become highly susceptible to preferential attack in an oxidizing environment such as HNO\(_3\). Intergranular corrosion associated with precipitates of a secondary phase becomes catastrophic if the precipitates form a continuous grain boundary network.

Localized attack at a precipitate-matrix interface resembles the case of alloy-depleted zones alongside grain boundaries. Precipitation of secondary phases enriched in important alloying elements within the matrix can leave behind a zone at the precipitate-matrix interface depleted in those elements. Such a zone becomes highly anodic (corroding) relative to the matrix (protected). A similar situation can be encountered in a weld heat–affected zone. Phases enriched in alloying elements, providing resistance to corrosion, can precipitate at grain boundaries in a weld heat–affected zone. If the part is used in the as-welded condition, it becomes susceptible to localized corrosion in alloy-depleted zones alongside grain boundaries, resulting in what is known as \textit{weld decay} or \textit{knife line attack}. Knife line attack differs from weld decay in that it is confined to a narrow band of the base metal immediately adjacent to the weld. In contrast, weld decay occurs in regions of the base metal at a larger distance from the weld.

Pitting is another type of localized corrosion where the attack is confined to specific areas relatively small in comparison with the total exposed area of the part. It occurs whenever localized surface couples, such as those described in Sec. 9.2.3, are developed. Engineering alloys which rely upon the formation of a passive surface oxide film for corrosion protection are usually more prone to pitting attack. In this case, pitting is promoted by localized disruption of the protective film, which can result from either the initial surface condition of the part.
and/or other conditions encountered during service. Localized surface deposits such as inclusions can preclude the formation of the protective film. Under this condition, pitting frequently occurs at locations below the surface deposit. If a large surface area is covered with mill scale or an applied coating, pitting can occur at discontinuities in those layers. Since chloride ions are particularly effective in disrupting the continuity of the passive film and penetrating into the metal, pitting is commonly encountered in chloride-containing environments. Because the pits retain the chloride-containing solution, they tend to grow along the direction of the gravitational force. Refractory transition elements, particularly Mo and to a lesser extent W, are found to be effective alloying elements in reducing the susceptibility to pitting of ferrous and nonferrous alloys in chloride-containing environments. It is possible that these elements promote the ability of the alloy to maintain a protective passive film. Although chloride ions are most common source of a pitting attack, it is important to realize that pitting can still occur even if the environment is free of chloride ions. Generally, if the alloy derives its corrosion resistance from a surface passive film, pitting may occur at those locations where the film is locally disrupted for any reason, e.g., mechanical effects, change in environmental conditions, an increase in temperature, etc.

Crevice attack is a form of localized corrosion related to pitting. It usually occurs in materials susceptible to pitting. However, it differs from pitting in that it is caused by localized changes in environmental conditions. As explained in Sec. 9.2.3, crevices are partially shielded areas of the part where the oxygen concentration of the environment becomes less than that of the environment present outside the crevice. This leads to the development of a concentration cell where the crevice site becomes anodic (corroding) relative the region outside the crevice (protected).

Localized corrosion can be judged from metallographic examinations at small magnifications, e.g., 40× or 50×. When corrosion rates are measured and expressed as mpy, they reflect both uniform and localized corrosion attack.

**Environmental Stress Cracking**

Environmental stress cracking is defined as the premature failure of materials under the combined effects of stresses (static or cyclic; applied or residual), and damaging environments (corrosive or noncorrosive). Various types of environmental stress cracking include (i) stress corrosion cracking, (ii) hydrogen damage, (iii) corrosion fatigue, and (iv) fretting corrosion.

Stress corrosion cracking results from the combined effect of a corrosive environment and a tensile stress. It is recalled that the surface energy required to extend a crack is considered to be provided totally by strain energy. However, if an aggressive environment is present at the crack tip, some or all of the crack surface energy can be provided by the free energy of the chemical reaction occurring at the crack tip. As a result, the fracture stress is considerably lowered in the...
presence of an aggressive environment. Usually, stress corrosion cracking is initiated at a corrosion pit. Stress corrosion cracks can propagate either intergranularly or transgranularly, depending up (i) nature of the material, (ii) the environment, and (iii) the stress level, as described below.

Experiment shows that a correlation exists between the crack propagation mechanism and the type of dislocation structure developed by materials having fcc structure. If dislocations cannot cross-slip with relative ease, they tend to form planar arrays. It is recalled that such a dislocation structure is promoted by short-range order and/or low stacking fault energy. Under this condition, the cracks tend to propagate intergranularly. Conversely, in the absence of short-range order or if the material has a high stacking fault energy, dislocations can cross-slip with relative ease, forming subgrain boundaries, and the material becomes resistant to intergranular stress corrosion cracking. However, in this case, the cracking can occur transgranularly. Sensitization of the material by grain boundary precipitates localizing corrosion attack alongside grain boundaries promotes intergranular stress corrosion cracking. Grain size is another material character influencing stress corrosion cracking. Whether the cracking occurs intergranularly or transgranularly, the material becomes more resistant to stress corrosion cracking as its grain size is reduced, which is related to the higher strength associated with a finer grain size.

For a given material, the crack propagation mechanism can change from transgranular to intergranular if the environment is altered. Also, the nature of the environment can have a strong influence on the stress level at which the cracks can propagate. In a given environment, the cracks can propagate under an applied stress as low as 10% of the yield strength of the material. However, in another environment, the stress required to propagate the cracks in the same material can be as high as 90% of the yield strength.

A distinguishing feature of stress corrosion cracking is that the cracks grow incrementally or in a stepwise manner, giving rise to branching as shown in Fig. 9.7. However, exceptions to this rule do exist. Although various models are proposed to explain the mechanism of stress corrosion cracking, the sequence of events leading to crack propagation according to a commonly accepted model are (i) A protective passive film at a crack tip is ruptured under the influence of the local stress; (ii) once the protective film is ruptured, anodic dissolution occurs at the crack tip; (iii) the material redevelops a protective passive film at the crack tip; and so on. There is accumulating evidence that transgranular stress corrosion cracks proceed by a series of cleavage events; i.e., the cracks propagate along certain crystallographic planes. For example, in the case of Ti, the cracking proceeds along a plane inclined at a small angle from the basal plane.

Stress corrosion cracking is pronounced in aggressive environments, particularly those containing chloride ions. Although austenitic stainless steels particularly the 300 series are susceptible to stress corrosion cracking in
Figure 9.7 Illustration of branching during stress corrosion cracking of a 321 stainless steel material exposed to H₂S environment. The EDS spectrum shows the presence of Fe-sulfide within the crack.
chloride-containing environments, most of the high-performance Ni-based alloys are resistant to cracking in a wide variety of chloride-containing environments. However, it is important to realize that no alloy is completely immune to this type of failure. For example, Ni-based alloys are susceptible to stress corrosion cracking in chloride and hydrogen sulfide environments, as well as in fluoride-containing environments. In general, it is rather difficult to predict the incidence of stress corrosion cracking particularly if it is realized that it can occur at an extremely low stress level.

Caustic cracking or embrittlement is a phenomenon related to stress corrosion cracking particularly encountered in boilers. It originates in riveted and welded components, where small leaks allow soluble salts to build up high local concentrations of caustic soda (NaOH) and silica. Although the concentration of soda required to cause this cracking is usually 15–30% in addition to a small concentration of oxygen, it can occur at smaller concentrations reaching 5%. Usually, caustic cracks proceed intergranularly and hence the name *caustic embrittlement*.

*Hydrogen damage* is a term reserved to describe a type of stress corrosion in an environment containing hydrogen. In applications involving hydrogen-rich environments, the resistance of the material to the effects of high-pressure hydrogen is of extreme importance in ensuring the safety of the product. It is possible to classify the damage produced by hydrogen into four main types: (i) hydrogen-induced cracking (HIC) or hydrogen embrittlement, (ii) hydrogen blistering, (iii) decarburization of steels, and (iv) hydrogen attack.

Hydrogen-induced cracking resembles stress corrosion cracking in that the corrodant (atomic hydrogen in this case) diffuses to the crack tip and lowers the stress required to propagate the crack. Usually, hydrogen-induced cracks propagate discontinuously in regions of high triaxial stress (small shear stress) at a rate dependent upon both the stress level and hydrogen concentration. Various chemical environments containing hydrogen atoms such as H2S plus moisture or high pressure H2, as well as poor welding or electroplating procedures, can charge hydrogen into the material.

Hydrogen blistering is a surface phenomenon, which occurs when atomic hydrogen is retained by the material as it solidifies from the melt or as a result of welding operations; acid pickling, electrolytic plating, and surface corrosion reactions. When the material becomes supersaturated in hydrogen, it precipitates as molecular gaseous hydrogen at internal interfaces producing voids. Under pressure, the material is blistered, as shown in Fig. 9.8. Such blisters are also referred to as holidays or fish eyes. They affect both the structural integrity of the part as well as its cosmetic appearance.

Decarburization of steels is a high-temperature process where hydrogen in the environment reacts with carbon in the steel, causing severe loss in strength. Hydrogen attack is another type of damage produced by hydrogen at elevated
temperatures, which is related to decarburization. In this case, however, atomic hydrogen diffuses into the steel and reacts with carbide phases to form gaseous methane. Generation of methane within the steel causes the formation of “bubbles” associated with high internal pressure, which can initiate and propagate cracks.

Another problem related to hydrogen pickup is the precipitation of brittle secondary phases, which in combination with dissolved hydrogen can lead to catastrophic failure. This occurs in alloy systems (e.g., Ti and Zr alloys) where metal halides are thermodynamically stable.

As its name implies, corrosion fatigue results from the combined effect of a corrosive environment and fatigue loading. Typically, this combination considerably shortens the fatigue life in comparison with the sum in whatever order of corrosive damage and mechanical damage incurred separately. Because of the lack of specificity to the corrosive environment, many industries consider corrosion fatigue as a more pervasive problem than stress corrosion cracking. In general, all engineering alloys susceptible to corrosion are expected to be prone to corrosion fatigue. A characteristic feature of cracks produced by corrosion fatigue is their wide-mouthed character and tendency to be tapered to a blunt tip with corrosion product filling much of the crack space. Some of the applications where corrosion fatigue can be a problem include (i) shafts, (ii) agitator blades, (iii) fans, and (iv) suspension wires and ropes.

Fretting corrosion is defined as the damage occurring at the interface of two contacting surfaces undergoing small-amplitude frictional vibrations and/or

Figure 9.8  Illustration of blistering/holidays formed due to the presence of gaseous hydrogen at the internal interface of a material.
slight relative displacement in a corrosive atmosphere. Sometimes, it is referred
to as friction corrosion. It occurs at contact areas of engine components, bolted
parts, automotive parts, etc. shipped by railroads or sea. Damage produced
by fretting corrosion is manifested by discoloration or pits filled with corrosion
product, which is usually an oxide. Since pits can act as stress raisers, fretting
corrosion can lead to fatigue failure.

**Erosion Corrosion**

Erosion corrosion is defined as the accelerated corrosion of a material as a result
of relative motion between the material and the corrosive environment. It is
recalled from earlier discussions that when a material is subjected to general cor-
rosion, it may develop a passive surface film, reducing the rate of attack. If, how-
ever, the corrosive fluid in contact with the metal surface is in a state of motion,
maintaining the surface protective film becomes dependent upon the speed of the
corrosive fluid. When the speed of the fluid reaches the turbulent regime, it can
locally remove the protective film. Such a process is referred to as erosion. Metal
surface exposed by erosion then becomes subject to accelerated corrosion.
Because of the high speed of fluid involved in erosion, complete repassivation
becomes rather difficult.

When a metal surface is damaged by erosion corrosion, it exhibits a distinc-
tive sculpted or carved morphology. Examples of parts which can fail by erosion
corrosion include (i) elbow in steam condensate line, (ii) exhaust or wet-steam
ends of steam turbine blades, (iii) external components of aircrafts, (iv) parts
in front of inlet pipes in tanks, and (v) bends.

**Selective Leaching**

By definition, selective leaching is the selective removal or dissolution of an
element from an alloy by means of a composition couple. A classical example
of selective leaching is dezincification of brass described in Sec. 9.2.3. Another
example is graphitic corrosion of gray cast irons, where its graphite flakes are
selectively removed in mildly corrosive environments.

**9.2.6 Corrosion Resistant Alloys**

Before describing the various classes of corrosion resistant alloys, it is instructive
to know the corrosion resistance of selected metals. Table 9.4 summarizes the
performance of various metals in specific media representing reducing environ-
ments (HCl), oxidizing environments (HNO₃), and alkaline environments
(NaOH/ NH₄OH).

It is possible to classify the corrosion resistant alloys into ferrous and
nonferrous alloys. Ferrous alloys can be either wrought or cast; they include
carbon steels, low-alloy steels, and stainless steels. Nonferrous alloys include
low-grade Ni-based alloys, high grade Ni-based alloys, Co-based alloys, Al-based alloys, Cu and Cu-based alloys, and various metals such as Ti, Zr, and Ta. A survey of various corrosion-resistant alloys and suitable areas of applications are given below.

**Wrought Ferrous Alloys**

It is recalled that carbon steels are Fe-based alloys containing carbon, manganese, silicon, sulfur, and phosphorus. All of these alloying elements are present in fractional percentages (<1%), with the exception of manganese which can be close to 1% or up to 1.5% in certain grades. Corrosion resistance of carbon steels is essentially dependent upon the formation of an oxide surface film acting as a barrier between the metal and the environment. However, in general, carbon steels have limited corrosion resistance. Atmospheric corrosion dependent upon the location is of particular importance in structural applications. In relatively clean air, the products of corrosion are either oxides or carbonates. Corrosion is accelerated in industrial areas where sulfuric acid is present in the atmosphere. A higher corrosion rate is encountered near cities as well as near the ocean. Because of the higher electrical conductivity of the rain and the tendency to form soluble chlorides or sulfates in such environments, the protective oxide film is destroyed. Low-alloy steels have better resistance to atmospheric corrosion. Elements such as copper, nickel, and chromium in small concentrations (<1% each) can considerably decelerate rusting. Phosphorus also provides protection against atmospheric corrosion. Presence of oxygen and/or acidic conditions generally promote atmospheric corrosion of carbon steels; however, corrosion is inhibited under alkaline conditions, e.g., steel embedded in concrete. Generally, carbon steels should not be used in contact with dilute acids. Although they can be used in concentrated sulfuric acid (90–98%) up to the boiling point, however, they are not suitable in the presence of hydrochloric, phosphoric, and nitric acids. Carbon steels are corroded at relatively slow rates in brines and sea water, and therefore they may be used in such environments. Also, they are slightly affected by neutral water and most organic chemicals.
Stainless steels are widely used in structural applications requiring corrosion resistance. All types of stainless steels are Fe-based alloys containing 12–30% Cr, 0–22% Ni, and minor concentrations of C, Nb, Cu, Mo, Ta, Ti, and Se. Depending upon their structures, stainless steels are classified into (i) austenitic, (ii) ferritic, (iii) martensitic, and (iv) duplex stainless steels. Table 9.5 summarizes the performance capabilities of various stainless steels in different environments.

It is to be noted that austenitic stainless steels cannot be hardened by heat treatment; however, their strength can significantly be increased by cold working.

<p>| Table 9.5  Chemical Stability of Various Grades of Stainless Steels |
|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Grade</th>
<th>Condition</th>
<th>Environments of good service performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Austenitic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>Annealed, and fully hardened</td>
<td>General atmosphere including milder industrial atmospheres</td>
</tr>
<tr>
<td>302</td>
<td>Annealed, and fully hardened</td>
<td>Industrial atmospheres, freshwater exposure</td>
</tr>
<tr>
<td>304</td>
<td>Annealed</td>
<td>Marine atmosphere, good stress corrosion cracking</td>
</tr>
<tr>
<td>316</td>
<td>Annealed</td>
<td>Sea water; acetic, nitric, and phosphoric acids, nitrates</td>
</tr>
<tr>
<td>347</td>
<td>Annealed</td>
<td>Dilute chemicals; excellent stress corrosion resistance</td>
</tr>
<tr>
<td><strong>Ferritic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>Annealed</td>
<td>General atmosphere</td>
</tr>
<tr>
<td>430</td>
<td>Annealed</td>
<td>Marine atmosphere, dilute acids, and nitrates</td>
</tr>
<tr>
<td>446</td>
<td>Annealed</td>
<td>Industrial atmospheres, high-temperature sulfur-bearing media</td>
</tr>
<tr>
<td><strong>Martensitic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>416</td>
<td>Hardened and tempered</td>
<td>General stability, fair chemical stability</td>
</tr>
<tr>
<td>431</td>
<td>Hardened and tempered</td>
<td>Same as 416</td>
</tr>
<tr>
<td>440B</td>
<td>Annealed, hardened, and tempered</td>
<td>Fair chemical stability</td>
</tr>
<tr>
<td><strong>Duplex stainless steels</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferralium alloy</td>
<td>Annealed</td>
<td>Sulfuric acid production, hot organic acids, crude oil treatments</td>
</tr>
</tbody>
</table>
All stainless steels derive their corrosion resistance from an adherent surface film of chromium oxide. Basic 18Cr-8Ni steels (301 and 302 grades) have good chemical stability under normal atmospheric conditions and they are weldable. Also, they are characterized by good corrosion resistance to industrial atmospheres as well as marine environments. However, pitting may occur with extended exposure. Salt water applications require higher grades. Solutions containing chlorine ions in addition to hydrochloric and hydrofluoric acid cause severe corrosion at all concentrations and temperatures. Another disadvantage of the lower grades of stainless steels is sensitization, referring to grain boundary precipitation of Cr-rich carbide phases. Sensitization promotes intergranular corrosion attack. This problem, however, is avoided in higher grades of stainless steels such as 304, 316, and 347. In these steels, the carbon content is kept at a very low level (0.08 maximum). Carbide stabilizing element such as Nb and Ta are added to grade 347, minimizing the tendency to form Cr-rich carbides at grain boundaries. A similar result is obtained by adding Ti to grade 321.

Ferritic grades of stainless steels derive their corrosion resistance from the presence of Cr requiring at least 12% Cr in the alloy. Grade 405, containing a little more than 12% Cr, has good corrosion resistance under common atmospheric conditions as well as in fresh water and even in the presence of mild acids. Higher Cr content increases the chemical stability of grade 430, and a better performance is obtained by using grade 440 particularly at moderately elevated temperatures.

Similar to ferritic stainless steels, martensitic steels derive their corrosion resistance from the presence of Cr. These grades develop maximum resistance to atmospheric conditions in the fully hardened condition.

Duplex stainless steels consist of a mixture of ferrite and austenite. They combine the high strength of ferritic steels and good ductility of austenitic steels. Such steels can find many applications in the chemical process, petrochemical, and oil industries.

**Cast Ferrous Alloys**

Some structural applications require that chemical stability in certain environments must be combined with good castability. Corrosion resistant cast ferrous alloys are available in three classes depending upon the main alloying element: (i) high Cr alloys, (ii) high Ni alloys, and (iii) high Si alloys.

High-chromium cast alloys contain 25–30% Cr and up to 3% C. Other alloying elements such as Ni are added to enhance their corrosion resistance. Although these alloys are rather expensive and have poor ductility, they are characterized by an outstanding corrosion resistance up to a temperature of about 1000 °C. Frequently, these alloys are used in applications requiring both chemical stability and oxidation resistance.

High-nickel cast alloys maintain an austenitic structure at room temperature. Their Ni content varies from 13 to 22%, and the C content can reach 3%
or slightly higher. Other alloying elements include Cr, Mn, and Si added in smaller amounts. Corrosion resistance of these alloys is comparable to that of wrought austenitic stainless steels in such environments as dilute acids, alkalies, and salts.

High-silicon alloys contain 13–17% Si; however, the C content is not more than 1%. Although these alloys have poor mechanical strength and fair castability, they have outstanding resistance to nitric, sulfuric, and phosphoric acids. However, they are readily attacked by hydrochloric acid. Resistance to hydrochloric acid is improved in some grades containing Mo and Ni.

**Nonferrous Metals and Alloys**

Nonferrous materials resistant to low-temperature corrosion include a wide variety of alloys and metals such as low-grade Ni-based alloys, high-grade Ni-based alloys, Cu and Cu-based alloys, Al-based alloys, Ti and Ti-based alloys, Zr, Ta, and Nb. A brief account for the characteristic corrosion properties of these materials is given below.

Commercially pure Ni is considered to be a general-purpose material used in corrosion applications where the special properties of Ni-based alloys described later are not required. It is known for its outstanding resistance to hot or cold alkalies as well as its resistance to stress corrosion cracking. Also, it is resistant to dilute nonoxidizing acids such as sulfuric, hydrochloric, and phosphoric, as well as anhydrous ammonia and ammonium hydroxide solutions of less than 1% concentration.

Low-grade Ni-based alloys such as Ni-Cu alloys (monel series) are known for their ability to handle sea and brackish water. Also, in comparison with commercially pure Ni, they have improved resistance to nonoxidizing acids. However, these alloys are not resistant to oxidizing media such as nitric acid, chromic acid, wet chlorine, and ammonia.

High-grade Ni-based alloys include a series of Ni-Mo, Ni-Mo-Cr, Ni-Mo-W-Cr, Ni-Cr-Fe, and Ni-Mo-Cr-Fe alloys. Most of these alloys are more resistant to stress corrosion cracking than the stainless steels. In general, Mo and W are important alloying elements for corrosion resistance in reducing media while Cr is essential for oxidizing media. Also, it is recalled that Mo promotes the resistance to pitting. Alloys based upon the Ni-Mo alloys are known for their outstanding resistance to reducing media such as HCl; however, they are not resistant to oxidizing media. Resistance to both reducing and oxidizing environments is provided by a series of Ni-Mo-Cr alloys. Other elements are added to Ni-Mo-Cr alloys to enhance their resistance to certain media. For example, addition of small concentrations of Cu promotes the resistance of Ni-Mo-Cr alloys to reducing acids. Some of the Ni-Mo-Cr alloys contain W for further improvement of corrosion resistance in reducing media. Alloys based upon the Ni-Cr-Fe system are particularly useful in handling water environments because of their improved
resistance to stress corrosion cracking in comparison with stainless steels. Also, these alloys are highly resistant to ammonia. Their resistance to oxidizing media is dependent upon the exact Cr concentrations. However, due to the lack of Mo, they are not resistant to reducing media.

Commercially pure Cu is highly resistant to various atmospheric environments, such as rural, marine, and industrial. However, if sulfur is present in the environment, it undergoes discoloration due to the formation of various oxides. Furthermore, Cu is resistant to dry gases; fresh and sea water; pure steam; most soils; deaerated nonoxidizing acids; organic acids such as acetic, formic, lactic, citric, and maleic; alkaline solutions except those containing ammonium compounds or cyanides; salt solutions such as sulfates and nitrates of sodium, potassium, magnesium, and calcium. An added advantage of Cu is that alloying elements added to improve its mechanical strength either have no effect on its corrosion resistance or improve this resistance in certain environments. For example, the addition of Zn to Cu improves its resistance to sulfur compounds. In addition to commercially pure Cu, commercial bronze (Cu-Pb-Zn alloys), red brass (Cu-Zn alloys), Muntz metal of the brass grades, and nickel silver are the most widely used in applications requiring resistance to general atmospheric environments. Both Cu and its alloys, however, are attacked by oxidizing acids, salts, moist ammonia, halogen gases, sulfides, high-velocity sea water, and a number of liquid metals such as tin and lead.

Commercially pure Al is characterized by excellent chemical stability under most atmospheric conditions owing to formation of a passive surface layer of Al₂O₃. Specifically, Al is resistant to rural, industrial, and marine atmospheres as well as neutral or nearly neutral fresh waters, sea waters, organic acids, oils, greases, waxes, ammonia and ammonium compounds, concentrated nitric acid (>82%), and many neutral aqueous inorganic salt solutions. Its mechanical strength, however, is inadequate for most structural and mechanical applications. Most of the alloying elements added to improve the mechanical strength of Al degrades its chemical stability. However, Mg, Mn, Cr, and Si are considered to be valuable alloying additions to Al. Addition of Mg to Al preserves its chemical stability, and therefore Al-Mg alloys combine fairly good strength and excellent corrosion resistance particularly in salt water and alkaline solutions. When these alloys, however, contain Mg-rich precipitates, the general corrosion resistance is lowered and the susceptibility to stress corrosion cracking is increased. A better combination of mechanical strength is found in Al-Mg-Si alloys. In this case, strengthening is provided by Mg₃Si, which does not affect the electrode potential of the alloy.

Commercially pure Ti has a chemical stability comparable to or better than that of the best grades of stainless steels. It has an outstanding corrosion resistance to most media including nitric acid in all concentrations and at all temperatures, inorganic acids, organic salts, and low-temperature dilute hydrochloric,
sulfuric, and phosphoric acids, and aqua regia. Also, it is resistant to pitting attack by chloride solutions (sea water), and to alkalies. However, Ti is attacked by aluminum chloride and boiling concentrated potassium hydroxide, and hydrofluoric acid. Although Ti alloys are generally less resistant to corrosion than the commercially pure metal, the high-strength, heat-treatable grades maintain adequate chemical stability in many environments. Industrial applications of Ti and its alloys are, however, limited to some extent due to their high cost.

Commerially pure zirconium (Zr) has excellent resistance to hydrochloric and nitric acids at all concentrations and temperatures up to boiling point. Also, it resists sulfuric acid up to 50% concentration as well as alkalies in all concentrations and at all temperatures. However, it is attacked by hydrofluoric acid and aqua regia. Commercially pure tantalum (Ta) is resistant to all acids except hydrofluoric and hot sulfuric acids. Also it is attacked by alkalies.

9.2.7 Corrosion Control and Protection

Design engineers can control the extent of damage caused by corrosion using four methods: (i) proper selection of corrosion-resistant materials, (ii) application of surface protective coatings, (iii) closer control of the environment, and (iv) proper design.

Material Selection

As described in Sec. 9.2.6, many materials are available to designers for applications requiring resistance to low-temperature aqueous corrosion. Selection of the proper material requires an exact knowledge of the environmental conditions to which the material is exposed to during service, particularly the composition of the corrosive medium and the operating temperature. Some of the corrosive environments encountered in practice and the corresponding materials requirements are described below.

In many industrial processes, HCl results from the hydrolysis of catalysts, i.e., their chemical decomposition by reaction with water. Evidently, structural materials for such applications require adequate resistance to the corrosive effect of HCl. It is recalled from earlier discussions that Mo is a very effective element in such an environment. Therefore, Mo-containing alloys must be considered for such applications. It is important to realize that the resistance to HCl is a function of its concentration in the environment, as well as the Mo content of the alloy. Stainless steels, in general, are not adequate because of their low Mo content. In a dilute HCl environment (1% concentration) up to the boiling point, a number of nonferrous alloys containing 5.5–27% Mo provide similar resistance. However, in concentrated HCl environments, only the high Mo-containing alloys must be considered. Frequently, the environment is not purely HCl. Rather, it contains other ionic species, e.g., aluminum chloride (nonoxidizing) from the
catalyst and ferric chloride (oxidizing) from the steel vessels containing the feedstock. In this case, alloys resistant to both oxidizing and reducing media can have better performance, e.g., alloys based upon the Ni-Mo-Cr system.

Environments containing hydrofluoric acid present a different case than those containing hydrochloric acid. In this case, it is rather difficult to correlate the resistance to hydrofluoric acid with certain alloying elements. Although the Mo-containing alloys are not necessarily the most resistant, Ni-based alloys usually have better performance than the stainless steels. In particular, Ni-Cu alloys exhibit a high resistance to hydrofluoric acid.

In practice, mixtures of different acids or acids and salts are usually encountered rather than a single corrosive species. Frequently, the corrosion resistance of a given alloy in a mixture of corrosive species cannot be predicted from its resistance in a single species. For mixtures of inorganic acids, the addition of HCl or other chloride salts to sulfuric acid tends to increase the corrosion rate. Under this condition, alloys containing a high concentration of Mo exhibit a relatively better performance. However, the effect of adding hydrofluoric acid is dependent upon the concentration of sulfuric acid. Addition of hydrofluoric acid to a dilute sulfuric acid increases the corrosion rate. Conversely, the corrosion rate is decreased in concentrated sulfuric acid. Severe corrosion results if Fe-based alloys are used in an environment consisting of dilute sulfuric acid and hydrofluoric acid. Some of the Ni-Mo-Cr-Fe alloys have, however, good resistance to this environment. Alloys containing Cu are recommended in environments consisting of hydrofluoric acid and concentrated sulfuric acid.

Mixtures of organic acids are encountered in such processes as crude oil distillation units. In this case, the acid mixture consists of saturated and unsaturated fatty acids. For such an application, the selection of material is based upon the acidity of the environment. Experience, however, shows that the presence of Mo in the alloy is highly beneficial in this application. Although many alloys are resistant to acetic acid, the addition of chloride as oxidizing chloride salts tends to increase the corrosion rate. In this environment, alloys containing low Mo content are more resistant to corrosion.

A corrosive environment usually encountered in oil refineries and petrochemical operations is polythionic acid (\( \text{H}_2\text{S}_n\text{O}_6 \), where \( n = 2–5 \)). It results from the combination of sulfur compounds, moisture condensation, and oxygen. This acid is found to result in stress corrosion cracking of both stainless steels and Ni-based alloys, particularly in the sensitized condition. In this case, the cracking is intergranular. However, it can also occur in some materials in the annealed condition, where it becomes transgranular. It is believed that polythionic acid forms in refinery units during downtime on surface covered with a sulfide in the presence of air (or oxygen) and moistures at ambient temperatures.
Surface Protection

Although there are many materials which are inherently resistant to corrosion in a wide variety of environments, their use in structural applications can be limited by high cost, inadequate strength, or fabrication difficulties. In applications requiring the use of materials whose inherent resistance to corrosion is inadequate, the vulnerable surface is covered with a thin layer of a highly corrosion-resistant material. Therefore, the underlying base material or substrate provides the desired mechanical or other functional requirements, and the surface coating provides the necessary environmental protection.

Surface protective coatings applied by hot dipping are the most commonly used for applications requiring resistance to low-temperature aqueous corrosion. This method of protection is based upon using metals of lower melting points in comparison with the base metal. Among those metals are Al, Zn, Sn, Pb, and their alloys. Zinc coating by hot dipping, commonly referred to as galvanizing, is the most important method used to protect ferrous alloys. The coating layer may be applied manually by dipping the part in molten Zn, particularly for small parts or irregularly shaped objects such as castings. Galvanizing of sheet and wire products is carried out automatically where the thickness of the coating is closely monitored. Although the intermetallic compound formed between Zn and the base iron or steel provides good adherence, it can also lead to brittleness, requiring a close control of the process variables to maintain a thin transition zone between the coating and base metal. Galvanized products are characterized by excellent resistance to general atmospheric attack, even in highly contaminated environments. Protective coatings produced by hot dipping in Sn are also used. Although Sn coatings provide chemical stability in severely corrosive environments such as dilute acids, alkalies, and several organic compounds, their use is rather limited by higher cost. It is important to realize that surface protection by hot dipping is limited to steels in the annealed condition. Because of the high temperature required to apply the coating, the mechanical strength of steels hardened by heat treatment or cold work can be seriously degraded.

In addition to hot dipping, surface protection of steels can be provided by formation of an oxide layer at the surface, which is particularly effective in mildly corrosive environments. Such an oxide layer acts as a barrier between the environment and the base metal. It is important that the oxide layer must be continuous and well adhered to the base metal. Since the natural iron oxide Fe₂O₃ (common rust) is porous and soft, it is not suitable for that purpose. A protective oxide layer of Fe₃O₄ (black oxide) is developed by immersing the steel in a salt baths at a temperature of about 150°C. A somewhat better protection can be achieved by forming a surface layer of iron phosphate at the surface of the steel part by dipping it into a processing solution at a temperature slightly above ambient.
Electroplating is another method of surface protection against corrosion attack. Both ferrous and nonferrous metals can be protected by this method. Cadmium, zinc, nickel, and chromium are the major plating elements for protection against corrosion attack. In an electrolytic cell, the plating metal is made the anode and the part to be plated acts as a cathode. Protection of steels against mildly corrosive environments, e.g., ordinary atmospheric conditions, is achieved by cadmium plating. Absorption of hydrogen by the steel during electroplating resulting in hydrogen embrittlement is a major problem associated with electroplating. However, this problem can be controlled most effectively by baking the component immediately after electroplating at a temperature of about 190°C.

Mechanical cladding is another method of surface protection. It consists of firmly attaching a surface layer of a corrosion-resistant material to the base metal by casting, forging, welding, or brazing followed by rolling, drawing, or extrusion operation. Therefore, various product shapes such as sheets, rods, wires can be protected by a solidly bonded thin surface layer of a corrosion-resistant material. Subsequent heat treatment can be used to optimize the mechanical properties of the base metal.

**Cathodic Protection**

Cathodic protection is a protective technique based upon reversing the chemical reaction leading to corrosion. It is recalled that electrochemical corrosion occurs as follows:

\[
\text{Oxidation reaction (anode)}: \ M \rightarrow Mn^+ + ne^- \\
\text{Reduction reaction (cathode)}: \ ne^- + nH^+ \rightarrow n/2H_2
\]

It is evident that supplying excess electrons to the above cell suppresses the anodic dissolution of the metal and accelerates the cathodic evolution of hydrogen. In practice, cathodic protection is accomplished by connecting a more active metal (anodic) to the metal to be protected. Therefore, the anodic metal dissolves preferentially and must be replaced periodically. Steels are cathodically protected by Zn and Mg anodes, and the method is particularly effective in marine structures, pipelines, bridge decks, equipment, and tanks of all types, particularly below water or underground.

Buried and submerged metal structures in the oil, gas, and waterworks industries are increasingly made resistant to corrosion by cathodic protection. Pipelines are now routinely designed to ensure the electrical continuity required for cathodic protection to function.

**Control of the Environment**

Since corrosion is very sensitive to the environmental conditions, a closer control of the environment provides an effective means for protection against corrosion.
attack. Various methods of controlling the environment include (i) a change in concentration of corrosive species, (ii) lowering the service temperature, (iii) decreasing the velocity of the corrosive fluid, and (iv) removing oxidizing agents.

To illustrate the role of changing concentration, it is recalled that the corrosion rate is dependent upon the concentration of corrosive species in the environment. Many acids become noncorrosive or inert at high concentrations, e.g., sulfuric acid and phosphoric acid. In such cases, it becomes possible to reduce the corrosion rate by increasing the concentration of the respective acid in the environment.

Lowering the service temperature as a means for corrosion protection is dependent upon the particular environment. Although, in many cases, lowering the temperature results in a decrease in corrosion rate, in other cases, changing temperature can have a little or no effect. Also, for certain environments, increasing the temperature lowers the corrosion rate, e.g., sea water.

A further reduction in corrosion rate can be obtained by decreasing the speed of the corrosive fluid. It is important to realize, however, that metals inherently protected by a surface passive layer exhibit higher corrosion rates in stagnant fluids. Extremely high fluid speed should always be avoided to minimize the incident of erosion-corrosion. Removing oxygen and oxidizing agents from the environment is another method contributing to a reduction in corrosion rate.

Addition of corrosion inhibitors to the environment is also used to reduce the corrosion rate. By definition, a corrosion inhibitor is a substance added to the environment in small concentration to reduce its corrosive effect. Examples of inhibitors include organic amines, arsenic and antimony ions, sodium sulfide, hydrazine, nitrates, chromates, and ferric salts.

**Design Considerations**

It is possible to control the extent of corrosion damage by better selection of plant layout and location, as well as by proper design of equipment. Whenever possible, coastal location of plants must be avoided. Also, the plant layout must take into consideration the direction of winds, and possible fallout of corrosive species. Both galvanic and crevice corrosion can be controlled by proper design of equipment.

### 9.3 High-Temperature Corrosion

High-temperature corrosion differs from low-temperature aqueous corrosion in that the corrosive species are primarily in the gaseous state. In this case, corrosion proceeds by a direct chemical reaction between gaseous species in the environment and the alloy. Corrosive environments responsible for materials degradation at elevated temperatures can be classified into two main types: (i) gas turbine environments and (ii) nongas turbine environments. Resistance
to high-temperature corrosive environments is provided either inherently or by means of applying a surface protective coating. Inherently protected alloys derive their resistance to high-temperature corrosion from a protective surface scale based upon chromia (Cr$_2$O$_3$), alumina (Al$_2$O$_3$), or a combination of both. Protection provided by Cr$_2$O$_3$ is only adequate up to a temperature of about 950°C. Above that temperature, Cr$_2$O$_3$ tends to convert into volatile CrO$_3$. Because of the higher thermodynamic stability of Al$_2$O$_3$, it provides better protection at extreme temperatures above 1000°C. Alloys with inadequate inherent resistance to high-temperature corrosion, such as those used in gas turbine blade applications, are protected by surface coatings. It is the primary function of the coating to develop a protective Al$_2$O$_3$ scale upon exposure to elevated temperatures.

High-temperature alloys used in gas turbine applications are commonly known as the superalloys, most of which are Ni-based. To satisfy the mechanical strength requirements of these applications, superalloys used in the hottest sections such as turbine blades must contain about 5% Al. Because of their insufficient ductility, these alloys are produced as castings. In a gas turbine where the environment is highly oxidizing, salts deposited on the surface of structural alloys and protective coatings can considerably accelerate the oxidation reaction, leading to a type of corrosion attack known as hot corrosion. It particularly affects the Ni- and Co-based superalloys and protective surface coatings used in gas turbine applications. Alkaline sulfates such as Na$_2$SO$_4$ are the main aggressive compounds formed in the environment of a gas turbine. Deposition of Na$_2$SO$_4$ can result from reaction between S in the fuel and NaCl in the combustion air. It is also possible that Na$_2$SO$_4$ results from combustion of fuels containing both S and Na. When Na$_2$SO$_4$ is deposited on the surface of the alloy or coating, it degrades the protective nature of the oxide scale resulting in sulfidation attack, which accelerates the oxidation reaction. Severe hot corrosion occurs at temperatures in the range of 700–900°C, and in the presence of sea salt it can occur at temperatures as low as 450°C.

Many of the nongas turbine applications at high temperatures require the use of wrought alloys. Except in a very few cases, the amount of Al required to develop a continuous protective layer of Al$_2$O$_3$ considerably reduces the ductility of the alloy precluding the processing of wrought products. Therefore, most of the high-temperature alloys used in nongas turbine applications, including the Fe-, Ni-, and Co-based alloys, are inherently protected against environmental degradation by Cr$_2$O$_3$ scale. Commonly encountered corrodatns in high-temperature environments of nongas turbine industrial processes are sulfur, chlorine, and carbon. Primary modes of corrosion attack include sulfidation, carburization, and chlorination. It is recalled that from a chemical point of view, any type of corrosion is an oxidation reaction.

At high temperatures, the presence of corrosive species in the environment, such as sulfur, carbon, and chlorine, can degrade the protective nature of the
oxide scale resulting in an accelerated oxidation attack. For both gas turbine and nongas turbine environments, the rate of attack varies with exposure time at a given temperature, as schematically illustrated in Fig. 9.9. Initially, the alloy develops a protective oxide scale corresponding to a very small rate of attack. Subsequently, the protective nature of the scale is degraded by various processes described later, which accelerates the attack. Since oxidation plays an important role in high temperature corrosion, a brief account for this reaction is given below.

9.3.1 High-Temperature Oxidation

When an alloy is exposed to a high temperature, the initial rapid up take of oxygen from the environment converts the surface of the alloy into an oxide whose composition is a function of the alloy composition. During the early stages of the oxidation reaction, various oxides of the reactive components of the alloy are formed, e.g., FeO, Cr$_2$O$_3$, NiO, CoO, Ti$_2$O$_3$, MnO, Al$_2$O$_3$, SiO$_2$, La$_2$O$_3$, Y$_2$O$_3$, etc. Each oxide of an element is formed in a proportion dependent upon the concentration of that element in the alloy. As the oxidation reaction progresses, the sequence of events becomes dependent on the relative stabilities of various oxides and their growth rate, as described below.

![Figure 9.9](image)

**Figure 9.9** Effect of exposure time at 1150°C on the oxidation behavior of Al$_2$O$_3$ forming (1) and Cr$_2$O$_3$ (2 and 3) forming alloys in still air.
Less stable oxides may be converted into more stable oxides, and oxides of higher growth rate overgrow those of slower growth rate. From a thermodynamic point of view, the relative stability of solid oxides is well reflected by their enthalpies of formation. Similar to free energy, enthalpy is a thermodynamic quantity representing the heat content of a substance. By convention, the heat of formation of a given compound is considered negative, and the more negative is the heat of formation or enthalpy of a given compound, the greater is its stability. Figure 9.10 summarizes the relative stability of various oxides. Among these oxides, NiO and CoO and FeO are the least stable. Alumina can be seen to have a relatively high stability, and Cr₂O₃ has an intermediate stability. Growth rate of various oxides can be determined from an equation of the type:

\[ K = K_0 \exp(-Q/RT) \]

where \( K \) is the temperature-dependent reaction rate constant, \( K_0 \) is a constant, \( Q \) is the activation energy of the reaction, \( R \) is the universal gas constant, and \( T \) is the temperature in degrees Kelvin (K). A plot of \( \log K \) vs. \( 1/T \) yields a straight line whose slope is proportional to the activation energy \( Q \), as shown in Fig. 9.11. Alumina has a comparatively slower growth rate. Oxides such as NiO, CoO, and FeO have higher growth rates, and Cr₂O₃ has an intermediate growth rate.
Consider the oxidation of a simple Ni-22Cr alloy that leads to the establishment of a protective scale. During the early stages of oxidation characterized by rapid kinetics, oxides of both Ni and Cr are formed on the surface. Since the concentration of Ni is higher than that of Cr, the proportion of NiO formed at the surface is greater than that of Cr$_2$O$_3$. Nickel oxide is expected to grow more rapidly than Cr$_2$O$_3$. With continued exposure, NiO overgrows Cr$_2$O$_3$. To increase its stability, i.e., lower its free energy, NiO reacts with Cr$_2$O$_3$ to form a solid solution by the following reaction:

$$\text{NiO} + \text{Cr}_2\text{O}_3 \rightarrow \text{NiCr}_2\text{O}_4$$

An oxide such as NiCr$_2$O$_4$, containing more than one metallic element, is called a spinel. Because the oxygen activity established at the alloy surface by oxides such as NiCr$_2$O$_4$ and NiO is less than that required to oxidize Cr in the alloy, oxygen continues to diffuse inward, precipitating Cr$_2$O$_3$. Eventually, particles of the slower growing Cr$_2$O$_3$ are spread laterally to cover the entire surface of the alloy forming a protective layer. Once the protective scale is formed, further growth of other oxides is blocked decelerating the kinetics of the reaction. At this stage, further growth of the oxide must occur by outward diffusional transport through
the protective scale, inward diffusion of oxygen or a combination of both. With extended exposure to elevated temperatures, the oxide scale can lose its protective nature by various processes accelerating the reaction. Examples of such processes include growth stresses leading to spallation of the scale, and diffusion of other corrosive species into the surface, converting the protective scale into nonprotective oxide.

Other than Cr, structural alloys for high-temperature service contain several alloying elements, complicating the simplified scheme of oxidation mechanics. However, the principles involved are similar. To be protective, the oxide scale developed by an alloy upon exposure to elevated temperatures must satisfy the following requirements:

1. Upon exposure to elevated temperatures, the oxide scale must initially form rapidly to cover the surface of the alloy, and its subsequent growth must proceed at a slow rate, satisfying a parabolic behavior.
2. It must be thermodynamically stable, i.e., maintain its structure and composition with continued exposure to elevated temperatures.
3. It must form a continuous tenacious layer, well adhered to the alloy and resistant to spallation upon thermal cycling.
4. It must have a low vapor pressure.
5. It must form an effective barrier to inward diffusion of oxygen and outward diffusion of metallic atoms.
6. It must be dense and free of cracks and interconnected pores.
7. It must be strong, tough, and resistant to abrasion.

Although Al$_2$O$_3$ has a greater thermodynamic stability and slower growth rate in comparison with Cr$_2$O$_3$, its adherence is rather inferior. However, the adherence of Al$_2$O$_3$ can be significantly improved by critical additions of active elements such as Hf and Y. Also, the protective nature of Cr$_2$O$_3$ can be improved by critical additions of La.

To summarize, during the initial stages of oxidation, the reaction is characterized by rapid kinetics. Commonly, this stage is referred to as primary oxidation. Upon the development of a protective oxide scale, the kinetics are decelerated and the reaction reaches the stage of steady-state oxidation. During this stage, the kinetics of the reaction at a given temperature are governed by a parabolic relationship of the type:

\[ x^2 = Kt \]

where \( x \) is weight gain per unit surface area, \( K \) is the temperature-dependent reaction rate constant, and \( t \) is the exposure time. When the oxide scale becomes nonprotective, the oxidation rate is accelerated during a stage called breakaway...
oxidation. Within this stage, the oxidation rate is usually governed by an equation of the type:

\[ x = Kt \]

### 9.3.2 Hot Corrosion

Hot corrosion is a term reserved to describe a type of high temperature corrosion encountered in gas turbine engines. As pointed out earlier, hot corrosion is induced by the formation of \( \text{Na}_2\text{SO}_4 \) and its deposition on the surface of the component. Two types of hot corrosion are encountered during service depending upon the service temperature. At relatively higher temperatures above about 800°C, where \( \text{Na}_2\text{SO}_4 \) is partially or totally molten, corrosion is referred to as type I. When the temperature is between 650 and 800°C, the salt deposit becomes solid, and type II corrosion is identified.

Depending upon the alloy composition and environmental parameters, type I hot corrosion can occur by two mechanisms: (i) basic fluxing and (ii) acidic fluxing. Basic fluxing occurs when the concentration or activity of oxide ions (\( \text{Na}_2\text{O} \)) in the molten salt deposited on the alloy is increased. Conversely, acidic fluxing is associated with a decreased concentration of the oxide ions.

Basic fluxing affects the alloys primarily protected by \( \text{Al}_2\text{O}_3 \) scale, which usually contains \(<15\%\) Cr. When the alloy surface becomes covered with an oxide scale, the oxygen pressure at the oxide-salt interface is reduced. As a result, sulfate ions (\( \text{SO}_4^{2-} \)) in the salt are partially dissociated by the following reaction:

\[ \text{SO}_4^{2-} \to \text{S} + \frac{1}{2}\text{O}_2 + \text{O}^{2-} \]

Protective scales on the alloy surface particularly \( \text{Al}_2\text{O}_3 \) react with \( \text{O}^{2-} \) to form anions soluble in the salt according to the following reaction:

\[ \text{Al}_2\text{O}_3 + \text{O}^{2-} \to 2\text{AlO}_2^{2-} \]

As the protective oxide scale is disrupted by the above reaction, S resulting from the dissociation of \( \text{SO}_4^{2-} \) diffuses into the alloy where it reacts with reactive elements particularly Cr to form Cr-rich sulfide phases. With continued loss of Cr, the attack becomes more severe.

Acidic fluxing occurs when the alloy contains refractory elements, e.g., Mo, W, and V forming acidic oxides (\( \text{MoO}_3 \), \( \text{WO}_3 \), and \( \text{V}_2\text{O}_5 \)). It can also occur if the fuel contains vanadium. When these oxides dissolve in the molten salt, the oxide ion concentration is decreased by the following reaction:

\[ \text{MoO}_3 + \text{O}^{2-} \to \text{MoO}_4^{2-} \]

and thus, the salt becomes acidic. When the salt becomes sufficiently acidic, it permits the dissolution of protective oxides such as \( \text{Al}_2\text{O}_3 \) and \( \text{Cr}_2\text{O}_3 \) by the
following reactions:

\[
\text{Al}_2\text{O}_3 \rightarrow 2\text{Al}^{3+} + 3\text{O}^{2-}
\]

\[
\text{Cr}_2\text{O}_3 \rightarrow 2\text{Cr}^{3+} + 3\text{O}^{2-}
\]

Type II hot corrosion occurs at temperatures between 650 and 800°C. Cobalt-based alloys and coating systems containing less than about 20% Cr are the most susceptible to this type of hot corrosion; however, it can also affect Ni-based alloys. At the relatively lower temperatures where type II hot corrosion occurs, the salt deposit is expected to be in the solid state (melting point of Na$_2$SO$_4$ = 884°C). However, the relatively high partial pressure of SO$_3$ in the salt at lower temperatures can promote the formation of mixed sulfates having low melting points. As a result, a mixed sulfate of the type CoSO$_4$-Na$_2$SO$_4$ in the liquid state is deposited on the surface of Co-based alloys. In the case of Ni-base alloys, the liquid deposit is of the type NiSO$_4$-Na$_2$SO$_4$. When the liquid sulfate is deposited on the oxide scale developed by the alloy, the partial pressure of O$_2$ is locally decreased, and the partial pressures of SO$_2$ and S$_2$ are increased. Under these conditions, the alloy cannot maintain a protective scale. Instead, either sulfides of Cr and/or Al are formed in the alloy, or a porous nonprotective oxide is formed in regions of high partial pressure of oxygen. Sulfidation occurs in Ni-based alloys or alloys containing a high concentration of Ni.

Protective scales based upon Cr$_2$O$_3$ are more resistant to hot corrosion attack than Al$_2$O$_3$ scale, particularly at lower temperatures where the environment is severely sulfidizing. Under these conditions, Al$_2$O$_3$ tends to be intermixed with deposited salt degrading its protective nature. However, at higher temperatures where the environment becomes more oxidizing, Al$_2$O$_3$ scale offers a greater protection. Therefore, alloys performing in hot corrosive environments must be capable of establishing Cr$_2$O$_3$ scale.

Each of the two types of hot corrosion is distinguished by a characteristic attack morphology. High-temperature or type I hot corrosion is distinguished by a network of a sulfide phase beneath an external oxide layer. In contrast, low-temperature or type II hot corrosion is distinguished by a pitting-type attack. Each pit consisting of Cr$_2$O$_3$, Al$_2$O$_3$, or a mixture of both is covered by either a Ni- or Co-rich oxide and is separated from the alloy or coating by an interfacial layer of a sulfide phase.

An example illustrating the attack morphology of type I hot corrosion is shown in Figure 9.12 where a Ni-based alloy protected by Cr$_2$O$_3$ scale is exposed to a hot corrosive environment for 100 h at 900°C. It is observed that the alloy has developed a surface layer of Cr$_2$O$_3$ scale. However, the attack can be seen to occur more rapidly at certain locations resulting in the precipitation of a Cr-rich sulfide phase at the grain boundaries of the alloy. When the composition of the oxide scale is examined, it is found to contain some S in the regions
where the attack occurs more rapidly. In contrast, at those locations where no or little attack occurs, the scale is found to be free of S. It is possible that localized pores or cracks developed in the scale have permitted S to diffuse into the alloy and initiate the attack.

In general, hot corrosion attack at a given temperature occurs after an incubation period. During this period, the alloy or coating system develops and maintains a protective oxide scale and therefore, the reaction rate is very slow and similar to that observed in simple oxidation. Initiation of the attack at the end of the incubation period occurs as a result of degradation of the protective nature of the oxide scale by various processes, e.g., spallation of the scale due to growth stresses and thermal cycling, erosion, fluxing of the scale into molten salt, and diffusion of S into the oxide. Subsequently, the reaction rate is accelerated during the propagation stage, and the attack becomes more severe.

It can be concluded from the above discussion that the length of the incubation period at a given temperature can serve as a criterion for the resistance of various materials to hot corrosion. This is influenced by a number of variables some of which are related to service conditions, and others are related to the material and design. Service conditions include such variables as temperature, thermal cycling, composition and velocity of gases, composition and deposition rate of

Figure 9.12  SEM micrograph of a Ni-based alloy upon exposure to a hot corrosive environment for 100 h at 900°C. The region labeled 1 developed relatively pure Cr₂O₃, region 2 formed Cr₂O₃ + S while region 3 exhibited a Cr-rich sulfide.
salts, physical state of the salts, and erosion. Variables related to the structural material include composition, fabrication method, and precipitation of secondary phases. Geometry of the part is the most important factor influencing hot corrosion attack. A brief account for each of the above variables is given below.

As explained earlier, temperature determines whether the hot corrosion attack is of type I or type II. Since the reactions involved in type II hot corrosion (low temperature) occur more rapidly than those of type I (high temperature), the time to initiate the attack can be decreased as the temperature is increased. Thermal cycling during service can cause cracking and spallation of the oxide scale, degrading its protective nature. Consequently, as the oxide scale becomes more resistant to spallation, the time to initiate the attack is increased. Gas composition can have a significant effect on hot corrosion particularly at lower temperatures. It is recalled that the presence of SO$_3$ promotes the formation of mixed sulfates in the liquid state, accelerating the kinetics of type II hot corrosion. However, as the temperature is increased, the effect of gas composition becomes increasingly less important. Since most of the acidic oxides such as MoO$_3$ are lost to the gas as its velocity is increased, the time to initiate hot corrosion by acidic fluxing is increased. Presence of certain substances in the liquid deposit can significantly affect the kinetics of hot corrosion. For example, if NaCl is present in Na$_2$SO$_4$, the time to initiate the attack at a given temperature is decreased. Chlorine can convert the protective scale into a porous non-protective oxide. Also, if present in larger concentrations, chlorine can promote the formation of volatile substances such as CrO$_2$Cl$_2$, depleting the alloy or coating system in Cr and in turn reducing its ability to maintain a protective scale. Salt deposition rate affects the kinetics of hot corrosion in a manner dependent upon the operative mechanism. When hot corrosion occurs by the fluxing process, the incubation period is increased with the deposition rate. In contrast, if the attack occurs as a result of accumulation of elements from the alloy in the deposit, a slower deposition rate decreases the time required to initiate the attack. Usually, a deposit in the liquid state accelerates the kinetics of hot corrosion. However, certain solid deposits such as dense layers of CaSO$_4$ can prevent gases from reaching the alloy surface accelerating the rate of attack by a sulfidation mechanism. Since particulate erosion can cause cracking and spallation of the protective oxide scale, it decreases the time to initiate the attack.

Effects produced by alloy composition are dependent on whether the hot corrosion attack is of type I or type II. Under conditions favoring type I hot corrosion, which primarily occurs by sulfidation, Co-based alloys are more resistant than Ni-based alloys. Increasing the Cr content of both Co- and Ni-based alloys improves the resistance to type I hot corrosion. Critical additions of Y also improves the resistance to type I hot corrosion by improving the protective nature of the oxide scale. Although Al can increase the resistance to this type of attack, relatively small concentrations of Al in Ni-based alloys below which a
continuous layer of Al\textsubscript{2}O\textsubscript{3} can promote sulfidation attack by increasing the sulfur activity in the melt. Alloys with smaller concentrations of Cr (\(<20\%\)) and relatively high concentrations of refractory elements such as Mo and W become susceptible to a very severe attack. All types of alloys are susceptible to considerable degradation by type II hot corrosion. Best resistance to this type of attack is obtained at higher Cr concentrations in the range of 30–35\%. In general, alloys produced as castings are more susceptible to hot corrosion attack in comparison with wrought alloys and those produced by powder metallurgy techniques.

Part geometry influences the kinetics of hot corrosion by its effect on the protective nature of the oxide scale. Edges and sharp corners can cause cracking of the oxide scale, decreasing the time to initiate the attack.

### 9.3.3 High-Temperature Protective Coatings

Both the severe mechanical and thermal conditions encountered in gas turbine applications are satisfied by a group of high-temperature alloys known as the superalloys. Most of the superalloys used in gas turbine applications are Ni-based, however, some Co-based alloys are also used. An inherent limitation of superalloys, particularly those used in the hottest sections such as turbine blades, is their inadequate environmental resistance at the service temperature, which is usually below about 1000°C, but can reach 1100°C due to localized hot spot conditions. To achieve an acceptable service life of gas turbine blades, they must be protected by surface coatings. Basically, the objective of applying the coating is to form an Al-rich layer at the surface of the alloy capable of developing and maintaining a protective Al\textsubscript{2}O\textsubscript{3} scale during service at elevated temperatures.

Various coatings used in practice can be classified into three main types: (i) diffusion aluminides, (ii) overlay coatings, and (iii) thermal barrier coatings. Simple diffusion aluminide coatings are applied by enriching the surface in Al and allowing the alloy to react with this layer to form an Al-rich compound based upon the NiAl composition known as the \(\beta\) phase. Characteristically, the coating layer contains various concentrations of alloy substrate elements. Other modifications of the simple aluminides to improve their protective nature involve the addition of Pt to produce a Pt-aluminide coating. Similar to the simple aluminide coating, a Pt-aluminide coating contains various concentrations of substrate elements; however, to a large extent, the presence of Pt excludes heavy or tramp elements such as W and Mo. In this case, the coating consists of a mixture of \(\beta\) phase and an intermetallic compound of Pt and Al, which can be PtAl, Pt\textsubscript{2}Al\textsubscript{3}, or PtAl\textsubscript{2} depending upon the type of coating. Since intermetallic compounds such as the \(\beta\) phase are rather brittle, diffusion aluminide coatings lack sufficient ductility. To improve the coating ductility, overlay coatings consisting of \(\beta\) phase dispersed in a ductile solid-solution matrix have been developed. They differ from diffusion aluminides in that the alloy does not participate in forming the coating...
layer. An alloy of predetermined composition is deposited on the alloy surface by various techniques. One of the most common types of overlay coatings is the MCrAlY where M is Ni, Co, or Ni + Co. Overlay coatings are characterized by higher temperature capabilities than diffusion aluminide coatings.

To further improve the temperature capabilities of turbine blades, the metallic coating whether a diffusion aluminide or an overlay coating, can be further protected by a ceramic layer acting as a thermal insulator. Conventionally, the simple aluminide coatings have been used as standard surface protection systems for gas turbine blades. However, the more advanced Pt-aluminide coatings and overlay coatings are used in certain applications and expected to be more widely used in future generations of gas turbine engines.

Typically, the thickness of a diffusion aluminide coating is 50–70 μm; however, the thickness of an overlay coating is about 120–150 μm. Figure 9.13a–c illustrates characteristic microstructures of diffusion aluminide and overlay coatings. As shown in Fig. 9.13a, the simple aluminide coating consists of an outer layer and inner zone about 15–20 μm in thickness and is called...

![Figure 9.13](image-url)

**Figure 9.13** Micrograph of (a) coating consisting of a ceramic top coat, outer bond coat layer and an inner interdiffusion zone, (b) Pt-aluminide coating divided into an outer zone consisting of a mixture of β phase and a Pt-Al compound and an inner zone that consists of β phase, and (c) CoNiCrAlY overlay coating that consists of one layer and a thin interdiffusion zone.
the interdiffusion zone. It results from interaction between the coating and alloy substrate during processing. In contrast, the Pt-aluminide coating layer is divided into two zones as illustrated in Fig. 9.13b: an outer zone consisting of a mixture of β phase and a Pt-Al compound (PtAl2 in this example), and the inner zone consisting of β phase. However, the interdiffusion zone is observed to be similar to that of a simple aluminide coating. Figure 9.13c illustrates the microstructure of a CoNiCrAlY overlay coating. It consists of one layer and a thin interdiffusion zone. Within the coating layer, particles of the β phase are dispersed in a ductile Co-rich solid-solution matrix phase.

Performance capabilities of a given type of coating can significantly vary from one alloy substrate to another depending upon its composition. In the case of diffusion aluminide coatings, elements in the alloy substrate can diffuse into the coating during heat treatment processes as well as during service. Consequently, the initial coating composition is a function of alloy substrate composition. Diffusion of substrate elements into overlay coatings during processing is insignificant; its initial composition is rather independent of alloy substrate composition. However, significant interdiffusion can occur between an overlay coating and the alloy substrate during service similar to the case of diffusion aluminides.

During service, protective coatings can be degraded by various mechanisms, the most important of which are oxidation and interdiffusion between the coating and alloy substrate. Oxidation plays an important role during the early stages of coating degradation. Initially, the coating develops a surface layer of protective Al2O3 scale. With extended exposure at elevated temperatures, the protective nature of the scale can be degraded by environmental factors, as well as service conditions as described in the previous section. Another important factor contributing to coating degradation is interdiffusion between the coating and alloy substrate. Elements from the alloy substrate can diffuse into the coating and influence its performance dependent upon the service temperature and exposure time at a given temperature. Interdiffusion degrades the thermal stability of the coating as well as the protective nature of the oxide scale. Although some elements can have beneficial effects on oxidation resistance, others can produce adverse effects as described below.

A given coating remains to be protective as long as Al is selectively oxidized to form a continuous layer of Al2O3 having a slow growth rate. Selective oxidation of Al, however, can be significantly affected by interdiffusion between the alloy substrate and coating. Outward diffusion of Ni and other elements from the alloy into the coating and/or inward diffusion of Al dilute the coating surface in Al. Consequently, the coating ability to maintain a continuous protective layer of Al2O3 is reduced, and other less protective and nonprotective oxides can be developed. However, these effects are dependent upon the extent of interdiffusion. If Cr remains in solid solution within the coating, it increases the Al activity...
near the surface, promoting the coating ability to develop a continuous layer of \( \text{Al}_2\text{O}_3 \) at smaller concentration of Al. A similar effect can be produced by critical concentrations of active elements such as Hf, Y, and Zr. However, excess Cr concentrations leading to precipitation of Cr-rich phases at the coating surface is detrimental because \( \text{Al}_2\text{O}_3 \) cannot develop upon phases containing little or no Al. Also, excess concentrations of active elements can degrade the adherence of the scale to the coating leading to its spallation. Titanium in concentrations greater than about 0.5% can have adverse effects by forming TiO\(_2\), which degrades the adherence of the scale to the coating. Refractory elements such as W, Mo, V, Ta, and Re can produce deleterious effects; however, the exact effect varies from one element to another and can be dependent upon concentration. Among the refractory elements, W produces the most adverse effects by reducing the diffusivity of Al to the surface, as well as by forming nonprotective oxides. To summarize, the net effect of interdiffusion on the performance of a given coating can significantly vary from one alloy substrate to another.

One of the most important factors affecting the diffusional stability of a given coating is the structure and composition of the interdiffusion zone separating the coating from the alloy substrate. Typically, for diffusion aluminide coatings, the interdiffusion zone consists of a mixture of \( \beta \) phase and sigma phase enriched in refractory elements such as W and Mo.

### 9.3.4 Sulfidation

One of the most common contaminants found in the combustion products generated by many high temperature industrial processes is sulfur. It is usually present in low-grade fuels, chemical additives, and fluxes. In a sulfidizing environment, elements in structural alloys, including those to be selectively oxidized to develop a protective scale, can react with sulfur and form sulfide phases. Some of these sulfide phases have relatively low melting temperature and high growth rate, degrading the structural integrity of the material. However, the severity of sulfidation attack is dependent upon a number of variables, the most important of which are (i) the exact nature of the environment, (ii) service temperature, and (iii) alloy chemistry. An account for each of these variables is given below.

Sulfidizing environments encountered in practice can be classified into three main types depending upon the relative activities of sulfur and oxygen. When S is present in the gas stream as SO\(_2\) and SO\(_3\), the activity of oxygen is higher than that of sulfur, and therefore the environment is predominantly oxidizing. In this type of an environment, sulfidation attack is less severe. However, if the environment consists of H\(_2\)S and H\(_2\), the activity of sulfur becomes considerably higher than that of oxygen, and the environment becomes predominantly reducing. Sulfidation attack is most severe in this environment because the alloy becomes incapable of developing a protective oxide scale. An intermediate
type of environment is that of a mixed gas, usually consisting of \( \text{H}_2, \text{H}_2\text{S}, \text{H}_2\text{O}, \text{CO}, \text{CO}_2, \) and \( \text{CH}_4 \), where the sulfur and oxygen activities are sufficient to cause both sulfidation and oxidation.

Predominantly oxidizing environments containing \( \text{SO}_2 \) and \( \text{SO}_3 \) are encountered in such processes where excess air is used to ensure complete combustion of fuel, as in the case of gas turbines discussed earlier. Another example is a coal-fired boiler. A typical example of a predominantly reducing environment is that generated in catalytic reforming units used in petroleum refineries to upgrade the octane number of gasoline. Mixed-gas environments are generated by combustion under conditions where no excess air or oxygen is present, as encountered in many industrial processes such as coal gasification, chemical and petrochemical processing, glass manufacturing, waste incineration, and fossil-fired power generation. It is to be noted that even in an oxidizing environment, sulfidation attack may occur because of a localized decrease in oxygen activity.

Because of the low oxygen activity in reducing environments, a competition exists between sulfidation and oxidation. Sulfidation dominates oxidation at reduced oxygen activity and lower temperatures. Conversely, at higher temperatures, where the environment becomes more oxidizing, sulfidation is dominated by oxidation. When the environment involves both sulfidation and oxidation, structural materials are selected on the basis of their sulfidation resistance, as described below. Usually, sulfidation attack is most severe at temperatures below about 900°C. However, at higher temperatures where oxidation dominates, the attack becomes less severe.

Sulfidation attack poses a serious problem when low-melting phases of high growth rate, particularly Ni and Ni-Fe sulfides, are formed on the alloy surface. In contrast, sulfide phases containing Co or Co + Cr have significantly higher melting temperature and a slower growth. Chromium is generally found to increases the sulfidation resistance of all classes of high-temperature alloys. Also, Al has beneficial effects on sulfidation resistance provided a protective \( \text{Al}_2\text{O}_3 \) scale in developed. However, Al can have adverse effects in the case of alloys protected by \( \text{Al}_2\text{O}_3 \) scale. In general, the performance capabilities of various alloys, and the effect of alloy chemistry, are dependent upon the exact environment as described below.

In a predominantly reducing environment consisting of \( \text{H}_2 \) and \( \text{H}_2\text{S} \), and the service temperature is close to 500°C as encountered in catalytic reforming processes in petroleum refineries, all classes of high-temperature alloys including Co-, Ni-, and Fe-based alloys, are susceptible to severe sulfidation attack because the very low oxygen activity precludes the formation of protective \( \text{Cr}_2\text{O}_3 \) scale. Alloys with higher Cr content usually have better sulfidation resistance.

When the environment is of a mixed character, Co-based alloys are the most resistant to sulfidation attack. Also, some of the Ni-based alloys containing a relatively high concentration of Co can have adequate resistance. With respect
to the major elements in the alloy, a correlation exists between the resistance to sulfidation and the ratio of (Co/Ni + Fe) at relatively lower temperatures below about 900°C where the environment is sulfidizing. A considerable improvement in sulfidation resistance occurs when this ratio exceeds unity. At temperatures higher than about 900°C, where the environment becomes more oxidizing, the above correlation is no longer valid. Instead, the resistance to sulfidation is considerably increased as the ratio (Cr/Ni) in the alloy exceeds unity. Minor elements which can have beneficial effects on sulfidation resistance include Si and Ti, which can improve the protective nature of Cr₂O₃ scale. In contrast, Mn can have adverse effects on sulfidation resistance because of its high diffusivity in Cr₂O₃ scale and tendency to form an external sulfide phase.

As pointed out earlier, sulfidation attack is less severe in an oxidizing environment. Under this condition, the alloy is capable of developing a protective oxide scale. At relatively lower temperature where the rate of sulfidation increases, sulfide phases can form. Generally, increasing the Cr content of all classes of alloys improves the resistance to sulfidation attack.

Sulfidation attack morphology is similar to that of type I hot corrosion. When the alloy is capable of developing a surface oxide scale, the rate of attack is very small as long as the scale remains protective. Any disruption of the oxide scale, however, can promote the sulfidation attack resulting in sulfide phases beneath the oxide layer, as shown in the example of Fig. 9.14 (see region marked 3), derived from an Fe-based alloy containing 33% Ni and 21% Cr. Sulfide phases containing Fe and Cr are observed beneath the Cr₂O₃ scale. In extreme cases where the low-melting sulfide phases overgrow the oxide scale, nodules of solidified sulfides are observed at the surface, as shown at the regions marked 2 and 3 in the inset of Fig. 9.14.

![Image](image-url)

**Figure 9.14** Micrograph showing the formation of Fe-Cr rich sulfide phases above (regions marked 1 and 2 in inset) and below (region marked 3) the Cr₂O₃ scale in Fe-33Ni-21Cr.
9.3.5 Carburization

Carbonaceous gases consisting of mixtures of CO, CO$_2$, and hydrocarbons in varying proportions such as CH$_4$ are generated in many high-temperature processes, particularly in the petrochemical industry. High-temperature carburization attack occurs by deposition of carbon on the alloy surface and its inward diffusion to react with carbide forming elements, particularly Cr-precipitating carbide phases. In addition to depleting the alloy in Cr, reducing its ability to develop a protective oxide scale, carburization can cause severe grain boundary embrittlement resulting in intergranular cracking. Another detrimental effect produced by carburization is called metal dusting. In this case, the attack is the form of pitting and/or thinning of the part resulting in metal wastage. Carburizing environments can be classified into two main types depending upon the relative activities of oxygen and carbon: (i) oxidizing environments and (ii) reducing environments.

In an oxidizing environment, C is generated by the following reaction:

$$2\text{CO} \rightarrow \text{CO}_2 + \text{C}$$

Carbon is generated in reducing environments containing H$_2$ and a hydrocarbon such as CH$_4$ by the following reaction:

$$\text{CH}_4 \rightarrow 2\text{H}_2 + \text{C}$$

Severely carburizing environments encountered in the petrochemical industry are of the reducing type. Similar to the case of sulfidation, a competition exists between oxidation and carburization in a reducing carburizing environment. If the environment is severely carburizing, i.e., the activity of C is considerably higher than that of O, Cr$_2$O$_3$-forming alloys cannot develop a protective scale. Instead, a carbide scale is formed. When massive coke is deposited on the surface of, e.g., pyrolysis furnace tubes, it can act as an insulator. In contrast, since Al$_2$O$_3$ scale can form even if the oxygen activity is very low, Al$_2$O$_3$-forming alloys offer a greater protection against carburization attack. Usually, carburization attack is accelerated at temperatures above about 900°C, and it proceeds as follows.

Carbon is first adsorbed on the alloy surface, and then it diffuses inward provided the temperature is sufficiently high. Since the carbon activity at the surface is rather high, a carbide of the form Cr$_3$C$_2$ may form. However, as the C activity decreases with distance from the surface, the carbide phase may change from Cr$_3$C$_2$ to Cr$_7$C$_3$ and Cr$_23$C$_6$. Precipitation of these Cr-rich carbide phases can have two detrimental effects on properties: (i) They considerably reduce the ductility of the alloy, and (ii) they reduce the overall corrosion resistance because of the localized depletion in Cr.

An example illustrating a typical carburization attack morphology derived from 310 stainless steel is shown in Fig. 9.15. It is observed that the surface scale
consists of a carbide phase determined to be of the type Cr$_2$C$_2$. In addition to Cr, the carbide contains a significant concentration of Fe (a carbide-forming element) and a very small concentration of Ni (Ni is not a carbide-forming element). Beneath the carbide scale, an oxide phase of the type (Cr$_x$Fe)$_2$O$_3$ is observed. Evidently, during the carburization attack, the carbide phase has overgrown the oxide phase. Below the surface scale a Cr-depleted zone is observed, which can be explained in terms of the Cr-rich carbide scale formed at the surface. As a result of Cr depletion near the surface, the alloy becomes incapable of developing a protective Cr$_2$O$_3$ scale. In the interior of the alloy, a high density of Cr-rich precipitates determined to be a carbide of the type Cr$_{23}$C$_6$ is observed. Initially, this carbide can be of the type Cr$_7$C$_3$; however, with extended exposure it can transform into the more stable Cr$_{23}$C$_6$ carbide.

For the alloys protected by Cr$_2$O$_3$ scale, a correlation is found to exist between the ratio (Ni/Fe + Cr) and the resistance to carburization. When this ratio is increased from 0.3 to 1, the resistance to carburization is improved by a factor of 3. Other elements which can improve the resistance to carburization are Si and refractory carbide-forming elements such as Mo, Nb, and W. However, these elements can degrade the weldability of the material. Silicon can either increase the stability of Cr$_2$O$_3$ and/or form a protective SiO$_2$ scale. As pointed out earlier, Al$_2$O$_3$-forming alloys are the most resistant to carburization.

9.3.6 Chlorination

Chlorinating environments are created in a number of high-temperature industrial processes such as incineration of municipal wastes, combustion of chlorine contaminated coal, and waste heat recuperation from chemical and metallurgical processes utilizing chlorine as a reactant. When Cl$_2$, HCl, or NaCl is present in the gas stream, it can significantly accelerate the oxidation kinetics. In the presence of Cl$_2$
and/or HCl, the surface oxide scale can lose its protective nature by forming volatile metal chlorides and/or metal oxychlorides accelerating the oxidation rate. Generally, alloys protected by Cr$_2$O$_3$ and free of Al exhibit a relatively low resistance to chlorination because of the tendency of Cr$_2$O$_3$ to transform into volatile CrO$_2$Cl$_2$. Least resistant alloys are those containing relatively high concentrations of refractory elements, such as Mo and W, due to the formation of oxychlorides of high vapor pressure and greater tendency to volatilize. Also, chlorine tends to convert protective scale into porous nonprotective oxides. Among the various classes of high temperature alloys, Fe-based alloys have the greatest resistance to chlorination. This is related to the lower vapor pressure of iron chloride in comparison with both nickel and cobalt chlorides. Alloys

![Figure 9.16](image)

**Figure 9.16**  (a) The scale formed on a Co-30Cr-4.5W-3Ni alloy after 400 h exposure at 900°C to a chlorinated environment. (b) EDS spectrum derived from the encircled region 1 showing the presence of Cl along with Cr and O.

*Copyright © 2004 by Marcel Dekker, Inc.*
protected by Al₂O₃ have higher resistance to chlorination than Cr₂O₃-forming alloys because of the very low-vapor pressure of AlCl₃ in equilibrium with Al₂O₃. Therefore, Cr₂O₃-forming alloys containing Al have improved resistance. 

Usually, the scale developed in a chlorinating environment consists of an oxide phase containing Cl as shown in Fig. 9.16a, b. It is observed that the scale consists of Cr₂O₃ containing a marked concentration of Cl. In some cases, it may also contain a chloride phase.
10

Metallurgical Aspects of Fracture and Fractography

10.1 Introduction

Chapter 7 dealt with various macroscopic aspects of fracture. Of particular importance is the recognition that the intrinsic fracture toughness of a material can be expressed in terms of the stress intensity factor $K$. It is recalled that under plane stress condition, catastrophic propagation of an existing crack occurs when $K = K_c$, and under plane strain condition, propagation occurs when $K = K_{ic}$. Similar to other mechanical properties such as yield strength, both $K_c$ and $K_{ic}$ are found to be sensitive functions of the microstructure of the material. Therefore, it becomes possible to enhance the fracture toughness of a material by microstructural control. In particular, $K_c$ and $K_{ic}$ provide quantitative means for evaluating the effect of microstructure on the tendency of engineering alloys to fracture by brittle propagation of cracks. Just as microstructure influences the tendency to fracture, the nature of microscopic events taking place during crack propagation is reflected on the morphology of the resulting fracture surface.

If fracture is involved in a failure case, it is essential to identify its mode before the cause of failure can be determined. By fracture mode is meant the mechanism of crack propagation eventually leading to fracture. In principle, it is possible to identify the fracture mode from the microscopic appearance or morphology of the resulting fracture surface. This is referred to as fractography. Scanning electron microscopy is the most powerful tool used in fractography.
It is the objective of this chapter to present some insight into the microscopic aspects of fracture. Particular emphasis is placed upon the relationship between microstructural features and resistance to fracture under various loading conditions. Also, the morphological features of fracture surfaces produced by various propagation mechanisms are described with illustrative examples.

It is recalled from earlier discussions that engineering products are never perfect; they inevitably contain some types of cracks, flaws, or discontinuities introduced by various fabrication techniques. However, it is important to understand how cracks are nucleated before they can propagate and eventually lead to fracture as described below.

10.2 Microscopic Aspects of Crack Nucleation

Almost all the models proposed to explain crack nucleation are based upon some localized plastic deformation processes. As discussed earlier, plastic deformation of crystalline materials occurs by motion and multiplication of dislocations under the influence of a shear stress. Eventually, however, the motion of dislocations is impeded by various types of obstacles, e.g., other dislocations, second phase particles, and various types of interfaces such as twin boundaries and grain boundaries.

When moving dislocations are intercepted by an obstacle, they tend to pile up against that obstacle creating a localized high level of tensile stresses. For example, Fig. 10.1 illustrates a row of dislocations in a grain piled up against a grain boundary. As the number of dislocations in the pileup increases, the degree of stress concentration increases. Such a localized stress concentration can be relieved either by motion of other dislocations, e.g., in the adjacent grain in Fig. 10.1, or by opening a crack at the grain boundary. If dislocation motion is restricted, the high level of local stresses is relieved by forming a crack.

It is recalled that an edge dislocation can be envisioned to be introduced into the crystal by inserting an extra half-plane of atoms above the slip plane. When a number of these dislocations are combined, they can form a small crack or void, giving rise to another possible mechanism for cavity formation. This can occur readily in regions where plastic deformation is highly localized, such as the necked section of a tensile specimen.

It can be concluded from the above discussion that a crack is nucleated in two stages. First, dislocations move and multiply. Second, dislocations are held by some type of an obstacle, creating a localized high stress concentration. Formation of a crack is one means for relieving the localized stress concentration. Although, measurable plastic deformation on macroscopic scale is realized at stresses exceeding the yield point as in a tensile test, experiment shows that on a microscopic scale, dislocations can move at stresses well below the yield strength. Microcracks can therefore be nucleated within the elastic range by
the residual stresses introduced into the material by various fabrication processes to relieve localized high stress concentrations.

Evidently, dislocations control two opposing mechanical properties, namely, strength and ductility. High strength, i.e., increasing the load carrying ability without fracturing, requires motion of dislocations to be restricted. In contrast, sufficient ductility associated with dislocation mobility is required to relieve high local stresses around notches and other discontinuities before fracture can occur.

Whether a crack nucleated by a dislocation mechanism as described above propagates by a brittle or ductile mechanism is dependant upon the mobility of dislocations at the crack tip, which is influenced by a number of microstructural parameters. Evidently, crack propagation by a ductile mechanism requires high dislocation mobility. Many crack nuclei can form by dislocation motion and eventually coalesce by plastically deforming and necking down the remaining connecting material. If it is required to prevent crack propagation by a ductile mechanism, it becomes clear that the distance dislocations can move freely before they are held by obstacles must be as small as possible. This requirement can be achieved by various means, such as (i) fine grain size, (ii) a fine dispersion of second phase particles within the material, and (iii) high strain hardening rate.
10.3 Microscopic Mechanisms of Crack Propagation

10.3.1 Crack Propagation by Cleavage and Quasi-Cleavage

By definition, cleavage is a term used to describe crack propagation by a brittle mechanism. A distinguishing feature of a cleavage crack is that it proceeds along certain crystallographic planes. Although the exact mechanism by which cleavage cracks are nucleated is not well understood, it is believed to involve some highly localized plastic deformation process. Once initiated, however, a cleavage crack propagates rapidly with very little or no plastic deformation at the crack tip, and therefore it is classified as brittle.

Usually, cleavage fracture is observed in body-centered cubic structures such as ferrite, where the cleavage plane is usually of the type \( \{100\} \), although it can also occur on \( \{110\} \) planes. Within a given grain of the material, a large portion of a cleavage crack proceeds on a certain plane, and this is referred to as primary cleavage. However, the primary crack may be intercepted by an obstacle within the grain, and in this case, the cleavage path is shifted to another parallel plane. This is called secondary cleavage, and it produces steps in the cleavage fracture surface. Another cleavage step is produced when the crack is intercepted by a grain boundary. Fracture by a cleavage mechanism is promoted by mode I loading corresponding to plane stress condition; i.e., it is favored by high tensile stresses.

Crack propagation by quasi-cleavage resembles cleavage in that it is brittle and it occurs transgranularly. However, it does not necessarily follow specific crystallographic planes. Typically, this is the type of crack propagation leading to fracture in low-carbon steels, which undergo ductile to brittle transition as described earlier.

10.3.2 Crack Propagation by Microvoid Coalescence

Similar to cleavage, crack propagation by microvoid coalescence is essentially transgranular; however, in some cases it can also occur at grain boundaries as described in the next section. In contrast with cleavage, fracture by microvoid coalescence occurs by a ductile mechanism. Microvoids usually nucleate at regions of localized high stress concentration such as second-phase particles and grain boundaries. As indicated earlier, these localized stresses can be relieved by opening a small crack or a microvoid. When the intensity of stresses increases, the microvoids grow in size and eventually coalesce to form the fracture surface. Coalescence of microvoids results in cuplike depressions at the fracture surface, which are called dimples. Some times, second-phase particles acting as nucleation sites for microvoids are observed within the dimples. Important information
about the fracture behavior of the material can be obtained from the observed size and shape of dimples, as described below.

It is evident from the mechanism by which dimples are developed at the fracture surface that their size is determined by the density and distribution of sites at which microvoids can be nucleated. Relatively small dimples result when the material contains a large number of sites at which microvoids can be nucleated (see Fig. 10.2a,b). In contrast, if the sites available for microvoid nucleation are rather limited and widely spaced, the dimples become large. Also, it is possible to observe various dimple sizes at the fracture surface if the nucleation sites of microvoids are nonuniformly distributed within the material.

Dimple shape is influenced by the state of stress leading to fracture. When the state of stress is of the uniaxial tensile type, the dimples on both fracture surfaces have equiaxed morphology. Each dimple is bounded by a lip and its depth is a function of both microstructure and ductility. For example, microvoids nucleated around large secondary phase particles can be expected to produce deeper dimples in comparison with smaller particles. Deeper dimples are also associated with higher ductility, while lower ductility contributes to the formation of shallower dimples. Sometimes small holes are observed at the bottom of dimples corresponding to connection with other microvoids below the final fracture.

Figure 10.2  Fracture surfaces produced by coalescence of micorvoids resulting in dimple-like structure. The dimple size is determined by the density and distribution of sites at which microvoids are nucleated.

Copyright © 2004 by Marcel Dekker, Inc.
surface. If the material has a relatively high ductility, some of the localized high stress concentration can be relieved by plastic deformation contributing to shallower dimples. Elongated dimples can result from the loading modes, i.e., tear or direct opening (mode I), forward shear (mode II), and sideways shear (mode III). As its name implies, an elongated dimple has one axis longer than the other. Unlike an equiaxed dimple which is completely surrounded by a lip, one end of elongated dimples is open. For a tear-type fracture (mode I loading), the dimples on both fracture surfaces have the same orientation, and their closed ends point at the origin of fracture. When these microscopic features are observed corresponding to a completely slant fracture mode on the macroscopic scale resulting from the combined effects of modes I and III of loading, they can be used to determine the direction of crack propagation. If the fracture results

Figure 10.2 Continued.
from a shear loading conditions, the elongated dimples on the two fracture surfaces are oriented in different directions.

10.3.3 Intergranular Crack Propagation

As its name implies, an intergranular crack propagation proceeds along grain boundaries. Typically, grain boundaries of a ductile polycrystalline material are strong at relatively lower temperatures; however, they become weaker at elevated temperatures. More often, these materials are fractured transgranularly at low temperatures. However, it is also possible that an intergranular fracture occurs at low temperatures for various reasons mostly related to changes in microstructure, e.g., precipitation of an embrittling grain boundary layer of a second phase, and a disorder-order transformation.

Two types of intergranular crack propagation at low temperatures are distinguished: (i) brittle intergranular propagation and (ii) ductile intergranular propagation. When propagation occurs by a brittle mechanism, the grains are separated without microvoid coalescence resulting in a “rock candy” morphology, as schematically shown in Fig. 10.3. In contrast, intergranular crack propagation by a ductile mechanism occurs by microvoid coalescence at grain boundaries similar to

Figure 10.3  SEM image of a fracture surface showing ‘rock-candy’ structure where grains are separated without microvoid coalescence
the case of cracks propagating transgranularly by a ductile mechanism described in the previous section. In this case, the fracture surface exhibits separated grain facets, as shown in Fig. 10.4. Each facet contains dimples resulting from microvoid coalescence as a result of localized plastic deformation in grain boundary zones as shown in Fig. 10.5a, b. Usually, this type of fracture is observed when certain precipitation reactions result in alloy depletion near the grain boundaries.

At elevated temperatures, the grain boundaries become rather weak, which can lead to intergranular crack propagation as encountered in creep tests. Most evidence points out that the mechanism responsible for crack propagation under creep conditions is grain boundary sliding. Because of the lattice misfit existing across a grain boundary, both dislocation density and vacancy concentration are substantially higher at grain boundaries in comparison with the bulk of the grains. This enhances diffusion rates at grain boundaries causing them to deform or slide past each other. Voids usually form at a triple point where three grain boundaries meet, because of the associated incompatibility. Eventually, the voids or cracks coalesce causing separation of the grains while the bulk deformation within the grains can only be a few percent. Most metallic materials are ductile when subjected to a short-term tensile test at the same temperature. However, under creep conditions (slow deformation), many metallic materials fracture after only a few percent elongation.

10.3.4 Crack Propagation by a Fatigue Mechanism

Under fatigue loading conditions, fracture occurs in three distinct stages. Initially, the material contains some types of cracks, as pointed out earlier. It is

Figure 10.4  SEM image of a fracture surface showing intergranular crack propagation by a ductile mechanism that occurs by microvoid coalescence at grain boundaries.

Copyright © 2004 by Marcel Dekker, Inc.
also possible that microcracks are nucleated by a localized dislocation mechanism during cyclic loading. Usually, crack nucleation occurs most readily at the surface of the part. However, at high stress levels, internal cracks may also be nucleated. Local plastic deformation at structural discontinuities such as grain boundaries or secondary phase particles is of a highly localized nature, and therefore it is not measurable as a macroscopic overall strain. Since plastic deformation occurs by motion of dislocation under the influence of a shear stress, initiation of fatigue cracks is governed by a maximum shear stress law, as described earlier in the case of a simple tensile test. Therefore, slip lines resulting from localized plastic deformation appear first in those crystals whose slip planes have the maximum resolved shear stress. With continued cycling, the number of slip lines increases, forming what is known as a persistent slip band. Since the strain is reversed in each cycle, localized deformation causes the formation of extrusions and intrusions at the surface. Usually, this surface disturbance appears as early as one tenth of the total life of the part.

It is evident from the above discussion that the fatigue life of a part can be extended by strengthening the surface or locking the dislocations in position. Conversely, weakening the surface, e.g., by a brittle layer, corrosion pits, increase in roughness, etc., accelerates the crack nucleation process, and in turn the fatigue life is reduced. Such a reduction in fatigue life is appreciable in the case of
low-stress high-cycle fatigue where the crack nucleation process is always initiated at the surface, constituting about 90% of the total life. In contrast, for high-stress low-cycle fatigue (less than $10^4$ cycles to failure), where the crack nucleation stage occurs at less than 50% of the total life involving nucleation of many cracks in the interior of the part, the surface condition has a lesser effect.

Following the initiation of fatigue cracks, propagation first occurs at a slow stable rate. One or more cracks start to propagate through the structure along a direction perpendicular to the tensile stress axis. Propagation of these cracks takes place slowly by a highly localized plastic deformation process. Microvoid coalescence can occur ahead of the main crack particularly at high stress levels. Because of the two surfaces created by the slow propagating crack rub against each other as the material is deformed back and forth, the corresponding fracture surface has a burnished or polished appearance (Fig. 10.6). In practice, however, the stress amplitude of cycling loading is variable, and therefore the crack may cease to spread when the stress level is low, and continues to propagate when rubbing action leaving behind characteristic striations or ripples on the fracture surface. Such striations result from the rubbing action of the two crack surfaces as the material is deformed back and forth through each stress cycle. This behavior changes the degree of rubbing resulting in a fracture surface of clamshell appearance. Usually, these markings, which are called fatigue

![Figure 10.6](image_url)  
**Figure 10.6** Illustration of a polished or burnished fracture surface formed as two surfaces rub against each other.
striations, are concentric with the origin of fracture. Occasionally, striations free regions are observed if the crack propagates some distance by either a sudden local shear fracture or by quasi-cleavage. In the third or final stage, the remaining cross section becomes too small to support the load, and the slow growing fatigue crack becomes large. Under these conditions, propagation occurs catastrophically at an unstable rate leading to complete separation of the part. Since no rubbing occurs at this stage, the surfaces developed have a rough and dull appearance. Fatigue fracture can also be assisted by corrosive species in the environment, and in this case it is called corrosion fatigue.

10.3.5 Crack Propagation by Stress Corrosion

It is recalled that a distinguishing feature of stress-corrosion cracking is that the cracks grow incrementally or in a stepwise manner, giving rise to branching. Crack propagation can occur either transgranularly or intergranularly. Fracture surfaces produced by transgranular crack propagation under stress-corrosion conditions can exhibit morphological features resembling a fracture produced by a fatigue mechanism. In this case, the fracture surface exhibits beach marks, or striations produced by differences in the depth of attack at the surface. A schematic illustrating the appearance of these marks in terms of cleavage is shown in Fig. 10.7a and a fracture surface produced by stress corrosion of steel is shown in Fig. 10.7b. In the schematic of Figure 10.7a, where the direction of crack propagation is normal to the tensile stress, the fracture surface contains parallel displaced facets separated by steps (beach markings). Frequently, each step has sawlike notches at its edge or is serrated. Parallel facets normal to the tensile stress follow the cleavage planes of the material, and when the edge of each facet is serrated, the resulting steps follow another set of crystallographic planes. Intergranular crack propagation can occur when the material is sensitized, i.e., when it contains a continuous network of grain boundary precipitates.

10.3.6 Hydrogen-Assisted Crack Propagation

Similar to the case of stress corrosion cracking, fracture can be produced by either intergranular or transgranular propagation of hydrogen-induced cracks depending upon the microstructure of the material. Whether the cracks propagates transgranularly or intergranularly, however, the resistance to hydrogen-induced cracking is improved by decreasing the grain size similar to the case of stress corrosion cracking. If the material is not sensitized, i.e., it is free of intergranular precipitates, the cracks propagate transgranularly. In this case, planar imperfections such as twin boundaries, stacking faults, and phase interfaces, as well as slip bands, are the most susceptible to hydrogen-induced cracking. If the material becomes sensitized, the grain boundaries become most susceptible to cracking, and in this case the material is said to be embrittled by hydrogen.
Figure 10.7  (a) Schematic illustration of striations produced due to stress corrosion cracking. (b) Fracture surface produced by stress corrosion of steel containing beach markings.

Copyright © 2004 by Marcel Dekker, Inc.
10.4 Fracture Modes and Fractography

As pointed out earlier, fracture mode refers to the mechanism by which cracks propagate to produce fracture. Before the cause of fracture can be determined, its mode must be characterized. Fractography refers to the metallographic technique used to characterize the fracture mode. However, it is important to realize that a great deal of information about the fracture mode can be obtained from the macroscopic appearance of the fracture surface. For example, a square fracture is indicative of a brittle mechanism. Also, if a failed component exhibits an obvious distortion, this indicates that fracture has proceeded by a ductile mechanism. In other cases of failure, a fractured component may not exhibit a macroscopic distortion giving the impression that the fracture is brittle. This situation can be resolved by microscopic examination of the fracture surface if it is available. Illustrative examples are given below to show characteristic microscopic appearances of fracture surface produced by different propagation mechanisms.

10.4.1 Fracture Surfaces Produced by a Cleavage Mechanism

Since most cleavage cracks are transgranular, the fracture surface appears granular or shiny at small magnifications, and cleavage steps become clearly visible at higher magnifications. In contrast, a quasi-cleavage fracture exhibits radial markings and chevrons at small magnifications, and quasi-cleavage steps are observed at high magnifications.

If cleavage occurs on a single crystallographic plane, the resulting fracture surface is smooth on an atomic scale. However, because engineering alloys are complex multiphase alloy systems, corresponding fracture surfaces produced by cleavage exhibit a variety of morphological features. Due to the presence of several microstructural features, cleavage cracks can nucleate at various locations within the grains. When these cracks propagate, the crack front can bypass various obstacles as described earlier producing a fracture surface of complex irregular morphology. An example is shown in Fig. 10.8.

10.4.2 Fracture Surfaces Produced by Microvoid Coalescence

Since plastic deformation is involved in producing a fracture surface by microvoid coalescence, the fracture surface shows evidence for slip lines or bands. When a void is created at the interface between a secondary phase particle and the matrix, plastic deformation by slip occurs in the vicinity of the particle. With continued deformation, the void increases in size, and eventually the interconnecting layers of the material are separated. This results in a fracture surface consisting of cusps or dimples. If voids are formed at the interface between the matrix and secondary
phase particles, the centers of some cusps contain particles. Figure 10.9 illustrates a fracture surface produced by microvoid coalescence. In some cases where the localized plastic deformation occurs in the vicinity of the voids, slip lines may be observed at the fracture surface. Straight slip lines are observed in fcc materials if

Figure 10.9 Fracture surface produced by microvoid coalescence.
dislocations are confined to move on single slip planes corresponding to high stacking fault energy. Conversely, if the stacking fault energy is low and dislocations can readily cross-slip, the slip lines become wavy.

Both the density, size, and depth of dimples observed on fracture surfaces produced by microvoid coalescence reflect the amount of plastic deformation, which has preceded fracture. For highly ductile materials, the fracture surface contains a smaller density of large deep dimples. As the ductility is reduced, the dimples become finer, shallower, and of larger density. Although in this case, localized plastic deformation can still occur in the vicinity of voids, the amount of macroscopic deformation required to complete the separation diminishes.

10.4.3 Fracture Surfaces Produced by Intergranular Separation

An intergranular fracture exhibits a smooth and shiny surface at small magnifications, and at higher magnifications a structure resembling the grain structure of the material is observed. At low magnifications a ductile fracture appears fibrous and rough. However, at higher magnifications, the fracture is found to result from the initiation, growth, and coalescence of voids, resulting in a pattern of dimples at the fracture surface as explained earlier.

10.4.4 Fracture Surfaces Produced by a Fatigue Mechanism

Fatigue fracture is distinguished by its characteristic striations. As described earlier, fracture mechanics can be used to estimate the number of cycles preceding fracture by a fatigue mechanism. However, this method is complicated by the strong dependency of the number of cycles on the starting crack size, as well as by a mathematical integration which can only be carried out numerically. A more convenient, easier, and perhaps even more accurate method for estimating the number of cycles is based upon measuring the spacing of fatigue striations observed on fracture surface. This can be done by using SEM or TEM. Fatigue striations will be more sharply defined in a TEM sample; however, SEM offers the advantage of relatively large area analysis. This enables identification of fracture origin area and differentiation between stage I and stage II of fracture propagation. The point of common intersection for all fracture marks will be the fracture origin. The fracture surface may contain more than one origin, indicating a high stress or discrete stress concentration sites. Multiorigin fracture is distinguished by several ratchet marks. The stage I fracture usually shows no sign of striations and extends only to a few grains around the site of origin. The stage II fracture is represented by a change in orientation of the fracture plane. The shape and
location of the crack front will indicate the type of loading (i.e., tension, bending, torsion, etc.) to which the part was subjected. Likewise, the location and size of the final fracture indicates the magnitude of nominal stress level.

10.4.5 Fracture Surface Produced by a Creep Mechanism

High-temperature creep is often initiated by nucleation of microvoids at grain boundary quadruple points due to localized melting. These voids enlarge and propagate along the grain boundaries resulting in intergranular cracking that ultimately leads to fracture. The grains revealed after fracture are rounded and elongated in contrast to smooth-faceted morphology typically shown by low-temperature intergranular fracture. Occasionally, evidence of oxidation at the fractured surface can also be found.

10.4.6 Fracture Surface Produced by Stress-Corrosion Cracking

As indicated earlier, crack propagation in stress-corrosion cracking can occur intergranularly or transgranularly. Since, stress-corrosion cracking takes place under a combined effect of stress and unfavorable environment, the exposed surface after fracture will normally be covered with a layer of corrosion/oxidation products. Examination under optical and SEM can also reveal extensive secondary cracking. The fracture surface will appear crystalline or flat depending on mode of crack propagation followed.

10.4.7 Fracture Surface Produced by Hydrogen-Induced Cracking

Hydrogen-induced cracking predominantly consists of a single crack with no branching. The crack propagation could be intergranular or transgranular or it can change from one mode to another. The fracture surface will not show any corrosion product unless the material is exposed to such an environment. The crack can initiate from within and at the surface of the material. Hydrogen damage can produce “flakes,” which are short, elliptical, and discontinuous cracks that take on a brightly shining appearance.
11

Failure Analysis Procedure

11.1 Introduction

Generally, the exact procedure for failure analysis is dependent upon the particular case. However, any successful and complete failure analysis investigation should provide answers to the following questions:

1. What is the mode of failure?
2. What is the most probable cause of failure?
3. How can failure be prevented in the future?

It is important to realize that any failure results from some sort of a mistake causing a weak link in the chain of the continuous process of engineering design-manufacturing-performance. Therefore, to provide clear and straight answers to the above questions, any failure analysis investigation must address three main issues: (i) service conditions, (ii) manufacturing, and (iii) design. By a process of elimination, it is possible to narrow down the most probable cause of failure and then identify the weakest link in the above process. Subsequently, it becomes possible to determine possible means by which future failures can be prevented.

Any effective failure analysis investigation must involve three main stages: (i) clear definition of the problem, (ii) complete search of the literature to develop a good technical background in the problem area as required, and (iii) development of an experimental program which must be well oriented and coordinated to address the problem. Clear definition of the problem is the most important of the above three aspects. It is evident that the uninformed engineer or scientist
cannot make a contribution to solving a problem that is not understood. Developing the proper technical background, as well as the experimental program necessary to solve the problem, is critically dependent upon a clear definition of the problem. Since the time required to complete a failure analysis investigation can be very important in many cases, the effective analyst must be capable of recognizing problems and communicating these problems to others with technical knowledge in a timely fashion.

It is the objective of this chapter to present a general procedure for failure analysis investigations. However, it is emphasized that this procedure can serve only as a guide. As demonstrated in the case studies of Chap. 12, a certain step in this procedure which can be of primary importance in one case may prove irrelevant in another case and so on. Any failure analysis investigation must begin with a clear definition of the problem, as described below.

### 11.2 Definition of the Problem

A failure analysis investigation begins by the proponent calling the analyst to investigate a given problem. Initially, the analyst meets with concerned personnel such as process engineers, maintenance engineers, etc. to discuss the problem. Typically, the discussions begin by the proponent giving some description of the problem; e.g. during regular maintenance, cracks are detected in a certain component, a part has undergone fracture, leaks are detected in storage tanks, etc. It is the responsibility of the analyst to ask relevant and important questions concerning the history of the component involved in the failure, such as (i) process involved, (ii) the application, (iii) type of material, its specifications, shape, dimensions, and processing technique, (iv) design parameters, (v) service conditions, (vi) maintenance history, (vii) sequence of events preceding the failure, (viii) frequency of failure and whether it is of localized nature, etc. Sometimes, some of this information may not be available. However, it is very important that the analyst collects as much information as possible about the history of the component or part involved in the failure.

Inspection of the failure site is a very important aid in defining the problem. A thorough examination of the condition of the failed component can provide a great deal of information. Depending upon the particular case, attention must be given to such features as location of cracks, fracture characteristics, perforations, deposits, wall thinning, sagging, detached pieces, etc. Also, whenever possible, the analyst must obtain as much information as possible from eyewitnesses about any abnormal conditions preceding the failure. In some cases, visual inspection of the failed component can reveal information which cannot otherwise be obtained. For example, if fracture is involved in the failure, visual inspection can provide information about the type of loading, load path, and fracture
path. From such information, it is possible to reconstruct the direction of stresses. Usually, the fracture path is normal to the direction of maximum principal stress.

Photographic documentation of the failed component is almost always necessary for future reference and further inspection after leaving the site. It is important to photograph all features of the failed component. A clear definition of the problem is largely dependent upon the depth of information obtained, site inspection, as well as the skill and judgment of the analyst.

11.3 Technical Background

Once the problem is clearly defined, the analyst must be able to identify the technical background required to solve the problem. Depending upon the skill and judgment of the analyst, it is possible to define the problem as environmental, mechanical, or a combination of both. Furthermore, through discussions with concerned personnel and visual inspection, the experienced analyst can narrow down the various possibilities. Also, in some cases, the failure may resemble a previous case. In this manner, the analyst can identify the technical literature relevant to the particular problem, e.g., high-temperature corrosion, fatigue, hydrogen embrittlement, stress corrosion cracking, etc. To illustrate this methodology with an example, one analyst determines that most evidence points out that the most probable cause of failure is high temperature carburization attack. It is by large deposits of coke. This identifies the technical background required to solve the problem.

11.4 Experimental Program and Analysis

Having defined the problem, an experimental program is developed to determine the cause of failure. Experiments must be coordinated and tailored so that they provide complementary information leading to identifying the cause of failure. Evidently, both the nature and number of experiments carried out can significantly differ from one case to another. Samples used to carry out those experiments must be carefully selected such that they represent as much as possible all features of the failed component. Deposits or pieces detached from the component can provide valuable information. Whenever possible, samples removed from sound sections, as well as samples never used in service, must also be collected for comparative purposes. It is also important that the analyst realizes that stress analysis and fracture mechanics are very valuable tools in failure analysis investigations. Given below is a guideline for the type of experiments and analysis involved in failure analysis investigations. Obviously, however, the exact experiments and analysis carried out are dependent upon the particular case.
Investigative tools used in failure analysis studies can be classified into two main types: (i) tools which can be used on site and (ii) laboratory tools.

11.4.1 Nondestructive Testing

It is recalled from earlier discussions that the primary objective of nondestructive testing is to detect both surface and internal cracks or discontinuities. Experimental stress analysis is used to evaluate stresses in the failed component, which can lead to failure by measuring the strain. Residual stresses are evaluated by X-ray diffraction.

11.4.2 Material Verification

Although rarely the material of the component turns out to be different from that specified, it is always recommended to verify the material(s) of the failed component. In most cases, microchemical analysis by energy dispersive X-ray spectroscopy in a scanning electron microscope provides sufficiently accurate results. However, in other cases where residual elements are of particular importance, other more accurate techniques must be used, such as electron microprobe analysis or one of the techniques used for bulk chemical analysis, particularly ICP depending upon the particular case.

11.4.3 Theoretical Stress Analysis

From knowledge of service conditions such as applied loads or pressure, and geometry and size of the component, it is possible to calculate the principal stresses developed in the part by the methods described in Chap. 5. Comparing the maximum principal stress with the strength of the material used in the application provides important information about the suitability of the material for the application and whether the design parameters were strictly followed during service.

11.4.4 Fracture Mechanics

Analysis using the principles of fracture mechanics provides valuable information about the stress developed in the failed component at the time of fracture in relation to the design stress, particularly in the case of crack propagation by a brittle mechanism. The outcome of this analysis determines the next course of action in the investigation. Although fracture mechanics can be used to determine the number of cycles preceding fracture by a fatigue mechanism, the method based upon measuring the spacing of fatigue striations can be more convenient provided the striations are clearly visible on fracture surfaces.
11.4.5 Macrostructure and Microstructure of the Material

A careful and thorough characterization of the macrostructure and microstructure of the material involved in the failure can frequently provide important clues. Whenever possible, it is particularly useful to compare the structural features of samples removed from the failed component with those representative of samples removed from sound sections, as well as of samples never used in service. It is also important to refer to any specific recommendations by the manufacturer about the use of the material.

Macrostructural characterization at small magnifications, e.g., $10 \times$, can provide further information about the fracture path and the location from which it was originated, e.g., internal surface or external surface. Selection of the proper technique for microstructural characterization is dependent upon the type of information required. However, it is always recommended to begin with light optical microscopy or scanning electron microscopy to characterize the overall microstructural features.

In examining the microstructure, particular attention must be given to such features as grain size and shape, secondary precipitates and their distribution, and microcracks and their location. An unusually large grain size is an indication that the material is exposed to an excessively high temperature. Grain shape and the shape of annealing twins provide information about the state of cold work. Elongated grains indicate that the material is severely cold worked. Plastic deformation can be reflected by bent twin boundaries. Fine slip lines within the grains also indicate that the material has been plastically deformed. If secondary precipitates are detected, particular attention must be given to their nature and morphology. Microchemical analysis by either energy dispersive or wavelength dispersive X-ray spectroscopy can be used to determine their composition. Provided the precipitates are present in sufficient amount (>10% by volume), they can be identified by X-ray diffraction. In certain cases, transmission electron microscopy may be required. While examining the gross microstructural features, particular attention must be given to whether microcracks are present. For example, grain boundary cracks may result from either creep deformation or embrittling the grain boundaries by precipitates of a secondary phase.

11.4.6 Characterization of Fracture Surfaces: Fractography

Although a great deal of information about fracture characteristics can be obtained by means of visual inspection, characterization of the mechanism responsible for fracture usually requires microstructural examination of the fracture surface, particularly in a scanning electron microscope. It is very important to realize that
any fracture results from an overloading condition. In contrast, cracks can propagate by various mechanisms. Therefore, the analyst must be extremely careful in reaching a conclusion about the mechanism responsible for crack propagation if the entire fracture surface is not available. Sometimes, cracks are not propagated to the extent causing fracture. If cracks extend through the entire thickness of the part, it is possible to induce complete separation by certain fixtures and examine the morphology of crack surface. In some cases, particularly when the component operates at high temperatures, the details of fracture or crack surfaces can be masked by oxidation. Descaling in a suitable acid, e.g., HCL, can be used to remove the oxide layer, revealing the morphology of the fracture surface.

11.4.7 Analysis of Corrosion Deposits

Whenever corrosion deposits are detected, they must be analyzed to determine their nature, e.g., oxide, sulfide, carbide, chloride, etc. Corrosion deposits may remain adhered to or separated from the surface of the component. By a combination of either microchemical analysis in a scanning electron microscope or bulk analysis by X-ray fluorescence, and X-ray diffraction, the nature of the deposit can be identified. In some cases particularly those where low-temperature aqueous corrosion is involved, the products of corrosion reactions may be washed away. Such cases require an exact knowledge of the environment composition and the corrosion resistance of the material in relevant environments.

11.4.8 Property Evaluation

Provided sufficient material is available, various tests can be carried out to evaluate the properties of the failed component, e.g., tensile, impact toughness, etc. However, in many cases, both surface hardness and microhardness tests are used as convenient means for evaluating the mechanical strength.

11.5 Mode of Failure vs. Cause of Failure

It is extremely important that the analyst distinguishes between the mode of failure and cause of failure. Sometimes the two terms are intermixed, leading to a great deal of confusion. To prevent future failures, it is essential to identify the cause of failure. To illustrate the difference between the two terms, consider the following examples.

An investigation reveals that a machine part has failed by a fatigue mechanism. In this case, the mode of failure is fatigue. However, this does not provide any information about why the part has failed by a fatigue mechanism. It is recalled from earlier discussion that if the part or component is not properly designed for fatigue, catastrophic failure can occur even if the best material known for its fatigue resistance is selected for the application. This can also
occur if the fabrication techniques introduce into the material flaws or discontinuities, weakening its fatigue resistance, or if the service conditions deviate from the design parameters.

In another case, the mode of failure is determined to be high-temperature carburization. This can be related to either a material with no adequate resistance to high-temperature carburization, or improper service conditions.

11.6 Data Interpretation and Terminology

Accurate and detailed interpretation of experimental data are key factors in completing a successful failure analysis study. Based upon this interpretation, the analyst can provide the proponent with proper recommendations to solve the problem. Evidently, this requires the analyst to have an adequate knowledge of materials engineering and material science as well as the capabilities and limitations of various investigative tools.

In interpreting the data, it is essential that the analyst use the proper technical terminology, which can be a key factor in determining responsibility. As pointed out earlier, improper terminology can lead to a great deal of confusion regarding the exact cause of failure. For example, it is reiterated that as much as possible, the term fracture must be used without an adjective. It is very important to distinguish between fracture and the mechanism of crack propagation. Also, in interpreting the data, the analyst must be very careful in making statements about the cause of failure. For example, the mode of failure in two separate cases can be the same, e.g., fatigue; however, the cause of failure can be entirely different as pointed out earlier. Statements about the cause of failure imply responsibility. If the cause of failure is improper design, it is the responsibility of the designer; however, if it related to processing, it is the responsibility of the manufacturer and so on.

11.7 Recommendations

A complete failure analysis investigation must conclude with a set of recommendations to prevent future failures. It is possible to cite recommendations for short-term solution as well as for a long-term solution. Any recommendation cited by the analyst must be consistent with the results of the investigation.

Since in many cases, microstructural characterization reveals some abnormal features, the analyst may recommend replacing the material. However, this should never be recommended unless the exact origin of the abnormality is identified. It may be due to improper processing or service conditions inconsistent with the design parameters. However, if it is determined that the material must be replaced, the analyst may recommend various alternatives, citing their capabilities and limitations as they relate to the application.
11.8 Failure Analysis Reports

Documentation of a failure analysis study in a detailed report is not only useful in providing a solution to a specific problem, information contained in the report can also be used to settle legal claims. If such information is made available to designers, it can be extremely helpful in selecting proper materials, minimizing or preventing future incidents of failure. Furthermore, results of failure analysis studies can be used in alloy developmental work. For these reasons, every effort must be made to produce a high-quality failure analysis report. Although the exact procedure for writing the report can differ from one case to another, as a guide, the report may include the following sections:

1. **Summary**, describing in concise manner the background of the problem, analysis procedure, the results of investigation and recommendations.

2. **Background**, which must include as much information as possible about the process, application, and the material involved. Whenever applicable or possible, both the maintenance and shutdown history must be included. Also, the sequence of events preceding the failure can be included in this section.

3. **Experimental procedure**, describing all the techniques used to analyze samples as well as the methods used for sample preparation.

4. **Visual inspection**, including photographs of the failed component illustrating specific features such as cracks, corrosion deposit, perforations, pits, thinning, etc. It is useful to illustrate the locations from which samples were removed for analysis.

5. **Material verification**, presenting the results of analyzing the chemical composition of the material in comparison with the nominal composition if provided by the proponent. This clearly shows whether the material is the same as that specified. It is useful to include in this section some material properties relevant to the application.

6. **Stress and fracture mechanics analysis**, presenting the results of analyzing the stresses in the failed component.

7. **Microstructural characterization**, presenting comparative microstructural features of failed and unfailed sections, as well as those of material never used in service. Particular attention should be given to such features as grain size, structure and morphology of grain boundaries, precipitation of secondary phases. Also, if applicable, this section may include results of characterizing fracture surfaces, corrosion deposits, etc. Results of property evaluation such as hardness measurement may also be included in this section.
8. **Mode of failure**, based upon the results presented. The mode of failure is clearly specified, e.g., fatigue, creep, carburization attack, etc.
9. **Cause of failure**, combining all above information. The most probable cause of failure is discussed in this section.
10. **Conclusions**, summarizing the mode and cause of failure.
11. **Recommendations**, presenting views about various possible solutions to the problem.
12. **References**, used for obtaining information during the investigation. Examples of failure analysis reports are presented in Chap. 12.
12

Case Studies

12.1 Introduction

It is the objective of this chapter to demonstrate how the knowledge developed in previous chapters can be applied to solve specific cases of failure. As explained earlier, the exact procedure can vary considerably from one case to another. This is demonstrated by the failure cases presented in this chapter. Also, the relative importance of various investigative tools are illustrated with specific examples. To draw the reader’s attention to various aspects of failure, two main types are considered in this chapter: (i) failure of engineering alloys as a result of improper processing practice and (ii) failure of engineering products.

Failures during processing of engineering alloys range from cracking of an ingot during forging to unsatisfying specifications of a finished alloy product. Understanding the origin of such processing-related failures can significantly contribute to developing alloys of better fabrication characteristics, as demonstrated later.

An engineering product can fail during the course of its manufacturing for various reasons, e.g., improper heat treatment practice. When a product is put into service, it can prematurely fail because of (i) errors in design, (ii) errors in manufacturing, or (iii) service conditions exceeding the design limits. Sometimes, it is implied that errors in design relate only to geometry and dimensions. However, it is very important to realize that material selection and tailoring the material to specific fabrication technique are integral parts of design. Therefore, failures which can be traced back to design include (i) improper geometry and/or dimensions resulting in localized stress concentration and (ii) incorrect selection of
material. Even if a proper material is selected on the basis of, e.g., mechanical strength, failure may occur because the material is incompatible with the prescribed fabrication technique, e.g., welding. In some cases, such failures are said to be caused by improper fabrication technique. However, as can be appreciated from the above discussion, they are more accurately caused by incorrect selection of material.

When a product of a sound design is put into service, it can fail as a result of manufacturing errors, such as introducing flaws or discontinuities beyond the limit set by design or exceeding certain material limitations set by the manufacturer, e.g., specific heat treatment requirements. A sound product in terms of design and manufacturing can still fail because of improper maintenance practice or service conditions exceeding the design limits.

It may be concluded from the discussion above that if the case studies presented in this chapter are classified by the mechanism leading to failure, e.g., fatigue-related failure, corrosion-related failure, etc., some of the above important aspects can be missed or overlooked. Instead, it is believed more useful to identify specific cases representative of various stages during the history of an engineering product, beginning with its building blocks, i.e., the engineering alloys. Furthermore, failures of engineering products during service are related to the industry, e.g., chemical process, petrochemical, oil, power generation, etc. This approach allows the reader to better appreciate the significance of developing the necessary background relevant to the particular stage where the failure has occurred. Also, it clearly demonstrates the importance of clear definition of the problem, which is essential in concluding a successful investigation.

By no means do the failure cases presented in this chapter cover all aspects of failure. However, they are carefully selected to serve as guides so that the reader develops the capabilities for conducting a failure analysis investigation. Specifically, the cases selected demonstrate how to focus on key issues and relate them to the relevant technical background required to determine the cause of failure and identify means for preventing future failure. Also, the complementary roles of materials science and materials engineering in solving failure cases are demonstrated.

12.2 Failure of Engineering Alloys Due to Improper Processing Practice

12.2.1 Failure of a Rolled Slab Due to Improper Cooling Practice

**Process**

Wrought products of a Ni-Mo alloy were processed by first casting 30.48 cm × 106.68 cm slab ingots for subsequent hot rolling into 14 cm thick slab.
Component
Hot-rolled slabs of 14 cm in thickness

Material
A Ni-Mo alloy of commercial grade

Application
Processing of wrought products

Processing Conditions
Rolling to final thickness (14 cm) consisted of three sessions each of which was performed at 1230°C. Subsequently, three slabs were stacked on top of each and air-cooled to room temperature before subsequent processing.

Problem
After a stack of three slabs were air-cooled to room temperature, a transverse crack as long as 76 cm and extending through the entire thickness was consistently observed in the middle slab. In contrast, no cracks were detected in either the top or bottom slabs.

Analysis
1. Visual Inspection
Visual inspection showed no evidence for measurable macroscopic plastic deformation of the failed slab. To examine the fracture mode, the failed slab was completely separated. The original crack surface had a texture different from that produced by the final separation. Chevron marks were distinguished on the original crack surface along the thickness of the slab. It is recalled that these marks run parallel to the crack growth direction, which is generally normal to the direction of maximum principal tensile stress.

2. Material Verification
According to specifications, the chemical composition of the alloy was Ni-(26–30)%Mo-2%Fe*-1%Cr*-1%Co*-1%Si*-1%Mn*-0.02%C*. Chemical analysis of the failed slab showed its composition to be Ni-26.94%Mo-0.7%Fe-0.31%Cr-0.1%Mn-0.02%Si-0.003C, which lies within the specifications. Similar results were obtained for the top and bottom slabs ruling out any deviation from the specified composition as a contributory factor to the failure.

*Asterisk indicates maximum throughout the chapter.
3. Stress Analysis

To take the above analysis one step further, it is required to determine the origin of the tensile stress leading to cracking. From a consideration of the stack of slabs, it can be concluded that the only applied stress is a bearing stress resulting from the slab weight. Evidently, only the middle and bottom slabs are acted upon by a bearing stress (compressive stress) and that the bearing stress acting on the bottom slab is twice that acting on the middle slab. If the alloy density is $\rho$ and the width, length, and thickness of the slab are $w$, $l$, and $t$, respectively, the bearing stress $\sigma_b$ acting on the middle slab is given by

$$\sigma_b = \frac{(wlt\rho)}{wl} = t\rho$$

where $wl$ is the bearing area. Given that $\rho$ of the alloy is 9.22 g/cm$^3$ and $t = 13.97$ cm, the bearing stress acting on the middle slab was

$$\sigma_b = (9.22)(13.97) = 129 \text{ g/cm}^2 = 12,900 \text{ g/mm}^2 = 18 \text{ ksi} = 124 \text{ MPa}$$

and the bearing stress acting on the bottom slab was $2\sigma_b = 248$ MPa.

To summarize, both the middle and bottom slabs were cooled under the influence of bearing stresses such that the bearing stress acting on the middle slab was one half of that acting on the bottom slab. However, because the middle slab was sandwiched between the top and bottom slabs, it would be expected to cool at a slower rate.

It is recalled that for ductile materials, the yield strength in tension is approximately equal to that in compression. During cooling, both the middle and bottom slabs tend to plastically deform at those temperatures where the applied bearing stress exceeds the yield strength. For example, at 1095°C, the tensile yield strength of the material is 55 MPa, while at 980°C, the yield strength is increased to 110 MPa. When the yield strength exceeds the bearing stress with continued cooling, the deformation becomes only elastic.

By analogy with a compression test, the constraint imposed upon the middle and bottom slabs by frictional forces would restrict deformation in the surface layers, inducing surface tensile stresses normal to their cross-sectional areas and reaching maximum values at midlength of the slabs. Because the stress distribution developed by the frictional force is rather complex, it is difficult to analyze. However, at some temperature during cooling where the yield strength becomes equal to the bearing stress, the maximum tensile stress $\sigma_t$ developed at the surface is approximately given by

$$\sigma_t = E \times \text{strain at the yield point (0.2%)}$$

where $E$ is the elastic modulus of the material ($E = 30 \times 10^3$ ksi = 207 \times 10^3 MPa), i.e.,

$$\sigma_t = 30 \times 10^3 \times 0.002 = 60 \text{ ksi} = 414 \text{ MPa}$$
A buildup of tensile stresses continues with further cooling reaching maximum values at midlength of the slabs. Therefore, it is plausible to assume that at room temperature, the maximum tensile stresses at mid-length of the middle slab had exceeded 414 MPa. However, it is important to realize that because of the greater bearing stress acting on the bottom slab, the induced surface tensile stresses would be greater relative to the middle slab. Since the bottom slab did not fail, it became evident that the magnitude of induced tensile stresses was not the sole factor responsible for the failure and that another contributory factor must have been involved.

It could be concluded from the above analysis that cracking of the middle slab was likely initiated at the surface with the aid of induced tensile stresses reaching maximum values at mid-length. Since the tensile stresses were diminished as the edges of the slab were approached, this could explain at least partially, why the crack was arrested near the edges. However, this could not explain why the bottom slab did not fail although it had experienced greater tensile stresses. Noting that the middle slab was cooled at a slower rate relative to the bottom slab, this seemingly contradictory behavior could only be resolved by microstructural characterization as described later.

4. Mechanism of Crack Propagation

Fractography revealed that the cracks developed in the middle slab had propagated by an intergranular mechanism, as shown in Fig. 12.1a–c. Secondary grain separation could be seen at the fracture surface. Evidently, the grain boundaries of the middle slab were severely weakened to the extent of grain separation under the influence of the induced tensile stress.

It is recalled that grain boundaries can be weakened at elevated temperatures by creep deformation. Another source of weakening the grain boundaries is precipitation of embrittling phases. However, since the final separation induced at room temperature also occurred intergranularly, it is more likely that the grain boundaries were weakened by a mechanism other than creep. A further evidence for this was provided by the results of surface hardness measurement. Both the top and middle slabs had an average surface hardness in the range of Rb 94–98. In contrast, the average surface hardness of the middle slab was found to be Rc 36, indicating that it was substantially hardened by some mechanism.

Based upon the above results, it could be concluded that the crack developed in the middle slab was propagated with very little or no plastic deformation caused by a severe hardening effect.

5. Microstructural Characterization

Even though the middle slab was extremely harder than either top or bottom slab, no marked difference in their microstructure could be detected on the scale of light optical metallography. Analysis of the slabs by X-ray diffraction revealed that both the top and middle slabs predominantly consisted of a face-centered...
cubic phase \( (a = 0.361 \text{ nm}) \). The diffraction pattern for the middle slab showed a mixture of fcc phase and another phase identified as Ni\(_4\)Mo (ordered tetragonal structure; \( a = b = 0.5727 \text{ nm}, c = 0.3566 \text{ nm} \)). It was estimated that Ni\(_4\)Mo constituted about 85% by volume of the middle slab. Evidently, the formation of Ni\(_4\)Mo in the middle slab occurred on a scale too fine to be detected optically requiring an investigative technique of a higher resolution.

Thin-foil specimens were prepared from the middle slab for examination by transmission electron microscopy. It is well known that in the binary Ni-Mo

Figure 12.1 Failure of a Ni-Mo alloy due to long-range ordering to Ni\(_4\)Mo. (a) A macrograph of fractured slab. (b) Secondary electron SEM image of region A, and (c) Secondary electron SEM image of region B.
system, an alloy containing 29.1 wt% Mo (25 atomic % Mo corresponding to the Ni₄Mo composition) exists as a disordered solid solution above a temperature of about 870°C. Below that temperature, the alloy undergoes a generalized long-range ordering reaction resulting in the transformation of the fcc phase into an ordered tetragonal structure (Ni₄Mo). If the alloy contains less than 29.1 wt% Mo, Ni₄Mo precipitates from the parent fcc phase during cooling.

Experiments show that an exposure of the slab material for only 10 min at 760°C reduces the tensile ductility from 55 to 3% corresponding to an increase in hardness from Rb 98 to Rc 34 and a corresponding increase in yield strength from 400 MPa (58 ksi) to 772 MPa (112 ksi). It is evident that precipitation of Ni₄Mo in the alloy can occur very rapidly. Prior to precipitation of Ni₄Mo where the alloy consisted of an fcc solid solution corresponding to 55% tensile ductility, fracture at the end of a tensile test occurred by microvoid coalescence. In contrast, after 10 min of exposure at 760°C (3% tensile ductility) resulting in the precipitation of Ni₄Mo, fracture occurred intergranularly.

It could be concluded from the observations above that the middle slab was severely embrittled by precipitation of Ni₄Mo as a result of slower cooling rate in comparison with the top and middle slab.

6. Stress Required for Crack Propagation

It is evident from the results above that the crack in the middle slab had propagated by a brittle mechanism. Therefore, the principles of linear-elastic fracture mechanics can be used to predict the onset of failure requiring (i) knowledge of the plane strain fracture toughness parameter of the material \( (K_{lc}) \) and (ii) modeling of the initial crack. Since the experimental value of \( K_{lc} \) of the material is not known, it can be estimated as follows:

\[
K_{lc} = \frac{[(0.05e_f n^2 E \sigma_y)/3]^{1/2}}
\]

where \( e_f \) is the true strain at fracture in a tensile test, \( n \) is the strain hardening exponent, \( E \) is the elastic modulus, and \( \sigma_y \) is the yield strength. A reasonable estimate of the nominal strain \( e \) at fracture is 3%, giving a value of \( e = \ln(1 + e) = 0.03 \), \( \sigma_y = 772 \) MPa, and the corresponding value of \( n = 0.1 \) as determined from tensile stress-strain data. Since \( E \) of the material is in the order of \( 30 \times 10^3 \) ksi = \( 207 \times 103 \) MPa, the value of \( K_{lc} \) is estimated as

\[
K_{lc} = \left(0.05 \times 0.03 \times 0.01 \times 207 \times 10^3 \times 772/3\right)^{1/2} = 28 \text{ MPa (m)}^{1/2}
\]

To determine whether the crack had propagated under plane strain conditions,

\[
t > 2.5(K_{lc}/\sigma_y)^2
\]

where \( t \) is the thickness of the slab (0.14 m). From the above data, \( 2.5 (K_{lc}/\sigma_y)^2 = 0.003 \) m, and therefore plane strain conditions are satisfied. Since
experimental results showed that the crack propagated by an intergranular mechanism, it is reasonable to model initial surface crack with a length equivalent to grain diameter (about 0.0003 m). In this case, the stress intensity factor $K$ is given by

$$K = C\sigma (\pi c)^{1/2}$$

where $C$ is a constant, $\sigma$ is the stress at the crack tip, and $c$ is the length of the crack. Using a value of $C = 1.15$,

$$K = 0.1\sigma$$

It is recalled that crack propagation occurs when $K = K_{\text{lc}}$, i.e.,

$$0.1\sigma = 28$$

and the stress $\sigma$ required to propagate the crack is

$$\sigma = 280 \text{ MPa}$$

Since the maximum tensile stress developed at midlength of the middle slab was estimated to be greater than 414 MPa, it is evident that the apparent stress at failure was significantly less than the induced tensile stress. Consistent with the results of microstructural characterization, this indicated that the middle slab became defective by slow cooling, resulting in the precipitation of Ni$_4$Mo and corresponding loss of ductility. Also, this could explain why the bottom slab did not fail although it had experienced greater tensile stresses because its higher cooling rate precluded extensive precipitation of Ni$_4$Mo.

**Mode of Failure**

It could be concluded from the results of this investigation that the mode of failure of the middle slab was intergranular embrittlement, resulting in cracking under the influence of induced tensile stresses.

**Cause of Failure**

Based upon experimental observations, the cause of failure was the relatively slow cooling rate of the middle slab as a result of being sandwiched between the top and bottom slabs. This made the middle slab defective by extensive precipitation of the embrittling Ni$_4$Mo phase.

**Solution**

Evidently, the solution of the problem above was to never stack slabs of the alloy on top of each other to avoid slow cooling of the middle slabs. In fact, when this practice was adopted, the problem was rectified.
Remarks

Clearly, the above investigation demonstrates the necessity of combining macroscopic and microscopic analyses to solve certain failure analysis cases. It is evident that macroscopic analysis could not provide an answer to why only the middle slab had failed. In fact this analysis alone could lead to the conclusion that the bottom slab was more likely to fail. Also, microstructural characterization provided a valuable aid in using the proper materials properties to estimate the value of the plane strain fracture toughness. Furthermore, the above case demonstrates the significance of selecting the proper investigative tools for microstructural characterization. In the meantime, microstructural characterization alone could not provide information regarding the mechanics of developing the crack and its propagation. To summarize, the above case illustrates the complementary nature of macroscopic and microscopic analyses to solve failure analysis cases. As a note added in proof, recent research aiming at solving the low-temperature embrittlement problem has led to the recognition that a significant improvement in ductility can be achieved by critical additions of boron without affecting the other properties of the alloy.

12.2.2 Failure of a Sheet Metal Due to Improper Heat Treatment

Process

A sheet product of a Ni-22Cr-14W-0.1C alloy was produced to have a creep strength corresponding to time to 0.5% creep strain at 40 MPa/925°C must be in excess of 15 h according to specifications.

Component

Sheet product of a Ni-22Cr-14W-0.1C alloy

Application

Processing of a sheet metal

Problem

After processing of a sheet metal product of the alloy (about 3 mm in thickness), it failed to pass the specification test. During creep testing, standard specimens from the sheet were ruptured after as short as 5 h of exposure at 40 MPa/925°C.
Analysis

1. Material Verification

According to specifications, the nominal chemistry of the alloy was Ni-22Cr-14W-3Co-3Fe-0.5Mn-0.4Si-0.3Al-0.1C-0.03La-0.005B. It is recalled from earlier discussions that elements such as C and B are of particular importance for resistance to high-temperature creep deformation. Therefore, in such a case of failure, it is essential that chemical analysis of the failed sheet must include the concentrations of C and B. Chemical analysis of the failed sheet showed its composition to be Ni-21.97Cr-13.82W-0.89Co-1.2Fe-0.48Mn-0.37Si-0.09C-0.026La-0.0046B. Evidently, the chemistry of the sheet including the concentrations of C and B conformed to specifications. This ruled out improper chemistry as a possible cause of failure.

2. Creep Behavior

Creep tests at 40 MPa/925°C showed that the failed sheet acquired a considerably large strain during the primary creep stage in comparison with the unfailed sheet. Also, the transition from the primary to the tertiary stage was abrupt in the failed sheet without a noticeable steady state. Rupture of the sheet occurred intergranularly.

It is recalled that deformation during primary creep is history dependent. Specifically, it is a function of the initial microstructure of the material. This suggested that processing of the failed sheet could have deviated from the normal practice.

3. Estimation of the Stress at Failure

From available data on the material at 925°C, the elastic modulus is 155 × 10³ MPa, the nominal strain at fracture is 0.7 corresponding to a true strain of 0.53, the strain hardening exponent is 0.3, and the yield strength is 182 MPa. Using this data, the plane strain fracture toughness parameter $K_{lc}$ can be estimated as follows:

$$K_{lc} = \left[ (0.05\varepsilon_f n^2 E \sigma_y)/3 \right]^{1/2}$$

where $\varepsilon_f$ is the true strain at fracture in a tensile test, $n$ is the strain hardening exponent, $E$ is the elastic modulus, and $\sigma_y$ is the tensile yield strength. Substituting the values of these parameters results in

$$K_{lc} = 150 \text{ MPa } (m)^{1/2}$$

To determine whether plane strain conditions are satisfied;

$$t > 2.5(K_{lc}/\sigma_y)^2$$

where $t$ is the thickness of the sheet = 0.003 m. From the estimated value of $K_{lc}$, $2.5 \left(K_{lc}/\sigma_y\right)^2 = 1.7 \text{ m}$, which is greater than $t$. Therefore, in this case, plane strain
conditions are not satisfied requiring the use of the more complex methods of elastic-plastic fracture mechanics to estimate the value of the plane stress fracture toughness \((K_c)\). However, from a safety point of view, it is still possible to use the value of \(K_{lc}\) as a design criterion, noting the value of \(K\) can be 2–10 times greater than that of \(K_{lc}\).

Since rupture occurred intergranularly, it is possible to model the initial crack with a length on the order of a grain diameter (about 0.00005 m). The crack length divided by half the specimen width is 0.004; the value of \(C\) is estimated to be about 1.15. However, because the material is not brittle, the crack length must be corrected to account for crack tip plasticity such that the stress intensity factor \(K\) is expressed as

\[
K = C\sigma\left[\pi(c + r_p)^{\frac{1}{2}}\right]
\]

where \(\sigma\) is the stress at the crack tip, \(c\) is the crack length, and \(r_p\) is the radius of the plastic zone. It is recalled that under plane stress conditions, \(r_p\) is given by

\[
r_p = \frac{1}{6}\pi\left(K_{lc}/\sigma_y\right)^2 = 0.04
\]

Therefore, \(K\) is given by

\[
K = 0.4\sigma
\]

Equating \(K_{lc}\) with \(K\) at failure results in \(\sigma = 375\) MPa. Evidently, if the greater value of \(K_c\) is applied, it results in a greater value of \(\sigma\). In either case, however, the stress at failure was considerably greater than the applied stress of 40 MPa. It is to be noted that stress intensification at the crack tip can exceed \(3\sigma_y\) at elevated temperatures. In conclusion, the results above suggested that the failed sheet behaved as if it was overloaded. Obviously, however, this was not the case for the unfailed sheet. To determine the cause of failure, it is then required to identify the reason for the apparent overloading condition.

4. Microstructural Characterization

On the scale of light optical metallography, the microstructure of the failed sheet was indistinguishable from that of the unfailed sheet. However, on the finer scale of transmission electron microscopy, a distinctive difference in microstructure was observed particularly at grain boundaries. Grain boundaries of the failed sheet contained a fine continuous network of the Cr-rich \(M_{23}C_6\) carbide precipitates assuming lamellar or cellular morphology, as shown in Fig. 12.2. A carbide of the type \(M_{23}C_6\) is identified by its characteristic reflections which appear at \(\frac{1}{4}\) and all equivalent positions of the matrix reflections. In contrast, the grain boundaries of the unfailed sheet were either free of precipitates or contained discrete particles of \(M_{23}C_6\) carbide.
A lamellar carbide morphology indicated that it had precipitated in the alloy by a discontinuous mechanism involving migration of grain boundaries. It is well known that a lamellar $\text{M}_2\text{C}_6$ carbide morphology weakens the grain boundaries and in turn shortens the creep rupture life. In contrast, discrete carbide particles can strengthen the grain boundaries. Discontinuous grain boundary precipitation of the carbide could only occur as a result of a rather slow cooling rate from the annealing temperature. However, according to specifications, the alloy must be rapidly cooled from the annealing temperature. It could then be concluded that slow cooling resulting in discontinuous precipitation of lamellar $\text{M}_2\text{C}_6$ carbide at grain boundaries had considerably reduced the creep strength of the material giving rise to an apparent overloading condition when it was tested at 40 MPa/925°C. In fact, when samples from the failed sheet were solution-annealed to redissolve the carbide, rapidly cooled, and then tested at 40 MPa/925°C, creep curves resembling that of unfailed sheets were obtained.

Mode of Failure

From the results of the investigation, the mode of failure was determined to be discontinuous grain boundary precipitation of lamellar $\text{M}_2\text{C}_6$ carbide leading to an apparent overloading condition, which had resulted in premature creep rupture.
**Cause of Failure**

Based upon the experimental results, the cause of failure was found to be slow cooling rate from the solution annealing temperature, deviating from standard practice which called for rapid cooling.

**Solution**

To realize the characteristic creep strength of the alloy sheet, it is essential to adhere to the standard practice requiring rapid cooling from the solution annealing temperature.

**12.2.3 Failure of a Rear Axle Shaft of an Automobile**

**Process**

An automobile turned over during a ride on a flat road. No information was available regarding speed, road conditions, load, etc., prior to the accident.

**Component**

Portion of a fractured rear axle shaft removed from the scene of accident

**Material**

Carbon steel. Typically carbon steels are used for automobile axis shaft applications.

**Analysis**

1. **Visual Inspection**

   There was no evidence for distortion near the fracture surface indicating that the fracture was not preceded by macroscopic plastic deformation. The morphology of the fracture resembled that produced by rotational-bending fatigue. Fatigue beach marks within the surface hardened layer (case) could be clearly distinguished. The beach marks converged at the axle surface where the fracture was originated, as shown in Fig. 12.3.

2. **Microstructural Features of the Fracture Surface**

   Fracture surface within the case exhibited an intergranular fracture. Fatigue striations were not clearly visible which could be related partially to the low ductility of the case. Fatigue crack propagation within the case occurred by a brittle mechanism; i.e., no plastic deformation occurred at the crack tip. In contrast, the fracture surface within the core exhibited features of a mixed intergranular and transgranular crack propagation.
3. Microstructural Characterization of the Axle Material

The case consisted of martensite, revealed by its characteristic platelet-type morphology, along with particles of manganese sulfide, as shown in Fig. 12.4.

4. Mechanical Strength

The average hardness of the case was Rc 57 corresponding to the martensitic structure. Average hardness of the core was about Rc 25. It appears that the case was excessively hardened, leading to poor ductility and higher notch sensitivity (low fracture toughness).

**Mode of Failure**

The axle shaft failed by rotational-bending fatigue.

**Cause of Failure**

Available evidence points out that the most probable cause of failure was low fracture toughness of the surface hardened layer (case).

**Conclusion**

Information regarding the specifications of the axle shaft should be obtained from the auto-maker before reaching a sound conclusion.

---

*Figure 12.3* A light optical micrograph showing beach marks within the case (surface hardened layer), converging at the surface where fracture had originated.
Figure 12.4  Typical microstructural features of the axle material within the case (specimen etched in Nital solution); secondary electron SEM images indicating that the case consisted of martensite. Also, EDS spectrum indicating presence of MnS particles.
12.2.4 Failure of a Beam Clamp

Process
Beam clamps used for mounting channels

Component
Beam clamps

Material
Carbon steel

Problem
The beam clamps had undergone galvanized coating and polyester coating for corrosion protection. They were installed at a location close to the sea front. They developed rust after 1 year in service while the expected life was 10 years.

Analysis
1. Visual Inspection
The failed sample exhibited regions where the coating had been separated from underlying substrate. These areas showed extensive rusting. The surface texture was relatively rough; however, the surface was free of scratches or indentations. An unused sample, examined for comparison, showed a sound coating with a smooth texture.

2. Microstructural Characterization
Microstructural examination of the failed sample confirmed the presence of Iron oxide at regions where coating had come off. Contaminants such as Ca, Si and Cl were also observed at the coating surface. Cross-sectional view of the samples confirmed that the coating was intact at regions showing no oxide formation.

   The coating on unused sample was uniform and free of defects such as voids, bubbles or scratches. The coating was organic in nature (C and O) and contained titanium compound and barium sulfate as additives. The cross-sectional sample revealed that the galvanized layer was only 5 μm in thickness at some regions. The maximum thickness was observed to be 20 μm.

Mode of Failure
The clamp material developed rust due to its exposure to environment after the polymer and Zn coat failed at localized regions.
Cause of Failure

As per ASTM specification for hot-dip galvanized coatings on steel products (A 123-89a), the thickness of Zn coating (for sheet material thickness of 3.2–8.4 mm) should be 85 μm, which is considerably higher than that observed on the beam clamp material. Polymer-based coating alone is not sufficient to protect the clamp material in a seashore environment. Improper galvanizing treatment is thought to be responsible for the rusting observed on the beam clamps.

Conclusion

Use steel material/component that has undergone a proper hot-dip galvanizing treatment.

12.2.5 Formation of White Rust and Black Stain on a Galvanized Steel Pipe

Process

Carbon steel pipes were produced by a seamless tubing process. To protect the pipes against corrosion attacks, they were galvanized by zinc using hot-dipping process. Prior to galvanization, pipe was prepared by first degreasing in a tank containing a sequestering and wetting agent at 60–80°C to remove grease, oil and other deposits, and then the pipes were rinsed in running water at 70–80°C. Subsequently, the pipes were pickled in sulfuric acid at 60–80°C and then rinsed in cold running water. After pickling, the pipes were immersed in a fluxing tank containing zinc chloride and ammonium chloride. Finally, they were dipped in molten zinc and then water quenched.

Component

Carbon steel pipe

Problem

White rust and black stains were observed to develop at the pipe surface after processing. These features appeared on the pipes within hours of galvanizing treatment.

Analysis

1. Visual Inspection

On-site inspection revealed that the white rust and black stain were developed on pipes stored either indoors or outdoors. Such white rust was typical of what is known as “white bloom,” usually developed on galvanized steel products stored in a humid atmosphere or near the sea. The white rust could be wiped off but the black stain adhered to the pipe surface.
2. Microstructural Characterization

The white rust was removed and analyzed by X-ray diffraction and SEM/EDS. It was identified as zinc oxide (ZnO, hexagonal structure: \( a = 0.3249 \text{ nm} \) and \( c = 0.1602 \text{ nm} \)). The major elemental constituent of the black stain was C with Zn, Al, S, Cl, and Fe. The thickness of the Zn layer was about 65 \( \mu \text{m} \), while the black stain existed as thin 1 \( \mu \text{m} \) layer at the outer surface.

**Origin of Black Stain**

A thin surface film enriched in C is thought to have originated from some residual grease (hydrocarbon). It is likely that this stain could have resulted from incomplete degreasing.

**Origin of White Rust**

When pipes are stacked together, anodic sites tending to corrode could develop at regions of intimate contact where the oxygen concentration is relatively low. Exposed regions where the oxygen concentration is relatively higher become cathodic. Corrosion could then occur by condensation of moisture (dew formation) according to the following reactions:

- **Cathode reaction**: \( \text{O}_2 + 4e^- + 2\text{H}_2\text{O} \rightarrow 4\text{OH}^- \) (in solution)
- **Anode reaction**: \( 2\text{Zn}_{\text{metal}} \rightarrow 2\text{Zn}^{2+} \) (ions in solution) + \( 4e^- \)
- **Reaction in solution**: \( 2\text{Zn}^{2+} + 4\text{OH}^- \rightarrow 2\text{Zn(OH)}_2 \)
- **Total reaction**: \( 2\text{Zn}_{\text{metal}} + \text{O}_2 + 2\text{H}_2\text{O} \rightarrow 2\text{Zn(OH)}_2 \)

Subsequently, zinc hydroxide \([\text{Zn(OH)}_2]\) could transform into zinc oxide \((\text{ZnO})\) by dehydration (drying), i.e.,

\[ \text{Zn(OH)}_2 \rightarrow \text{ZnO} + \text{H}_2\text{O} \]

**Solution**

The surface quality of the pipes could be improved by chromating to avoid the ZnO rust formation. In this process, the galvanized part is dipped in a bath containing hexavalent chromium ions and mineral acid. The formation of black stains can be avoided by proper degreasing.

12.2.6 Failure of a Furnace Outlet Pipe

**Process**

A furnace in a benzene plant used to handle a gaseous hydrocarbon

**Component**

An outlet pipe of the furnace
Material
The furnace outlet pipe was made of Incoloy alloy 800H (internal diameter = 110 mm, wall thickness = 14 mm).

Application
The furnace had four anchored outlets to separate the inside and outside piping for independent thermal expansion. Each outlet pipe consisted of welded sections. Anchoring of the outlets was carried out by fillet welding of two stoppers made of 304 stainless steel at the top and bottom of the outlet pipes. The stoppers were placed inside grooves made in the tube support bracket.

Service Conditions
The gas composition was 64.27% H₂, 18.92% C₁, 3.48% C₂, 0.96% C₃, 0.03% C₄, 0.24% C₅, 1.05% C₆-C₈ NON-ARO, 5.91% benzene, 1.45% C₈ aromatics, 0.04% C₉ and 0.15% diphenyl. During operation, the temperature at the furnace outlet was 595°C.

Problem
A furnace outlet pipe developed transverse cracks about 10–20 mm from the weld region.

Analysis
1. Visual Inspection
The transverse crack occurred near the toe of the fillet weld. It followed a zigzag path suggesting that it occurred along the grain boundaries.

2. Material Verification
Microchemical analysis verified that the outlet pipe was made of Incoloy alloy 800H. The nominal composition (wt%) of the alloy is given below:

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
</tr>
<tr>
<td>Ni</td>
<td>33</td>
</tr>
<tr>
<td>Cr</td>
<td>21</td>
</tr>
<tr>
<td>Ti</td>
<td>0.40</td>
</tr>
<tr>
<td>Al</td>
<td>0.40</td>
</tr>
<tr>
<td>Mn</td>
<td>0.80</td>
</tr>
<tr>
<td>Si</td>
<td>0.50</td>
</tr>
<tr>
<td>C</td>
<td>0.08</td>
</tr>
</tbody>
</table>
3. Microstructural Characterization

Considerable grain growth was observed in the weld heat-affected zone. It is well known that appreciable grain growth may occur in this alloy at temperatures above 980°C.

4. Mechanical Strength

The average microhardness at regions remote from the weld was HV 224 (Rb 95) typical for this alloy. However, the hardness within the heat-affected zone near the crack was HV 268 (Rc 25), indicating that the material was somewhat hardened probably due to the residual stresses associated with the welding.

5. Examination of Fracture Surfaces

Cracking of the pipe was observed to initiate at its outer surface, ruling out that it was process related. The crack had propagated along the grain boundaries. Fracture surface revealed dimples on separated grain facets formed due to microvoid coalescence in grain boundary regions. Presence of both equiaxed and elongated dimples suggested a mixed fracture mode.

6. Analysis of Heat-Affected Zone

The heat-affected zone showed layers of grain boundary precipitates about 15–20 μm in thickness. These precipitates were oxide phases based upon Fe and Cr. The grain boundary was found to be Ni rich. Most evidence suggested that internal oxidation occurred during welding which led to intergranular embrittlement. Under the influence of a residual and applied stress, plastic deformation would be concentrated within the relatively soft Ni-rich phase at the grain boundary, eventually leading to ductile intergranular failure as observed here.

Mode of Failure

The pipe had failed by developing cracks along the grain boundaries within the heat-affected zone. This was caused by internal oxidation leading to the formation of Cr-, Fe-rich oxides enveloping a Ni-rich phase at the grain boundaries. Under the influence of a residual or applied stress, plastic deformation could be highly localized within the soft Ni-rich phase at the boundary promoting intergranular failure.

Cause of Failure

Improper welding procedure

Solution

It was recommended to check the welding parameters. Particular attention must be given to the atmosphere so that internal oxidation could be avoided.
12.2.7 Failure of a Second-Stage Turbine Blade

Process
Electric power generation by land-based gas turbine engine.

Component
Second stage gas turbine blade that had been refurbished after 40,000 h of service.

Material
INCO alloy 700

Service Conditions
The service temperature was 700°C. Information on exact history of the fracture blade, including the materials specifications, as well as the proprietary refurbishing process, was not available.

Problem
After refurbishing, the blade fractured after 900 h of operation.

Analysis

1. Visual Inspection
The general appearance of the fractured blade is shown in Fig. 12.5a. No macroscopic evidence of overheating or creep damage was visible. The extent of the discoloration of the fractured surface indicated that the cracking initiated at the trailing edge. Light photography revealed secondary cracks nearly parallel to the fracture surface, as shown in Fig. 12.5b. The orientation of these cracks indicated that the maximum tensile stress was orientated along the blade axis.

2. Material Verification
The blade material was identified as INCO alloy 710. The nominal composition of the alloy is given below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Inco Alloy 710</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni</td>
<td>Balance</td>
</tr>
<tr>
<td>Co</td>
<td>30</td>
</tr>
<tr>
<td>Cr</td>
<td>15</td>
</tr>
<tr>
<td>Al</td>
<td>3.2</td>
</tr>
<tr>
<td>Ti</td>
<td>2.2</td>
</tr>
<tr>
<td>Mo</td>
<td>3</td>
</tr>
<tr>
<td>Fe</td>
<td>1</td>
</tr>
<tr>
<td>Si</td>
<td>0.25</td>
</tr>
<tr>
<td>C</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
3. Microstructural Characterization

The matrix contained a fine dispersion of γ’ phase and the grain boundaries contained discrete particles of Cr-rich M₂₃C₆ carbide and Ti-rich carbide. The average hardness was found to be Rc 42, indicative of no overaging effect. However, specifications call for a hardness equivalent to Rc 34, suggesting that the material was improperly treated during refurbishing. The blade was protected by a Cr-rich coating typically used in land-based engines.

4. Examination of Fracture Surface

The fracture occurred intergranularly; however, the fracture surface appeared to consist of numerous fine dimples superimposed on separated grain facets, as

Figure 12.5  (a) Photograph of fractured blade where three zones 1, 2, and 3 can be distinguished. (b) Optical micrograph showing cracks nearly parallel to the fracture surface.
Figure 12.6  (a) SEM image showing characteristic microstructural features of the airfoil and root. (b) SEM image showing morphology of fracture surface; arrows indicate dimples at grain facets.

Copyright © 2004 by Marcel Dekker, Inc.
shown in Fig. 12.6b. This type of fracture results from microvoid growth and coalescence alongside grain boundaries in precipitation-hardened alloys containing grain boundary denuded zones.

**Cause of Failure**

The blade fractured by intergranular mode associated with grain boundary denuded zones. Since the formation of these zones requires dissolution of $\gamma'$-phase by a solution heat treatment followed by reprecipitation of the $\gamma'$ phase by a heat treatment schedule involving moderate cooling, and that there was evidence that these processes could have occurred during service, it is possible that the $\gamma'$-denuded zones were present in the as-refurbished condition.

**Solution**

It was recommended to consider the practice followed during the refurbishing operation, particularly heat treatment schedule.

**12.2.8 Failure of a Welded Joint between Carbon Steel Pipe and SS304 Elbow**

**Process**

As part of its operations, a petrochemical facility utilized a process by which reformed gas passed through a water cooler followed by compression in a three-stage centrifugal synthesis gas compressor.

**Processing Condition**

The reformed gas had the following condition at the suction of the compressor:

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Mol%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_4$</td>
<td>3.53</td>
</tr>
<tr>
<td>CO</td>
<td>14.50</td>
</tr>
<tr>
<td>CO$_2$</td>
<td>7.48</td>
</tr>
<tr>
<td>H$_2$</td>
<td>73.71</td>
</tr>
<tr>
<td>N$_2$</td>
<td>0.21</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>0.57</td>
</tr>
</tbody>
</table>

**Materials**

The suction drum was made of SS304 clad carbon steel. The outlet piping of the suction drum was comprised of carbon steel straight pipe (8 mm wall thickness/24 in. diameter) and type 304 stainless steel elbow fittings. The carbon steel pipe
was joined to SS304 elbows through butt weld joints by arc-welding process using E309/ER309 rods. These joints were made at a number of points in the installation. Material specifications for the pipe and the elbows were as follows:

Pipe: ASTM A 53 Grade B
Elbow: ASTM A 403 WP304

Problem
On-site radiographic tests revealed a number of cracks at various pipe-elbow weld joints in the piping system. These cracks were circumferential and always appeared along the weld seam region adjacent to the carbon steel pipe. The length of these cracks varied between 120 and 600 mm. Some of these joints had already undergone weld repair in the past. One of these joints failed again after 2 \(\frac{1}{2}\) years of service following weld repair. The length of the crack formed at this point was 300 mm and it had occurred at the weld seam.

Analysis
1. Visual Inspection
Visual examination of the failed sample indicated a separation of the weld bead from the carbon steel pipe. Circumferential cracking was evident at areas adjacent to the weld region. There was no evidence of branching in the primary crack. Corrosion deposits were not observed at the pipe surface.

2. Material Verification
ASTM A53 pipe material can have as much as 0.3wt% C. Chemical compositions (wt%) obtained from the pipe and the weld by inductive coupled plasma (ICP-AES) confirmed material specifications:

<table>
<thead>
<tr>
<th></th>
<th>Fe</th>
<th>Cr</th>
<th>Ni</th>
<th>Mn</th>
<th>Mo</th>
<th>Si</th>
<th>Cu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipe</td>
<td>99.3</td>
<td>0.05</td>
<td>0.04</td>
<td>0.55</td>
<td>0.003</td>
<td>0.02</td>
<td>—</td>
</tr>
<tr>
<td>Weld</td>
<td>67.8</td>
<td>19.7</td>
<td>11.2</td>
<td>1.66</td>
<td>0.003</td>
<td>0.18</td>
<td>0.04</td>
</tr>
</tbody>
</table>

3. Light Optical Microscopy
Light optical microscopy at low magnification revealed primary circumferential crack in the carbon steel pipe close to the weld region. Separation between the weld and the carbon steel pipe was evident in the sample. The fracture at the inner surface of the pipe was continuous through the pipe circumference, indicating that the cracking originated at the inner surface of the pipe near the root weld.
The primary crack was circumferential and the leakage occurred when this crack branched and traversed through the wall thickness of the pipe at different areas.

4. Analysis of Weld Metal

The EDS analysis obtained from the regions showed lower concentrations of Cr (13–15wt%) and Ni (5–6wt%) compared to the composition of E309 type welding rods used, indicating that the weld was diluted during welding.

5. Examination of Carbon Steel Pipe Fracture Surface

The region was located near the inner surface of the pipe and exhibited plastic deformation. A dimpled structure due to microvoid coalescence was evident.

6. Microstructure of Weld Cross Section

Microstructural examination of the weld region was conducted in order to determine the nature of cracking. The crack originated from the inner surface and extended toward the outer surface of the pipe. The crack did not exhibit branching and mainly traversed along the weld line. It was clear that its origin was located within the weld zone. The crack propagated intergranularly.

The microstructure of the carbon steel pipe constituted ferrite and pearlite regions. It was observed that the carbon steel adjacent to the interface between the steel pipe and the SS weld was devoid of pearlite, as shown in Fig. 12.7a. The chemical compositions obtained by EDS analysis of different regions of this area are also shown in Fig. 12.7b. It can be observed that the carbon steel-weld interface is predominantly composed of Fe with a small amount of Cr. Carbon was not included in the analysis due to difficulties in quantification. Micro cracks were observed within the pearlite depleted ferrite region.

7. Mechanical Strength

The results of Vickers microhardness tests carried out at different regions of the weld crosssection sample are summarized as follows:

<table>
<thead>
<tr>
<th>Area</th>
<th>VHN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stainless steel weld</td>
<td>321</td>
</tr>
<tr>
<td>Carbon steel–SS weld interface</td>
<td>700</td>
</tr>
<tr>
<td>Pearlite denuded zone near CS-SS interface</td>
<td>137</td>
</tr>
<tr>
<td>Ferrite and pearlite near denuded zone</td>
<td>148</td>
</tr>
<tr>
<td>Carbon steel pipe away from weld</td>
<td>183</td>
</tr>
</tbody>
</table>

The hardness measured at the interface of the carbon steel and the weld corresponds to Rockwell C 60, whereas other microhardness values corroborate the microstructure observed in the weld cross section. Unsurprisingly, the lowest
Figure 12.7  (a) Optical micrograph of the weld cross section showing decarburized zone. (b) SEM micrograph of the weld cross section showing the general microstructure, interface layer and stainless steel weld along with EDS compositions of different regions.
hardness was exhibited by the pearlite-denuded zone due to a lack of carbon in the region.

**Mode of Failure**

The carbon steel pipe material cracked from the internal surface at the root weld region which was in contact with the hydrogen-rich gas. The crack propagated primarily circumferentially and traversed through the pipe wall thickness, resulting in leakage. The crack originated at the interface between the carbon and stainless steel weld and its propagation within the carbon steel pipe was intergranular. The carbon–stainless steel fusion zone was decarburized resulting in a low hardness pearlite-denuded–ferrite region. The decarburization occurred during the welding process where the carbon in the pipe segregated toward the SS weld. This resulted in a carbon-rich segregated layer at the CS-SS interface. A hardness value of Rockwell C 60 indicated presence of martensite at this region. The fracture is thought to have initiated in this martensite zone.

The presence of martensite indicates that sufficient heat was generated during welding to fuse an excessive amount of carbon steel that resulted in dilution of weld deposit locally. As a consequence, a thin layer of low alloy steel was produced between the carbon steel and the stainless steel weld which possessed sufficient hardenability to form martensite during cooling from welding temperature. The layer was thin enough to have undergone quenching as it was cooled while in contact with a large mass of carbon steel pipe.

The presence of a thin layer of brittle martensite at the interface between carbon steel pipe and stainless steel weld sets up a metallurgical notch which is undesirable and can lead to cracking. Large amounts of hydrogen present in the medium being transferred through the pipe can result in hydrogen damage at martensite region during service. The cracks formed in this manner will initiate at the inner surface of the pipe, as observed in this study. Maximum sensitivity to hydrogen cracking occurs at or near room temperature, which is comparable to the service temperature encountered in the piping system under consideration here. The lack of multiple branching of the crack is also a characteristic of hydrogen-induced damage. The decarburized area adjacent to the carbon steel pipe can also result in cracking under tensile stress, as corroborated by micro-cracks observed in the region.

**Cause of Failure**

Experimental results indicate that the joint between the carbon steel pipe and the stainless steel weld failed due to the development of a thin martensite layer at its interface during cooling from welding temperature. The hydrogen-rich gas transported through the pipe can possibly initiate cracking at the martensitic region. The presence of a decarburized layer adjacent to the carbon steel pipe can also induce cracking under stress.
Solution

Employing low-heat input (i.e., small-diameter electrode at a low current and relatively higher voltage) during welding can eliminate dilution. The welding procedure should be controlled to hold penetration into the surface of the carbon steel to a minimum.

Contamination by hydrogen by any source (e.g., electrodes, moisture) should be avoided during welding.

As a long-term measure, the carbon steel pipe should be replaced by a higher grade material, e.g., stainless steel. This should provide sound welds as well as adequate protection against hydrogen damage.

12.3 Failure of Engineering Products During Manufacturing

12.3.1 Failure of a Scrubber Head Due to Improper Practice of Cold Working/Welding

Process

A scrubber head was made of a Ni-Mo alloy by cold working (bending) to produce the desired circular shape, followed by welding to handle hydrochloric acid in a chemical plant.

Component

A scrubber head

Material

A Ni-Mo alloy of a commercial grade

Application

Fabrication of a scrubber head

Problem

Nondestructive evaluation after welding using the liquid dye-penetrant test revealed that the scrubber head had developed cracks in the weld heat-affected zone.

Analysis

1. Visual Inspection

Cracks perpendicular to the weld were detected only at the outer surface of the scrubber. Since the scrubber was formed by bending, tensile stresses would be
developed above the neutral axis reaching maximum value at the outer surface and compressive stresses would be developed below the neutral axis reaching a maximum value at the inner surface. Evidently, the surface of each crack was perpendicular to the maximum tensile stress. Apparently, as the crack propagated inward through the thickness, it was arrested by the compressive stresses.

2. Material Verification

According to specifications, the nominal chemical composition of the scrubber alloy was Ni-26-30Mo-2*Fe-1*Cr-1*Co-1*Mn-1*Si-0.02*C. Chemical analysis showed the composition to be Ni-28.06Mo-1.03Fe-0.55Cr-0.241Mn-0.04Si-0.004C which lies within the above specifications.

3. Evaluation of Mechanical Strength

Surface hardness measurements in the vicinity of the cracks showed an average hardness of Rc 37. This is to be compared with Rb 93 of the same alloy in the annealed condition. Evidently, the material of the scrubber head was severely hardened resulting in substantial loss of ductility.

4. Cracking Mechanism

Using a fixture, a crack was made to propagate until complete separation occurred. Examination of the exposed surface showed that the initial crack was formed along the grain boundaries. Also, the remainder of the fracture was intergranular. This indicated that the scrubber head was embrittled by some mechanism, which was revealed by microstructural characterization as described below.

5. Microstructural Characterization

On the scale of light optical metallography, there were no specific microstructural features which could account for the observed hardening. However, on the finer scale of transmission electron microscopy, the material was found to contain platelets of Ni$_3$Mo identified by their characteristic diffraction patterns. X-ray diffraction analysis showed that about 80% of the material by volume consisted of Ni$_3$Mo. Formation of Ni$_3$Mo in Ni-Mo alloys is known to cause a substantial loss of ductility. Evidently, the Ni$_3$Mo phase had formed in the weld heat-affected zone during cooling.

Mode of Failure

Experimental results showed that the mode of failure was intergranular embrittlement caused by precipitation of Ni$_3$Mo in the weld heat-affected zone. As a result, the material could only accommodate the residual stresses of cold working by opening cracks along grain boundaries normal to the direction of maximum tensile stresses at the surface of the scrubber.
Cause of Failure

In this case, there are two factors that had contributed to the failure: (i) the residual stresses of cold working and (ii) the precipitation of Ni$_3$Mo in the weld heat-affected zone during cooling that embrittled the material. Evidently, the primary cause of failure was carrying out the welding operation immediately after cold working without an intermediate annealing treatment to relief the residual stresses.

Solution

Because of the susceptibility of the material to precipitating embrittling phases, it is necessary to relieve the residual stresses of cold working by an annealing treatment prior to welding.

12.3.2 Failure of Copper Cables Due to a Defective Protective Coating

Process

Copper wires (1 mm diameter) were produced from copper rods initially 8 mm in diameter by (i) wire drawing using an organic compound as a lubricant, which also provides protection against rust and (ii) stress relief annealing of drawn wires at 600°C, followed by quenching in a soluble oil maintained at a temperature below 200°C, which resulted in the formation of a transparent coating layer on the Cu wires for color protection. (iii) Coated wires were then twisted to form a cable wound into a bobbin and stored for subsequent insulation by cross-linked polyethylene (XLPE). Application of the insulator is carried out in an extrusion machine where a polyethylene melt at about 200°C is extruded onto the moving twisted Cu cable. Subsequently, the insulated cable is cooled in a water bath and finally wound into a bobbin for a curing process of the insulating material at a later date. Prior to curing, the exposed ends of the cable are tightly covered with a plastic tape. Curing consists of completely immersing the bobbin in a water bath maintained at about 90°C for about 4 hr. Its function is to induce a cross-linking reaction by means of water and a catalytic substance. After curing, the cable is removed from the water bath and allowed to cool to room temperature in air.

Component

Copper wires

Material

Commercially pure Cu
**Application**

Electric cables

**Problem**

For a period of about 10 years of production, no problems were encountered in producing the copper cables by the above procedure. However, a problem involving discoloration of the copper wires was subsequently encountered, leading to rejection by customers. According to the proponent, the extent of discoloration was observed to increase with time after curing in water. Also, the problem became less pronounced as the temperature of the curing water was decreased.

**Analysis**

1. **Visual Inspection**

   Only wires of cured cables exhibited black discoloration visible to the naked eye particularly at the inner section of the cable. Furthermore, the surface of these wires appeared to be tarnished. Smaller black spots could also be detected by means of a 15× magnifying lens. Although the smaller spots were also detected in the uncured wires, their density was found to be smaller in comparison with the cured wires.

   Known processes which can lead to black discoloration of Cu include oxidation into cupric oxide (CuO) and tarnishing in the presence of sulfur compounds. Oxidation of Cu into CuO can occur either by heating to a temperature of about 600°C or by creation of an aeration cell at ambient temperature. Although the wires were annealed at 600°C, the fact that the black discoloration was confined to localized spots could exclude high-temperature oxidation of Cu as a possible cause.

   Based upon the above observations and the information provided by the proponent, the process leading to discoloration appeared to result from localized corrosion attack limited to the inner section of the cable, which was initiated prior to curing; however, it was accelerated during curing.

2. **Material Verification**

   Commercially pure Cu used in cable applications must contain a minimum of 99.3% Cu, which was verified by chemical analysis. Therefore, deviation of composition from standard specifications could be ruled out as a contributory factor in the observed discoloration.

3. **Nature of the Spots**

   When observed at higher magnifications, the surface spots in the uncured wires exhibited a texture different from that of the cured wires. In the cured wires, the localized regions of spots were relatively flat and thin; however, they became rather irregular and thick in the cured wires, as shown in Fig. 12.8a,b. Microchemical
Figure 12.8  Secondary electron SEM images illustrating the morphology of black surface spots in the Cu wires. (a) Uncured sample. (b) Cured sample.
analysis by energy dispersive X-ray spectroscopy and electron spectroscopy for chemical analysis (ESCA) showed that the surface spots of both the uncured and cured wires consisted of CuO.

Morphologically, the region of the spot consisted of an outer layer, an intermediate layer, and an inner layer. Energy dispersive X-ray spectroscopy revealed that the only elemental constituent of the outer layer was C, indicating that it corresponded to the organic protective coating. The intermediate layer consisted of Cu and O, which was confirmed by ESCA to be CuO. However, the inner layer predominantly consisted of Cu. The crystals of CuO became well developed in the cured wire, suggesting that the oxidation reaction was promoted by the curing and/or subsequent storage.

It is possible to conclude from the preceding results that a continuous protective layer of the organic coating could not be developed by the quenching process described earlier, leaving behind localized spots of exposed unprotected Cu. A likely cause for this behavior was a defective coating of low-emulsifier content resulting in the formation of pinholes.

4. Discoloration Mechanism of Uncured Wires

Having determined that the organic coating was nonprotective, it is possible to explain the localized oxidation of Cu into CuO leading to discoloration of uncured wires in terms of the formation of aeration cell. At those spots, liquid water would be expected to form at the surface as a result of condensation of water vapor in the atmosphere (dew formation) during storage prior to curing. It is well known that dew formation results from a difference of 2.6°C in temperature between stagnant air of 85% relative humidity in contact with a cooler surface at 20°C. At a lower relative humidity, dew is formed even if the temperature difference is very high. It is recalled that oxygen-enriched regions at a metal surface behave as cathodic relative to those depleted in oxygen which act as anodic. In this case, the cathodic region corresponds to a water droplet-air interface, and the interior of the water droplet behaves in an anodic manner, leading to the establishment of an aeration cell involving the following reactions:

\[
\text{Cathode: } O_2 + 2H_2O + 4e^- = 4OH^- \\
\text{Anode: } Cu + 2OH^- = Cu(OH)_2 + 2e^- 
\]

Summation of the above two reactions results in

\[
2Cu + 2H_2O + O_2 = 2Cu(OH)_2 
\]

Copper hydroxide can readily transform into CuO by dehydration, i.e.,

\[
Cu(OH)_2 = CuO + H_2O 
\]
It is recalled from Chap. 9 that electrochemical corrosion requires the presence of an electrolyte, which could likely be CuSO$_4$ explaining the occasional presence of a S-rich compound at the surface. Sulfur could be present in many industrial atmospheres.

5. Discoloration Mechanism During Curing and Subsequent Storage

To determine any possible role of curing in the observed discoloration, it is essential to relate the characteristic corrosion properties of Cu to the composition of the curing water. The composition of the curing water (ppm) and its pH value follows:

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Composition (ppm)</th>
<th>Constituent</th>
<th>Composition (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barium</td>
<td>0.056</td>
<td>Silicon</td>
<td>119.7</td>
</tr>
<tr>
<td>Calcium</td>
<td>75.03</td>
<td>Strontium</td>
<td>0.82</td>
</tr>
<tr>
<td>Iron</td>
<td>&lt;0.03</td>
<td>Zinc</td>
<td>0.02</td>
</tr>
<tr>
<td>Potassium</td>
<td>17.45</td>
<td>Sulfate</td>
<td>289</td>
</tr>
<tr>
<td>Magnesium</td>
<td>31.3</td>
<td>Chloride</td>
<td>163</td>
</tr>
<tr>
<td>Manganese</td>
<td>&lt;0.01</td>
<td>Bicarbonate</td>
<td>115.8</td>
</tr>
<tr>
<td>Sodium</td>
<td>225.3</td>
<td>pH = 8.26</td>
<td></td>
</tr>
</tbody>
</table>

One of the most common corrosion-related failures of Cu is pitting in cold water. Hot water pitting is usually associated with soft water of (i) a pH value below 6.5, (ii) a CO$_2$ content in the range of 10–50 ppm, and (iii) a low content of chloride, sulfate, and nitride ions. Since the pH value of the water was greater than 6.5, and the sulfate and chloride contents were rather high, it is unlikely that corrosion attack by the curing water had contributed to discoloration. However, it is possible that further oxidation could occur during curing by the same mechanism described above as follows.

Moist air trapped in the cable during insulation by XLPE would be expected to flow to the inner section of the cable because of the adherence of the insulator to the outer section. As a result, oxidation of Cu into CuO could occur by establishing an aeration cell as described earlier. This could explain, at least partially, the increased extent of discoloration at the inner sections of the cable. Also, this mechanism appears to be consistent with the observation that the extent of discoloration decreased with reducing the temperature of the curing water, which could be related to decelerating the oxidation kinetics. Oxidation promoted by the presence of moist air could also continue during subsequent storage after curing.
Mode of Failure
Based upon the results of this study, it could be concluded that the mode of failure was localized corrosion attack at the surface of the wires resulting in oxidation of Cu into black CuO as a result of establishing aeration cells by dew formation.

Cause of Failure
Experimental results suggested that the most probable cause of failure was the use of a defective protective coating, leaving behind localized exposed surface regions which subsequently became susceptible to oxidation.

Solution
In view of the results of this investigation, it was recommended to consider the quality of the organic coating and its conformation to specification to ensure adequate surface protection.

12.4 Effect of Variations in Design on Service Performance
To demonstrate the effect of design on performance of engineering products, the following two case studies are presented. Each case involved the same process and application; however, as described below, the product design varied in dimensions and in material selection.

Process
Hydrogen gas (H₂) is produced by catalytic steam reformer process. In this process, methane (CH₄) saturated with steam is cracked over a catalyst, usually a nickel base, by the following reaction:

\[ \text{CH}_4(g) + 2\text{H}_2\text{O}(g) = 2\text{CO}_2(g) + 4\text{H}_2(g) \]

where \( g \) denotes gas. As schematically illustrated in Fig. 12.9, the feedstock of the process predominantly consisting of methane is fed into the reforming section consisting of a series of vertical tubes. Since the above reaction is endothermic, the heat of reaction is provided by heating the reforming section from the outside using gas burners fired vertically downward. Products of the above reaction predominantly consisting of \( \text{H}_2(g) \) are carried through a common outlet manifold pipe welded to the vertical tubes.

Product (Component)
Outlet manifold pipe of a catalytic steam reformer. To produce the desired tube length, segments of tubes are joined by welding.
Service Conditions

During service, the internal pressure can reach 30 atm (3 MPa = 0.44 ksi), and the temperature ranges from 750 to 950°C. Usually, the environment in the tubes is moderately carburizing, and therefore coke is not deposited at the inner surface of the tubes.

Materials Requirements

Since the stresses generated by the internal pressure at the service temperature are rather significant, it is customarily to calculate the dimensions of the tubes on the basis of circumferential or hoop stress and the 100,000 h creep-rupture strength. Provided the correct design parameters are adhered to during service, the tubes fail by longitudinal creep rupture after 100,000 h of service. In practice, however, a service life of 100,000 h is rarely, if ever, achieved. One possible reason is carburization attack accelerating failure by creep rupture.

Figure 12.9  A schematic illustration of the steam reformer indicating the location of failure in the manifold pipe.
Design 1

In this case, the designer had selected a heat-resistant steel casting grade A 606 for the application considering a service temperature of about 873°C and an internal pressure of 1.88 MPa (0.27 ksi). Nominal composition of the steel in wt% was Fe-32Ni-20.5Cr-1.9Mn-1.05Si-1Nb-0.105C-0.2Mo. INCONEL alloy 82 was used as filler metal in assembling the manifold pipe by welding. Its nominal composition in wt% was 67Ni + Co**-18.22Cr-2.3Nb + Ta^*-3Fe-2.5-3.5Mn-0.75Ti-0.5Si-0.5Cu-0.1C-0.015S-0.5Other. Based upon this design, the dimensions of the tube were 254 mm in internal diameter and 24 mm in wall thickness.

Frequently, weld filler metals are selected such that they produce a weld metal composition either the same or similar to that of the base metal. Since the heat-resistant steel castings have high strength levels at elevated temperatures, the solidifying weld metal becomes highly stressed requiring the use of a filler metal of sufficient ductility. For many of these steels, HASTELLY alloy W is highly recommended for use as a filler metal because of its relatively high strength at elevated temperatures combined with sufficient ductility preventing cracking during solidification. Although this filler metal is not age hardenable, its mechanical strength is adequate for many applications. As shown later, using the Nb-containing INCONEL alloy 82 could have contributed to the failure.

Design 2

Unlike the design above, the wrought INCOLOY alloy 800H tubing was selected for the application considering a service temperature of 820°C and an internal pressure of 2.94 MPa (0.43 ksi). Based upon this design, the dimensions of the tube were 102 mm in internal diameter and 20 mm in wall thickness.

As demonstrated below, the preceding variations in design have resulted in significant differences in service performance.

12.4.1 Failure of an Outlet Manifold Pipe of a Steam Reformer: Design 1

Problem

Although the expected service life of the manifold tube was 100,000 h (11.2 years), after about 22,000 h of service (2.5 years), one of the weld joints in the
outlet manifold pipe developed a leak. Subsequent radiographic tests in the field revealed the presence of cracks in the other five weld joints.

**Analysis**

1. **Visual Inspection**

Detection of cracks in all weld joints suggested the possibility of a defective weld joint. Visual inspection revealed the absence of macroscopic cracks. Also, there was no evidence for high-temperature corrosion attack.

2. **Materials Verification**

Both the pipe material and filler metal were verified by chemical analysis to be HP steel + Nb and INCONEL alloy 82, respectively, as specified in the design.

3. **Stress Analysis**

Since the ratio of internal diameter to wall thickness \( (d/t) \) of the manifold pipe \( = 254/24 = 10.58 < 15 \), the stresses generated in the pipe as a result of internal pressure can be calculated on the basis of a thin-walled cylinder. It is recalled that in this case, the principal stresses \( \sigma_1, \sigma_2, \) and \( \sigma_3 \) acting on an elemental cube (Fig. 12.10a) are

\[
\begin{align*}
\sigma_1 \text{ (circumferential or hoop stress)} &= \frac{pd}{2t} \\
\sigma_2 \text{ (longitudinal stress)} &= \frac{pd}{4t} \\
\sigma_3 &= 0
\end{align*}
\]

where \( p \) is the internal pressure, \( d \) in the internal diameter of the pipe, and \( t \) is the wall thickness. Given the internal pressure \( p = 1.88 \text{ MPa} \), \( d = 254 \text{ mm} \), and \( t = 24 \text{ mm} \), the principal circumferential stress \( \sigma_1 \) tending to produce longitudinal rupture is calculated to be \( \sigma_1 = 9.9 \text{ MPa} \) (1.44 ksi). A rough estimate of the 100,000 h creep-rupture strength was obtained by extrapolating short-term data and found to be about 30 MPa (4 ksi) at 873°C, corresponding to a factor of safety \( N = 30/9.9 = 3 \).

Combining all of the above results suggested that the process of material selection was not well coordinated and tailored to be compatible with welding as a joining technique. This was confirmed by microstructural characterization as described below.

4. **Microstructural Characterization**

Intergranular cracks were detected in the weld. It was noted that the extent of cracking increased as the weld-metal interface was approached, as shown in Fig. 12.10b. Voids were aligned along the interface. Near the weld-metal interface, the material of the pipe contained a continuous network of coarse grain...
Microchemical analysis showed that the main elemental constituents of the precipitates were Ni, Nb, and Si. Both the composition and morphology of the grain boundary phase were characteristic of a topologically close-packed Laves phase of the type Ni$_3$Nb$_2$Si. Although a binary Ni-Nb Laves phase is not thermodynamically stable, it is known to be stabilized by the Si consistent with the above results. Particles of the Laves phase were occasionally enveloped by a Cr-rich carbide phase. Since the Laves phase is relatively depleted in Cr, excess Cr rejected from the regions where the phase was precipitated could lead to precipitation of the carbide. Lamellae of Laves phase were also observed at the weld-metal interface. In general, there was no evidence for significant carburization attack consistent with the moderately carburizing environment.

It is well known that precipitation of Laves phase particularly in coarse lamellar morphology at grain boundaries leads to a considerable embrittlement and loss of rupture ductility at elevated temperatures. Typically, precipitation of Laves phase in lamellar morphology at grain boundaries occurs very rapidly at elevated temperatures. Based upon the above results, it could be concluded that precipitation of Laves phase at grain boundaries in the weld heat-affected zone was promoted by the relatively high Si content of the pipe material.
Associated loss of rupture ductility combined by the insufficient ductility of the filler metal could have promoted cracking in the weld.

**Mode of Failure**

Experimental results indicated that the mode of failure was considerable loss of rupture ductility in the weld heat-affected zone caused by precipitation of Si-stabilized Laves phase combined with the use of a filler metal lacking a sufficient ductility.

**Cause of Failure**

Most evidence pointed out that the cause of failure was two fold: (i) selection of a pipe material incompatible with welding as a joining technique and (ii) selection of a filler material lacking sufficient ductility.
Solution

Since the environment was only moderately carburizing, the selection of a pipe material containing a relatively high Si concentration was not justified. Therefore, it was recommended to replace the pipe material with a steel grade low in Si and also replace the filler metal with HASTELLOY alloy W.

12.4.2 Failure of an Outlet Manifold Tube of Primary Steam Reformer: Design 2

Problem

As in the case of design 1 described above, the manifold pipe was designed to last 100,000 h. However, in this case, after about 83,000 h of operation corresponding to 83% of the service life, longitudinal cracks were detected at various locations. This is to be compared to only 22% of service life corresponding to design 1.

Analysis

1. Visual Inspection

All cracks were observed to extend inward from the outer surface of the tube normal to the direction of the circumferential tensile stress, reaching a maximum value at the tube surface. As explained in the previous case, if the tubes used in primary reformer applications are properly designed, they should fail by longitudinal creep rupture after a design service life of about 100,000 h, suggesting that the above cracks were caused by creep rupture.

Surface scale was observed at the inner surface of the tube. However, the scale appeared to be well adhered to the surface. Similar to the previous case corresponding to design 1, there was no evidence for significant corrosion attack, suggesting that the failure was mechanically related.

2. Material Verification

Chemical analysis (wt%) confirmed that the pipe material was INCOLOY 800H as summarized below.

It is recalled that the carburization resistance of Fe-Ni-Cr alloys is a function of the (Ni/Fe + Cr) ratio and an increase in carburization resistance by a factor of 3 is obtained as this ratio is increased from 0.3 to 1. According to chemical composition stated above, the Ni/Fe + Cr ratio = 0.4. However, in the previous case of design 1, Ni/Fe + Cr ratio of HP steel = 0.5. However, because of the greater Si content of HP steel, it would be expected to have a higher resistance to carburization in comparison with INCOLOY alloy 800H.
### Table 1: Elemental Composition

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>Balance</td>
</tr>
<tr>
<td>Ni</td>
<td>30–35</td>
<td>30.09</td>
</tr>
<tr>
<td>Cr</td>
<td>19–23</td>
<td>21.29</td>
</tr>
<tr>
<td>Mn</td>
<td>1.5*</td>
<td>0.67</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.10</td>
</tr>
<tr>
<td>Cu</td>
<td>0.75*</td>
<td>0.35</td>
</tr>
<tr>
<td>Al</td>
<td>0.15–0.60</td>
<td>0.37</td>
</tr>
<tr>
<td>Ti</td>
<td>0.15–0.60</td>
<td>0.31</td>
</tr>
<tr>
<td>C</td>
<td>0.1*</td>
<td>0.07</td>
</tr>
<tr>
<td>S</td>
<td>0.015*</td>
<td>ND</td>
</tr>
</tbody>
</table>

\*Maximum.  
ND: Not detected.

3. Stress Analysis

Similar to the case of design 1 described earlier, the tube could be treated as a thin-walled cylinder since the \( \frac{d}{t} = \frac{102}{20} = 5.1 < 15 \). Given the internal pressure \( p = 2.94 \text{ MPa} \), the circumferential stress \( \sigma_1 \) is

\[
\sigma_1 = \frac{pd}{2t} = 7.5 \text{ MPa}
\]

compared to 9.9 MPa in the case of design 1. From available data, the 100,000 h creep-rupture strength of INCOLOY alloy 800H at 820°C is about 25 MPa, corresponding to a factor of safety \( N = \frac{27}{7.5} = 3.3 \). Therefore, the 100,000 h creep-rupture strength has served as a reliable base for calculating the tube dimensions as in the case of design 1.

4. Analysis of Surface Scale

The scale formed on the inner surface of the tube appeared to consist of two layers. The scale consisted of a mixture of Cr2O3 oxide and Cr2C carbide phases; however, the oxide phase appeared to be the major scale component. Energy dispersive X-ray spectroscopy showed that the outer scale layer was a carbide and the inner layer was an oxide. Evidently, the carbide phase had overgrown the oxide phase. It is recalled that in a reducing environment, carbon be generated by the following reaction:

\[
\text{CH}_4 = 2\text{H}_2 + \text{C}
\]
5. Crack Propagation Mechanism

The cracks were observed to propagate from the outer surface of the tube along grain boundaries of the material. Alongside the grain boundaries, the material contained a continuous layer of a Cr-rich carbide phase identified to be Cr$_7$C$_3$. Evidently, the moderate carburization attack had resulted in precipitation of Cr-rich carbide at grain boundaries near the surface, which could reduce the rupture life of the material.

Mode of Failure

Based upon the above analysis, it could be concluded that the tube had failed by longitudinal creep rupture along grain boundaries by the circumferential stress.

Cause of Failure

Most evidence pointed out that the cause of failure was moderate carburization attack reducing the creep-rupture life of the material.

Recommendation

It is recalled that a design service life of 100,000 h is rarely if ever achieved in tubes used in catalytic steam reformer applications. Although the main emphasis in design is the 100,000 h creep-rupture strength, this can be reduced by other factors, particularly high-temperature carburization attack. Since about 83% of the expected service life was realized, there was no apparent reason for changing the design. However, a modified version of alloy 800H (alloy 800HT) with about the same resistance to carburization attack, but of improved creep-rupture strength, could be considered as a replacement when the tubes become due for a change.

12.4.3 Comparison Between Designs 1 and 2

Although HP steel + Nb was employed in design 1, its higher creep-rupture strength in comparison with INCOLOY alloy 800H was not fully utilized. This could be attributed to its rather poor weldability. Also, the advantage of Si in improving the carburization resistance of HP steel + Nb was evidently outweighed by its detrimental effect on weldability. Furthermore, the use of a material more resistant to carburization attack may have not been justified on the basis of the moderately carburizing environment. It is therefore clear that design 2 based upon using INCOLOY alloy 800H was more effective than design 1 using HP steel + Nb, as evidenced by the realized service lives of 83,000 and 22,000 h, respectively.

The two cases above demonstrate the extreme importance of making the results of failure analysis investigations available to designers, as indicated earlier. Prior experience with various designs and corresponding service performance can
be extremely useful to designers in coordinating and tailoring the various steps in
the continuous process of design-manufacturing-performance to deliver more
reliable and durable products.

12.5 Failure of Engineering Products During
Service Because of Unanticipated Service
Conditions

12.5.1 Failure of a Contactor by Low-Temperature
Aqueous Corrosion

Process
Vinyl chloride monomer (VCM) is produced by cracking ethylene dichloride
(EDC) resulting in HCl as a by-product. Hydrochloric acid is assumed to be
consumed by reaction with ethylene and air to form EDC. After VCM is con-
verted into a powdery resin, the resulting liquid waste consisting of chlorinated
hydrocarbons is disposed by first burning it in incinerators at about 985–900°C.
Subsequently, the flue gas containing CO, CO₂, N₂, Cl₂, H₂O, and HCl is cooled
to about 270°C. Finally, the flue gas at 270°C is carried into a contactor where it is
condensed by quenching from 270°C to about 90°C using circulating water jets at
3–4 kg/m² pressure.

Component
A contactor in the form of a welded cylinder 3 m in internal diameter, 4 mm wall
thickness, and 2 m in height

Material
According to the proponent, the contactor was made of 4 mm thick plate of HAS-
TELLOY alloy C.

Application
Handling the condensates of the flue gas produced by burning the liquid waste
resulting from VCM production

Service Conditions
During service, the contactor was in contact with the condensates of CO, CO₂,
Cl₂, N₂, H₂O, and HCl, and the temperature ranged from about 270 to 90°C.
Problem
After about 3 weeks of operation, the contactor was subject to severe corrosion attack particularly in the weld heat-affected zone, which caused line shutdown for repair.

Analysis
1. Visual Inspection
On-site visual inspection of the failed contactor revealed that it had lost its structural integrity by extensive perforation near the weld, leading to complete separation. Also, considerable wall thinning occurred in regions remote from the weld. The wall thickness was reduced from 4 to 0.4 mm, i.e., by 1 order of magnitude. However, there was no evidence for the formation of corrosion deposits. It was evident from the inspection that the material came into contact with a solution in which it could readily dissolve, and that the dissolution rate was accelerated in the weld heat-affected zone.

2. Sample Collection
Because of the severe loss of structural integrity in the weld heat-affected zone, no samples were available to represent that region. Samples were removed from the nearest possible regions, as well as from regions remote from the weld.

3. Material Verification
Microchemical analysis by energy dispersive X-ray spectroscopy in a scanning electron microscope verified that the contactor was made from HASTELLOY alloy C, as summarized below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni</td>
<td>Balance</td>
<td>Balance</td>
</tr>
<tr>
<td>Co</td>
<td>2.5*</td>
<td>1.2</td>
</tr>
<tr>
<td>Cr</td>
<td>14.5–16.5</td>
<td>15.38</td>
</tr>
<tr>
<td>Mo</td>
<td>15–17</td>
<td>15.20</td>
</tr>
<tr>
<td>W</td>
<td>3–4.5</td>
<td>3.88</td>
</tr>
<tr>
<td>Fe</td>
<td>4–7</td>
<td>5.77</td>
</tr>
<tr>
<td>C</td>
<td>0.08*</td>
<td>0.06</td>
</tr>
<tr>
<td>V</td>
<td>0.35*</td>
<td>0.21</td>
</tr>
<tr>
<td>P</td>
<td>0.04*</td>
<td>ND</td>
</tr>
<tr>
<td>S</td>
<td>0.03*</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.
4. Evaluation of Mechanical Strength and Stress Analysis

Surface hardness measurements at various locations showed an average hardness of Rb 89 closely approaching the typical hardness of alloy C in the annealed condition (Rb 90). Since the ratio of internal diameter \( d \) to wall thickness \( t \) is \( d/t = 750 > 15 \), the contactor could be considered as a thin wall cylinder. Given an internal pressure of 3–4 kg/m\(^2\), the maximum tensile principal stress \( \sigma_1 \) (circumferential stress) corresponding to a wall thickness of 4 mm is given by

\[
\sigma_1 = \frac{pd}{2t} = (0.03–0.04) \times 10^{-3} \times 375 \text{ MPa} = 0.01–0.015 \text{ MPa}
\]

At a wall thickness of 0.4 mm, \( \sigma_1 \) becomes

\[
\sigma_1 = 0.1–0.15 \text{ MPa}
\]

Even if the thickness had locally reached a value as small as 0.01 mm, the corresponding stress \( \sigma_1 \) would be in the range of 4.5–6 MPa. Evidently, the stresses generated by the internal pressure were too small to be involved in the failure.

It could be concluded from the above analysis that the failure had most likely resulted from environmental effects.

5. Microstructural Characterization

Although some scattered voids were observed, there was no evidence for marked intergranular attack in the material. Voids as well as surface irregularities and intergranular separation were distinguished, suggesting that the material was subject to a severe corrosion attack.

Among the various species present in the liquid waste, which could attack alloy C is hydrochloric acid (HCL). Although most of the HCl was assumed to be consumed by reaction with ethylene before the flue gas was condensed in the contactor, it appeared likely that a significant concentration of HCl was retained. It is well known that alloy C is highly susceptible to corrosion attack by HCl in a weld heat-affected zone due to precipitation of Mo-rich carbides, and therefore cannot be used in the as-welded condition. Also, the uniform corrosion rate of alloy C in boiling HCl is rather high, e.g., 8 mm/year in boiling 10% HCl. It is, however, likely that the concentration of HCl was greater than 10%.

Mode of Failure

Most evidence pointed out that the mode of failure was corrosion attack by hydrochloric acid.

Cause of Failure

Based upon the results of the investigation, it could be concluded that the cause of failure was unanticipated environment leading to improper material selection. Although most of the HCl was assumed to be consumed prior to condensing the flue gas in the contactor, the evidence suggested otherwise.
Solution

Either adjust process parameters to ensure that most of the HCl is consumed by reaction with ethylene or replace the contactor material with another one more resistant to HCl such as the high Mo HASTELLOY alloy B2.

12.5.2 Failure of a Control Rod of an Aircraft

Background

A military aircraft encountered a malfunction after unspecified number of flying hours. It was suspected that the incident could be related to failure of a control rod of the flying control system caused by overheating.

Component

Control rod of the flying control system of an aircraft. The rod was located adjacent to fuel and hot gas lines, reaching temperature of about 530°C.

Material

The rod was made of an age hardenable Al-Cu alloy protected by a chromate conversion coating system.

Analysis

1. Visual Inspection

Visual inspection revealed that the failed rod had sustained considerable damage, as shown in Fig. 12.12a. Relatively small hot spots and a longitudinal crack were visible. The path of the crack indicated that the rod was subjected to a circumferential tensile stress. The above observations suggest that the control rod was overheated during service, leading to extensive plastic deformation terminated by rupture. However, there was no evidence that any portion of the rod was melted and re-solidified indicating that the maximum possible temperature which the rod could have reached was below its melting point, i.e., 500°C.

2. Material Verification

Microchemical analysis by energy dispersive X-ray spectroscopy in a scanning electron microscope verified that the control rod was made from Al-Cu alloy as shown below:

<table>
<thead>
<tr>
<th>Element</th>
<th>Composition (wt%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>94.22</td>
</tr>
<tr>
<td>Cu</td>
<td>5.78</td>
</tr>
</tbody>
</table>

Copyright © 2004 by Marcel Dekker, Inc.
Figure 12.12  (a) Photograph showing small hot spots observed at the failed rod surface.  
(b) Secondary electron SEM image of the inner surface of the control rod showing Cu-rich precipitates within the matrix and at the grain boundaries.
3. Analysis of the outer surface of the failed rod

Hot spots observed at the outer surface of the rod were rich in C. Moreover, the outer surface of the rod was contaminated with S while the inner surface was relatively free of S.

4. Microstructural Characterization

A typical microstructure of the failed rod material is shown in Fig. 12.12b. A distinctive feature is the presence of continuous layers of grain boundary precipitates 1–2 μm in size. These precipitates were Cu rich. These precipitates are thought to be Cu2Al which form when the alloy is overaged. In contrast, the grain structure of the sound section of the rod was relatively free of precipitates.

5. Mechanical Strength

Surface Brinell hardness number of the failed rod was 18 compared to 27 of the sound section of the rod. It is evident that the failed rod had lost some of its mechanical strength as indicated by the lower hardness.

6. Fracture Mode

Fracture surface of the failed rod exhibited numerous small dimples superimposed on separated grain facets. This type of fracture results from microvoid growth and coalescence alongside grain boundaries in alloys containing denuded zones.

Failure Mechanism

Initially, the material was heat-treated to the desired strength level. However, during service, the rod was overheated by an apparent leakage in the adjacent fuel/hot gas lines. As a result, the rod was heated to a temperature above about 200°C but below 500°C as indicated by its microstructural features. Therefore, the material was first transformed into a solid-solution phase. Upon moderate cooling, precipitation of Cu rich phase occurred more rapidly at the grain boundaries resulting in a denuded zone accelerating fracture by intergranular separation.

Solution

Take necessary measures to prevent any leakage in fuel/hot gas lines.

12.5.3 Failure of Low Tension Cables

Background

A long cable connecting a plant with a substation was causing short circuit faults. The cable was located in a trench and covered by concrete slabs. The cables failed
at a point where there is a road above the trench. Previously, a road accident above this trench had caused it to sag. The cables had been in use for at least 8 years.

**Problem**

The insulation of the cables had transverse cracks at the top surface. Infrared scanning studies had revealed that the temperature of the cable in this section was higher than the rest of the region. The cables were 31/2 core type and were rated for 650 V, although used at 440 V ac.

**Analysis**

1. **Visual Inspection**

   The outer insulation of the failed cable had transverse macro-cracks along with tiny cracks orientated along the longitudinal direction of its whole length. Surface of copper wires within the insulation core was found to be black. Optical microscopy revealed bubblelike surface, as shown in Fig. 12.13, probably caused by the degradation of the PVC insulation due to overheating.

![Figure 12.13](image)

**Figure 12.13** Optical micrograph of the top surface of failed cable insulation showing bubblelike structure and micro-cracks oriented along the length of the cable.
2. Microstructural Characterization

Cross-sectional view of the cable showed the outer insulation to be made up two layers: possibly outer PVC and an inner PTFE layer. The two layers appeared to have fused together. Small cracks were observed around the bubbles, indicating that the former had formed after the PVC insulation became brittle.

An unused cable, examined for comparison, showed two concentric layers that were not bonded. Also, an additional white thin plastic layer was observed underneath these two layers. This layer, probably polypropylene used as a filler material in the cable core, was loose and damaged in the failed cable.

3. Soil Analysis

The phases identified in the soil were quartz (SiO$_2$) 61%, calcite (CaCO$_3$) 28%, gypsum (CaSO$_4$.2H$_2$O) 7%, and halite (NaCl) 4%. The soil composition did not appear to be a major factor in the deterioration of the cable.

Cause of Failure

The results indicate that the cable insulation had suffered fatigue/stress concentrations at regions which had experienced impact at the time of the above-mentioned road accident. This had caused leakage current between the cable cores, resulting in slight heating of the cable. The prolonged heating resulted in migration of plasticizers, thus reducing PVC flexibility with time. Subsequently, the glass transition temperature of PVC was approached rendering it brittle and resulting in its cracking. Once cracked, it allowed moisture into the affected regions, causing arcing and short circuit.

Conclusion

Most evidence pointed out that the failure of cables had occurred due to the road accident.

12.5.4 Failure of a Furnace Outlet Header Flange

Process

A furnace was used to preheat isobutene prior to cracking in a downstream reactor. The isobutene temperature at the furnace inlet was 180–200°C. By means of a heating coil, the isobutene temperature at the furnace outlet was raised to 630–635°C. Although every effort was made to prevent cracking of isobutene inside the heating furnace, cracking could have occurred to a small extent. Preheated isobutene was carried from the furnace by means of an insulated pipe containing a flange welded to a T joint at the outlet. During service, the operating pressure ranged from 0.97 to 1.2 kg/cm$^2$. 

Copyright © 2004 by Marcel Dekker, Inc.
Component
The component involved in the failure was the flange welded to the T joint.

Material
Both the flange and the T joint were made of 321 stainless steel plates. The thickness of the plate used to manufacture the flange was 15.1 mm and that of the T joint was 22 mm.

Environment
The composition (in wt%) of the process fluid was: 96.97% isobutene, 0.68% propane, 0.51% isobutylene, 0.50% normal butane, 0.43% T-2-butane, 0.33% C-2 butane, 0.32% 1-butane, 0.15% propylene, 0.09% 1,3 butadiene, and 0.02% methane.

Problem
After 5 years of trouble-free service of the heating furnace, a leakage was detected at the flange. Upon removal of the insulating material, cracks were detected between the flange and T joint in the weld heat-affected zone.

Analysis
1. Visual Inspection
   The general appearance of the as-received section of the flange is shown in Fig. 12.14. Cracks followed a zigzag path, suggesting that they occurred at grain boundaries. In general, the cracks were initiated at the inner surface indicating a process irregularity. Also, the inner surface contained a black layer typical of a coke deposit.

2. Material Verification
   The material was confirmed to be 321 stainless steel by SEM/EDS analysis. The nominal and measured composition (wt%) of the steel is given below:

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>balance</td>
<td>72.39</td>
</tr>
<tr>
<td>Cr</td>
<td>17–19</td>
<td>17.54</td>
</tr>
<tr>
<td>Ni</td>
<td>9–12</td>
<td>9.05</td>
</tr>
<tr>
<td>Mn</td>
<td>2⁺</td>
<td>ND</td>
</tr>
<tr>
<td>Si</td>
<td>1⁺</td>
<td>0.36</td>
</tr>
<tr>
<td>C</td>
<td>0.08⁺</td>
<td>ND</td>
</tr>
<tr>
<td>Ti</td>
<td>5XC⁺⁺</td>
<td>0.67</td>
</tr>
</tbody>
</table>

   *Maximum.  **Minimum.  ND: Not detected.
3. Microstructural Characterization

The gross grain structure of the weld and flange material is shown in Fig. 12.15a. Elongation of the grains near the weld region indicated that the material was cold-worked. Ti-rich particles of the primary MC-type carbide were also observed. Fine striations typical of slip lines were also observed within some of the grains. However, near the weld region and somewhat away from the cracks, there was a considerable increase in the density of the slip lines and the grains became coarser. In comparison with the grain size of 321 stainless steel in the annealed condition, it appeared that the material was subjected to localized grain growth, degrading its tensile strength. Considerable grain growth suggested the possibility of localized overheating.

4. Fracture Mode

In the as-received condition, the surface contained a relatively thick layer of deposit which was Cr rich with marked concentration of Fe and trace amounts of Ni. The crack propagation was intergranular, as shown in Fig. 12.15b. A precipitate layer of carbide (7–10 mm) delineated the grain boundaries.
Figure 12.15  (a) Secondary electron SEM image showing the gross grain structure of the weld and flange; elongation of the grains near the weld is indicative of cold work. (b) Secondary electron SEM image showing intergranular cracks in the flange material as indicated by arrows.
5. Evaluation of Mechanical Strength

Measured average hardness near the surface was about HV 250 compared to HV 160 in an annealed 321 stainless steel. The hardness decreased with depth until it reached the typical value of the base metal. This demonstrated the considerable hardening effect produced by carburization.

Mode of Failure

Based upon the results presented above, it could be concluded that the mode of failure was high-temperature carburization attack leading to intergranular embrittlement. As a result of differences in thermal expansion characteristics of the carburized and uncarburized zones, high levels of stresses were generated in those regions. Moreover, the insulating effect of coke deposits requires a higher tube wall temperature to maintain the same fluid temperature, which accelerates the rate of carburization and causes a considerable localized grain growth, as observed.

Cause of Failure

Stainless steel type 321 does not have adequate resistance to high temperature carburization. Cracking of isobutene under the specified service conditions led to this attack and material failure.

Solution

Partial cracking of the isobutene should be prevented. However, if this proves to be impractical, the SS 321 should be replaced by Fe-based 800H or 314 stainless steel, which show adequate resistance to carburization.

12.5.5 Failure of a Cathodically Protected Pipe

Background

The surface of a newly constructed pipeline buried underground was protected by an insulating organic coating layer. The pipeline was cathodically protected. During initial hydraulic tests where the pipe was filled with water, a leakage was detected. Inspection showed that the pipe had developed a hole. The hole was formed at the bottom of the pipe.

Material

Type 304 stainless steel
Analysis

1. Visual Inspection

Photographs illustrating the outer and inner surfaces of the as-received section of the pipe are shown in Fig. 12.16a,b. At the outer surface (Fig. 12.16a), the insulating coating layer in the vicinity of the hole was detached, exposing the pipe material. Closer inspection showed that the pipe material appeared to have been etched in the region surrounding the hole. A similar behavior was observed at the inner surface (Fig. 12.16b); however, as indicated by the arrows, the hole appeared to be encircled by a hot spot of yellowish color. The surface of the pipe material at the outer surface was rather rough in comparison with the inner surface.

Based upon the preceding observations, it could be concluded that the pipe was subject to localized corrosion attack leading to the formation of the hole. However, it is possible that two forms of corrosion were involved, as indicated by the appearance of the outer and inner surfaces as explained later.

2. Material Verification

The pipe material was verified to be SS304 type stainless steel by SEM/EDS (wt%) analysis as shown below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Observed</th>
<th>304</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>70.78</td>
<td>Balance</td>
</tr>
<tr>
<td>Cr</td>
<td>18.72</td>
<td>19</td>
</tr>
<tr>
<td>Ni</td>
<td>9.95</td>
<td>10</td>
</tr>
<tr>
<td>Mn</td>
<td>ND</td>
<td>2*</td>
</tr>
<tr>
<td>Si</td>
<td>0.55</td>
<td>1*</td>
</tr>
<tr>
<td>C</td>
<td>ND</td>
<td>0.08*</td>
</tr>
<tr>
<td>S</td>
<td>—</td>
<td>0.03*</td>
</tr>
<tr>
<td>P</td>
<td>ND</td>
<td>0.045*</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.

It is well known that 304 stainless steel has an excellent resistance to atmospheric corrosion. However, the corrosion resistance of 316 stainless steel including the resistance to pitting is known to be the best among the standard stainless steels.

3. Microstructural Characterization

The thickness of the pipe was somewhat reduced as the hole was approached. Although the outer surface was rather irregular, the finer surface was relatively flat. This could be related to the type of corrosion attack at both surfaces, as
Figure 12.16  Photographs illustrating the general appearance of the as-received section of the pipe. (a) Outer surface. (b) Inner surface.
shown later. The grain structure at the outer surface was indistinguishable from that at the inner surface. A typical microstructure is shown in Fig. 12.17. At both surfaces, the pipe had the same grain size independent of the location. Evidently, any localized heating effect as manifested by the hot spot at the inner surface was not sufficient to cause any appreciable grain growth.

4. Mechanical Strength

Microstructural observations were reflected by the values of microhardness representative of different regions of the pipe. In the vicinity of the hole the average Vickers microhardness number was found to be HV 164, corresponding to Rb 83 and an ultimate tensile strength of 76 ksi. At regions away from the hole, the average microhardness number was HV 175 corresponding to Rb 86 and an ultimate tensile strength of 81 ksi. It is possible to relate the slightly less microhardness in the vicinity of the hole to the localized heating effect. However, the observed hardness values and corresponding ultimate tensile strengths were consistent with those typical of 304 stainless steel in the annealed condition (hardness: Rb 80, ultimate tensile strength: 82–85 ksi).

![Figure 12.17](image-url)

**Figure 12.17** Secondary electron SEM image showing typical grain structure of the material.
On the basis of the preceding observations, localized intense heating could be excluded as the cause of failure.

5. Corrosion Behavior of the Pipe

Extensive pitting was observed at the inner surface of the pipe. Pitting was particularly more pronounced in the vicinity of the hole. An Fe-rich oxide was observed in the hot spot. At some locations, the oxide layer was spalled off revealing the underlying pipe material. Generally, steel pipe buried underground can be corroded by a mechanism where regions at the bottom of the pipe can become anodic (dissolving) in moist soil, while localized corrosion at the top of the pipe can occur by stray currents associated with cathodic protection.

Based upon the results above, it is possible that pitting at the inner surface of the pipe could have resulted from water used in the hydraulic test. Also, stray current associated with cathodic protection appeared to have contributed to forming the hole. Localized heating associated with passage of the current could explain the development of hot spot and formation of an Fe-rich oxide.

Mode of Failure

Experimental results indicated that the mode of failure was of a purely environmental nature. However, two types of corrosion appeared to have contributed to forming the hole.

Cause of Failure

Most evidence pointed out that the failure was caused by a combination of pitting resulting from water used to fill the pipe during hydraulic tests, as well as stray currents associated with cathodic protection.

Conclusion

It could be concluded from the results of this study that the pipe had developed the hole as a result of accelerated localized corrosion. Most evidence pointed out that pitting resulting from water used to fill the pipe during hydraulic tests and stray currents associated with cathodic protection had contributed to forming the hole.

Solution

Based upon the results of this study, it is recommended to closely examine the quality of water used in the hydraulic tests and/or use stainless steel more resistant to pitting. Since the development of stray currents is related to the location of other buried constructions in the region, attention must be given to the particular location where the pipe was constructed.
12.5.6 Failure of a Piping System

Process

A three-branch piping system was used to transfer demineralized water at ambient temperature to various locations in a newly constructed polypropylene plant.

Application

The application involved in the failure was 2, 3, and 4 in welded piping system.

Material

All pipes were manufactured from 304 type stainless steel.

Service Conditions

Water was circulated by the piping system at ambient temperature and a pressure ranging from 60 to 100 psi.

Problem

During the initial pressurization of the system under normal operating conditions, the system was inspected and no leaks could be detected. Subsequently, the system was drained for unspecified reason leaving water residue for 2 months. Following this period, the system was put into service, and shortly after startup leaks were detected. Inspection revealed the presence of localized pinholes. Only few pinholes were detected in the welds, suggesting that the failure was not weld related.

Although the water circulated by the pipes had met the specifications calling for a chloride content less than about 30 ppm, it was indicated that the water could have been contaminated by salty water in the transfer trucks.

Analysis

1. Visual Inspection

Close examination revealed that the entire internal surface including top and bottom sections of each pipe contained localized pits. This suggested that the piping system was subjected to a localized form of corrosion. At various locations, corrosion was accelerated to the extent of penetrating the entire pipe thickness causing the formation of pinholes, as shown in Fig. 12.18a. In the vicinity of each pinhole, a surface scale well adhered to the surface was observed.

In general, the frequency of observing both the pits and pinholes was similar in the weld region and away from the weld, suggesting that the corrosion attack was not weld related. However, it was evident that the formation of the pits was initiated only at the inner surface, suggesting the possibility of a process
Figure 12.18  (a) Secondary electron SEM image of a pipe cross-section illustrating a hole. (b) Backscattered electron image showing the grain structure of inner pipe surface.
irregularity weakening the protective nature of the surface scale developed by the pipe material.

2. Material Verification

The composition (wt%) observed by SEM/EDS was in agreement with that of 304 stainless steel, as shown below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>70.77</td>
</tr>
<tr>
<td>Cr</td>
<td>18–20</td>
<td>18.43</td>
</tr>
<tr>
<td>Ni</td>
<td>8–12</td>
<td>10.30</td>
</tr>
<tr>
<td>Mn</td>
<td>2*</td>
<td>ND</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.49</td>
</tr>
<tr>
<td>C</td>
<td>0.08*</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.

3. Microstructural Characterization

A typical grain structure of the pipe material is shown in the secondary electron SEM image of Fig. 12.18b. By comparing the observed grain structure with those published for 304 stainless steel (Metals Handbook, 1973), it appeared that the pipe material was not fully recrystallized, as indicated by the relatively low density of annealing twins. For all three sections of the piping system, the average surface hardness was measured to be about Rb 78, in agreement with that typical of 304 stainless steel (Bittence, 1982). These results suggested that the pipes were stress-relief annealed. Detailed examination of the grain structure revealed the presence of fine precipitates at slip lines as well as at grain boundaries. However, due to the apparent nonuniformity in the distribution of residual cold work, the precipitates were present in selected grains. It is likely that they were of a Cr-rich carbide of the type M23C6 (Metals Handbook, 1973), suggesting that the material was sensitized during the stress-relief anneal. It is to be noted that 304 stainless steel is recommended for general-purpose uses including welded equipment (Bittence, 1982).

4. Corrosion Properties of Stainless Steels

Generally, austenitic stainless steels such as the 304 type are characterized by high resistance to oxidizing environments dependent upon the formation of a protective surface oxide film. However, reducing conditions as well as chloride ions degrade the protective nature of the oxide film causing rapid attack (Jones, 1982).

*References may be found at the end of each section.
1992; Schweitzer, 1983). Localized corrosion under conditions preventing restoration of passivity can cause rapid penetration at the point of initiation (Metals Handbook, 1977). This can occur in the presence of any acid chloride phases such as FeCl₃, other halide salts, and some sulfates leading to pitting (Metals Handbook, 1977).

5. Analysis of Surface Scale

The surface scale had a layered structure. The external layer consisted of discrete particles and an underlying scale. However, this layer was spalled off or cracked to reveal the innermost layer. Microchemical analysis showed that the composition of the outer scale layers was rather similar to the alloy composition, indicating a spinel-type oxide; however, the scale also contained small amounts of Cl and Ca. Similarly, the innermost scale layer had a composition similar to that of the alloy; however, it was relatively free of Cl and Ca.

It is evident from the results above that during service, the material of the pipe had developed a protective surface oxide film as expected. However, this film was contaminated by other elements such as Cl and Ca. As indicated earlier, both reducing conditions and/or the presence of Cl could destroy the protective oxide film accelerating the corrosion rate. It is to be noted that chlorine could aid in developing reducing conditions by forming hydrochloric acid.

6. Analysis of Pinholes and Pits

Extensive corrosion attack occurred locally at grain boundaries leading to removal of entire grains. A common feature of such pits was the presence of Cr-rich regions near the surface, which could be associated with carbide precipitates as described earlier. Under reducing conditions or in the presence of Cl, the Cr-rich regions would be expected to act as anodes in comparison with the Cr-depleted zones acting as cathodes. This appeared to be consistent with the observed intergranular attack, which could be related to the presence of Cr-rich carbides at grain boundaries.

Mode of Failure

Based upon the results of this study, the mode of failure appeared to be localized corrosion attack under conditions preventing restoration of passivity. This could be due to the development of reducing conditions aided by the presence of Cl.

Cause of Failure

Experimental evidence pointed out that the most probable cause of failure was contamination of the water with Cl. However, another contributory factor appeared to be the residual cold work accelerating the corrosion rate.
Solution

As an immediate measure, it was recommended to exert a closer control on possible sources of water contamination by Cl as well as the stress-relief annealing practice. However, for a long-term solution, the possibility of replacing the material with a grade of stainless steel containing Mo and of a lower C content such as type 316L should be considered. The presence of Mo improves the corrosion resistance under reducing conditions, as well as in the presence of Cl, and the lower C content reduces the tendency of the steel toward sensitization.

References


12.5.7 Failure of a Caustic Header

Background

Sodium hydroxide (NaOH) was produced in a petrochemical plant by means of caustic cells. Both salt (NaCl) and water were fed into the cell where the following basic reaction occurred:

\[ \text{NaCl} + \text{H}_2\text{O} \rightarrow \frac{1}{2} \text{Cl}_2 + \frac{1}{2} \text{H}_2 + \text{NaOH} \]

Caustic soda produced by the above reaction was loaded into a crude tank by means of a caustic header line consisting of 10 and 16 in diameter pipes. Some of the water fed into the cell was consumed in the above reaction and the remainder left the cell as saturated vapor along with chlorine and hydrogen gases.

After an unspecified fraction of the expected service life of the 16 in header, severe corrosion occurred particularly at two locations. At the top section, corrosion started in the flange and extended about 1.5 m into the pipe. However, at the bottom section, corrosion occurred in the flange only. The design specifications of the caustic header were as follows:

Size: 16 in
Material of the pipe and flange: API 5LB STD
Operating temperature: 110°C
Design pressure: 1035 KPa
11–13% caustic with 12–15% NaCl, 0.4–0.8% sodium sulfate and
5–50 ppm sodium hypochlorite

Flow: 2542 M3/D
Velocity: 0.74 ft/s

Analysis

1. Visual Inspection

Visual inspection revealed that the corrosion attack was localized. It was evident that the 16 in diameter header was subjected to an attack by a highly corrosive environment leading to loss of its structural integrity. Metal wastage occurred to the extent of perforation. Apparently, however, the nature of the attack at the top section was somewhat different from that at the bottom section. This was evident from the formation of massive corrosion deposit at the top section.

Restriction of the corrosion attack to pipe connection at the flanges could suggest either galvanic corrosion and/or caustic embrittlement (Metals Handbook, 1977).

2. Material Verification

The SEM/EDS analysis of the pipe and flange material exhibited similarity in their chemical compositions. The composition was consistent with that of carbon steel. The amount of pearlite in the microstructure of carbon steels is directly proportional to the C content. Secondary electron SEM images illustrating characteristic microstructural features of the pipe and flange materials are illustrated in Fig. 12.19a,b. It was observed that the amount of pearlite in the pipe material, approaches that of the flange material, suggesting similar C contents. To summarize, both spectral and microstructural data appeared to be consistent with the grades of carbon steels stated in the design specifications.

Based upon the results above, it could be concluded that a galvanic effect was not a contributory factor in the corrosion attack.

Also, the materials of the 10 in diameter pipe and its flange were found to have similar chemical composition.

3. Analysis of the Corrosion Deposit

Phase analysis of the corrosion deposit removed from the top section of the 16 in diameter header revealed that it consisted of a mixture of iron oxides (FeO(OH) and Fe₂O₃) and chlorides (FeCl₂ and FeCl₃). Analysis of several samples of the deposit suggested that chlorides were mostly concentrated in the outer sections of the deposit. It is evident from the above results that the top section of the header
was attacked by chlorine. Initially, the material of the header developed a surface oxide scale which was subsequently attacked by chlorine. As a result, the surface oxide was converted into a porous nonprotective scale which accelerated the oxidation attack. By analogy to caustic embrittlement (Metals Handbook, 1977),

Figure 12.19 Secondary electron SEM images illustrating gross microstructural features of the 16-in. header parallel to the surface. (a) Pipe. (b) Flange. (c) Pipe along transverse direction.
there is also a possibility that reactions due to accumulation of concentrated solutions at the pipe connection (flange) could lead to the generation of chlorine. This may explain why the corrosion attack started at the flange.

4. Microstructural Characterization

Typical microstructural features of the 16 in diameter pipe and flange are illustrated in the secondary electron SEM image of Fig. 12.19c. Elongated pearlite, particularly near the inner surface of the pipe, indicated a cold-worked layer possibly resulting from improper installation of the header. Cold working is known to accelerate the corrosion rate (Metals Handbook, 1977). In contrast, the flange was free of cold working. It is possible that the cold working played an important role in extending the corrosion attack from the flange into the pipe as observed. Similar results were obtained for the 10 in-diameter pipe and flange.

5. Analysis of Surface Scale

Analysis of the surface scale formed at the top section of the 16 in header revealed the presence of iron chloride. However, there was no evidence for the presence of chlorine-rich phases in the surface scale of the flange. Possibly, this behavior could be related to the severity of the attack initiated at the flange and leading to spallation of chlorine-rich phases.

Internal surface pits were observed in the flange of the 16 in header. Predominantly, the pits consisted of iron-rich oxides. However, it is possible, that these pits contained chlorine-rich phases which were washed away. When observed along the transverse direction, these pits appeared to be associated with intergranular attack.

Surface scale formed on the 10 in header consisted essentially of iron-rich oxides. In contrast, to the 16 in header, there was no evidence for the formation of chlorine-rich phases.

Mode of Failure

Based upon the results of this study, the mode of failure of the 16 in header appeared to be localized attack by chlorine and/or hydrochloric acid resulting from accumulation of concentrated solutions at pipe connections (flanges).

Probable Cause of Failure

Most evidence pointed out that there could be more than one cause for the observed failure. First, misalignment of the pipe at the flange leading to a cold worked layer at the surface could accelerate the corrosion attack particularly in the absence or insufficiency of inhibitors (Metals Handbook, 1977). Secondly, condensation of HCl could significantly contribute to the failure since carbon steels are not resistant to it.
Solution

As a short-term solution to the problem, it is recommended to consider proper alignment of the pipe and avoidance of cold work as well as the addition of inhibitors. However, for a long-term solution, it was recommended to consider replacing the header material by one of the higher grade Ni-based alloys such as the Ni-Cr-Mo HASTELLOY alloy C-276.

Reference


12.5.8 Failure of Boiler Tubes

Background

Two steam boilers were used as power plants at a production facility. Each boiler used natural gas as the source of heat. One boiler was always in operation and the other was on a stand-by basis. Boiler operation had an expected service life of 15–20 years. However, after 9 years in service, a cracking problem was encountered with the boiler tubes as described below.

Hot spots were detected in the economizer section after 1 year and were weld repaired. After 3 years, a leak was detected in one of the tubes and was also repaired by welding. An overhaul of the system was conducted after 4 years of operation. Immediately afterward, one of the tubes leaked and was repaired by welding. Two overhaul operations were conducted after 5 and 7 years. Again, a leak in one of the tubes was detected a few months later. After inspection, three tubes were replaced.

The capacity of the boiler was 86,000 kg/h with a design pressure of 75 kg/cm². However, the operating pressure is 65 kg/cm². At the drum heater inlet, the feed water temperature is 126°C and the steam temperature at the superheater outlet is 460°C. Each tube is made of A178 grade C low carbon steel with an outer diameter of 76.2 mm and a thickness of 4 mm.

Analysis

1. Visual Inspection

Visual inspection revealed the absence of large macroscopic cracks in both samples. Also, their surfaces appeared to be free of blisters or voids. However, slight scaling was observed suggesting high temperature oxidation. It is to be noted that burning of natural gas results in a purely oxidizing environment.
As illustrated in Fig. 12.20a, the tube at the top bend section near the steam drum was patch welded to repair a puncture crack. However, the straight midlength section of the tube (Fig. 12.20b) was free of such cracks.

Both the above observations suggested that the mode of failure was related to an elevated temperature phenomenon.

2. Material Verification

Microchemical analysis confirmed the tube material to be carbon steel. The chemical composition (wt%) of the tube material as determined from wavelength dispersive X-ray spectroscopy in comparison with the nominal composition of A178 grade C carbon steel is given below.

![Diagram of top bend and straight midlength section of a water wall tube]

**Figure 12.20** Continued
<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>Balance</td>
</tr>
<tr>
<td>Mn</td>
<td>0.8*</td>
<td>0.75</td>
</tr>
<tr>
<td>C</td>
<td>0.35*</td>
<td>0.44</td>
</tr>
<tr>
<td>Si</td>
<td>—</td>
<td>0.11</td>
</tr>
<tr>
<td>P</td>
<td>0.05*</td>
<td>0.01</td>
</tr>
<tr>
<td>S</td>
<td>0.06*</td>
<td>0.01</td>
</tr>
</tbody>
</table>

*Maximum.

Typically, carbon steels such as the A178 grade C are recommended for boiler tube applications (Metals Handbook, 1977). Therefore, the proper material was selected for the application.
3. Properties of Surface Scale

The scale developed by the internal tube surface at the top bend section was identified as Fe$_3$O$_4$. Considerable scale spallation was observed at this region. For the straight midlength section of the tube, the internal surface scale was found to also consist of Fe$_3$O$_4$. However, the extent of scale spallation appeared to be less in comparison with the bend section which could be related to the lower exposure temperature.

For the bend section, the scale thickness was about 15 $\mu$m and the maximum depth of oxidation attack was about 20 $\mu$m. The straight midlength section exhibited a similar behavior. However, the scale layer was relatively thicker (about 30 $\mu$m thick) and there was no evidence for internal penetration.

Loss on ignition value for a scale specimen was found to be $\leq$0.001 (specimen heated for 2 h at 1000°C). Such a value corresponds to no loss on ignition indicating that the scale was free of C and H$_2$O. Analysis of the scale by ICP-AES showed the major elemental constituent to be Fe with trace amounts of other elements ($<0.1\%$). As can be seen, these results are consistent with those of X-ray diffraction and SEM-EDS.

Both the nature and extent of surface scale as determined from the above results suggested that it was not a contributory factor in the failure.

4. Tube Thickness

The initial tube wall thickness was 4 mm. The average wall thickness at the bend section was only 3.8 mm. This corresponds to a 5% reduction in wall thickness. However, the wall thickness at the straight midlength section was about 4 mm.

Reduction in wall thickness at the top bend section could result from swelling due to localized overheating which reduced the yield strength of the material (Metals Handbook, 1975). Apparently, however, the extent of overheating at the top bend section was greater in comparison with the straight midlength section. A similar result was derived from microstructural observations as described below.

5. Microstructure of Tube Material

Microstructure of the outer and inner surfaces of the top bend section tube exhibited pearlite colonies (mixture of ferrite and cementite) dispersed in a ferrite matrix, as shown in Fig. 12.20c. A distinctive microstructural feature, however, was the presence of precipitate particles exhibiting nodular morphology. Both the morphology and contrast of these precipitates are typical of graphite resulting from decomposition of pearlite (Metals Handbook, 1975). The predominant constituent of the tube material was ferrite ($\alpha$-Fe: body-centered cubic; $a = 0.2866$ nm). Generally, there was no evidence for a cold-worked structure. The pearlite exhibited a typical lamellar morphology. Nodules of graphite were observed to form networks delineating the ferrite grain boundaries. However,
the extent of graphitization appeared to vary from one location to another, suggesting nonuniform distribution in temperature.

Graphite (carbon) can result from decomposition of pearlite in carbon steels by a process known as graphitization (Metals Handbook, 1975). Typically, graphitization occurs during prolonged exposure to temperatures below about 550°C. When the graphite particles form a continuous network, they can embrittle the steel (2). It is to be noted that graphitization can occur during welding of carbon steels. However, the observation of graphite particles at the straight midlength section of the tube indicated that local graphitization of the bend section resulted at least partially from prolonged exposure at temperatures below about 550°C. In addition to graphitization, there was also evidence for spheroidization of pearlite resulting in inhomogeneous microstructure, as described below.

Pearlite can also decompose by another process known as spheroidization. In this case, the resulting carbon is present in the form of spheroidal carbide particles. However, spheroidization which weakens the steel occurs during exposure to temperatures above about 550°C. An evidence for spheroidization was found at the top bend section of the tube.

Two important conclusions could be derived from the above results. First, during service, the maximum operating temperature exceeded the stated 470°C as evidenced from spheroidization which occurs above about 550°C. Secondly, the temperature was nonuniformly distributed as evidenced from graphitization which occurs below about 550°C.

Only graphitization appeared to occur at the straight midlength section of the tube indicating that it was exposed to a lower temperature in comparison with the top bend section. However, the extent of graphitization was found to be greater at the internal surface. There was a greater pearlite density at the outer surface. It could be concluded from these observations that the straight midlength section was exposed for long periods at a temperature below about 550°C.

Macroetching in HCl solution revealed the absence of weld seams at the top bend section.

6. Mechanical Strength

Mechanical strength was evaluated from surface hardness measurements. For the top bend section of the tube, the hardness was found to be Rockwell B 89 to 91. Also, there was no significant difference between the hardness of the inner and outer surfaces. In contrast, the hardness of the straight midlength section was found to be Rockwell B 76 to 77. It is possible that the higher hardness value at the top bend section was due to the presence of residual stresses. All of the above hardness values were consistent with the absence of low-temperature transformation products such as bainite and martensite, which exhibit Rockwell C hardness.
It is evident from the above hardness values that although the tube was overheated, the maximum temperature reached remained to be below the eutectoid temperature, i.e., 727°C.

Mode of Failure

Although neither cracked nor ruptured samples were provided for examination, the results presented above suggested that the failure was related to overheating.

Two types of failures related to overheating can be distinguished. Rapid overheating results in tensile failure because of the considerable decrease in yield strength of carbon steels with increasing temperature. In contrast, gradual overheating promotes formation of voids at grain boundaries, which ultimately leads to grain separation, i.e., stress rupture.

It is evident from the results of the investigation that the tube was overheated. Since no grain boundary voids could be detected in the samples provided, it appeared that failure of the tube at the bend section resulted from rapid overheating, i.e., tensile-type failure. However, since hot spots were sometimes detected particularly in the economizer section, it is possible that the mode of failure varied from one component of the boiler to another.

Cause of Failure

In general, overheating of boiler tubes can result from restriction of flow within a heated tube or from localized hot spots in a tube wall (Metals Handbook, 1975). Heat extracted by flowing water or steam results in cooling of the tube metal; however, flow irregularities can result in overheating as described below.

Slight flow restriction causes a small degree of overheating which promotes failure by stress-rupture. Among the factors which contribute to slight flow restrictions are local imbalance of flow in tubes joined by a common header, localized deposits near a tube inlet, and local variations in the inner tube diameter resulting from various installation techniques.

Sudden loss of circulation which may result from feed-pump failure can cause rapid overheating. This can occur during various stages of operation such as startup, full power operation, or rapid fluctuation in steam demand (Metals Handbook, 1975).

To summarize, available evidence suggested that the cause of failure was overheating resulting from operational irregularities, particularly poor circulation.

Recommendations

It is essential to avoid excessive overheating in order to maintain a material strength commensurate with the working pressure. To avoid excessive overheating, it is recommended to consider a better control of furnace temperature. A lower furnace temperature can significantly contribute to combating the problem. It is also of equal importance to maintain a better boiler circulation practice.
References


12.6 Failure of Engineering Products During Service Because of Improper Material Selection

12.6.1 Failure of a Ti Floater Gage by Hydrogen Damage

Process

Methyl Tertiary Butyl Ether (MTBE) is used as a gasoline additive to replace tetraethyl lead in automobile fuels. It is produced in a plant consisting of three units: (i) deisobutanizer/butaner unit, (ii) dehydrogenation unit, and (iii) MTBE unit. Feedstock for the production process consists of liquid butanes mixed with hydrogen gas and fed into the deisobutanizer/butaner unit.

Component

Magnetic level floater assembled by welding. A schematic of this component is shown in Fig. 12.21.

Material

Commercially pure Ti sheet about 0.35 mm thick was used to manufacture the floater.

Application

The floater was used in the deisobutanizer/butaner unit as a gauge to monitor the liquid level both at the top and bottom of the unit.

Service Conditions

The floaters were operated at ambient temperature. Compositions (mol%) of the inlet and outlet feedstock as provided by the proponent are summarized below.

<table>
<thead>
<tr>
<th>Inlet liquid phase</th>
<th>Outlet liquid phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$-Butane: 0.38</td>
<td>$H_2$: 0.03</td>
</tr>
<tr>
<td>$n$-Butane: 99.58</td>
<td>$i$-Butane: 0.38</td>
</tr>
<tr>
<td>Neo Pentane: 0.04</td>
<td>$n$-Butane: 99.55</td>
</tr>
<tr>
<td>Sulfur: &lt;1 ppm</td>
<td>Neo Putane: 0.04</td>
</tr>
<tr>
<td></td>
<td>Sulfur: &lt;1 ppm</td>
</tr>
</tbody>
</table>
After an unspecified period of operation, the Ti gages failed to provide the correct liquid level in the deisobutanizer/butaner unit. Field inspection of the floaters removed from the unit indicated that the floaters were subject to corrosion attack; however, the attack was

<table>
<thead>
<tr>
<th>Inlet gas phase</th>
<th>Outlet gas phase</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{H}_2$: About 100</td>
<td>$\text{H}_2$: 100</td>
<td></td>
</tr>
<tr>
<td>$\text{CO}$: $&lt; 1$ ppm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{CO}_2$: $&lt; 1$ ppm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_4$: $&lt; 1$ ppm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
more severe in the bottom floaters where the environment consisted of the liquid phase.

**Analysis**

1. **Visual Inspection**

Visual inspection revealed that the attack was heavily concentrated at the weld-metal interface. Surface flaws, however, were observed at scattered locations within the base metal. Also, complete penetration or puncturing of the base metal near the weld was noted at a few locations. In general, there was no evidence for overall metal thinning ruling out the possibility of uniform or overall corrosion attack.

   It is recalled that Ti is resistant to most media except those containing halogens such as chlorides and fluorides as well as fuming nitric acid containing less than 2% water or more than 6% nitrogen dioxide, and liquid oxygen on impact. Also, stress corrosion cracking may occur if chloride salts come into contact with stressed parts, which are subsequently exposed to elevated temperatures. All of these possibilities, however, could be ruled out on the basis of the compositional analysis listed earlier.

   Hydrogen absorption by Ti and its alloys, however, is well known to have a detrimental effect on their mechanical strength; however, the effect is dependent upon the microstructural constituents. Below a temperature of 882°C, pure Ti undergoes a phase transition resulting in a change in its crystal structure from body-centered cubic (bcc β phase) into a hexagonal close-packed (hcp α phase). It is known that the α phase degrades the resistance of Ti and its alloys to hydrogen damage and stress-corrosion cracking. Since the outlet liquid phase in contact with the mostly affected bottom floaters contained 0.03% H₂, the possibility of some type of hydrogen damage could not be ruled out.

2. **Material Verification**

Microchemical analysis by SEM/EDS verified that the floater was made of commercially pure Ti. Microstructural characterization combined with X-ray diffraction analysis showed that the floater material predominantly consisted of platelets of α-Ti (hexagonal structure; \( a = 0.295 \text{ nm} \) and \( c = 0.468 \text{ nm} \)). The α phase was in the form of platelets assuming a Widmanstätten-type morphology revealed by etching in 10 ml HF + 5 ml HNO₃ + 85 ml H₂O. This observation suggested that hydrogen damage was a rather strong possibility.

3. **Characterization of the Attack Morphology**

When examined in the as-received condition, the base metal near the weld was found to contain transgranular cracks. Transgranular cracks were also observed around punctured regions near the weld, as shown in Fig. 12.22a.
In the as-polished condition, the weld was observed to contain a high density of voids, as shown in Fig. 12.22b. Blisters or “holidays” were observed in the base metal near the weld. Characteristic Widmanstatten-type morphology of $\alpha$ phase in the weld was revealed by etching where voids and cracks can be seen. It is recalled that these morphological features are characteristic of hydrogen blistering.

The observations above suggested that the floater material was subject to hydrogen damage including blistering and cracking. It is recalled that hydrogen
can be retained by a molten metal during welding operations, or it can be absorbed by a solid metal from the environment. If the problem was limited to hydrogen retention by the weld, the behavior of the top and bottom floaters would be expected to be similar. However, as pointed out earlier, the bottom floaters in contact with the liquid phase were more severely attacked, indicating that another factor promoting hydrogen absorption in the solid state must have been involved.

It is known that sulfur acts as a catalyst promoting hydrogen absorption in the solid state by deterring the formation of harmless bubbles of hydrogen at the surface of a metal. Although the inlet and outlet liquid phases contained a trace amount of S, it is possible that it could promote hydrogen absorption by the bottom floaters. In contrast, the absence of S in the gaseous phase could have facilitated the formation of harmless bubbles of hydrogen at the surface of the top floaters.

Mode of Failure
Most evidence pointed out that the mode of failure of the bottom floaters was hydrogen damage, including blistering and cracking.

Cause of Failure
Based upon the results of the investigation, it could be concluded that the most probable cause of failure was the presence of floater material in the form of \( \alpha \)-Ti phase susceptible to hydrogen damage.

Solution
Structural applications of Ti requiring resistance to hydrogen damage call for the use of Ti alloys consisting of the bcc \( \beta \) phase, which can be stabilized by the addition of elements such as Mo, V, Nb, and Ta. Therefore, it was recommended to replace the floater material with a Ti alloy consisting of the \( \beta \) phase.

12.6.2 Failure of Impeller Vane of an Exhaust Blower by Corrosion Fatigue

Process
Polyvinyl chloride (PVC) is produced by polymerization of vinyl chloride monomer (VCM) in polyreactors. Polymerization results in a wet slurry (powder) of VCM, which is transferred into three trains for drying in a centrifuge. One of the trains differs in design from the other two in that the blowers are used to transfer the wet slurry into the scrubbing unit (dryer). Blowers of the other two trains, however, located downstream are used to transfer only air.
Component
Impeller vane of a blower used to transfer wet slurry of VCM into the drying unit.

Material
According to design specifications, the impeller vanes were made of $\frac{1}{2}$ in thick plate of 304 stainless steel.

Service Conditions
According to the proponent, the rotational speed of the blower was 1200 rpm and the temperature was about 100°C.

Problem
After about 6 months of continuous service, the impeller vanes of the blower used to transfer wet slurry of VCM were fractured into several pieces. However, no problem was encountered in the other blowers used to transfer air.

Analysis
1. Visual Inspection

It was evident from the visual inspection that the impeller vanes were subject to severe conditions leading to complete loss of structural integrity, as illustrated in the Fig. 12.23a. As can be seen, the surface of the failed vane contained many cracks and appeared to have been subjected to severe corrosion attack. Also, a corrosion deposit was observed at the inlet suction piping of the blower. Some of the fractured pieces of the vane, however, were unavailable, precluding a reconstruction of the propagation direction of the cracks observed in Fig. 12.23a.

2. Material Verification

Microchemical analysis (wt%) confirmed that the vane material was 304 stainless according to design specifications as summarized below:

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>71.25</td>
</tr>
<tr>
<td>Cr</td>
<td>18–22</td>
<td>20.36</td>
</tr>
<tr>
<td>Ni</td>
<td>8–12</td>
<td>8.06</td>
</tr>
<tr>
<td>Mn</td>
<td>2*</td>
<td>1.01</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.33</td>
</tr>
<tr>
<td>C</td>
<td>0.08*</td>
<td>0.06</td>
</tr>
</tbody>
</table>

*Maximum

It is recalled that austenitic stainless steels such as the 304-type exhibit their best performance in oxidizing environments because their corrosion...
Figure 12.23  (a) Photograph illustrating the general appearance of the as-received piece containing a high density of cracks indicated by arrows. Secondary electron SEM images illustrating (b) a transgranular crack while arrows indicate slip lines, and (c) typical morphology of fracture surface after descaling in hydrochloric acid.
resistance is dependent upon a surface oxide film. However, this protective film
can be destroyed, leading to an accelerated corrosion attack in reducing environ-
ments as well as in the presence of chloride ions. Also, chloride ions combined
with high tensile stresses are known to cause stress corrosion cracking of austen-
itic stainless steels, and under alternating stresses they become subject to cor-
rosion fatigue.

3. Analysis of Corrosion Products

Analysis of corrosion deposit removed from the inlet suction piping of the blower
revealed that it predominantly consisted of iron chloride. Scale at the surface of
the vane was found to consist of an inner layer of iron oxide containing Cr, which
could be of the type \((\text{Fe,Cr})_2\text{O}_3\), and an outer layer of iron chloride phase.
It could be concluded from the above observations that the initial protective
oxide was attacked by chloride ions during service, converting it into a nonpro-
tective scale. This could accelerate the corrosion attack.

5. Cracking Mode

Cracks observed at the surface of the vane were found to be transgranular as
dicating that plastic deformation was involved in crack nucleation.

6. Fracture Mode

Since the available fracture surfaces were covered with corrosion products, dec-
aling was required to reveal the fracture mode. A characteristic morphology of
the fracture surface is shown in Fig. 12.23c. Striations characteristic of crack
propagation by a fatigue mechanism can be seen. It is recalled that a similar mor-
phology can result from stress corrosion cracking. However, since the part was
rotating, it is more likely that the striations observed in Fig. 12.23c were produced
by fatigue.

Mode of Failure

Since no problem was encountered with the air blowers subject to the same stress
conditions, the results of this investigation indicated that the mode of failure of
the blower handling the wet slurry was corrosion attack by chloride ions, redu-
cing the fatigue life of the vane.

Cause of Failure

It could be concluded that the cause of failure was the selection of a material with
inadequate resistance to corrosion attack by chloride ions.
Recommendation

It was recommended to either coat the vane material with one of the high grade Ni-based alloys more resistant to attack by chloride ions or replace it with one of those alloys.

12.6.3 Failure of a Gas Turbine Seal Ring by Thermal Fatigue

Process

In a gas turbine engine, a seal ring is installed for each turbine stage in such a manner that the tips of the blades rotate in close proximity to the seal ring. It is the primary function of the seal ring to minimize the loss in pressure of the gas stream between the blades and the engine casing and thus maintain engine efficiency. Seal rings are fabricated into hoops from plate and bar products. Typically, they are designed as parts of brazed honeycomb assembly where abradable materials are brazed to the seal rings.

Component

A honeycomb brazed assembly of a gas turbine seal ring

Materials

A seal ring made of HASTELLOY alloy S brazed to a honeycomb made of HASTELLOY alloy X

Service Conditions

During engine operation, the seal ring assembly was subject to rapid temperature cycling with a maximum temperature of about 760°C.

Problem

During regular maintenance, cracks were detected in the seal ring.

Analysis

1. Visual Inspection

Figure 12.24a illustrates a section of the honeycomb brazed assembly of the failed seal ring. Cracks in the seal ring are shown in Fig. 12.24b after removal of the honeycomb. It is observed that the cracks extended inward from the surface.
2. Material Verification

Microchemical analysis verified the materials of the seal ring and honeycomb. The nominal compositions (wt%) of these alloys are given below:

<table>
<thead>
<tr>
<th>Element</th>
<th>Hastelloy X</th>
<th>Hastelloy S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni</td>
<td>Balance</td>
<td>Balance</td>
</tr>
<tr>
<td>Cr</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>Fe</td>
<td>18</td>
<td>3*</td>
</tr>
<tr>
<td>Mo</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Co</td>
<td>1.5</td>
<td>2*</td>
</tr>
</tbody>
</table>

Figure 12.24 (a) A photograph illustrating a section of a gas turbine seal ring assembly. (b) Appearance of seal ring after removal of honeycomb structure. Arrows indicate the cracks. (c) Optical micrograph showing the transgranular crack extending from the surface. (d) Secondary electron image of the fracture surface showing typical fatigue striations.
<table>
<thead>
<tr>
<th>Element</th>
<th>Hastelloy X</th>
<th>Hastelloy S</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>0.6</td>
<td>1*</td>
</tr>
<tr>
<td>C</td>
<td>0.1</td>
<td>0.02*</td>
</tr>
<tr>
<td>Mn</td>
<td>1*</td>
<td>0.5</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.4</td>
</tr>
<tr>
<td>B</td>
<td>0.008*</td>
<td>0.015*</td>
</tr>
<tr>
<td>Al</td>
<td>—</td>
<td>0.25</td>
</tr>
<tr>
<td>La</td>
<td>—</td>
<td>0.02</td>
</tr>
</tbody>
</table>

HASTELLOY alloy S is known for its resistance to thermal fatigue associated with its relatively low coefficient of thermal expansion. Also, both alloys are capable of being brazed to each other because of the absence of significant amounts of Al and Ti. Oxidation of these elements is rather difficult to control, and in most furnace brazing operations their oxides inhibit braze alloy wetting. Therefore, it is apparent that proper materials were selected for the application. However, as explained later, another factor related to the use of alloy X as a honeycomb material could be the direct cause of failure.

3. Cracking Mode

Figure 12.24 is a light optical micrograph derived in the as-polished condition to illustrate a crack. Fine striations in the vicinity of the crack near the surface appeared to be slip line, suggesting that plastic deformation was involved in nucleating the crack. It is observed that the crack had propagated by a transgranular mechanism.

4. Mechanism of Crack Propagation

To characterize the mechanism of crack propagation, an existing crack was made to propagate until complete separation had occurred using a fixture. Because the surfaces of the initially present crack were covered with an oxide layer masking the fracture surface morphology, descaling was used to remove the oxide. Figure 12.24d illustrates the surface morphology of the crack after descaling. Fine striations characteristic of crack propagation by a fatigue mechanism can be seen.

5. Estimation of the Number of Cycles to Failure

The number of cycles preceding the development of fatigue cracks was estimated from the spacing of striations observed at the fracture surface. Measurements conducted at five locations at crack surface yielded an average number of N cycles of 3781.

Mode of Failure

Based upon the results of this investigation, it could be concluded that the seal ring had failed by thermal fatigue.
Cause of Failure

Most evidence pointed out that the cause of failure was the incompatibility between the materials of the seal ring and honeycomb. Diffusional transport of Fe from the honeycomb material into the surface of the seal ring had resulted in precipitation of embrittling $\mu$ phase. Since the embrittlement associated with precipitation of $\mu$ phase is particularly more pronounced at lower temperatures, cracks at the surface could have developed at the lower temperature of the cycle and subsequently propagated by a fatigue mechanism.

Solution

Since the seal ring material used in the application is known for its resistance to thermal fatigue, it was recommended to replace the honeycomb material with a low Fe alloy to avoid the possibility of precipitating $\mu$ phase.

12.6.4 Failure of Hydrocracker Heat Exchanger Tubes by Stress Corrosion Cracking

Process

Hydrocracking of waxy distillates resulting from fractionation of crude oil is conducted by a two-stage process in the unit schematically illustrated in Fig. 12.25. As can be seen, the unit consists of two reactors and seven heat exchangers. In the first stage of the process which is exothermic, nitrogen is removed. Hydrocracking, which consumes significant amounts of hydrogen, takes place in the second stage. Recycled gas flowing in the heat exchanger tubes was stated to consist of liquid petroleum gas, naphtha, kerosene, about 5.6% of uncracked material, about 400,000 ppm of $H_2S$, significant but unspecified amount of $NH_3$, and about 40–50 ppm of fluoride. Due to the proprietary nature of the process, no further details could be given.

Component

Heat exchanger tubes 1 in. in outer diameter and $\frac{1}{4}$ in. in wall thickness.

Material

According to design specifications, the heat exchanger tubes were made of 321 stainless steel.

Service Conditions

During operation, the tube temperature reached about 400°C and the internal pressure was 120 barg (1 barg = 105 N/m² = 0.1 MPa) in comparison with the design values of 435°C and 158 barg.
Figure 12.25 A schematic illustration of the hydrocracking process.
Problem

To replace the catalyst used in the process, the unit was shutdown and then restarted. During startup, the temperature was about 100°C. After about 4 months, the unit was shut down again to remove H₂S by the addition of soda and sodium carbonate, and then restarted. However, after about 48 h of startup, a leakage was detected in the heat exchanger tubes. Nondestructive evaluation using a dye-penetrant test revealed cracks only at the bent section of the tubes. According to the proponent, the problem was limited to tubes put into service in the cold-worked state.

Analysis

1. Visual Inspection

On-site inspection revealed that the cracks were initiated at the inner surface of the tubes, suggesting a process irregularity. Some of the cracks had extended through the entire thickness. Also, the cracking was detected only at the bent sections of tubes put into service in the cold-worked state. No cracking was developed in tubes which were stress-relief annealed prior to service, suggesting that the residual stresses of cold work could have contributed to the failure.

A photograph illustrating cracks at the outer surface of a tube is shown in Fig. 12.26a. A crack inclined at 45° to the tube axis is shown in Fig. 12.26b. Since the maximum principal tensile stress generated by the internal pressure (circumferential stress) was normal to the tube axis, the above observation indicated that the

Figure 12.26  (a) Photograph illustrating cracks at the outer surface of the tube. (b) Cracks can be seen inclined at an angle of 45° to the axis of the tube.
cracks were propagated under the influence of a shear stress corresponding to mode II cracking. Corrosion deposits were detected at the inner surface of the tubes.

The above observations suggested that the failure could have resulted from the combined effects of mechanical stresses and damaging environments.

2. Material Verification and Requirements

Microchemical analysis (wt%) confirmed that the tube was made of 321 stainless steel as summarized below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>Balance</td>
</tr>
<tr>
<td>Cr</td>
<td>17–19</td>
<td>17.89</td>
</tr>
<tr>
<td>Ni</td>
<td>9–12</td>
<td>10.02</td>
</tr>
<tr>
<td>Mn</td>
<td>2*</td>
<td>1.02</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.36</td>
</tr>
<tr>
<td>C</td>
<td>0.08*</td>
<td>0.07</td>
</tr>
<tr>
<td>Ti</td>
<td>5C**</td>
<td>0.52</td>
</tr>
</tbody>
</table>

*Maximum
**Minimum

In general, stainless steels used in oil refinery applications are selected on the basis of their corrosion resistance (Metals Handbook, 1961). Most of the corrosion problems encountered in oil refineries are known to be caused by inorganic substances such as water, H₂S, CO₂, sulfuric acid, and sodium chloride (Fontana and Green, 1978). Water is usually present in crude oils and complete removal is difficult. It acts as an electrolyte causing corrosion and also tends to hydrolyze other substances, particularly chlorides forming acidic environments. Presence of H₂S can lead to formation of sulfide scales at elevated temperatures. However, aqueous condensates during downtime in refineries could promote the conversion of sulfide scale into polythionic acid causing stress-corrosion cracking of stressed components (Logan, 1968; Sridhar et al., 1989; Crum et al., 1988). Although H₂S can be removed by reaction with sodium hydroxide, lime or sodium carbonate, complete removal is rather difficult particularly at the early stages of the process (Fontana and Green, 1978). In general, Fe-based alloys particularly those containing small additions of Ni, have adequate resistance to H₂S (Ref. 2).

3. Analysis of Corrosion Deposit

The corrosion deposit formed in the inner surface of the tube consisted of two layers. X-ray diffraction and SEM/EDS revealed that the outer layer was Fe₂S while the inner was a Cr-rich sulfide. This shows that during service at relatively high temperatures, the tube material had developed a surface scale consisting predominantly of Fe- and Cr-rich sulfides, indicating a highly sulfidizing
environment. Evidently, the material was incapable of developing a protective scale due to very low oxygen activity.

4. Grain Structure and Mechanical Strength of the Tube Material
The tube exhibited a grain structure typical of a cold-worked and recrystallized material. The recrystallization of the used tube material is thought to have occurred during service since the unused tube exhibited a coarser grain structure. Also, the average surface hardness of the unused tube was greater than that of used tubes.

5. Fracture Mode
Cracking initiated at the inner tube surface. The cracks were predominantly transgranular and extensively branched. The cracks also contained an Fe-rich sulfide at various locations. All evidence indicates that the tubes were fractured transgranularly as a result of stress corrosion cracking.

Mode of Failure
Sulfidation attack at relatively high temperatures during service results in the formation of sulfide phases on the inner surface of the tubes. Water vapor condensates on the inner surface of the tubes (dew) in the moist air of the environment during downtime. Following the formation of the dew, conditions become favorable for the formation of polythionic acid that leads to cracking in the presence of internal stresses.

Cause of Failure
Three main factors contributed to the observed failure: (i) sulfidation attack at high temperature due to the presence of H₂S, (ii) condensation of water vapor in moist air during downtime, and (iii) residual stresses resulting from cold work and/or the corrosion product. Due to the absence of cracking in the tubes that were stress relief annealed prior to service, it is possible that the internal stresses of the cold-worked tube could be the major factor. In conclusion, the tubes had failed due to stress corrosion cracking in polythionic acid.

Solution
Since it is rather difficult to remove H₂S in the early stages of the process, a more viable option is to stress relief anneal the tubes prior to service. For a long term solution, replacement of tube material on the basis of existing corrosive conditions during downtime could be considered.

References
12.6.5 Failure of a Caustic Cooler Tube

Process
A spent caustic cooler installed at an ethylene plant was in operation for 3 years. During operation, cooling water was passed through the tube, and the shell inside contained 1–5% spent caustic (NaOH), 10,000–20,000 ppm sodium sulfate, less than 1% sulfide, 5000 ppm sodium thiosulfate, 10,000–20,000 ppm sodium carbonate, and some hydrocarbon.

Service Conditions
The fluid temperature was 130°C.

Material
A carbon steel was used to manufacture the cooler tube (internal diameter = 12.7 mm, wall thickness = 2.5 mm).

Problem
Severe corrosion attack occurred at the external surface of the tube (shell side).

Analysis
1. Visual Inspection
Most of the corrosion attack at the external surface occurred locally as there was no evidence for significant wall thinning. At certain locations, the tube was perforated. Also, the corrosion product consisted of an inner and an outer discontinuous layer.

2. Material Verification
Microchemical analysis verified that the material was carbon steel with small amounts of Mn and Si.

3. Microstructural Characterization
Characteristic microstructural features of the tube cross section near the outer surface are shown in Fig. 12.27. It consisted of a mixture of ferrite and pearlite. The tube was attacked at the outer surface especially severely at
the grain boundaries, leading to grain separation highlighting localized nature of attack.

4. Analysis of Corrosion Products

Main elemental constituents of the corrosion product were Fe and O along with smaller concentrations of S, Na, and Ca. Evidently, S and Na were due to the presence of Na-S compounds at the shell side and Ca could be transported from the water.

Cause of Failure

Experimental results indicated that the tube had failed by severe localized corrosion attack leading to loss of its structural integrity. Most evidence pointed out that the cause of failure was direct contact between the steel and hot caustic.

Recommendation

It was recommended that the tube material be replaced with either a steel grade containing Ni or a Ni alloy, e.g., monel. Nickel alloys, however, offer the best resistance in such environments.
12.6.6 Failure of a Lube Oil Pipe in a Power Plant

Background
Lube oil (grade 320) for turbine bearing was transported by means of a pipe connected to the turbine support. Therefore, during operation, the pipe was subjected to mechanical vibration, i.e., alternating stresses. The pipe was fractured after 166,701 h of service. The working pressure was 1 Bar, the operating temperature was 50–55°C, and the pipe was made of carbon steel.

Analysis
1. Visual Inspection
The fracture surface was normal to the direction of maximum tensile stress developed within the pipe as a result of the internal fluid pressure exerted on the pipe walls. In general, there was no evidence for macroscopic deformation of the pipe, indicating that the fracture was preceded by very little or no deformation.

2. Material Verification
The material was identified as carbon steel with the presence of Mn and Si. The microstructure was a mixture of ferrite and pearlite. Surface hardness measurements indicated an average hardness Rb 94 consistent with the microstructure. This hardness corresponds to a tensile strength of about 95,000 psi, characteristic of a low-strength steel.

3. Fracture Mode
Secondary transgranular cracks normal to the direction of maximum tensile stress were observed along cross sections of the pipe as shown in Fig. 12.28. The exposed fracture surface was covered with a thick layer of oxide deposit. Stria-tions were detected at the fracture surface, suggesting that the mode of failure was fatigue consistent with service conditions.

Cause of Failure
Cause of failure was lack of adequate strength for resistance to fatigue.

Recommendation
It was recommended that means for improving the resistance to fatigue such as carburizing the surface be adopted, or that use of a heat-treated steel possessing a tempered martensite structure be considered.
12.6.7 Failure of a Catalytic Steam Reformer Tube

Process

Hydrogen was produced in a steam reformer unit by cracking of methane. Cracking took place in vertical catalysis tubes externally heated by means of burners normally using natural gas as a fuel. However, a waste liquid formed as a by-product of the process was certified by the manufacturer to be used as a fuel instead of natural gas.

Component

Catalysis tubes 102 mm in internal diameter and 9–10 mm in wall thickness.

Material

The tubes were made of a heat-resistant steel casting nominally containing 24% Cr, 24% Ni, and 1.5% Nb all in wt%.

Service Conditions

Internal gas pressure was about 30 atm, and the flame temperature of the burner was about 1100°C. Average external temperature of the tubes was 950°C;
however, the exact temperature varied from the top to bottom of the tube with the top section being hotter.

**Problem**

Natural gas was used to fire the burners. The tubes were designed to last 10,000 h (11.4 years). However, after 2 weeks of using the waste liquid as a fuel, a corrosion product appeared to develop at the external tube surface, and most of it could be washed away. An investigation was conducted to determine if the service life of the tubes could have been affected by using liquid fuel.

**Analysis**

1. **Visual Inspection**

Visual inspection of the as-received sections of the tube showed no evidence for a significant wall thinning. Also, the scale developed by the tube material appeared to be well adhered to the outer and inner surfaces. However, the outer tube surface was more blackish and rough, suggesting a greater extent of metal wastage to a corrosion product in comparison with the inner surface. The grain structure was large and elongated typical of as-cast material. Above observations suggested that the replacement of natural gas with liquid waste has accelerated the corrosion rate at the outer tube surface.

2. **Material Verification**

The material was verified by SEM/EDS analysis and both the measured and nominal compositions (wt%) are shown below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>46.72</td>
</tr>
<tr>
<td>Ni</td>
<td>23–25</td>
<td>25.01</td>
</tr>
<tr>
<td>Cr</td>
<td>24–26</td>
<td>24.59</td>
</tr>
<tr>
<td>Nb</td>
<td>1.4–1.8</td>
<td>2.11</td>
</tr>
<tr>
<td>Si</td>
<td>0.5–1.5</td>
<td>1.57</td>
</tr>
<tr>
<td>Mn</td>
<td>1(^*)</td>
<td>ND</td>
</tr>
<tr>
<td>C</td>
<td>0.25–0.35</td>
<td>ND</td>
</tr>
<tr>
<td>S</td>
<td>0.03(^*)</td>
<td>ND</td>
</tr>
<tr>
<td>P</td>
<td>0.03(^*)</td>
<td>ND</td>
</tr>
<tr>
<td>Pb</td>
<td>0.01(^*)</td>
<td>ND</td>
</tr>
<tr>
<td>Sn</td>
<td>0.01</td>
<td>ND</td>
</tr>
</tbody>
</table>

\(^*\)Maximum.
ND: Not detected.

Proper material was selected for the application on the basis of the 100,000 h creep-rupture strength. However, it is evident that the Si content of
the heat alloy used in the application approached the upper limit of the specification. As shown later, this was found to stabilize the Laves phase.

3. Materials Characterization

Depth of corrosion attack below the outer surface was greater in comparison with the inner surface of the tube. Internal attack occurred primarily at the grain boundaries.

The scale formed at the outer tube surface exhibited ridges indicative of its non-protective nature. The surface scale was about 40 μm in thickness and it consisted of three oxides: outer layer of a spinel-type \((\text{Fe,Ni})\text{Cr}_2\text{O}_4\) containing particles of a Nb-rich oxide, and an inner layer of a Cr-rich oxide expected to be \(\text{Cr}_2\text{O}_3\). The outer oxide layer was detached and voids were observed at the scale-alloy interface as well as at the grain boundaries. The precipitate particles were Fe, Cr, and Nb rich with a smaller concentration of Ni suggesting that they could be sigma-type phase.

The above results suggest that the use of liquid waste as a fuel had accelerated the oxidation rate at the outer surface, resulting in a nonprotective scale. Apparently, reaction between the products of combustion and the outer scale had resulted in the formation of a soluble corrosion product.

As expected of a high Ni steel containing Nb and relatively high Si concentration, the tube material was susceptible to precipitation of a Laves phase of the type \(\text{Nb}_2\text{Ni}_3\text{Si}\), as shown in Fig. 12.29. The Laves phases assumed a lamellar morphology which is known to be detrimental to creep-rupture strength. Evidently, the grain boundary voids could have resulted from creep damage accelerated by precipitation of Laves phase.

Long Term Service Performance of the Tube

Most evidence suggests that the interim use of liquid waste fuel had no significant effect on the long-term service performance of the tube. However, the susceptibility of the tube material to precipitation of lamellar Laves phase at grain boundaries would be expected to shorten. Previous experience with similar materials suggest that the actual life could be as short as 50,000 hours or even less.

Conclusions

The use of liquid waste fuel could accelerate the oxidation rate of the tube material. Provided that the use of this fuel was discontinued at the stage represented by the condition of the as-received sections, no long-term detrimental effect could be expected. However, the use of a high Ni steel strengthened by Nb and containing a relatively high Si concentration could substantially reduce the creep rupture.
Recommendations

Since the environment was moderately carburizing, there was no need to use a steel with a high Si content. Most of the expected service life could be realized by using one of the Fe-based superalloys such as INCOLOY alloy 800H or 800HT.

12.6.8 Failure analysis of Heat Exchanger Components

Background

In a methanol production plant, natural gas was fed into a desulfurization section for sulfur removal. Subsequently, the desulfurized gas entered into the reforming section where it was converted by steam maintained at 530°C into CO, CO₂, and H₂ according to the following reactions:

\[
\begin{align*}
\text{CH}_4 + \text{H}_2\text{O} & \rightarrow 3\text{H}_2 + \text{CO} \\
\text{CO} + \text{H}_2\text{O} & \rightarrow \text{H}_2 + \text{CO}_2
\end{align*}
\]

Copyright © 2004 by Marcel Dekker, Inc.
Upon completion of the above reactions, the reformed gas at a temperature said to be 850°C was fed into a heat exchanger to reduce its temperature down to 325°C. Boiler feed water for cooling purposes was introduced into the heat exchanger through tubes made of a low alloy steel (2.25Cr, 1Mo). Type 310 stainless steel was used to manufacture the shroud plate which covered the heat exchanger, the baffle plates used to deflect the flow of gases and the spacer tubes. Either type 310 or type 316 stainless steel was used to manufacture the tie rods. Gas pressure in the heat exchanger was estimated to be 14.5 kg/cm². The inlet gas composition was:

- H₂: 74%
- CO: 14%
- CO₂: 8%
- CH₄: 3.5%
- N₂: 0.5%

After being cooled to 325°C in the reforming section, the reformed gas entered the compression section and then the synthesis section where crude methanol was produced. Finally, the crude methanol was fed into the distillation section.

After about 40,000 h of plant operation, several components of the heat exchangers suffered premature failure that led to an unscheduled shutdown despite a life expectancy of 15 years. Failed components included the shroud plate, spacer tubes and tie rods. They exhibited extensive pitting combined with blackish surface deposit. Thinning of the shroud plate was severe in certain regions to the extent that the material was perforated. A large volume of blackish deposit was observed at the bottom of the tower. In addition to environmental degradation, the tie rods appeared to have undergone mechanical failure. Most of the failure was restricted to a vertical zone about 3 m long and located about 3 m downstream from the inlet gas. Although it is not possible to determine the exact temperature at the failed zone, it can be expected to be below 850°C and above 325°C.

Analysis

1. Visual Inspection of Failed Components

Visual inspection of failed components revealed blackish surface deposit. Also, pits filled with the blackish deposit were observed particularly in the spacer tube and tie rod. Most of the pits were large enough to be observed with naked eye. Perforations were observed in certain regions of the shroud plate due to extensive thinning of material. Appearance of a blackish deposit at the surface of the failed components and at the bottom of the tower is indicative of high-temperature carburization.
2. Materials Verification

Results of microchemical analysis by energy dispersive X-ray spectroscopy in a scanning electron microscope are summarized below along with the nominal chemical composition of type 310 stainless.

<table>
<thead>
<tr>
<th>Element</th>
<th>Shroud</th>
<th>Spacer</th>
<th>Tie rod</th>
<th>Water pipe</th>
<th>SS310</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>55.14</td>
<td>54.54</td>
<td>54.28</td>
<td>97.46</td>
<td>Balance</td>
</tr>
<tr>
<td>Cr</td>
<td>24.62</td>
<td>25.31</td>
<td>26.45</td>
<td>2.28</td>
<td>25</td>
</tr>
<tr>
<td>Ni</td>
<td>19.85</td>
<td>18.90</td>
<td>18.54</td>
<td>—</td>
<td>20</td>
</tr>
<tr>
<td>Si</td>
<td>0.4</td>
<td>0.66</td>
<td>0.73</td>
<td>0.08</td>
<td>1.5*</td>
</tr>
<tr>
<td>Mo</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.18</td>
<td>—</td>
</tr>
<tr>
<td>Mn</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>2.0*</td>
</tr>
<tr>
<td>C</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>0.25*</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.

The preceding analysis indicates that type 310 stainless steel is the material of the shroud plate, spacer tubes, and tie rods. Also, the observed composition of the water tubes is typical of a low alloy steel. Thus, it can be concluded that the materials utilized in constructing the heat exchanger were as stated.

3. Analysis of the Blackish Deposit

The predominant phase in the blackish deposit collected from the bottom of the tower was identified by X-ray diffraction as graphite along with small amounts of iron oxides. The chemical composition of the deposit is given below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Composition (Wt%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>51.70</td>
</tr>
<tr>
<td>Ni</td>
<td>27.56</td>
</tr>
<tr>
<td>Cr</td>
<td>16.43</td>
</tr>
<tr>
<td>Mn</td>
<td>1.60</td>
</tr>
<tr>
<td>Co</td>
<td>1.18</td>
</tr>
<tr>
<td>Ca</td>
<td>0.66</td>
</tr>
<tr>
<td>Al</td>
<td>0.39</td>
</tr>
<tr>
<td>Si</td>
<td>0.27</td>
</tr>
<tr>
<td>P</td>
<td>0.23</td>
</tr>
<tr>
<td>C</td>
<td>ND</td>
</tr>
</tbody>
</table>

It can be concluded from the above results that the observed blackish deposit at the bottom of the tower and on the surfaces of the failed components is carbonaceous in nature. This suggests that the primary mode of failure was
high temperature carburization which is most marked at 800–1000°C (Betteridge and Heslop, 1974).

4. Microstructural Characterization

Carburized zone was observed in the shroud plate sample (type 310 stainless steel) located at low-temperature zone. A Cr denuded zone was also present below the carbide scale. Microhardness measurements confirmed the existence of a surface carbide scale as the Vickers hardness number changed from 329 at the surface to 252–110 within the grains. Variations of microhardness within the grains appears to be associated with the precipitation of Cr$_{23}$C$_{6}$ carbides.

The shroud plate obtained from the hot zone suffered severe carburization associated with internal oxidation and absence of a surface carbide scale. At the surface of the specimen, the Vickers microhardness number was measured to be 200. Blocky precipitates observed below the denuded zone are identified to be of the CrFe sigma phase. It is possible that competition for Cr between the sigma phase and formation of surface carbides during carburization had resulted in the formation of the denuded zone.

Extensive arrays of fine precipitates are observed within the grains in the thinned section of the shroud plate. This is a typical morphology of Cr$_{23}$C$_{6}$ precipitates in stainless steels. Also, extensive decohesion along the grain boundaries was observed which is indicative that the material in this section had lost its integrity. Extensive thinning of material in this section appeared to have occurred by grain separation. Microhardness measurements revealed extensive hardening effect as a Vickers hardness number as high as 724 was measured. The above observations are typical of a “metal dusting” effect due to severe carburization.

Examination of the water tube (2.25Cr, 1Mo low alloy steel) revealed no significant corrosion on either the water side or the gas side. This was confirmed by microhardness measurements. At the surface, the Vickers microhardness number was measured to be 150 and in the interior it was found to be 146.

Extensive precipitation of the CrFe sigma phase in platelet-type morphology was observed in the tie rod (type 310 stainless steel). It is well known that sigma phase in platelet-type morphology is extremely detrimental to ductility particularly at low temperatures. Thus, it is possible that the tie rod had fractured due to the precipitation of sigma phase.

Examination of the spacer tube (type 310 stainless steel) near the baffle revealed catastrophic failure typical of metal dusting. Microhardness measurements confirmed the extensive carburization attack. For example in a region of the tube free of pits the Vickers microhardness number varied from 190 to 210. However, in a pitted region the hardness number was as high as 723.
**Cause of Failure**

Based upon the experimental results presented above, it can be concluded the primary mode of failure of the heat exchanger was high temperature carburization. Among the group of stainless steels resistant to carburization, type 310 is considered the best. However, when compared with nickel- and cobalt-based alloys, type 310 stainless steel lags behind (Lai, 1984). This may lead to the conclusion that type 310 stainless steel was not the best choice for the present heat exchanger.

**Recommendations**

In principle, there are three methods for reducing the corrosive attack of the environment: (i) replacing the alloy with another more resistant alloy, (ii) application of an appropriate surface coating, or (iii) modification of the environment to reduce its corrosive nature, e.g., by filtering (Betteridge and Heslop, 1974). Of the above three methods, one of the first two appears to be the most appropriate. Previous work on carburization has shown that alloys where the surface oxide scale is predominantly Al₂O₃ rather than Cr₂O₃ are more resistant to carburization (Lai, 1984). The nickel-based HAYNES alloy 214 is the alloy known to be most resistant to carburization. Alloys having above-average resistance to carburization are alloy 800H and HAYNES alloy 556. Alloy replacement should be considered for a long-term solution to the problem. However, application of a surface coating, particularly Al₂O₃, or poisoning the surface with H₂S to reduce the diffusivity of carbon may be considered.

**References**


12.6.9 Failure of Transfer Line Exchange Cone

**Background**

An ethylene (C₂H₄) producing plant consisted of eight pyrolysis furnaces where ethylene was produced at about 840°C from feed stock. During operation, the transfer line exchange cones were exposed to an environment consisting of H₂, CH₄, CO, H₂O and other carbonaceous gases (Lai et al., 1986). A schematic illustrating a typical transfer line exchange cone is shown in Fig. 12.30a. It consists of welded components made of INCOLOY 800H and INCONEL alloy 625 as a weld metal.
Figure 12.30 Continued
Every 40 days of continuous service, a decoking operation was performed where the temperature inside the furnace tube rose to 890°C for about 8–12 h. Although the cone assembly material were designed for a service life of 100,000 h, a cracking problem was encountered at certain locations after only 16,000 h of service, i.e., at about 20% of the life expectancy. Two types of cracking, designated as types I and II, were encountered. Referring to Fig. 12.30a, type I cracking was observed in the weld heat-affected zone transverse to the weld joint. Type II cracking was found to occur longitudinally along the weld joint (weld toe). The cracking was not apparently preceded by plastic deformation which typifies creep rupture.

### Analysis

1. Visual Inspection of Failed Components

Type I cracking is shown in Fig. 12.30b. The crack had penetrated the entire thickness. Branching of the cracking at the location marked A is typical of intergranular failure. Microstructural characterization confirmed this conclusion as
described later. A blackish deposit was observed at the inside surface of the tube which is typical of high-temperature carburization attack.

Type II cracking is demonstrated in Fig. 12.30c. As indicated by the arrows, the cracks are longitudinal along the weld joint.

2. Material Verification

Microchemical analysis results are shown in comparison with the nominal chemical composition (wt%) of alloy 800H.

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>39.5**</td>
<td>47.42</td>
</tr>
<tr>
<td>Ni</td>
<td>30–35</td>
<td>30.58</td>
</tr>
<tr>
<td>Cr</td>
<td>19–23</td>
<td>20.64</td>
</tr>
<tr>
<td>Mn</td>
<td>1.5*</td>
<td>ND</td>
</tr>
<tr>
<td>Si</td>
<td>1.0*</td>
<td>0.39</td>
</tr>
<tr>
<td>Cu</td>
<td>0.75*</td>
<td>ND</td>
</tr>
<tr>
<td>Al</td>
<td>0.15–0.60</td>
<td>0.30</td>
</tr>
<tr>
<td>Ti</td>
<td>0.15–0.60</td>
<td>ND</td>
</tr>
<tr>
<td>C</td>
<td>0.1*</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
**Minimum.
ND: Not detected.

It is evident that the tube material is alloy 800H. A comparison between the nominal and measured chemical compositions of alloy 625 is given below:

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni</td>
<td>Balance</td>
<td>56.12</td>
</tr>
<tr>
<td>Cr</td>
<td>20–23</td>
<td>22.12</td>
</tr>
<tr>
<td>Mo</td>
<td>8–10</td>
<td>10.56</td>
</tr>
<tr>
<td>Fe</td>
<td>5*</td>
<td>10.41</td>
</tr>
<tr>
<td>Nb + Ta</td>
<td>3.15–4.15</td>
<td>ND</td>
</tr>
<tr>
<td>Co</td>
<td>1*</td>
<td>ND</td>
</tr>
<tr>
<td>Mn</td>
<td>0.5*</td>
<td>ND</td>
</tr>
<tr>
<td>Si</td>
<td>0.5*</td>
<td>0.73</td>
</tr>
<tr>
<td>Al</td>
<td>0.4*</td>
<td>ND</td>
</tr>
<tr>
<td>Ti</td>
<td>0.4*</td>
<td>ND</td>
</tr>
<tr>
<td>C</td>
<td>0.1*</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.
3. Analysis of Type I Cracking

A blackish deposit was present on the fracture surface obtained from the sample that exhibited type I cracking. The surface deposit was composed of following:

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Wt%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maghemite (Fe$_2$O$_3$)</td>
<td>11</td>
</tr>
<tr>
<td>Hematite (Fe$_2$O$_3$)</td>
<td>13</td>
</tr>
<tr>
<td>Bunsenite (NiO)</td>
<td>14</td>
</tr>
<tr>
<td>Graphite (C)</td>
<td>24</td>
</tr>
<tr>
<td>Corundum (Al$_2$O$_3$)</td>
<td>37</td>
</tr>
</tbody>
</table>

The failure was unrelated to the presence of Al$_2$O$_3$, which could have been incorporated into the surface during handling and/or inspection of failed components. Detection of a large concentration of graphite is typical of high-temperature carburization attack. Usually, oxidation occurs concurrently with carburization, which explains the presence of iron oxides in the deposit. It is to be noted that carburization/oxidation is a typical mode of high-temperature corrosion of pyrolysis furnace tubes in the ethylene cracking process (Lai et al., 1986).

In an environment characterized by a low-oxygen potential, carburization can proceed as follows. First, methane is decomposed according to the following reaction:

$$\text{CH}_4 \rightarrow 2\text{H}_2 + \text{C}$$

As excess H$_2$ is generated in the environment, the above reaction proceeds to the left to reestablish equilibrium conditions. Second, carbon generated by the above reaction is deposited on the alloy surface and subsequently reacts with carbide-forming elements, particularly Cr, to form a surface carbide scale. Third, carbon diffuses into the alloy at a rate dependent on the alloy composition and results in the precipitation of carbide phases. Precipitation of these carbide phases cause marked embrittlement of the alloy and loss of corrosion resistance due to localized depletion in Cr. If the oxygen potential of the environment is relatively high, carbon is generated by the following reaction:

$$2\text{CO} \leftrightarrow \text{CO}_2 + \text{C}$$

The fracture mode was intergranular, which reflects an extensive embrittlement of the material. This is a typical consequence of high-temperature carburization. The material suffered extensive intergranular cracking to the extent of complete grain separation. No protective scale (Cr$_2$O$_3$) could be detected at the inside surface of the tube. This indicated that the environment was characterized by a high-carbon activity and relatively low-oxygen potential. Typically, under these conditions, a carbide of the type Cr$_3$C$_2$ forms at the surface where the carbon activity
is high and it changes to Cr$_7$C$_3$ and Cr$_2$$_3$C$_6$ with depth due to the reduction in carbon activity. Surface pits signifying the initial stage of “metal dusting” typically encountered in a severely carburizing environments and grain boundary carbide precipitates were observed.

An excessively large alloy grain size clearly indicated that the tube was exposed to a temperature considerably higher than 840°C. It is typical of alloy 800H to undergo an appreciable grain growth if exposed to temperature higher than about 980°C (Inco Alloys International, 1986). Carburization kinetics in high-temperature alloys is known to increase considerably with temperature above about 900°C (Lai, 1985). A considerable loss of ductility is associated with carburization at 980°C, which provides an indirect evidence that the alloy was exposed to a temperature higher than 925°C.

4. Analysis of Type II Cracking

Similar to the case of type I cracking, alloy 800H cracked intergranularly. No crack propagation from alloy 800H into the weld was observed. A crack in the weld metal was observed, yet it appeared to be independent of the crack in alloy 800H. Excessive grain growth also occurred in alloy 800H at this region similar to region examined earlier.

Conclusion

Based upon the results above, it can be concluded that the observed cracking was due to high-temperature carburization which results in severe embrittlement. Available carburization data indicates that of the various classes of high temperature alloys, alloy 800H has above average resistance to carburization (Lai, 1985; Lai and Rothman, 1984). As pointed out earlier, the results obtained indicate that the material was exposed to an excessively high temperature. Thus, the cracking problem may not necessarily be due to selection of wrong material. However, it is important to note that alloys protected by Al$_2$O$_3$-based scale provide the highest resistance to carburization. Of such wrought alloys commercially available is a Ni-16Cr-4.5Al-4Fe-0.01Y alloy known as HAYNES alloy 214. Another Fe-based alloy whose carburization resistance is either better than or comparable to that of alloy 800H, is HAYNES alloy 556 (Lai, 1985; Lai and Rothman, 1984).

Recommendations

In view of the results of this investigation, an immediate attention must be given to the temperature of the environment. If the temperature could be maintained below about 900°C, alloy replacement may not be necessary. However, if it is proved not possible to avoid excessive rise in temperature, even for relatively short periods of time, alloy replacement must be considered. In this case, alloy 214 may provide the solution.
References

12.6.10 Failure of Furnace Tubes
Background
A petrochemical industry produced vinyl chloride monomer (VCM) and poly vinyl chloride (PVC) which are used as raw materials by the plastic industry. Vinyl chloride monomer (C₂H₃Cl) was produced by a pyrolysis process where ethylene dichloride (EDC, C₂H₄Cl₂) is chemically decomposed by the action of heat according to the following reaction:

\[ \text{C}_2\text{H}_4\text{Cl}_2 = \text{C}_2\text{H}_3\text{Cl} + \text{HCl} \]

HCl was recycled to produce EDC by reaction with ethylene (C₂H₄) and air. In the PVC plant, VCM was converted into a powdery resin by a polymerization process.

In EDC cracking unit of the VCM plant, EDC feedstock was heated from a temperature of 50°C to about 489°C in a furnace to convert it into VCM and HCl by the above reaction. At 489°C, all the reactants and products are in the gaseous state. After 5 years of operation the four finned tubes of the furnace evaporation zone developed cracks which caused leakage of the gases. As per design specifications, the material of the finned tube was 347H wrought stainless steel.

Analysis
1. Visual Inspection
Visual inspection revealed that the furnace tube was exposed to a highly corrosive environment which led to a considerable metal wastage. Substantial thinning of the tube was observed at various locations. Metal wastage occurred to the extent of decohesion of grains and loss of structural integrity. It was apparent that cracking of the tube was initiated at the inner surface.
2. Verification of the Tube Material

Microchemical analysis (wt%) of the tube material by SEM/EDS is compared with the nominal chemical composition of 347H stainless steel below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>68.96</td>
</tr>
<tr>
<td>Cr</td>
<td>17–19</td>
<td>17.41</td>
</tr>
<tr>
<td>Ni</td>
<td>9–13</td>
<td>9.71</td>
</tr>
<tr>
<td>Nb</td>
<td>&gt;10C</td>
<td>3.58</td>
</tr>
<tr>
<td>Mn</td>
<td>2*</td>
<td>ND</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.34</td>
</tr>
<tr>
<td>C</td>
<td>0.08*</td>
<td>ND</td>
</tr>
<tr>
<td>P</td>
<td>0.045*</td>
<td>ND</td>
</tr>
<tr>
<td>S</td>
<td>0.030*</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.

Chemical composition of the tube material was consistent with that typical of 347H stainless steel.

3. Mechanical Strength

Surface hardness measurements at various locations indicated that the tube material had an average room temperature hardness of Rb 84. It is typical of annealed 347H stainless steel to have a room temperature hardness of Rb 85 (Bittence). Thus, it could be concluded that the tube material was not subject to any hardening effect.

4. Microstructural Characterization

A closer examination of the grain structure revealed grain boundary precipitates assuming the morphology of blocky particles particularly at triple points (location at which three grain boundaries meet), as illustrated in Fig. 12.31. Microchemical analysis revealed that the main elemental constituents of the precipitate particles were Cr and Fe with relatively small concentrations of Nb and Si. From this observation, the grain boundary precipitate could be identified as sigma phase essentially of the CrFe type. Both Nb and Si are known to be stabilizers of sigma phase (Wallwork and Crop, 1976).

It is typical of many Fe-based alloys to precipitate the CrFe sigma phase upon exposure to elevated temperatures (Wallwork and Crop, 1976). Due to the effect of alloy composition on the kinetics of sigma phase precipitation (Wallwork and Crop, 1976), long-term exposure may be required to precipitate the sigma phase at relatively low temperatures as it appears to be the case for the tube alloy.
Generally, the extremely hard and brittle sigma phase can precipitate both in the matrix and at grain boundaries. When it precipitates in the matrix, it can assume either a blocky- or platelet-type morphology. However, there was no evidence for matrix precipitation of sigma phase in the tube material. Possibly, this behavior could be related to the relatively low-exposure temperature which favors grain boundary precipitation. It is to be noted that the absence of sigma phase precipitates in the matrix is consistent with the measured hardness. However, since the sigma phase is enriched in Cr, its precipitation is expected to degrade the chemical stability of the alloy in oxidizing environments.

5. Analysis of Corrosion Product

The corrosion deposit collected from the furnace was highly concentrated in Cl. However, the concentrations of metallic elements particularly Fe and Ni were found to change considerably from one location to another, suggesting that the deposit consisted of more than one compound.

Structural analysis by X-ray diffractometry revealed that the deposit consisted of FeCl$_3$, Fe$_2$O$_3$, and a spinel isomorphous with NiFe$_2$O$_4$. It is likely that the oxide phases, i.e., Fe$_2$O$_3$ and NiFe$_2$O$_4$, were contaminated with chlorine. The spinel could be of the type Ni(Fe,Cr)$_2$O$_4$. It is also possible that this spinel could contain Mn.

In contrast to the corrosion deposit collected from the furnace, the scale which was adhered to the inside surface of the tube was found to contain a considerably less concentration of Cl. The main elemental constituents of the scale...
were Fe and Ni with relatively small concentrations of Cl and Cr. It could be concluded that the scale was a spinel of the type Ni(Fe,Cr)\(_2\)O\(_4\) contaminated with Cl. Since Cl is known to convert protective scales into porous nonprotective oxides (Oh et al., 1986; Rhee et al., 1985; Baranow et al., 1984), the deposit collected from the furnace was initially an oxide scale which was highly contaminated with Cl to the extent that it spalled off the surface. This is consistent with the observation of a proportion of oxide scale in the deposit.

Based upon the results above, it could be concluded that the tube furnace was subject to high-temperature corrosion by oxidation and chlorination. It is well known that the rate of oxidation attack is greatly accelerated in the presence of chlorine (Hossain, 1981). Also, it is known that VCM production is one of the industrial processes where severe high-temperature corrosion problems are encountered due to the presence of chlorine in the environment (Oh et al., 1986). Generally, in the production of chlorinated organic compounds where HCl bearing gas environments are encountered, pyrolysis furnace tubes are known to corrode by oxidation and chlorination (Lai et al., 1985).

**Mode of Corrosion Attack**

In addition to overall metal wastage to corrosion products (oxides and chlorides), the tube material was heavily attacked along grain boundaries. During exposure to the operating temperature, the inside surface of the tube developed a surface scale consisting of granular oxide particles and an underlying oxide. The main metallic constituent of the particles was Fe with a relatively small concentration of Cr, suggesting that the particles were an oxide of the type (Fe,Cr)\(_2\)O\(_3\). The Fe-rich oxide was free of Cl. This behavior could be due to the relatively high stability of Fe-rich oxides in Cl-contaminated environments due to the high vapor pressure of Fe-rich chlorides (Oh et al., 1986;). In addition to Ni as the main metallic constituent, the underlying scale contained a relatively high concentration of Fe, small concentrations of Mn and Cl, and trace amount of Cr. From this observation, the scale could be identified as a spinel of the type Ni(Fe,Cr,Mn)\(_2\)O\(_4\). However, the scale was contaminated with Cl. Cracking of the spinel scale could be due to the reduced fracture toughness caused by Cl contamination.

Examination of specimens near the ruptured regions revealed extensive intergranular attack. Cracking was initiated in the inside of the tube and extended outward. However, branching of the crack indicated that it followed grain boundaries. The grain boundaries appeared to contain a corrosion product.

In many locations, the intergranular attack was so severe to cause complete grain separation and loss of structural integrity. Grain boundaries contained a Cr-rich discontinuous phase (possibly M\(_{23}\)C\(_6\)) and an Fe-rich continuous phase. Typically, in a chlorine-contaminated environments, metallic components first develop a protective oxide scale. Subsequently, the protective scale is attacked.
by chlorine which converts it into a porous nonprotective scale of very low resistance to spallation (Oh et al., 1986; Rhee et al., 1985; Baranow et al., 1984). Chlorination results in the formation of chlorides and oxychlorides, which are volatile at relatively high temperatures above about 500°C. Observation of chlorides in the corrosion product appeared to be consistent with the relatively low temperature in the furnace. Due to contamination by chlorine, the oxide scale formed on the tube became porous and fragile and thus, incapable of protecting the metal underneath from continued chlorination attack until the tube was ruptured.

Alloys that contain a relatively high concentrations of refractory elements, such as Mo, W, Nb, and Ta, are known to be the least resistant to a chlorine-contaminated environment due to the formation of volatile chlorides and oxychlorides (Oh et al., 1986; Rhee et al., 1985; Baranow et al., 1984). The heat of 347H stainless steel used to manufacture the tube contained about 3.58wt% Nb, suggesting that improper alloy selection contributed to the failure.

Conclusion

Based upon the results presented earlier, it could be concluded that the cause of failure of the furnace tube was high temperature corrosion caused by a chlorine-contaminated environment, which greatly accelerated the oxidation rate.

Recommendation

Based upon the results presented above, it is recommended to replace the tube material. Generally, Fe-based alloys free of refractory metals such as 310 stainless steel, INCOLOY alloy 800H and HAYNES alloy 556 have adequate resistance to chlorine-contaminated environments. Among these alloys, HAYNES alloy 556 is the most resistant to such environments (Rhee et al., 1985; Baranow et al., 1984).

References

12.6.11 Failure of Dryer Support

Background

Another problem was encountered in the dewatering and drying unit of the PVC plant mentioned in Sec. 12.6.10. After unspecified length of time, a dryer support plate made of wrought 304 stainless steel fractured. The inlet gases temperature was 155–190°C. Both the outlet gases and liquids had a temperature in the range of about 50–80°C.

Analysis

1. Visual Inspection

Fracture occurred near a weld and was preceded by very limited macroscopic deformation. Also, the fracture surface was relatively flat suggesting a brittle-type failure. However, the fracture surface appeared to be covered with oxide phases. No macroscopic cracks could be detected near the fracture section.

2. Material Verification

Microchemical analysis (wt%) by SEM/EDS confirmed that the dryer support material was 304 stainless steel as shown below:

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal Range</th>
<th>Measured Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>72.56</td>
</tr>
<tr>
<td>Cr</td>
<td>18–20</td>
<td>18.11</td>
</tr>
<tr>
<td>Ni</td>
<td>8–12</td>
<td>8.82</td>
</tr>
<tr>
<td>Mn</td>
<td>2*</td>
<td>ND</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.51</td>
</tr>
<tr>
<td>P</td>
<td>0.045*</td>
<td>ND</td>
</tr>
<tr>
<td>S</td>
<td>0.030*</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.

Surface hardness measurements indicated that the hardness was essentially the same at locations near and away from the fracture section. The average hardness was found to be Rb 82, which approaches that of 304 stainless steel in the annealed condition (Rb 80). It could be concluded from this observation that the observed fracture was not purely mechanical in nature.

3. Fracture Mode

Detailed examination of the fracture surface revealed that it was predominantly covered with oxide phases. Microchemical analysis revealed that the surface layer was enriched in Fe and Cr, suggesting the presence of Fe- and Cr-rich
oxide phases. The presence of these oxides precluded a direct determination of the fracture mode. It is possible that the surface was oxidized during crack propagation.

Figure 12.32a–c illustrates typical microstructural features near the fracture section. Fine striations were observed in some grains, as illustrated in Fig. 12.32a,b. The striations changed orientation from one neighboring grain to another. Voids at grain boundaries were also observed as indicated by the arrows in Fig. 12.32b. Another important microstructural feature observed near the fracture section was the presence of transgranular and intergranular cracks, as shown in Fig. 12.32c. Also, voids within the grains and at grain boundaries can be distinguished in Fig. 12.32c as indicated by the arrows. However, no corrosion product could be detected in the vicinity of the cracks.

It is likely that the striations correspond to slip lines, which suggested that the material was subject to localized plastic deformation. Another important feature that can be derived is that the slip was of the planar type, which is promoted by a low stacking fault energy. Localization of the plastic deformation appeared to be consistent with the result of surface hardness measurement presented earlier where no overall hardening effect was observed. Thus, the dryer support appeared to fracture both intergranularly and transgranularly.

Cause of Failure

Based upon the observations above, the failure of the dryer support appeared to result from degradation of its mechanical strength. Both the localized nature of plastic deformation and the observed voids tend to suggest that the degradation in mechanical strength was caused by absorption of hydrogen which can induce transgranular and intergranular cracking (Bernstein and Thompson, 1976). According to the proponent, the outlet gas contains 107 ppm hydrogen. Absorption of hydrogen by a metal can occur either in the liquid or solid state. During melting or welding practices, hydrogen can be retained during solidification. In the solid state, absorption of hydrogen can be facilitated by a catalyst, particularly S, even if it is present in trace amounts (Bernstein and Thompson, 1976). Because of the considerable decrease in hydrogen solubility with decreasing temperature, the solid crystal becomes supersaturated with hydrogen. Supersaturation in hydrogen can be relieved by precipitation of molecular hydrogen (H₂) which creates voids. Internal stresses created by hydrogen absorption can also induce localized plastic deformation. In general, hydrogen tends to segregate to interfaces such as twin and grain boundaries. Thus, both transgranular and intergranular cracks can result from hydrogen absorption.

Conclusion

Based upon the results obtained, it could be concluded that the dryer support failed by hydrogen-induced cracking.
Figure 12.32  Secondary electron SEM images illustrating characteristic microstructural features near the fracture section of the dryer support. (a) Fine striations in adjacent grains. (b) Same area as observed at a higher magnification. (c) Transgranular and intergranular microcracks.
Recommendation

It is generally believed that there is a direct correlation between the susceptibility to hydrogen-induced cracking (HIC) and the stacking fault energy of austenitic stainless steels (Bernstein and Thompson, 1976). Due to a lower stacking fault energy, slip tends to be of the planar-type as observed here. As a result, plastic deformation becomes highly localized which reduces the resistance to HIC. When stainless steels are used in a hydrogen-producing environment, elements such as S and P must be avoided and the C content must be kept below 0.06% (Bernstein and Thompson, 1976). Another important requirement is that the 18% critical Cr concentration must be avoided and the Ni content must be above 10%. Clearly, these requirements are not satisfied by the 304 stainless steel used to manufacture the dryer support. Available data indicates that this steel is one of the least resistant to HIC. One of the steels which satisfy the above requirements is 309 stainless steel.

Reference


12.6.12 Failure of Steel Piping System

Process

A heat recovery equipment was located at the downstream of a steam reformer. Heat stored in the reformed gas was recovered at boilers and then economizers. A piping system was used to transfer the reformed gas from the economizer identified as E-206 in the schematic of Fig. 12.33 into a reboiler (E-505 in Fig. 12.33). Typically, the reformed gas composition in mol% was 2.71 CH₄, 13.59 CO, 9.08 CO₂, and 74.62 H₂.

Operating Conditions

The operating conditions at the reformer outlet were 14.5 kg/cm² and 845°C. Within the interconnecting piping system handling the reformed gas, the operating conditions were 14.9 kg/cm² and 200°C.

Materials

As shown in Fig. 12.33, the interconnecting piping system was made of C-1/2Mo steel (low-alloy steel) and 304 stainless steel.
Problem

A galvanic corrosion problem had been frequently encountered on the C-1/2Mo steel side to the SS304 piping for over 10 years. Opening up of weld seam of SS304 piping and fitting was encountered after 6 years. The piping system was partially replaced by 16 mm thick 304 SS after 9 years of service. At that time, samples were removed from the locations shown in Fig. 12.33 for a failure analysis study. Pitting of the C-1/2Mo steel piping was observed (samples 1 and 3 in Fig. 12.33). Also, the 304 SS welded joint developed a crack (sample 2A in Fig. 12.33); however, no cracking was detected in sample 2B.

Analysis

1. Visual Inspection

Light optical macrographs illustrating the macroscopic features of as-received samples are shown in Fig. 12.34a–d. Samples 1 and 3 (Fig. 12.34a,b) exhibited similar features in that the C-1/2Mo steel pipe contained relatively large round pits about 2 mm in diameter. It is observed from Fig. 12.34c that the crack in

Figure 12.33  A schematic illustration of the interconnection between the economizer and the piping system, the locations from where samples were removed for analysis are indicated.
sample 2A appeared to originate at the inner pipe surface within the heat-affected zone (304 SS). Figure 12.34d illustrates the sound section of the pipe corresponding to sample 2B where no cracking occurred.

Pits observed in Fig. 12.34a,b (samples 1 and 3) were observed to contain a mixture of yellowish and black sooty deposit. Also, the black deposit was observed at the inner surface near the pits. Similarly, the pits in sample 3 were
found to contain black sooty deposit. Such features appeared to be characteristic of a type of corrosion attack encountered in carbonaceous environments and known as metal dusting (Lai, 1990). In this case, the corrosion product within the pits consisted of carbon soot and powders of metal, oxides, and carbides. For a reducing carbonaceous gas such as that handled by the failed unit, carbon can be deposited at the inner pipe surface according to the following reaction:

\[ \text{CH}_4 \rightarrow \text{C} + 2\text{H}_2 \]

It was observed in sample 2A (304 SS pipe) that the fracture surface contained chevron marks, which appeared to originate at the inner pipe surface, suggesting that the crack was propagated by a fatigue-type mechanism as further confirmed later. The cracking occurred at a sharp corner (90° elbow), which could promote the susceptibility to fatigue failure.

2. Material Verification

The chemical compositions (wt%) of the four samples as measured by ICP-AES are summarized below.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Pipe (C-1/2Mo)</th>
<th>304SS (Patch weld)</th>
<th>Pipe (304SS)</th>
<th>Weld (304SS)</th>
<th>Pipe (304SS)</th>
<th>Weld (304SS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98.55 0.03 0.03 0.44 0.76 0.19 &lt;0.01 &lt;0.05</td>
<td>71.10 18.12 8.33 0.17 0.98 0.35 0.33 0.07</td>
<td>97.90 1.07 0.08 0.47 0.39 0.67 0.07 &lt;0.05</td>
<td>71.70 18.10 8.15 0.28 0.95 0.40 0.27 0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>97.90 1.07 0.08 0.47 0.39 0.67 0.07 &lt;0.05</td>
<td>71.70 18.10 8.15 0.28 0.95 0.40 0.27 0.08</td>
<td>69.00 18.10 8.64 0.31 1.19 0.38 0.15 &lt;0.05</td>
<td>67.10 20.18 9.37 0.09 1.50 0.56 0.06 0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2A</td>
<td>69.00 18.10 8.64 0.31 1.19 0.38 0.15 &lt;0.05</td>
<td>67.10 20.18 9.37 0.09 1.50 0.56 0.06 0.08</td>
<td>70.80 17.49 8.39 0.30 1.16 0.37 0.15 &lt;0.05</td>
<td>67.90 19.46 9.10 0.11 1.44 0.58 0.07 0.07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample 2A

2B

The nominal compositions of C-1/2Mo steel and 304 SS are also shown below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Fe</th>
<th>Cr</th>
<th>Ni</th>
<th>Mo</th>
<th>Mn</th>
<th>Si</th>
<th>C</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2Mo Steel</td>
<td>Bal.</td>
<td>—</td>
<td>—</td>
<td>0.5</td>
<td>0.55</td>
<td>0.30</td>
<td>0.14</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>304 SS</td>
<td>Bal. 18–20</td>
<td>8–12</td>
<td>—</td>
<td>2*</td>
<td>1*</td>
<td>0.08*</td>
<td>0.045*</td>
<td>0.030*</td>
<td></td>
</tr>
</tbody>
</table>

It is evident from a comparison of the above data that the compositions of the pipes and welds were well within the specifications. Therefore, deviation from specified compositions could be ruled out as a contributory factor to the failure.

The pipe material consisted of a mixture of ferrite and pearlite typical of carbon and low-alloy steels (see Fig. 12.35a). Within the detection limits of
Figure 12.35  Continued
energy dispersive spectroscopy (about 0.2wt%), the elemental constituents of the pipe material were Fe, Mo, Mn, and Si.

3. Analysis of Corrosion Deposit (Samples 1 and 3)

In sample 1 carbon was a major constituent of the deposit; however, the deposit could consist of a mixture of oxide and carbide consistent with that produced by metal dusting. It is observed that regions deficient in oxygen were enriched in carbon, and vice versa, suggesting that the deposit consisted of oxides and carbides of iron. Most of the yellowish deposit, however, was found to consist of rather porous oxide.

It could be concluded from the above observations that initially, the steel pipe developed a surface layer of protective oxide. Subsequently, however, this layer was converted into a nonprotective scale by the deposition of carbon at various scattered locations accelerating the corrosion rate.

Figure 12.35 Secondary electron SEM images illustrating (a) typical microstructure of the pipe material. Fracture morphology of sample # 2A near the (b) inner and (c) outer surface.
Sample 3 was found to exhibit similar features to sample 1 in that the corrosion product deposited at the inner surface was found to consist of a mixture of oxide and carbide. Microchemical analysis showed that carbon was the major constituent of a region, suggesting that the major component of this region was a carbide. However, in contrast with sample 1, the deposit was found to contain a small concentration of Cr, which could have been transported into the C-1/2Mo steel from 304 SS during welding. Most of the deposit in contact with the metal was an oxide similar to the case of sample 1, although a small amount of carbide was also detected.

4. Cracking Mode (Sample 2A)

Both the weld and base metal exhibited normal microstructural features. Also, the average microhardness value near the major crack and in regions far removed from the crack were essentially the same (HV 155 corresponding to Rb 80) and almost identical to the known hardness of 304 SS in the annealed condition ranging from Rb 80 to Rb 97 (2). This is consistent with the observed transgranular nature of the cracks. Characteristic morphological features of the fracture surface are shown in Fig. 12.35b,c. Near the inner pipe surface (Fig. 12.35b) the fracture surface exhibited striations characteristic of a fatigue-type failure. However, near the outer pipe surface (Fig. 12.35c), the fracture surface exhibited features resembling a fracture produced by a ductile-type mechanism similar to an overload fracture in a standard tensile test. As expected, the initial crack was propagated by a fatigue-type mechanism until the cross-sectional area was unable to support the stresses resulting from the internal gas pressure leading to overload fracture (Fig. 12.35c).

Examination of sample 2B (no cracking occurred) showed that its average microhardness value was also HV 155 corresponding to Rb 80. This further confirmed that cracking of sample A was not related to a defective material. Most likely, however, the section of the pipe was improperly designed for fatigue (90° elbow; sharp corner).

5. Modes of Failure

C-1/2Mo Steel Pipe. Experimental results suggested that the mode of failure of the C-1/2Mo steel pipe was metal dusting. This is known to be a type of carburization attack encountered in environments containing carbonaceous gases. In a reducing atmosphere, carbon can be deposited at the inner pipe surface by decomposition of methane degrading the protective nature of the surface oxide developed by the material.

304 SS Pipe. In contrast with the C-1/2Mo steel pipe, cracking of the 304 SS pipe was not environmentally related. Rather, it was found that the mode of
failure was fatigue as indicated by the characteristic features of the fracture surface.

Causes of Failure

*C-1/2Mo Steel Pipe*. Carbon and low-alloy steels are, among other materials, not resistant to metal dusting. Therefore, the most probable cause of failure was the poor resistance of C-1/2Mo steel to metal dusting. However, it is important to realize that no clear understanding of the effects of alloying elements on resistance to metal dusting has been developed. Therefore, it is rather difficult to select a material resistant to this type of corrosion attack. Although solved studies suggested that the susceptibility to metal dusting can be reduced by using an alloy in which the total percent of chromium plus 2 times the percent of silicon is in excess of 22%, it was shown that many alloys which satisfy this rule suffer metal dusting (Lai, 1990).

*304 SS Pipe*. Most evidence pointed out that the likely cause of failure was improper design of the elbow for fatigue resistance. A 90° elbow would be expected to increase the notch sensitivity of the material (higher level of stress concentration at the sharp corner), reducing its resistance to fatigue.

Recommendations

Based upon the results of this study, it is rather difficult to completely combat the pitting problem of the C-1/2Mo steel pipe by recommending a replacement since many types of engineering alloys are prone to metal dusting. However, the results indicated that the cracking problem of the 304 SS pipe could be solved by a more proper design for fatigue.

Conclusions

It was concluded from this study that the C-1/2Mo steel pipe was pitted by metal dusting resulting from the deposition of carbon at the inner pipe surface. However, since many of the engineering alloys are susceptible to this type of attack, it is rather difficult to find another material with significantly more resistance to metal dusting. Fatigue was identified as the mode of failure of the 304 SS pipe. In this case, the problem could be solved by a proper design avoiding sharp corners.

References

12.7 Failure of Engineering Products During Service Because of Improper Service Conditions

12.7.1 Failure of Ethane Pyrolysis Furnace Tubes Due to Overheating

Process
Production of ethylene by cracking of ethane

Component
Ethane pyrolysis furnace tubes

Material
HP 45 steel casting (Fe-33.5 Ni-21.5 Cr-2.5 Si maximum-2 Mn maximum- < 2.5 Nb-0.55 C).

Service Conditions
Temperature was 850 ± 20°C, pressure approached 1 atm, and the conversion ration of ethane feedstock was about 0.65. Decoking was carried out at unspecified intervals.

Problem
After about one third of the expected service life, the furnace tubes developed cracks.

Analysis
1. Visual Inspection
Visual inspection revealed that the cracks were longitudinal typical of creep rupture. Sagging of the tube was noted near the cracks, as shown in Fig. 12.36a, indicating localized plastic deformation. At various locations, the cracks had penetrated the entire thickness of the tube (Fig. 12.36b). There was no evidence for either general or localized wall thinning, i.e., metal wastage by high temperature corrosion. It was apparent from the visual inspection that the tube had failed by creep rupture.

2. Material Verification
Energy-dispersive X-ray spectroscopy in a scanning electron microscope confirmed that the tube material was HP 45 steel as shown below (in wt%). Such a steel is austenitic, i.e., it consists entirely of austenite. Chemical composition
(in wt%) of the alloy, as measured by SEM/EDS, is compared with the nominal composition below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>40.50</td>
</tr>
<tr>
<td>Cr</td>
<td>20–23</td>
<td>23.67</td>
</tr>
<tr>
<td>Ni</td>
<td>32–35</td>
<td>35.23</td>
</tr>
<tr>
<td>Si</td>
<td>2.5*</td>
<td>0.60</td>
</tr>
<tr>
<td>Mn</td>
<td>2*</td>
<td>ND</td>
</tr>
<tr>
<td>C</td>
<td>0.35–0.75</td>
<td>ND</td>
</tr>
<tr>
<td>Nb</td>
<td>≤2.5</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.
3. Analysis of Scale

Analysis of the external scale developed at the inner surface of the tube by energy dispersive X-ray spectroscopy and X-ray diffraction showed that it consisted of an outer granular spinel of the type Ni(Fe,Cr)₂O₄, and an inner layer of Fe-rich oxide containing Cr of the type (Fe,Cr)₂O₃. Consistent with the above results, the scale in contact with the tube material (subscale) consisted of an Fe-rich oxide containing Cr; however, it also contained Ni. Evidently, under the service conditions, the tube material was incapable of developing a protective Cr₂O₃ scale.

4. Analysis of Tube Material

Underneath the scale, the tube material contained Cr-rich massive thick precipitates typical of a Cr₇C₃ carbide within the alloy matrix resulting from carburization attack. When viewed at higher magnifications, the subscale contained voids. Although it is evident from the above observations that the oxygen activity in the environment was sufficiently high, permitting the tube material to develop an oxide scale, it was not sufficiently protective as the Cr₂O₃ scale. It is possible, however, that a layer of Cr₂O₃ had initially formed and then transformed into a carbide by reaction with CO in the environment:

\[
7\text{Cr}_2\text{O}_3 + 33\text{CO} = 2\text{Cr}_7\text{C}_3 + 27\text{CO}_2
\]

Detailed microstructural analysis of the tube material revealed a high density of voids at grain boundaries near macroscopic cracks indicating an advanced stage of creep deformation, as illustrated in Fig. 12.37a. A magnified image of a grain boundary void is shown in Fig. 12.37b. Massive Cr-rich carbide precipitates were observed at grain boundaries. Separation of grain boundaries occurred at the austenite-carbide interface, suggesting that the effect of creep damage was compounded by the carburization attack.

5. Evaluation of Mechanical Strength

For a tube never used in service, the microhardness of austenite at the inner surface was measured to be HV 352 corresponding to Rc 38. Typically, the hardness of heat-resistant steel castings in this category is in the range of HV 170–180 (Rc 87–89). Evidently, the surface hardened layer had resulted from machining to remove bore porosity. Surface hardness of the failed tube was found to be HV 480 (Rc 47), although recrystallization of the initially present hardened layer would be expected to decrease the surface hardness. Obviously, this was outweighed by the hardening effect produced by carburization, which also precluded the formation of a protective Cr₂O₃ scale.

Copyright © 2004 by Marcel Dekker, Inc.
Failure Mode

Most evidence pointed out that the mode of failure was a creep rupture accelerated by carburization attack.

Cause of Failure

Based upon the results of the investigation, it was concluded that the most probable cause of failure was exposure to an excessively high temperature accelerating the creep rate as well as the rate of carburization.

Recommendations

Better control of furnace temperature to avoid overheating is recommended; however, if this solution is not viable, replacing the tube material with a steel more resistant to carburization such as grade HP + W should be considered.

12.7.2 Failure of Blade Bolts Because of Overtightening

Process

A fin fan unit consisting of blades fastened by bolts was used in a chemical plant.

Component

Bolts 8.8 mm (0.346 in) in diameter
Application
Mechanical fasteners of rotating blades of a fin fan

Material
An unspecified grade of steel

Service Conditions
Ambient temperature and atmospheric conditions

Problem
After an unspecified period, the bolts shown in Fig. 12.38 were all fractured catastrophically.

Figure 12.38  Schematic illustrating a fin fan and fractured blade bolt. Locations from where specimens from the bolt were removed for analysis are indicated.
Analysis

1. Visual Inspection

Visual inspection indicated that the bolts were fractured with very limited macroscopic plastic deformation. Also, it was evident that the bolts were galvanized (coated with zinc), as indicated by the yellowish surface appearance.

Macroscopic examination of the fracture surface indicated that it consisted of two regions differing in texture, as marked by A and B in Fig. 12.39. Region A appeared to be burnished while region B exhibited a granular texture. Usually, such features typify a fracture produced by a fatigue mechanism, which could be expected in a rotating part.

2. Identification of the Bolt Material

Galvanization of the bolts was confirmed by microchemical analysis revealing a Zn-rich surface layer. The galvanized layer contained a marked concentration of Cr and a trace amount of Fe. Since the depth resolution of energy dispersive spectroscopy in SEM is 1 μm considerably less than the typical thickness of a galvanized layer (tens of micrometers), the observed Cr in the galvanized layer was more likely due to alloying with the base metal. Both microchemical analysis and microstructural characterization were used to identify the bolt material. The bolt appeared to be made of a low-alloy steel grade containing Ni and Cr. Quantification of the spectral data indicated that the steel contained about 0.58% Cr, 0.60% Ni, 0.43% Si, and 0.35% C, suggesting that the bolt was made of a medium carbon low-alloy steel of
AISI grade 86XX. Metallographic examinations of the bolts revealed microstructural features typical of a hardened and tempered steel.

Quenched and tempered steels are specified for bolt applications, where prestressing to a sufficiently high value is used to prevent slip of the joined surfaces by friction. Prestressing is defined in terms of a proof load corresponding to the applied tensile load which the bolt can support without evidence of any permanent deformation. Also, high-strength values of steels for fatigue applications can be achieved by a quench-and-temper heat treatment resulting in a homogeneous microstructure.

Based upon the above results, it can be concluded that a suitable material was selected for the application. Also, in principle, the proper heat treatment used to achieve high strength values was employed.

3. Evaluation of Mechanical Strength and Loading Conditions

Comparison of the above results with bolt specifications suggested that the bolt material was selected on the basis of either grade 5 or grade 6 of the Society of Automotive Engineers (SAE) specifications. Both grades call for the use of a medium-carbon steel (0.28–0.55% C) quenched and tempered at a minimum temperature of 425°C (800°F). Since galvanization requires dipping in molten Zn (melting point of Zn = 419.5°C), it is apparent that the bolt material was tempered during galvanization.

Surface hardness measurements indicated that the hardness of the bolt varied from an average of Rc 21 at the surface to Rc 23 at the center, roughly corresponding to an approximate ultimate tensile strength in the range of 765–815 MPa (111–118 ksi). Evidently, there was no significant difference between the surface bulk hardness, ruling out inhomogeneous strength as a contributory factor in the failure. It is recalled that for steels consisting of tempered martensite, the ratio of endurance limit and ultimate tensile strength (endurance ratio) is about 0.6; i.e., the endurance limit of the bolt material was approximately in the range of 459–489 MPa (67–71 ksi).

A tensile strength of 765–815 MPa appeared to conform to SAE grade 5 specifications, calling for a minimum tensile strength 827 MPa (120 ksi) for bolts of a maximum diameter of 19 mm. In this case, the proof load is specified as 6600 lb, corresponding to a stress of 480 MPa (70 ksi).

4. Fractography

The region of the surface where fracture was initiated exhibited fatigue striations as shown in the Fig. 12.40a, b. Striations were formed under a cyclic stress where slip or deformation bands group into packets. Dimples were observed between regions containing striations, indicating highly localized plastic deformation (see Fig. 12.40c). Generally, in the granular or rough surface region, the fracture was brittle as shown in Fig. 12.40d.
Mode of Failure

Failure of the bolts occurred by fatigue.

Cause of Failure

Galvanizing treatment resulted in overtempering of the steel which caused considerable softening, thus leading to failure.
Solution
Use nongalvanized bolts of the same steel.

12.7.3 Failure of Gas Turbine Blades Because of Overheating

Process
Electric power generation by land-based gas turbine engine.

Component
First-stage gas turbine blade that was refurbished after 35,000 h of service

Material
Udimet Alloy 710 protected by a CoCrAlY-type surface coating

Service Conditions
The firing temperature was in the range of 1065–1085°C (1950–1985°F). However, the firing temperature could vary depending on the power requirements. A high-grade fuel was used in the process.

Figure 12.40 Secondary electron SEM images showing (a) typical fatigue striations, (b) magnified image of the enclosed region in (a), (c) dimples due to localized deformation between striations while tensile tears are indicated by arrows and (d) brittle fracture in the bolt region.
Problem

After refurbishment, the blade developed an open crack at the leading edge of the airfoil after 8100 h of service. A blade never used in service after refurbishment was not available to determine if the blade was properly heat-treated.

Analysis

1. Visual Inspection

The crack developed by the blade consisted of segments oriented at $90^\circ$ and $45^\circ$ from the stacking axis of the blade, suggesting that the maximum applied tensile stress was oriented parallel to the stacking axis (see Fig. 12.41a). The blades showed no evidence for corrosion attack suggesting that the failure was mechanically related. It was also noted that the surface coating had lost its adherence to the alloy substrate at various locations. However, there was no macrostructural evidence for distortion or overheating. On a macroscopic scale, the exposed fracture surface was of a granular nature (see Fig. 12.41b) typical of a fracture produced by crack propagation along grain boundaries. Secondary intergranular cracks were also observed near the open crack. Absence of marked macroscopic deformation suggested that the plastic deformation involved in crack propagation was of a highly localized nature.

2. Material Verification

The blade material was identified as Udimet alloy 710. The nominal composition (wt%) of the alloy is given below.

<table>
<thead>
<tr>
<th>Element</th>
<th>Udimet 710</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni</td>
<td>Balance</td>
</tr>
<tr>
<td>Co</td>
<td>15</td>
</tr>
<tr>
<td>Cr</td>
<td>18</td>
</tr>
<tr>
<td>Al</td>
<td>2.5</td>
</tr>
<tr>
<td>Ti</td>
<td>5</td>
</tr>
<tr>
<td>Mo</td>
<td>3</td>
</tr>
<tr>
<td>W</td>
<td>1.5</td>
</tr>
<tr>
<td>B</td>
<td>0.02</td>
</tr>
<tr>
<td>C</td>
<td>0.07</td>
</tr>
</tbody>
</table>

This alloy is suitable for the application provided that the elevated temperature creep resistance is not of a particular significance.

3. Microstructural Characterization

A typical microstructure of the center of airfoil section is shown in Fig. 12.41c. In addition to strengthening $\gamma'$ phase assuming cubical morphology, the alloy
Figure 12.41  Continued
4. Mechanism of crack propagation

First, intergranular cracks were developed at grain boundaries as a result of creep deformation. Second, the coating could only accommodate deformation in the substrate by opening cracks, eventually leading to its spallation, and exposing the alloy to the environment. Finally, oxygen from the environment diffused inward along the grain boundaries causing oxidation.
Mode of Failure

Mode of failure was intergranular creep rupture of the alloy. Creep rupture was accelerated by the development of cellular M 23C 6 carbide structure at the grain boundaries within the leading edge. Creep deformation in the alloy was accommodated by opening cracks in the coating due to lack of sufficient ductility leading to spallation.

Cause of Failure

It is unlikely that localized overheating such as that encountered during heat treatment could lead to observed microstructural features. Instead, variations in firing temperature exceeding the temperature capability of the alloy particularly within the leading edge is thought to be the most probable cause of failure.

Solution

If the same operating temperatures are to be maintained, it is recommended to consider replacing Udimet 710 with another first-stage alloy of higher creep strength, such as a directionally solidified alloy.

12.7.4 Failure of a Cracked Gear

Process

An extruder assembly was used for processing resin in a petrochemical plant. Power for the extruder was transmitted through a gear box designed for a gear ratio of 14:5:1. The extruder had been in service for 17 years.

Application

The application involved a forged helical gear that is used to transmit motion between parallel or crossed shafts. It can also be used to transmit motion between a shaft and a rack. Characteristically, two or more teeth of each gear are always in contact.

Material

According to specification, the gear was made of Japanese S45C grade carbon steel.

Service Conditions

The gear had been operating at 40°C and at an output speed of 222 rpm.
Problem

After 14 years of operation, the gear box produced an abnormal sound. Although the gear box was put under surveillance, it was not until 6 months later that the gear box was opened. At that time, two broken teeth were found.

Analysis

1. Visual Inspection

There was no evidence for the presence of fragments or shavings of steel in the gear box, suggesting that wear could be ruled out as a possible cause of failure. Also, there was no evidence for macroscopic deformation or distortion indicating that the stress level remained below the macroscopic yield strength of the material at the time of failure. Therefore, plastic flow could be ruled out as a possible cause of failure.

Substantial portion of the tooth was fractured. Examination of cross section of the as-received gear revealed the presence of secondary cracks extending from the root, as shown in Fig. 12.42a. Both of these features are characteristics of breakage as a failure mechanism.

2. Material Verification

According to standard specifications, the majority of gears are made of carbon steels and are case-hardened to promote their resistance to fatigue. The composition (wt%) of gear, as shown below, was consistent with that of a carbon steel.

<table>
<thead>
<tr>
<th>Element</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>97.50</td>
</tr>
<tr>
<td>Si</td>
<td>1.19</td>
</tr>
<tr>
<td>Mn</td>
<td>0.73</td>
</tr>
<tr>
<td>C</td>
<td>0.57</td>
</tr>
</tbody>
</table>

3. Microstructural Characterization

The microstructure of the alloy was composed of a mixture of ferrite and pearlite typical of carbon steel. The hardness of this region was Rb 97, characteristic of an annealed material. The carburized layer at the surface was maintained at the time of failure and exhibited high hardness (Rc 55).

In conclusion, proper material was selected for the application. Also, the material maintained its expected structure and strength at the time of failure. There was no apparent evidence that could correlate the failure with defects in the material.
Figure 12.42  (a) A photograph illustrating macroscopic cracks at the root of a gear tooth in the as-received condition. (b) Secondary electron SEM image of the tooth fracture surface illustrating fatigue striations.
Mode of Failure
The morphology of the fracture surface indicated that the cracks leading to fracture had propagated at least partially by a fatigue mechanism, as shown by typical fatigue striations in Fig. 12.42b. The striations had a uniform spacing suggesting that crack growth was rather continuous. Occasionally, striations free regions were observed at the fracture surface. These relatively flat facets suggested that cracks were also propagated by a quasi-cleavage mechanism, which is common in body-centered cubic metals (ferrite). In this case, the cracks follow certain crystallographic planes. This may explain the branching of secondary cracks observed along the cross section of the gear below the fracture surface.

To summarize, most evidence pointed out that the failure mode of the gear was fatigue breakage, which could have been initiated at the roots. Exact location at which the fracture initiated could not be determined due to unavailability of the entire fracture area.

Cause of Failure
Fatigue breakage of gears is caused by repeated bending stress, which exceeds the endurance limit (fatigue limit) of the material. It is thought that at or prior to the time of failure, the gear was subjected to a local stress level exceeding the endurance limit. Poor design, surface or subsurface defects, and overload could be ruled out as possible causes that could lead to stress raiser. Therefore, it appears that the breakage was caused by misalignment, which could at least partially explain the noise produced by the gear 6 months prior to the failure.

Conclusion
Most evidence pointed out that the gear had failed by fatigue breakage, which could have been caused by misalignment. Although there is an indication that the crack growth had propagated at a relatively rapid rate, it is not possible to correlate exact fatigue life with crack growth rate.

12.7.5 Failure of Copper Tapes Used as Metallic Screening for Medium-Voltage Copper Cables

Process
Copper tape covered with an organic transparent coating was used as metallic screening for medium voltage copper cables. Such tapes are used to carry out the earth fault current.

Service Conditions
The tape was exposed to both heat and moisture. During the summer, the temperature reached a maximum of about 55°C.
Problem

After about 10 years of service, the copper tape was discolored. The tape developed a dark brown/purple color at its outer surface. The inner surface developed a dark blue/dark purple/reddish color, and in one area the tape was perforated. Surrounding the perforation was a black deposit. Also, a greenish deposit was observed at the inner surface.

Analysis

1. Visual Inspection

Visual inspection suggested that the tape was oxidized likely by exposure to both heat and moisture. However, the tape appeared to have been exposed to higher temperature, as indicated by localized melting leading to perforation.

2. Microstructural Characterization

Analysis by SEM/EDS of the discolored region at the outer surface of tape indicated the presence of Cu and O with a smaller concentration of C. X-ray diffraction corresponding to the region showed the presence of Cu and Cu$_2$O. The green deposit at the inner surface of the tape showed a high concentration of S. X-ray diffraction, however, could not detect any S-bearing compound. In addition, Cu and O were also present. X-ray diffraction showed Cu$_2$O and CuO. Carbon was found to be the major elemental constituent of the black region surrounding the perforation.

Mode of Failure

It could be concluded from the results above that the discolored regions contained a mixture of Cu$_2$O and CuO; however, a high S concentration was observed, particularly at regions containing the green deposit. To summarize, the tape was discolored as a result of oxidation.

Copper oxides could form by direct reaction between Cu and O or by condensation of moisture (dew formation) and decomposition of copper sulfate. Although the nominal temperature during service was 55°C, it was evident that the tape was locally overheated to the extent of melting (hot spot), as indicated by the observed perforation and resolidified Cu. This behavior could result from excessive electric current carried by the tape. Therefore, overheating appeared to be one cause of oxidation leading to discoloration.

Dew formation could also lead to oxidation as shown below:

- Cathode reaction: $O_2 + 2H_2O + 4e^- \rightarrow 4OH^-$
- Anode reaction: $2Cu$(metal) $\rightarrow 2Cu^{2+}$(ions in solution) + $4e^-$
- Reaction in solution: $2Cu^{2+} + 2(OH)^- \rightarrow Cu(OH)_2$
Copper hydroxide resulting from the above reaction could readily transform into CuO by dehydration (drying), i.e.,

\[ \text{Cu(OH)}_2 \rightarrow \text{CuO} + \text{H}_2\text{O} \]

Establishing an aeration cell as described above requires the presence of an electrolyte, which could be CuSO\(_4\). It is also possible that CuSO\(_4\) could decomposed by heating resulting in the formation of CuO. Therefore, it is possible that the remaining volume fraction of CuSO\(_4\) was too small to be detected by X-ray diffraction.

Since various colors were observed at the surface of the tape, it is likely that the exact color at a given region was developed as a result of mixing different types of oxides.

**Cause of Failure**

The tape was discolored by oxidation to various types of oxides, including CuO and Cu\(_2\)O. The oxidation could have taken place by direct reaction between copper and oxygen as a result of overheating as well as by contamination with sulfur leading to the formation of copper sulfate and its subsequent decomposition into an oxide. However, dew formation could have also contributed to oxidation. The tape was also subjected to localized overheating leading to melting.

**Recommendation**

Checking the current load carried by the tape to avoid overheating was recommended. Checking the exact source of sulfur contamination was also recommended.

**12.7.6 Failure of Ethylene Furnace Tubes**

**Process**

In an ethylene (C\(_2\)H\(_4\)) production plant, the process stream in a furnace consisted of a mixture of steam and ethane, which was passed through a coil of reaction tubes externally heated to a temperature of 950–1150°C. The gas temperature was raised quickly and passed through the coil at a high velocity with short residence time. The temperature in the reaction tube was approximately 850°C. Cracking of ethane produced free C according to the reaction:

\[ \text{C}_2\text{H}_4 \rightarrow \text{CH}_4 + \text{C} \]

Carbon produced by the above reaction was deposited at the internal surface of the tube wall as adherent coke. The coke was removed by shutting off the hydrocarbon feed and passing air and steam through the coil. Such a process is known as decoking. Frequent decoking accelerates the thermal damage of the tubes, while less frequent decoking increases the rate of carburization attack.
Problem
Three tubes, which formed part of a coil in the furnace, sagged.

Material
The nominal compositions of the tubes were 25Cr-35Ni-Fe.

Analysis
1. Visual Inspection
Visual inspection revealed that the tubes were sagged to an extent that they had lost their circular shapes to a considerable degree. The scale formed at the surface of the tubes was adherent and there was no evidence of flaking. Effects of general or localized wall thinning were not observed in any of the tube samples. Moreover, cracking in tube walls was also not evident.

2. Material Verification
The material composition was confirmed as stated above. Based on the nominal composition, these alloys are generally suitable for the kind of service conditions encountered in an ethylene furnace.

3. Microstructural Characterization
A typical microstructure obtained from an uncarburized region of the tube material is shown in the secondary electron SEM image of Fig. 12.43a. The carbide precipitation appears relatively coarse both at the austenite grain boundaries and within the matrix itself. This is corroborated by EDS analysis where the Cr concentration was determined to be as low as 13 wt% at various regions within the austenite. The depletion of Cr within the matrix indicates a tendency of Cr to diffuse and form Cr-rich carbides at the austenite grain boundaries and/or coarsen preexisting carbides within the austenite. Diffusion of Cr is enhanced at an elevated temperature. The higher the temperature, the greater the carburized zone and the coarser the carbide precipitates.

The carburized region showed a high degree of carbide precipitation at the grain boundaries and the appearance of blocky carbide particles in the matrix as shown in Fig. 12.43b. The grain boundary precipitate was predominantly comprised of Cr and was present in the form of a continuous network. The white precipitates adjacent to the grain boundaries were Nb rich. In contrast, the carburized region of an undeformed tube showed a low density of precipitation.

Examination of cross sections of the tubes revealed carburization attack. The average depth of carburized zones in the tubes was 58% of the total wall thickness, as compared to 22% for an undeformed tube obtained from a different location of the furnace.
Figure 12.43  Secondary electron SEM images obtained from (a) uncarburized and (b) carburized regions of the tube.
Relatively small additions of Nb to the heat resistant castings can increase their resistance to thermal shock. Furthermore, Nb acts as a carbide stabilizer by forming MC-type carbides which prevent massive carbide precipitation at the grain boundaries. Presence of Nb was not detected during material verification by SEM/EDS, probably due to its low concentration.

Massive carbide precipitation at the austenite grain boundaries, forming a continuous network and the presence of carbides as coarse blocky particles within the matrix, indicates exposure of the furnace tubes to an excessively high temperature leading to heavy carburization attack.

4. Analysis of the Scale

The scale was discontinuous and broken, thus providing minimal protection to the underlying alloy from the diffusing species. The oxide formed at the surface was Cr-rich while Fe and Ni could also be detected in the scale. In addition, Si and to some extent Cr rich oxide was detected at the austenite grain boundaries immediately beneath the alloy surface. The Cr content of the austenite grains in this region was as low as 4.2 wt%, which is not adequate to sustain a continuous protective Cr₂O₃ scale.

Based upon the above observations, it could be concluded that the alloys were incapable of developing a protective oxide scale. This could partly be due to the relatively low oxygen potential environment typically encountered in ethylene production. The presence of a protective oxide scale can impede C diffusion into the alloy and play an important role in reducing the extent of carburization attack.

**Effect of Carbide Precipitation**

The microhardness within the carburized zone was found to be around HV 316 which is relatively high. Carburized zones depicted higher hardness values than the uncarburized regions. Increased hardness at the carburized zone can be attributed to the high degree of carbide precipitation and also to C pickup due to the carburizing environment. Hardness measurements combined with the microstructural observations suggest that the carburization attack led to the formation of a surface hardened layer of reduced ductility. Formation of this layer further corroborates the inability of the alloy to form a continuous protective oxide scale at its surface.

**Mode of Failure**

The tubes examined in this study exhibited carburization attack. The average depth of carburized zones is 58% of the total wall thickness. Moreover, visual examination also revealed that the tubes were clearly sagged, indicating significant plastic deformation during service. This was also corroborated by SEM/EDS examination, where the microstructures exhibited relatively coarse blocky
carbides within the austenite matrix and continuous carbide networks at the grain boundaries. These observations suggest that the alloy had been exposed to an excessively high temperature during service. The hardness of carburized zone was also found to be relatively high, indicating higher precipitation and C pickup by the alloy. Deposition of coke at the inner pipe wall also promotes C diffusion and precipitation of secondary carbides within the alloy.

**Cause of Failure**
Experimental data shows that all the tubes had undergone carburization attack due to an exposure to excessively high temperature during service.

**Recommendation**
It was recommended that the temperature of the furnace should be controlled closely in order to avoid overheating during ethylene production as well as decoking.

### 12.7.7 Failure of a Cracked Heat Exchanger Tube

**Process**
A tubular heat exchanger was used to cool a kerosene pump mechanical seal in an oil refining installation. A slip of kerosene was acquired from the pump, discharged at 244°C and was cooled by secondary cooling water at ambient temperature. The heat transfer took place through the heat exchanger such that kerosene flowed through the tube while the cooling water was in contact with the external surface of the tube. The secondary water that was circulated in the refinery for cooling purposes had been potable water containing a nitrite inhibitor. The secondary water in turn was cooled by seawater. Leakage of seawater into the secondary water had taken place in the past.

**Problem**
The heat exchanger was opened after a leak was detected. The exchanger tubes were heavily fouled at the external surface (secondary cooling water side). The fouling was removed by 5% phosphoric acid solution with water and the leak source was found to be cracking of the heat exchanger tubes.

**Component**
Heat exchanger coiled tube

**Material**
The heat exchanger tube material was of type 304 stainless steel.
Analysis

1. Visual Inspection

Visual examination of the damaged area of the tube revealed a primary circumferential crack. The crack had penetrated through the wall thickness of the tube. The primary crack had branched off into the longitudinal direction of the pipe giving rise to other cracks. Some of these cracks had also penetrated through the thickness of the tube wall. The cracks originated at the external surface of the tube (secondary cooling water side) and traversed through the wall towards its internal surface.

2. Analysis of Deposits at the Tube Surface

Remnants of deposits present at the external surface of the tube were examined in a scanning electron microscope. The elemental composition of the deposit comprised of Ca, Mg, O and C. Minerals such as calcium and magnesium are relatively insoluble in water and can form scale deposits on heat transfer surfaces when exposed to conditions commonly found in cooling water systems. Calcium carbonate and magnesium salts are typical deposits observed at these surfaces. No evidence of pitting was observed at any region of the damaged tube.

3. Examination of Fracture Surface

The fracture surface appears flat and brittle in nature. In addition to the alloy elements (Fe, Cr, Ni, Si) and scale deposit constituents (Ca, Mg, S, C, O), the spectrum indicates the presence of small amounts of chlorine, which is attributed to a possible leakage of seawater into the secondary cooling water.

4. Analysis of Tube Cross Section

Microstructural examination of the damaged section was conducted through the fracture surface in order to determine the nature of cracking. The cracks originated from the outer surface of the tube and exhibited multiple branching. Moreover, the transgranular nature of cracking was also evident. Transgranular cracking indicates that the material has received proper heat treatment and microstructure is not sensitized. Evidence of corrosion was found within the cracks. EDS spectra obtained from the cracks indicated that the deposit present within the cracks was a combination of the scale deposit observed earlier at the external tube surface and probably an oxide of Cr. A significant observation was the presence of Cl peaks in each of the EDS spectra obtained from the cracked regions.

Mode of Failure

The tube material cracked from the external surface which was in contact with the cooling water. The cracks penetrated through the tube wall thickness, resulting in leakage. The fracture surface was relatively flat and the cracking was
transgranular exhibiting multiple branching. Elemental analysis of the crack showed the presence of chlorine. Pitting was not observed at any region in the damaged section. This evidence indicates that the type 304 stainless steel tube failed due to stress corrosion cracking which was initiated due to the presence of Cl in the system and probably aided by fouling deposits present at the external surface of the tube.

Cause of Failure

The deposits formed at the external surface of the tube consisted primarily of Ca and Mg salts which had precipitated at the hot surface. The deposits could have provided sites for Cl concentration which was introduced into the system by a leakage of seawater. The presence of Cl is thought to have initiated chloride stress corrosion cracking in the tube material which resulted in failure.

Austenitic stainless steels can crack in aqueous solutions that contain as little as 2 ppm of chloride at 200°C. Sources of stress in the system can be varied. In heat exchangers, thermal gradients can create strains of much greater equivalent stresses than the residual or applied stresses. These thermal stresses can produce cracking if the cold side of the heat exchanger has an environment (Cl in this case) conducive to stress corrosion.

Conclusions

Experimental results indicate that the type 304 stainless steel heat exchanger tube failed due to stress corrosion cracking. The cracking was induced due to a combination of Cl and stresses generated by thermal gradients in the system. The source of Cl is thought to be leakage of seawater into the secondary cooling water. Presence of deposits on the tube surface is also thought to have contributed toward material failure.

Recommendations

It was recommended to take measures that would prevent foulant formation at the tube surface. Fouling deposits already formed should be removed on a regular basis. Scale formation at the surface should also be prevented by the use of appropriate inhibition chemicals and scale conditioners.

Steps should be taken to eliminate the presence of Cl from the system by preventing the leakage of seawater into the secondary cooling water. Moreover, any deviation from normal operating conditions should be closely monitored since it can introduce impurities even in controlled environments.
12.7.8 Failure of a Boiler Superheater Tube

Process
Boilers at a fertilizer plant were used to produce superheated steam at about 413°C. However, the actual temperature could have been 455°C.

Application
A steel superheater tube was involved in the failure.

Material
The failed superheater tube was made of A213 Grade T22 low-alloy steel.

Service Conditions
Natural gas is combusted at the outer surface of the tube to maintain the temperature of the steam inside the tube at about 455°C.

Problem
Rupture in superheater tubes was detected 6 months into service.

Analysis
1. Visual Inspection
Swelling of the tube was apparent which could result from localized plastic deformation. Scaling was observed on both the outer (fire side) and inner (steam side) surfaces of the tube. On both surfaces a black deposit was found suggesting that the tube was subject to high-temperature oxidation. Inspection of the tube cross section revealed localized wall thinning corresponding to the swollen portion, suggesting that the tube was overheated (Metals Handbook, 1975).

Most evidence gathered from the results of visual inspection described above suggested that the tube was ruptured as a result of prolonged overheating. Metallurgical evaluation confirmed this conclusion.

2. Verification of Tube Material
A typical grain structure of the as-received section as viewed at different magnifications is shown in Fig. 12.44a. In comparison with published grain structures of carbon and tube steels (Metals Handbook, 1973), it could be concluded that the grain size corresponding to Fig. 12.44a is about ASTM No. 9. It is evident from Fig. 12.44a that the microstructure was essentially free of pearlite (lamellar mixture of ferrite and cementite). Body-centered cubic structure of α-Fe (ferrite) with a lattice constant of about 0.297 nm was identified by X-ray diffraction technique. The nominal chemical composition (wt%) of A213 Grade T22 steel in comparison with the SEM/EDS are shown below.
<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>96.22</td>
</tr>
<tr>
<td>Cr</td>
<td>1.90–2.60</td>
<td>2.19</td>
</tr>
<tr>
<td>Mo</td>
<td>0.87–1.13</td>
<td>1.10</td>
</tr>
<tr>
<td>Si</td>
<td>0.50*</td>
<td>0.49</td>
</tr>
<tr>
<td>Mn</td>
<td>0.30–0.60</td>
<td>ND</td>
</tr>
<tr>
<td>C</td>
<td>0.15*</td>
<td>ND</td>
</tr>
<tr>
<td>P</td>
<td>0.03*</td>
<td>ND</td>
</tr>
<tr>
<td>S</td>
<td>0.03*</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.
The observed composition can be seen to be in agreement with the nominal composition confirming that the tube material was A213 Grade T22 steel tubing. Both carbon and low-alloy steels such as grade A 213 Grade T22 are recommended for use in boiler applications including superheater tubes (Metals Handbook, 1977), ruling out selection of wrong material as a probable cause of failure.

3. Analysis of Surface Scale

Scaling was observed on both the inner and outer surfaces of the tube. However, as shown in Fig. 12.44b, the thickness of scale layer developed on the outer surface (fire side) was about 4 times that developed on the inner surface (steam side). However, there was no evidence for any significant internal penetration. The scale developed on both surfaces consisted of an Fe-rich oxide. X-ray diffraction showed that the scale predominantly consisted of Fe_3O_4.

It is evident from the results of Fig. 12.44b that the oxidation kinetics at the outer surface of the tube were accelerated in comparison with the inner surface, which could be related to a higher temperature and/or a greater oxygen potential at the fire side (outer surface). Both the nature and extent of scaling suggested that it was not a major factor contributing to the failure as further confirmed in the next section.

Figure 12.44  (a) Secondary electron SEM image illustrating the grain structure as viewed at different magnifications. (b) Secondary electron SEM image of the tube cross-section illustrating comparative thicknesses of the surface scale developed at the inner and outer surfaces. (c) Backscattered electron image illustrating intergranular cracking associated with a grain boundary phase of a larger average atomic number in comparison with the matrix phase (phase exhibiting bright contrast).
4. Metallurgical Structure of the Tube

The as-received section of the tube consisted essentially of ferrite. Near the outer and inner surfaces of the tube, the grains had about the same size suggesting that the temperature was rather uniform across the tube wall. Precipitates within the grains and at grain boundaries were observed near both surfaces. Detailed analysis revealed that the tube material contained more than one type of precipitate as demonstrated below.

As an example, Fig. 12.44c illustrates the presence of a rather continuous layer of a grain boundary phase. The main metallic constituents of the phase were Fe and Mo, along with Si. Intergranular cracking is also clearly visible. Based upon this data, the grain boundary phase could be a carbide of the type \( M_6C \) (\( M = \text{metal} \)) such as \( \text{Fe}_3\text{Mo}_3\text{C} \). However, since C could not be detected, an intermetallic mu phase such as \( \text{Fe}_7\text{Mo}_6 \) could be present. It is to be noted that the \( M_6\text{C} \) carbide and mu phase are closely related and an \( M_6\text{C} \) carbide particle could act as a preferred nucleation site for mu phase (Tawancy, 1993). Also, Si could act as a stabilizer of both the \( M_6\text{C} \) carbide and mu phase (Tawancy, 1993). Typically, during exposure at elevated temperatures, precipitation of mu phase is preceded by precipitation of the \( M_6\text{C} \) carbide. Both phases, however, particularly when present as a continuous grain boundary layer could lead to intergranular embrittlement.

Analysis of a darker phase enveloping the Mo-rich grain boundary phase revealed that it consisted of Fe, suggesting that it was an oxide phase. Evidently, some intergranular oxidation had taken place noting that either the carbide or the mu phase could act as an easy path for oxidation.

In addition to the Mo-rich phase at the grain boundaries, the grain boundaries also contained a mixture of two phases which most likely were \( M_23\text{C}_6 \) carbide such as \( (\text{Fe,Cr})_21\text{Mo}_2\text{C}_6 \) and an intermetallic \( \text{CrFe} \) sigma phase. A relationship similar to that between the \( M_6\text{C} \) carbide and mu phase also exists between the \( M_23\text{C}_6 \) carbide and sigma phase (Tawancy, 1993). Some \( M_23\text{C}_6 \) carbide particles were also observed within the grains.

Both the topologically close-packed sigma and mu phases are extremely hard and brittle particularly at relatively low temperatures (Ref. 4). When present at grain boundaries, they could lead to severe intergranular embrittlement. Also, a continuous grain boundary layer of \( M_6\text{C} \) carbide could promote intergranular embrittlement. Surface hardness measurements indicated that the average hardness was \( \text{Rb} 75 \), consistent with the observation that the hard embrittling phases were mostly present at grain boundaries. Therefore, most evidence pointed out that oxidation, particularly along grain boundaries, was a secondary effect not significantly influencing the failure.

Mode of Failure

Based upon the results described above, it could be concluded that the mode of failure was intergranular embrittlement. Experimental results indicated that the
embrittlement was primarily caused by grain boundary precipitation of a continuous layer of M₆C carbide and/or mu phase as well as sigma phase.

**Most Probable Cause of Failure**

Most evidence pointed out that the most probable cause of failure was prolonged overheating of the tube during service. Observation of the metallurgical structure suggested that the tube was overheated to a temperature below the respective Ac₁ (temperature at which austenite begins to form during heating). Although it is not possible to determine the overheating temperature, the results suggested that it could be between 700 and 800 °C; however, it is likely that it was on the lower side of this range.

Overheating is a common cause of failure of steam boilers components (Metals Handbook, 1975). It could result from restriction of flow within the heated tube or form localized hot spots in the tube wall. Mild flow restriction promotes a small degree of overheating favoring stress rupture as observed in the present case. However, a sudden or severe restriction causes rapid overheating leading to tensile failure. Local imbalance of flow among tubes joined to a common header, localized deposits near the tube inlet, and local variations in the inner diameter associated with installation techniques are among the causes of mild flow restriction (Metals Handbook, 1975).

**Recommendations**

Based upon the results of this study, it was recommended to exercise measures to avoid overheating of the tube during service. Since the results suggested that the degree of overheating was relatively small, various causes for mild restriction of flow described above must be considered to combat the problem.

**References**


Ref. 4, p. 185.

12.7.9 Failure of a Corrugated Metal Gasket

Process

A polyethylene production plant contained hydrocarbon transfer lines assembled by bolted flange joints. An asbestos-filled corrugated metal gasket was installed as a part of the flange assembly.

Service Conditions

The service temperature was less than 100°C. The pressure in the transfer line was 17 kg/cm²; however, it was designed for a pressure up to 63 kg/cm².

Component

The component involved in the failure was an asbestos-filled corrugated double gasketed metal gasket. Its dimensions were outer diameter = 1114.6 mm, inner diameter = 927.1 mm, and thickness = 3.5 mm.

Material

According to specification, the metal gasket was made of 304 stainless steel.

Problem

The bolted joint was disassembled for maintenance. Reinstallation involved tightening the bolts, increasing the pressure, retightening, increasing the pressure, etc. until the desired pressure was reached. Initially, no leak was detected at the joint. However, when the pressure was increased to 7 kg/cm², a leak was detected. Additional tightening of the bolts stopped the leak. Therefore, the pressure was subsequently increased to 12 kg/cm². Since no leak was then detected, the pressure was further increased to 17 kg/cm². Again, a leak was detected at the same location but at an increasing rate. When the joint was disassembled and checked, the gasket was found to have ruptured at about the 7 O’clock position, as schematically shown in Fig. 12.45. It is to be noted that although up to 10 bolted joints were used, only 1 gasket was ruptured.

Analysis

Before presenting the results of this study, it is useful to review the loading conditions of bolted joints as well as some design considerations.

1. Bolt Preload and Gasketed Joints

In pressure vessels, the performance of a bolted joint in relation to the degree of tightness depends to a large extent upon the bolt spacing, flange thickness, and properties of gasket materials. For a given fluid in the pipe, the rate of leakage
dependent upon the internal pressure is another factor influencing the joint performance (Blake, 1990).

When designing a flange connection, the available load from bolting must be sufficient to perform three specific functions (Holmes, 1973). First, it causes the gasket material to flow into flange imperfections and overcome any surface irregularities in the gasket itself. When expressed as a minimum load to cause the necessary degree of yield per unit area, a value of minimum initial seating stress is derived. Second, the load derived from bolting compensates for any hydrostatic end thrust normally expressed as a multiple of maximum internal pressure in the pipe and the closed area bounded by the inside sizes of the gasket.
Third, the load allows for sufficient residual stresses in the gasket to maintain its sealing function.

Asbestos is well known to be an effective sealing material. However, metal gaskets have the advantage of high mechanical strength and they are usually selected for rigorous applications to extend the usefulness of asbestos. Steels are among the commonly used metals. Corrugated metal gaskets are particularly useful in that the concentric peaks tend to intensify the seating stress, and the areas in between are filled to advantage with asbestos (Holmes, 1973). These gaskets are recommended for such applications as pressure vessels where the pressure is low or moderate (Holmes, 1973). Among the common causes of gasket failure are improper selection of materials, incorrect gasket thickness, corrosion, and insufficient, excessive, or uneven tightening of bolts (Holmes, 1973).

Based upon the discussion above, it could be concluded that proper materials were selected for the application. Asbestos is well known for its sealing action, and 304 stainless steel is characterized by excellent resistance to atmospheric corrosion. Also, based upon the information provided about the service performance of similar bolted joints, it could be concluded that the joint was properly designed.

2. Visual Inspection

Visual inspection showed no evidence for any corrosion attack. Parts of the gasket in contact with the flange were blown away. Although the remaining sections were relatively flat, the free part of the gasket not in contact with the flange was severely deformed. It was suggested by these observations that the failure was of a purely mechanical origin.

A photograph illustrating the two ends of the ruptured section in the as-received condition is shown in Fig. 12.46a. As can be seen, portions of the inner section in contact with the flange were completely separated and blown away. In comparison with the outer free section which had undergone severe deformation, the inner section was relatively flat. Secondary cracks in the outer section are indicated by the arrows. Photographs of the two ends of the ruptured section at higher magnifications are shown in Fig. 12.46b, c. Closer examination of the ruptured ends indicated that the material had lost its structural integrity by propagation and joining of several cracks, leading to complete separation.

3. Material Verification

The results of EDS analysis (wt%) are shown below in comparison with the nominal composition of 304 stainless steel. Also shown are the results of ICP-AES analysis.
<table>
<thead>
<tr>
<th>Element</th>
<th>Nominal</th>
<th>Measured EDS</th>
<th>Measured ICP-AES</th>
<th>Measured CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>Balance</td>
<td>72.40</td>
<td>74.00</td>
<td>—</td>
</tr>
<tr>
<td>Cr</td>
<td>18–20</td>
<td>18.69</td>
<td>19.09</td>
<td>—</td>
</tr>
<tr>
<td>Ni</td>
<td>8–12</td>
<td>8.43</td>
<td>8.48</td>
<td>—</td>
</tr>
<tr>
<td>Mn</td>
<td>2*</td>
<td>ND</td>
<td>1.17</td>
<td>—</td>
</tr>
<tr>
<td>Si</td>
<td>1*</td>
<td>0.48</td>
<td>0.47</td>
<td>—</td>
</tr>
<tr>
<td>C</td>
<td>0.08*</td>
<td>ND</td>
<td>—</td>
<td>0.059</td>
</tr>
<tr>
<td>P</td>
<td>0.045*</td>
<td>ND</td>
<td>0.049</td>
<td>—</td>
</tr>
<tr>
<td>S</td>
<td>0.030</td>
<td>ND</td>
<td>&lt;0.001</td>
<td>—</td>
</tr>
</tbody>
</table>

*Maximum.
ND: Not detected.
—: Not measured.

It is evident that the measured composition lies within the specified composition, verifying that the gasket metal was 304 stainless steel.
4. Metallurgical Structure of the Gasket Metal

A typical grain structure of the gasket metal derived from the unfailed section is shown in Fig. 12.47a. It is evident from the high density of annealing twins that the material was fully recrystallized. Comparison with published microstructures characteristic of 304 stainless steel (Davis, 1985) showed no unusual features. However, the gasket metal had a much finer grain size (ASTM grain size no. 8–9) in comparison with 304 stainless steel annealed at 1065°C (ASTM grain size no. 5) (Davis, 1985). It is likely that the gasket metal was processed to have such a fine grain size to improve the low-temperature mechanical strength.

Characteristic grain structure of the failed section is shown in Fig. 12.47b, where a secondary crack can be seen. A comparison of the microstructures shown in Fig. 12.47a, b indicated the microstructure of the unfailed section was indistinguishable from that in the failed section.

It could be concluded from the observations above that the failure was not related to any apparent defects in the material.

5. Fracture Mode

Cracks in the failed section of the gasket were mainly transgranular; however, some intergranular cracks were also observed. When viewed at relatively small magnifications, the fracture surface was relatively flat. At higher magnifications, however, beach marks or striations characteristic of fatigue failure were observed, as shown in Fig. 12.47c. The above observations suggested that the metal gasket had failed by a fatigue mechanism. It is to be noted that 304 stainless steel is not recommended for use in fatigue applications.

6. Evaluation of Mechanical Strength

It was observed that the average microhardness of the unfailed section was HV 227 (Rb 96), corresponding to an approximate ultimate tensile strength of 105 ksi (Lyman, 1977). In contrast, the failed sections had microhardness values in the range of about HV 285–300 (Re 28–30), corresponding to ultimate strengths of 133–144 ksi.

It could be concluded from the observations above that the failed sections were subject to a considerable hardening effect. It is likely that the observed hardening had resulted from cold work caused by improper tightening of the bolts. If the bolts are excessive tightened, the gasket is plastically deformed.

Failure Mode

Combining the results above could lead to the conclusion that the failure mode of the metal gasket was fatigue. Cyclic loading conditions could be established when the external applied load exceeds the bolt load. However, since the gasket was plastically deformed, it is apparent that it acquired a finite amount.
Figure 12.47 Continued
of macroscopic strain during each loading cycle corresponding to low-cycle fati-
gue. It is possible that this behavior could result from uneven tightening of the
bolts, i.e., nonuniform loading conditions.

**Cause of Failure**

Most evidence pointed out that the cause of failure was improper tightening prac-
tice of the bolts, subjecting the gasket to nonuniform loading conditions.

**Conclusion**

It could be concluded from the results of this study that the metal gasket had
failed by a low-cycle fatigue mechanism resulting from improper tightening prac-
tice of the bolts.

*Figure 12.47* Secondary electron SEM images of (a) a sound section of the metal gas-
ket. (b) failed section of the metal gasket indicating a secondary crack. (c) fracture surface
illustrating beach marks or striations indicated by arrows.

Copyright © 2004 by Marcel Dekker, Inc.
Recommendations

Based upon the results of this study, it is recommended that the bolts must be tightened according to the prescribed practice ensuring uniform and proper loading conditions to prevent future failures.

References

Tables A.1 and A.2 summarize the AISI classification of carbon and low-alloy steels, and the concentrations of Mn, C, S, and P. In Table A.2, the range of C content is given, and therefore the last two digits of the steel grade are replaced by XX.

Table A.3 illustrates the major alloying elements in some stainless steels and their concentrations, as well as the carbon contents. Some of these steels, however, have accepted designation other than the AISI number as indicated in the table.
Table A.1  Continued

<table>
<thead>
<tr>
<th>AISI no.</th>
<th>Mn</th>
<th>C</th>
<th>S (maximum)</th>
<th>P (maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1018</td>
<td>0.60–0.90</td>
<td>0.15–0.20</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1019</td>
<td>0.70–1.00</td>
<td>0.15–0.20</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1020</td>
<td>0.30–0.60</td>
<td>0.18–0.23</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1021</td>
<td>0.60–0.90</td>
<td>0.18–0.23</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1022</td>
<td>0.70–1.00</td>
<td>0.18–0.23</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1023</td>
<td>0.30–0.60</td>
<td>0.20–0.25</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1024</td>
<td>1.35–1.65</td>
<td>0.19–0.25</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1025</td>
<td>0.30–0.60</td>
<td>0.22–0.28</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1026</td>
<td>0.60–0.90</td>
<td>0.22–0.28</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1027</td>
<td>1.20–1.50</td>
<td>0.22–0.29</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1029</td>
<td>0.60–0.90</td>
<td>0.25–0.31</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1030</td>
<td>0.60–0.90</td>
<td>0.28–0.34</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1031</td>
<td>0.30–0.90</td>
<td>0.28–0.34</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1033</td>
<td>0.70–1.00</td>
<td>0.30–0.36</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1035</td>
<td>0.60–0.90</td>
<td>0.32–0.38</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1036</td>
<td>1.20–1.50</td>
<td>0.30–0.37</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1037</td>
<td>0.70–1.00</td>
<td>0.32–0.38</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1038</td>
<td>0.60–0.90</td>
<td>0.35–0.42</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1039</td>
<td>0.70–1.00</td>
<td>0.37–0.44</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1040</td>
<td>0.60–0.90</td>
<td>0.37–0.44</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1041</td>
<td>1.35–1.65</td>
<td>0.36–0.44</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1042</td>
<td>0.60–0.90</td>
<td>0.40–0.47</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1043</td>
<td>0.70–1.00</td>
<td>0.40–0.47</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1045</td>
<td>0.60–0.90</td>
<td>0.43–0.50</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1046</td>
<td>0.70–1.00</td>
<td>0.43–0.50</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1049</td>
<td>0.60–0.90</td>
<td>0.46–0.53</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1050</td>
<td>0.60–0.90</td>
<td>0.48–0.53</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1051</td>
<td>0.85–1.15</td>
<td>0.45–0.56</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1052</td>
<td>1.30–1.50</td>
<td>0.47–0.55</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1053</td>
<td>0.70–1.00</td>
<td>0.48–0.55</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1055</td>
<td>0.60–0.90</td>
<td>0.50–0.60</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1060</td>
<td>0.60–0.90</td>
<td>0.55–0.65</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1070</td>
<td>0.60–0.90</td>
<td>0.65–0.75</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1078</td>
<td>0.30–0.60</td>
<td>0.72–0.85</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1080</td>
<td>0.60–0.90</td>
<td>0.75–0.85</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1084</td>
<td>0.60–0.90</td>
<td>0.80–0.93</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1085</td>
<td>0.70–1.00</td>
<td>0.80–0.93</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1086</td>
<td>0.30–0.50</td>
<td>0.80–0.93</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1090</td>
<td>0.60–0.90</td>
<td>0.85–0.98</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>1095</td>
<td>0.30–0.50</td>
<td>0.90–1.03</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>AISI no.</td>
<td>Major alloying elements</td>
<td>Nominal concentration</td>
<td>Carbon range</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------</td>
<td>-----------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>13XX</td>
<td>Manganese</td>
<td>1.75</td>
<td>0.30–0.45</td>
<td></td>
</tr>
<tr>
<td>23XX</td>
<td>Nickel</td>
<td>3.50</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>25XX</td>
<td>Nickel</td>
<td>5.00</td>
<td>0.14–0.18</td>
<td></td>
</tr>
<tr>
<td>31XX</td>
<td>Nickel</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>0.65</td>
<td>0.20–0.40</td>
<td></td>
</tr>
<tr>
<td>33XX</td>
<td>Nickel</td>
<td>3.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>1.55</td>
<td>0.10–0.16</td>
<td></td>
</tr>
<tr>
<td>40XX</td>
<td>Molybdenum</td>
<td>0.25</td>
<td>0.23–0.68</td>
<td></td>
</tr>
<tr>
<td>41XX</td>
<td>Chromium</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.20</td>
<td>0.18–0.50</td>
<td></td>
</tr>
<tr>
<td>43XX</td>
<td>Nickel</td>
<td>1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.25</td>
<td>0.20–0.40</td>
<td></td>
</tr>
<tr>
<td>46XX</td>
<td>Nickel</td>
<td>1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.28</td>
<td>0.15–0.40</td>
<td></td>
</tr>
<tr>
<td>47XX</td>
<td>Nickel</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>48XX</td>
<td>Nickel</td>
<td>3.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.25</td>
<td>0.12–0.20</td>
<td></td>
</tr>
<tr>
<td>50XX</td>
<td>Chromium</td>
<td>0.30 or 0.60</td>
<td>0.15–0.46</td>
<td></td>
</tr>
<tr>
<td>51XX</td>
<td>Chromium</td>
<td>0.80/0.95/1.05</td>
<td>0.17–0.60</td>
<td></td>
</tr>
<tr>
<td>61XX</td>
<td>Chromium</td>
<td>0.80/0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vanadium</td>
<td>0.10/0.15 (min)</td>
<td>0.17–0.50</td>
<td></td>
</tr>
<tr>
<td>81XX</td>
<td>Chromium</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.12</td>
<td>0.15–0.27</td>
<td></td>
</tr>
<tr>
<td>86XX</td>
<td>Nickel</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.20</td>
<td>0.15–0.60</td>
<td></td>
</tr>
<tr>
<td>87XX</td>
<td>Nickel</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.25</td>
<td>0.15–0.50</td>
<td></td>
</tr>
<tr>
<td>92XX</td>
<td>Silicon</td>
<td>2.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>0.15/0.30</td>
<td>0.55–0.60</td>
<td></td>
</tr>
<tr>
<td>93XX</td>
<td>Nickel</td>
<td>3.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>1.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.12</td>
<td>0.10–0.14</td>
<td></td>
</tr>
<tr>
<td>98XX</td>
<td>Nickel</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>0.25</td>
<td>0.40–0.50</td>
<td></td>
</tr>
<tr>
<td>AISI no./Other*</td>
<td>Major alloying elements</td>
<td>Nominal concentration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>Chromium 17.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 17.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.15 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>302</td>
<td>Chromium 18.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 9.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.15 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>302B</td>
<td>Chromium 18.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 9.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.15 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>303</td>
<td>Chromium 18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 9.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.15 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>304</td>
<td>Chromium 19.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 10.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.08 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>Chromium 18.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 16.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.12 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>308</td>
<td>Chromium 20.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 11.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.08 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>309</td>
<td>Chromium 23.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 13.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.20 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>310</td>
<td>Chromium 25.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 20.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>316</td>
<td>Chromium 17.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 12.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum 2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.08 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>316L</td>
<td>Chromium 17.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 11.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum 2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.03 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>317</td>
<td>Chromium 18.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 13.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum 3.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.08 (max)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>AISI no./Other*</th>
<th>Major alloying elements</th>
<th>Nominal concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>321</td>
<td>Chromium 18.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 10.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Titanium 5 × %C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.08 (max)</td>
<td></td>
</tr>
<tr>
<td>347</td>
<td>Chromium 18.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 11.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Niobium or Tantalum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.08 (max)</td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>Chromium 13.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.08 (max)</td>
<td></td>
</tr>
<tr>
<td>429</td>
<td>Chromium 15.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.12 (max)</td>
<td></td>
</tr>
<tr>
<td>430</td>
<td>Chromium 16.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.12 (max)</td>
<td></td>
</tr>
<tr>
<td>434</td>
<td>Chromium 17.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum 1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.12 (max)</td>
<td></td>
</tr>
<tr>
<td>436</td>
<td>Chromium 17.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum 1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nb-Ta 5 × %C (min), 0.7 (max)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.12 (max)</td>
<td></td>
</tr>
<tr>
<td>446</td>
<td>Chromium 25.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.20 (max)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nitrogen 0.25 (max)</td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>Chromium 12.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.15 (max)</td>
<td></td>
</tr>
<tr>
<td>416</td>
<td>Chromium 13.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.15 (max)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sulfur 0.15 (min)</td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>Chromium 13.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.15 (min)</td>
<td></td>
</tr>
<tr>
<td>431</td>
<td>Chromium 16.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nickel 1.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbon 0.20 (max)</td>
<td></td>
</tr>
<tr>
<td>440</td>
<td>Chromium 17.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molybdenum 0.75 (max)</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>AISI no./Other*</th>
<th>Major alloying elements</th>
<th>Nominal concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH 15–7*</td>
<td>Carbon</td>
<td>0.70/0.85/1.10</td>
</tr>
<tr>
<td></td>
<td>Chromium</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>Nickel</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td>Aluminium</td>
<td>1.15</td>
</tr>
<tr>
<td>17–4 PH*</td>
<td>Chromium</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>Nickel</td>
<td>4.25</td>
</tr>
<tr>
<td></td>
<td>Copper</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>Niobium</td>
<td>0.25</td>
</tr>
<tr>
<td>AM-356*</td>
<td>Chromium</td>
<td>15.5</td>
</tr>
<tr>
<td></td>
<td>Nickel</td>
<td>4.25</td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>Nitrogen</td>
<td>0.10</td>
</tr>
<tr>
<td>18% nickel</td>
<td>Nickel</td>
<td>18.0</td>
</tr>
<tr>
<td>(maraging steel)</td>
<td>Cobalt</td>
<td>7.75</td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>4.85</td>
</tr>
<tr>
<td></td>
<td>Titanium</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Aluminium</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Carbon</td>
<td>0.03 (max)</td>
</tr>
</tbody>
</table>
APPENDIX B

Units of Measurements in Mechanics

It is evident from the preceding discussions that once units of measurements are assigned to the three basic concepts of mechanics (length, mass, and time), the units of all derived concepts such as force, energy, and stress or pressure are defined. Three systems of units are distinguished: (i) the SI (System International) where the units of measuring length, mass, and time are the meter, kilogram, and second, respectively; (ii) the centimeter-gram-second system (CGS); and (iii) the foot-pound-second (FPS) system. Tables B.1 and B.2 summarize the conversion factors for length and mass among the three systems, respectively.

It follows from Secs. 3.2 and 3.3 that force $F = (\text{mass} \times \text{length})/\text{(time)}^2$. In the SI system, $F = (\text{kg} \times \text{m})/\text{s}^2$, which is called the Newton (N), i.e., if a force of 1 N acts on a body whose mass is 1 kg, the body is accelerated by 1 m/s$^2$. Similarly, in the CGS system, $F = (1 \text{ g} \times 1 \text{ cm})/\text{s}^2 = 1$ dyne, which is

<table>
<thead>
<tr>
<th>Table B.1 Conversion Factors of Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
</tr>
<tr>
<td>1 m</td>
</tr>
<tr>
<td>1 cm</td>
</tr>
<tr>
<td>1 in</td>
</tr>
<tr>
<td>1 ft</td>
</tr>
</tbody>
</table>

1 cm = 10 mm
the force required to accelerate a 1 g body by 1 cm/s². In the FPS, the unit of force is the poundal. A 1 kg force and 1 lb force are also used as units of force in the CGS and FPS systems, respectively. Table B.3 summarizes the conversion factors of force.

Units of pressure and stress can be obtained by dividing the unit of force by the unit of area; i.e., the units of pressure and stress in the SI, CGS, and FPS systems are N/m², dyne/cm², and lb/in², respectively. Table B.4 summarizes the conversion factors of pressure and stress.

Both work and energy have the units of force multiplied by length. In the SI system the unit of work and energy = 1 N × 1 m = 1 J. One erg is the unit of work and energy in the CGS system, where 1 erg = 1 dyne × 1 cm. In the FPS system, the unit of work and energy is the 1 lb ft. It is recalled from Eq. (3.9) that the unit of torque or moment of a force is the same as that of work and energy. Table B.5 summarizes the conversion factors of work, energy, and torque.

It is evident from the definition of strain [Eq. (3.17)] that it has no units, i.e. dimensionless, because it is defined as length/length. Also, Poisson’s ratio [Eq. (3.18)] is dimensionless. Since the elastic moduli [Eqs. (3.32), (3.33), and (3.34)] are defined as stress divided by strain, they have the same units as stress.

### Table B.2 Conversion Factors of Mass

<table>
<thead>
<tr>
<th></th>
<th>kg</th>
<th>g</th>
<th>lb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 kg</td>
<td>1</td>
<td>1000</td>
<td>2.20462</td>
</tr>
<tr>
<td>1 g</td>
<td>0.001</td>
<td>1</td>
<td>0.00220462</td>
</tr>
<tr>
<td>1 lb</td>
<td>0.453592</td>
<td>453.592</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table B.3 Conversion Factors of Force

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>kg</th>
<th>dyne</th>
<th>poundal</th>
<th>lb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 N</td>
<td>1</td>
<td>101972</td>
<td>105</td>
<td>7.23300</td>
<td>0.224809</td>
</tr>
<tr>
<td>1 kg force</td>
<td>9.80665</td>
<td>1</td>
<td>9.80665</td>
<td>70.9316</td>
<td>2.20462</td>
</tr>
<tr>
<td>1 dyne</td>
<td>10⁻⁵</td>
<td>1.01972</td>
<td>1</td>
<td>7.233</td>
<td>2.24809</td>
</tr>
<tr>
<td>1 poundal</td>
<td>0.138255</td>
<td>1.40981</td>
<td>1.38255</td>
<td>0.031081</td>
<td></td>
</tr>
<tr>
<td>1 lb force</td>
<td>4.44822</td>
<td>0.453592</td>
<td>4.44823</td>
<td>32.174</td>
<td>1</td>
</tr>
</tbody>
</table>
**TABLE B.4** Conversion Factors of Pressure and Stress

<table>
<thead>
<tr>
<th>N/m²</th>
<th>kg/cm²</th>
<th>lb/in² (psi)</th>
<th>atmos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 N/m²</td>
<td>1</td>
<td>$1.01972 \times 10^{-5}$</td>
<td>$1.45038 \times 10^{-4}$</td>
</tr>
<tr>
<td>1 kg/cm²</td>
<td>$980.665 \times 10^2$</td>
<td>1</td>
<td>14.2234</td>
</tr>
<tr>
<td>1 lb/in²</td>
<td>$6.89476 \times 10^2$</td>
<td>0.0703068</td>
<td>1</td>
</tr>
<tr>
<td>1 atm.</td>
<td>$1.01325 \times 10^5$</td>
<td>1.03323</td>
<td>14.6959</td>
</tr>
</tbody>
</table>

1 N/m² = 10 dynes/cm²
1 torr = 133.322 N/m² = 1/760 atm.
1 N/m² = 1 Pa
1 psi = 0.704 g/mm²
106 Pa = 1 MPa
1 ksi = 103 psi = 6.985 MPa

**TABLE B.5** Conversion Factors of Work, Energy, and Torque

<table>
<thead>
<tr>
<th>J</th>
<th>erg</th>
<th>ft lb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 J</td>
<td>1</td>
<td>$10^7$</td>
</tr>
<tr>
<td>1 erg</td>
<td>$10^{-7}$</td>
<td>1</td>
</tr>
<tr>
<td>1 ft lb</td>
<td>$1.3559$</td>
<td>$0.7375 \times 10^{-7}$</td>
</tr>
</tbody>
</table>

1 cal = 4.1868 J
APPENDIX C

Moment of Inertia of Selected Cross Sections

Table C.1 gives the moment of inertia of selected cross sections.

<table>
<thead>
<tr>
<th>Cross section</th>
<th>Moment of inertia $I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square $(a \times a)$</td>
<td>$a^4/12$</td>
</tr>
<tr>
<td>Rectangle $(w = b, \text{ height } = d)$</td>
<td>$bd^3/12$</td>
</tr>
<tr>
<td>Circle</td>
<td>$\pi d^4/64$</td>
</tr>
<tr>
<td>Square inclined at $45^\circ$ $(a \times a)$</td>
<td>$a^4/12$</td>
</tr>
</tbody>
</table>