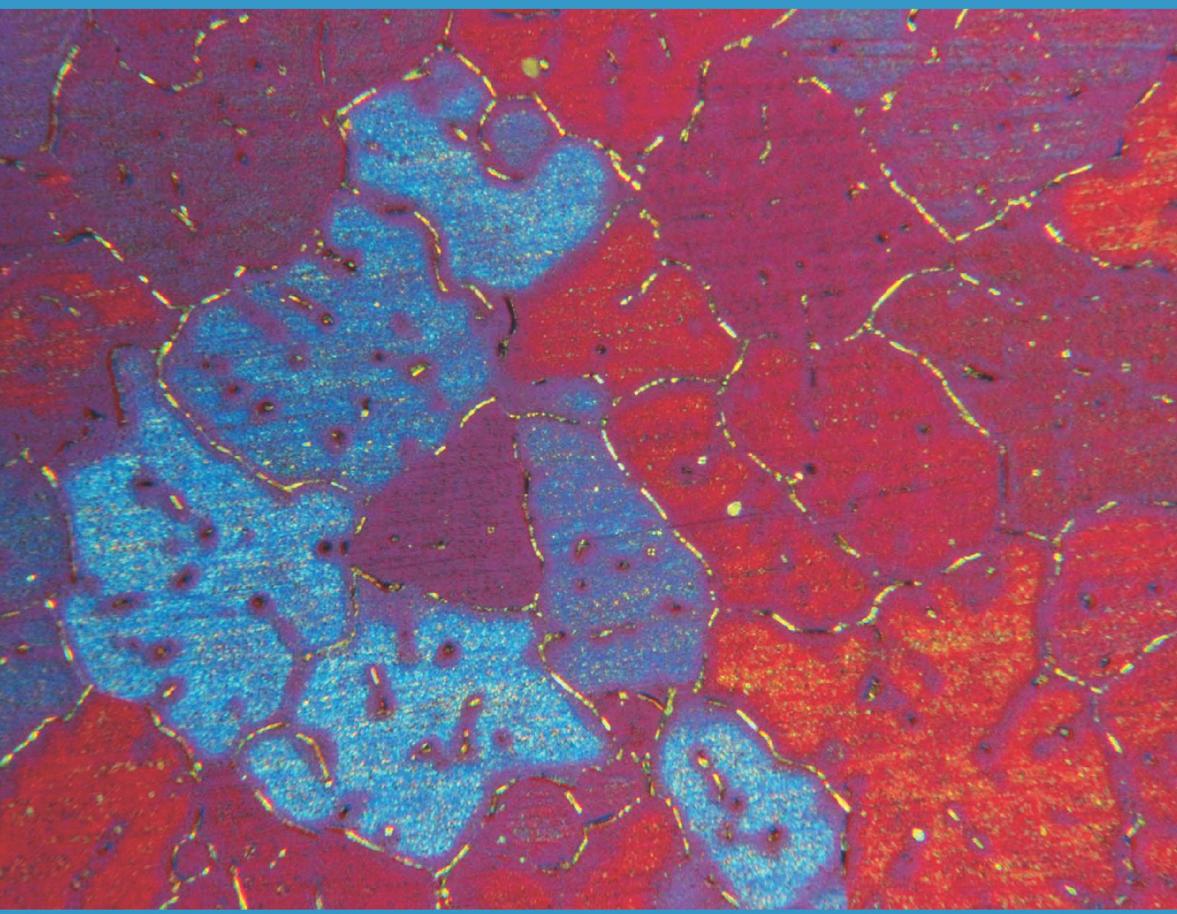


PHYSICAL METALLURGY

PRINCIPLES AND DESIGN



GREGORY N. HAIDEMENOPoulos



CRC Press
Taylor & Francis Group

PHYSICAL METALLURGY

PRINCIPLES AND DESIGN



Taylor & Francis
Taylor & Francis Group
<http://taylorandfrancis.com>

PHYSICAL METALLURGY

PRINCIPLES AND DESIGN

GREGORY N. HAIDEMENOPoulos



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Cover photo is courtesy of Dr. Panagiota Sarafoglu

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2018 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper
Version Date: 20180111

International Standard Book Number-13: 978-1-1386-2768-0 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Dedication

*Dedicated
to all four*



Taylor & Francis
Taylor & Francis Group
<http://taylorandfrancis.com>

Contents

Preface.....	xiii	
Chapter 1	Introduction	1
1.1	What is physical metallurgy	1
1.2	The aim of the book.....	2
1.3	Who should read this book	2
1.4	Book structure	3
1.5	How to read the book.....	4
Chapter 2	Structure of metals.....	5
2.1	Introduction	5
2.2	Crystalline vs. amorphous materials.....	5
2.3	The crystal lattice	6
2.4	The crystal structure of metals	11
2.5	Allotropy.....	16
2.6	Crystal structure effects	17
2.7	Solid solutions	18
2.7.1	Interstitial solid solutions.....	18
2.7.2	Substitutional solid solutions	22
2.8	Intermetallic compounds and intermediate phases.....	23
2.9	A first look at the microstructure of alloys.....	24
2.10	Thermodynamics and kinetics of structure.....	26
2.10.1	Thermodynamic equilibrium	27
2.10.2	Internal energy	28
2.10.3	Enthalpy	29
2.10.4	Entropy.....	29
2.10.5	Gibbs free energy and thermodynamic equilibrium	34
2.10.6	Kinetics of structure.....	35
2.11	Synopsis.....	38
2.12	Review questions	40
Chapter 3	Structural imperfections	41
3.1	Introduction	41
3.2	Point defects	41
3.2.1	Vacancies and interstitials	41
3.2.2	Vacancies and diffusion	45
3.3	Linear imperfections – Dislocations.....	45
3.3.1	The plastic deformation of crystals.....	45
3.3.2	The strength of perfect crystals.....	47

3.3.3	Dislocations and plastic deformation in metal crystals	49
3.3.4	Geometrical characteristics of dislocations	55
3.3.5	Mixed dislocations and dislocation loops	56
3.3.6	Energy and density of dislocations	60
3.3.7	Perfect and partial dislocations, stacking faults.....	60
3.3.8	Dislocation movements: glide, cross-slip, and climb	64
3.3.9	Jogs and kinks, dislocation reactions	65
3.3.10	Plastic deformation due to dislocation glide	66
3.3.11	Energy and stress fields of dislocations	69
3.3.12	Forces on dislocations.....	71
3.3.13	Observation of dislocations.....	74
3.3.14	Dislocation effects	75
3.4	Interfaces	75
3.4.1	Introduction – Interfacial energy.....	75
3.4.2	Free surfaces	77
3.4.3	Grain boundaries.....	77
3.4.4	Interphase boundaries	81
3.4.5	Interface effects.....	84
3.5	Synopsis.....	84
3.6	Review questions	87
Chapter 4	Alloy thermodynamics and phase diagrams.....	91
4.1	Introduction	91
4.2	Free energy of one-component systems (pure metals)	92
4.3	Free energy of solid solutions.....	96
4.4	Chemical potential and thermodynamic equilibrium	101
4.5	The Gibbs phase rule	105
4.6	Equilibrium phase diagrams in binary systems	108
4.6.1	Binary system with complete solid solubility	109
4.6.2	Binary system with a miscibility gap.....	110
4.6.3	Binary system with eutectic point.....	111
4.6.4	Binary system with peritectic point	118
4.7	Examples of phase diagrams	119
4.7.1	The binary system Cu – Zn.....	119
4.7.2	The binary system Fe-C.....	120
4.8	Case study: Solder alloys – The Pb-Sn phase diagram	123
4.9	Synopsis.....	130
4.10	Review questions	131
Chapter 5	Diffusion	135
5.1	Introduction	135
5.2	Diffusion mechanisms	135

5.3	Fick's first law of diffusion – The diffusion coefficient	137
5.4	Random walk and diffusion.....	139
5.5	Fick's second law of diffusion.....	140
5.6	Temperature dependence of diffusion	142
5.7	Thermodynamics and diffusion	146
5.7.1	Driving force for diffusion	146
5.7.2	Concentration dependence of the diffusion coefficient – The thermodynamic factor	149
5.8	Substitutional diffusion.....	151
5.8.1	Self-diffusion in pure metals.....	152
5.8.2	Self-diffusion in a homogeneous solid solution....	152
5.8.3	Substitutional diffusion in a solid solution with concentration gradients – Interdiffusion	153
5.8.4	Review of diffusion coefficients	158
5.9	Irreversible thermodynamics and diffusion	158
5.9.1	Diffusion in multicomponent systems	160
5.9.2	Diffusion by non-chemical forces: Curvature and stress.....	163
5.10	Effects of diffusion	166
5.11	Analytical solutions to the diffusion equation	167
5.11.1	Types of diffusion problems	168
5.11.2	Short diffusion times or semi-infinite media	168
5.11.3	Long diffusion times or finite media.....	173
5.11.4	Application of the Laplace transform in diffusion problems.....	175
5.11.5	Moving boundary problems.....	176
5.12	Numerical methods – Computational kinetics	179
5.13	Synopsis.....	180
5.14	Review questions	181
Chapter 6	Phase transformations.....	185
6.1	Introduction	185
6.2	Nucleation and growth transformations (NGT).....	188
6.3	Nucleation.....	189
6.3.1	The chemical driving force for nucleation.....	189
6.3.2	Homogeneous nucleation.....	190
6.3.3	Heterogeneous nucleation	192
6.3.4	Nucleation strain energy and the morphology of the new phase	196
6.4	Growth	200
6.4.1	Growth processes	200
6.4.2	Interface-controlled growth.....	202
6.4.3	Diffusion-controlled growth	205
6.5	Overall rate of concurrent nucleation and growth	207

6.6	Coarsening	212
6.7	Continuous transformations.....	215
6.7.1	Sharp and diffuse interfaces	216
6.7.2	Spinodal decomposition.....	218
6.7.3	Order-disorder transformations.....	226
6.8	Martensitic transformations.....	230
6.8.1	Basic characteristics of martensitic reactions	230
6.8.2	The shape deformation.....	232
6.8.3	Habit planes and orientation relationships.....	239
6.8.4	Morphology of martensite.....	239
6.8.5	Martensitic interfaces.....	241
6.8.6	Nucleation of martensite	243
6.8.7	Mechanical effects	248
6.9	Effects of phase transformations	254
6.10	Synopsis.....	255
6.11	Review questions	258
Chapter 7	Plastic deformation and annealing.....	261
7.1	Introduction	261
7.2	Mechanisms of plastic deformation.....	262
7.3	Deformation of single crystals by slip	263
7.4	Deformation in polycrystals	270
7.5	Strain hardening.....	272
7.5.1	Dislocation multiplication.....	273
7.5.2	Mechanisms of strain hardening	274
7.5.3	Strain hardening in polycrystals.....	277
7.6	Mechanical twinning	279
7.7	Annealing	283
7.7.1	Annealing processes.....	283
7.7.2	Stored energy	284
7.7.3	Property changes during annealing	284
7.7.4	Recovery	285
7.7.5	Recrystallization	285
7.7.6	Grain growth	289
7.7.7	Grain boundary pinning and grain refinement	292
7.8	Texture in polycrystalline metals.....	295
7.8.1	Development of anisotropy	295
7.8.2	Deformation texture	295
7.8.3	Recrystallization texture	296
7.8.4	Texture effects	296
7.9	Synopsis.....	297
7.10	Review questions	300

Chapter 8	Strengthening mechanisms	301
8.1	Introduction	301
8.2	Slip as a thermally activated process	301
8.3	Overview of strengthening mechanisms.....	306
8.4	Lattice resistance	307
8.5	Solid solution strengthening	310
8.5.1	Substitutional solutes – Symmetric fields.....	311
8.5.2	Interstitial solutes – Non-symmetric fields	313
8.5.3	Cottrell atmospheres, yield point, and strain aging.....	316
8.6	Grain boundary strengthening	319
8.7	Precipitation strengthening.....	322
8.7.1	Aging.....	322
8.7.2	Precipitation strengthening mechanisms.....	327
8.8	Implications of strengthening mechanisms	329
8.9	Synopsis.....	332
8.10	Review questions	333
Chapter 9	Fracture, fatigue, and creep of metals	335
9.1	Introduction – Mechanical behavior of metals	335
9.2	Fracture	336
9.2.1	Introduction – The problem of brittle fracture	336
9.2.2	Ductile and brittle fracture	338
9.2.3	Elements of fracture mechanics	344
9.3	Fatigue	349
9.3.1	Introduction.....	349
9.3.2	Cyclic behavior	351
9.3.3	Fatigue without pre-existing cracks – Strain-life approach.....	355
9.3.4	Fatigue with pre-existing cracks – Damage tolerance.....	363
9.4	Creep.....	368
9.4.1	Introduction.....	368
9.4.2	Dislocation creep	372
9.4.3	Diffusional flow	375
9.4.4	Creep fracture and creep life	377
9.4.5	Design of creep resistant alloys	380
9.5	Synopsis.....	383
9.6	Review questions	385
Chapter 10	Physical metallurgy of steels	389
10.1	Introduction	389
10.2	Phases in steels	390

10.2.1	Solid solutions – Ferrite, austenite, and martensite.....	390
10.2.2	Carbides and nitrides.....	390
10.2.3	Phase mixtures – Pearlite and Bainite.....	391
10.3	The Fe-C phase diagram.....	391
10.4	Alloying elements in steels.....	395
10.5	Phase transformations in steels.....	398
10.5.1	Isothermal and continuous cooling transformations	398
10.5.2	The transformation of austenite to ferrite ($\gamma \rightarrow \alpha$)	399
10.5.3	The transformation of austenite to pearlite ($\gamma \rightarrow P$).....	401
10.5.4	Transformation of austenite to martensite ($\gamma \rightarrow M$).....	404
10.5.5	Transformation of austenite to bainite ($\gamma \rightarrow B$).....	407
10.5.6	Isothermal transformation diagrams	408
10.5.7	Continuous cooling transformation diagrams	413
10.6	Hardenability	415
10.7	Tempering of martensite	420
10.8	Heat treatment of steel.....	423
10.9	Case studies in steels	429
10.10	Synopsis.....	433
10.11	Review questions	435
Chapter 11	Alloy design.....	439
11.1	Introduction	439
11.2	The Alloyneering methodology for alloy design.....	440
11.3	Simulation framework	441
11.3.1	Computational alloy thermodynamics	441
11.3.2	Computational kinetics	444
11.3.3	Phase-field modeling	446
11.4	Simulation examples	452
11.4.1	Intercritical annealing of a medium-Mn steel	452
11.4.2	Simulation of homogenization in Al-Mg-Si alloys	454
11.5	Alloy design: Medium Mn steels	456
11.6	Process design: Multi-pass hot rolling of steels	461
11.7	Synopsis.....	464
Index.....	467	

Preface

The discovery of several, industrially important, advanced metallic alloys, has been based on the fundamental understanding of the interrelations between processing, structure and properties. Physical Metallurgy has evolved, over the years, as the discipline, which aims to provide the fundamental principles of these interrelations. Due to the development of advanced materials characterization techniques and materials modeling on the macro, meso, micro and nano scales, the level of our understanding of the basic principles has risen. Physical Metallurgy is ready to go to the next level. From a descriptive science, used to explain or provide a rationale for the behavior of metallic systems it assumes a more active role: the design of specific metallic materials tailored for specific applications. This is a breakthrough step in the evolution of metallic materials. The current book, highlighted *principles and design*, aims to provide both: the fundamental principles of Physical Metallurgy and introduce the reader to the design methodologies for alloys and processing.

The book has been written primarily for students of engineering. However I believe it will also serve as a useful guide for engineers either in the processing or use of metals and alloys. In preparing the book I have sought to focus on the fundamentals: Thermodynamics and kinetics. They provide the foundation for the study of structure, defects, phase transformations, but also plastic deformation and strengthening mechanisms and, of course, alloy design.

Always the first edition of a book contains numerous errors and omissions, for which I would like to apologize in advance.

I would like to acknowledge several people who contributed, directly or indirectly, to the creation of this book. Prof. Alex Tzavaras, who inspired me as a young student to the field of materials. My teachers at MIT, the late Professors Koichi Masubuchi and Morris Cohen, but also Professor Greg Olson, who inspired me to Alloy Design. My colleague at the University of Thessaly, Prof. Nick Aravas who provided insight, from the mechanics point-of-view on the martensitic transformation. My co-worker in several steel research projects, Prof. Wolfgang Bleck of the Institute of Ferrous Metallurgy of RWTH-Aachen for stimulating discussions. I am also indebted to my former and current doctoral students and coworkers at the Laboratory of Materials: Anna Zervaki, Helen Kamoutsi, Antonis Katsamas, Apostolis Vasilakos, Gioula Sarafoglou, Margianna Tzini and John Aristeidakis for providing teaching support of the subject in class.

I would like to particularly express my gratitude to my co-worker Dr. Helen Kamoutsi for her endless hours on the preparation of the Latex manuscript and the figures of this book. Without her motivation this book would not be made possible. Special thanks to Margianna Tzini for preparing the solution manual and to John Aristeidakis for the proofreading of the manuscript. I would also like to acknowledge the assistance of the Taylor and Francis / CRC people, especially Allison Shatkin, who has been very supportive and provided guidance during the process.

The book would not have been made possible without the understanding and encouragement of my great and wise wife Yiota, who is always there to support new initiatives.

1 Introduction

1.1 WHAT IS PHYSICAL METALLURGY

Physical Metallurgy is the part of metallurgical science, which deals with the shaping of the microstructure of metals and alloys, so that they can obtain desirable properties required in technological applications. The microstructure evolves through different thermal, thermo mechanical or thermochemical treatments, which are applied to the metal at the solid state. The central issue in physical metallurgy is the correlation between processing, structure and properties of engineering alloys. This correlation is shown schematically in Figure 1.1. Processing, structure and properties occupy the

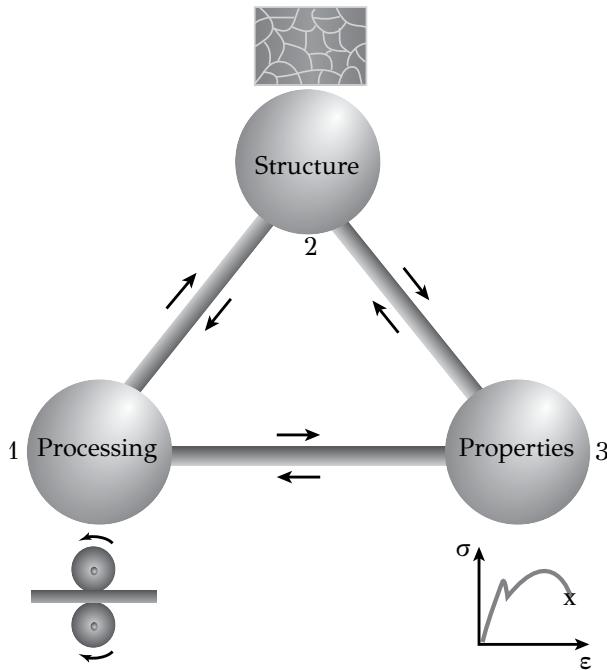


Figure 1.1: Correlation between processing, structure and properties of engineering alloys.

corners of a triangle connected with two-way arrows, signifying the interaction between the corners. As a result, the heat treatment of an alloy modifies and shapes its structure, which then determines the mechanical properties of the alloy (path 1 → 2 → 3). The correlation can be reversed: Suppose that a new alloy is sought with a specified strength. Then the required microstructure could be identified and the chemical composition and processing of the alloy could be designed, following

the reverse path ($3 \rightarrow 2 \rightarrow 1$). This reverse path defines the scientific framework of knowledge-based alloy design, which follows a completely different route from the traditional empirical alloy development approaches. It also explains the word *design* in the title of the book.

The objectives of physical metallurgy can be best understood by an example. It will be shown in the next chapter that metals exhibit a crystal structure, where the atoms are periodically arranged in space, forming crystals. It will then be shown that a metal is made up of numerous crystals, called grains, arranged in a polycrystalline aggregate. It appears that the grain size plays a key role in the mechanical properties of the metal. The finer the grain size the stronger the metal. This means that the use of a fine-grained alloy can lead to significant weight reductions in metallic structures. This is especially important in many applications. One example is steel pipes for the transfer of oil and gas. The demand for transport of large quantities of oil and gas generated the requirement of larger diameter pipes. In order to keep the weight of these pipes low, they should be made of strong steel, so that the thickness and the weight can be kept within reasonable limits. The answer of physical metallurgy to this was the development of high-strength low alloy steels (*HSLA* steels). With microalloying and thermomechanical control processing (*TMCP*), these steels can be made with an ultra fine grain size ($2 - 5 \mu\text{m}$) and obtain high strength to cover the requirements discussed above. The correlation between processing (microalloying and *TMCP*), structure (fine grain size) and properties (strength) made the development of these materials possible.

There is another, equally important, part of metallurgical science, called extractive or chemical metallurgy, which deals with the extraction of the metal from its ore as well as with the processing of the metal in the liquid state (refining, cleaning, deoxidation, alloying). It can be said that chemical metallurgy refines the metal and determines its chemical composition. This is where physical metallurgy starts. Processing to shape the microstructure and achieve the required properties.

1.2 THE AIM OF THE BOOK

The main aim of the book is to present the fundamental principles of physical metallurgy and the mechanisms by which processing can shape the microstructure and properties of metals and alloys. The aim of the book is to help the reader understand the structure of metals and what processes he can apply to modify structure in order to achieve certain properties. The aim of the book is to help the reader understand the behavior of metals subjected to mechanical or thermomechanical loads. Finally the aim of the book is to present the basic principles of computational alloy and process design, which is an emerging and dynamic part of physical metallurgy.

1.3 WHO SHOULD READ THIS BOOK

The book is mainly addressed to students of engineering schools worldwide, where engineering materials or physical metallurgy is a part of the course structure. The book is also addressed to professional engineers seeking a deeper understanding

of the behavior of metals and alloys. It should be noted that the book has a scientific character and deals with the fundamental principles of physical metallurgy. As such, the book does not contain data on specific chemical compositions, processing parameters or properties of commercial alloys. The reader is advised to seek this information in specific handbooks or databases.

1.4 BOOK STRUCTURE

The structure of the book can be realized with the aid of Figure 1.2 depicting two concentric circles. Metals and alloys are used in a large range of applications such

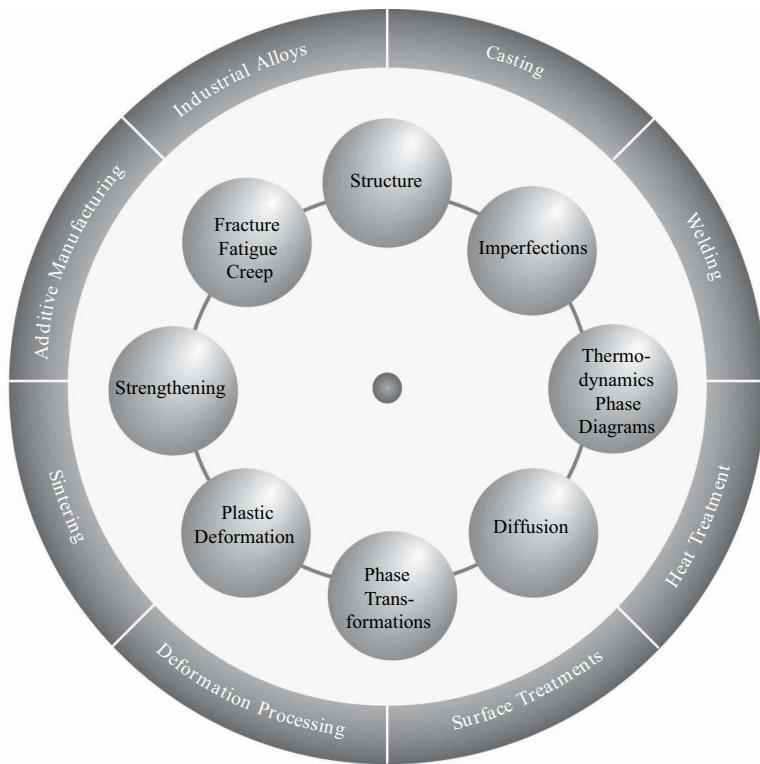


Figure 1.2: Process technologies (external circle) and fundamental principles (internal circle) of physical metallurgy.

as structural, transportation, energy and manufacturing engineering. In order for the metals to obtain the required properties they are subjected to specific processing technologies. These process technologies occupy the external circle. Now imagine yourself at the center of the circle, trying to master these technologies, with which you can design and produce specific alloys, to shape alloys into useful engineering components, to cast, form, weld or heat treat metals or to form engineering compo-

nents by sintering or by novel additive manufacturing methods. In order to achieve this, you should first pass through the internal circle, which is the circle of Physical Metallurgy. You should first understand the basic principles of structure, defects, thermodynamics and kinetics, diffusion and phase transformations, plastic deformation and strengthening mechanisms, mechanical behavior of alloys such as fracture, fatigue and creep. The solid understanding of the internal circle is a prerequisite for the successful application of the technologies of the external circle. The elements of the internal circle correspond to Chapters 2–9 of this book. Two more chapters have been included. Chapter 10 deals with the physical metallurgy of steels. Steels are studied for two reasons: (a) they are still the most important class of metallic alloys for industrial applications and (b) steels are used in this book as a case study for the application of the fundamental principles of physical metallurgy, discussed throughout the book. Finally Chapter 11 deals with the fundamental concepts of alloy design, an emerging part of materials science and engineering, which aims at the knowledge-based design of materials and processes.

1.5 HOW TO READ THE BOOK

There are three ways to read the book. The first is to follow the regular chapter sequence of the book. In this way the reader can get the complete picture of physical metallurgy. The second way is to follow the sequence 1-2-3-4-5-6-10-11, which is focused on phase transformations and the way processing affects the structure of metals and alloys. The third way is the sequence 1-2-3-7-8-9, which is focused on the structure-property relationships, the mechanical behavior and strengthening of metals and alloys.

2 Structure of metals

2.1 INTRODUCTION

The study of structure of metals is the first essential step towards the understanding of their behavior in service and during processing. Metals are crystalline solids, i.e., they exhibit a *crystal structure*. Atoms in a crystal structure occupy positions in an arrangement, which is characterized by *periodicity*. The atomic arrangement is systematically repeated in three-dimensional space. Despite the qualitative descriptions of the crystal structure of metals in the past centuries, it was not until 1912, when X-rays were first used for the study of structure, that a proof of the crystal structure of metals was obtained. The discovery of X-ray diffraction by Laue and the observation, by Bragg, of X-ray scattering from evenly spaced planes within the crystal, set the foundations for the development of Physical Metallurgy. The scientific field concerned with the study of crystal structure is termed *Crystallography*. For the first time scientists had a way to determine the shape and size of the building block of the crystal, the *unit cell*, as well as the arrangement of atoms within the unit cell. The way towards creating links between structure and properties of metals was established. A complete coverage of the science of crystallography is out of the scope of this chapter. Only fundamental elements, especially those which highlight the correlation between structure and properties of metals, will be discussed here. For a more complete treatment, the reader is advised to study the bibliography at the end of this chapter.

2.2 CRYSTALLINE VS. AMORPHOUS MATERIALS

As discussed above the crystal structure is an orderly and periodic array of atoms. Not all materials exhibit a crystal structure. In crystals, when the liquid is cooled below the freezing point, it transforms into a crystalline solid. Some liquids, however, due to either a complex atomic configuration or limited atomic mobility, do not organize into a crystalline arrangement, but instead they form a disordered array of atoms. Actually this disordered state resembles the atomic arrangement in the liquid state and is called a *glass*. Glasses lack long range order and are therefore *amorphous materials*. The transition from the liquid to the glassy or amorphous state is gradual and takes place at the *glass transition temperature* T_g . On the other hand the transition from the liquid to the crystalline state for crystalline materials is abrupt, and takes place at the *melting point* T_m . Although metals are generally crystalline, it is possible, under certain conditions, to produce amorphous metals, called *metallic glasses*. Metallic glasses can be produced by rapid solidification of metal droplets or by vacuum deposition on a substrate at low temperatures and exhibit interesting properties. Glasses are isotropic materials, i.e., they exhibit the same physical and chemical properties in all directions. On the other hand, crystalline materials are

anisotropic properties, such as thermal or electrical conductivity, thermal expansion and elastic constants depend on the direction of measurement. The anisotropy of crystals is exhibited during unconstrained crystal growth, where the external free surface of the growing crystal develops facets, associated with the crystal structure in the bulk of the material.

2.3 THE CRYSTAL LATTICE

As discussed above, a crystal is a periodic array of atoms in three-dimensional space. This structure can be described with respect to a three-dimensional net of straight lines, called the *lattice*. An example is depicted in Figure 2.1. The lattice divides

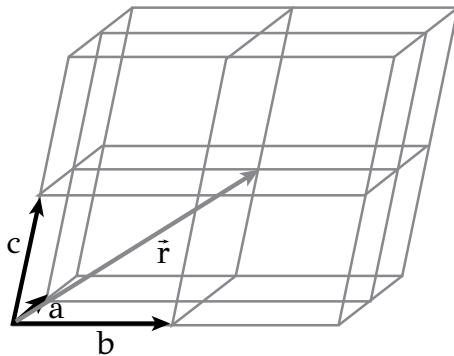


Figure 2.1: The crystal lattice. Periodicity is defined by unit vectors \mathbf{a} , \mathbf{b} and \mathbf{c} .

space into cells, defined by the unit vectors \vec{a} , \vec{b} and \vec{c} . The cells are in contact on all faces and fill space entirely. The cell is then called the *unit cell*, since it plays the role of the building block for the whole crystal. Its repetition in three directions generates the crystal structure. The unit vectors define the periodicity of the crystal in the three directions. Every node of the lattice can be defined by the position vector \vec{r} given by

$$\vec{r} = u\vec{a} + v\vec{b} + w\vec{c}$$

where u , v and w are integers corresponding to the coordinates of the point. The unit vectors form the reference axes of the crystal. Atoms in crystalline materials can be arranged in fourteen possible lattices, called the *Bravais lattices*. The unit cells of these lattices are depicted in Figure 2.2. Each Bravais lattice is associated with a set of reference system of axes, which are not necessarily orthogonal while the unit vectors may not have the same magnitude. In crystallography, seven systems of axes are used for the description of lattices. The seven crystal systems are shown in Table 2.1 together with the relation between length and angles between axes.

For the study of the structure of metals, it is important to describe crystallographic directions and crystal planes. A system of indices is employed. Since the vast majority of metals have either a cubic or a hexagonal structure, only the indices for

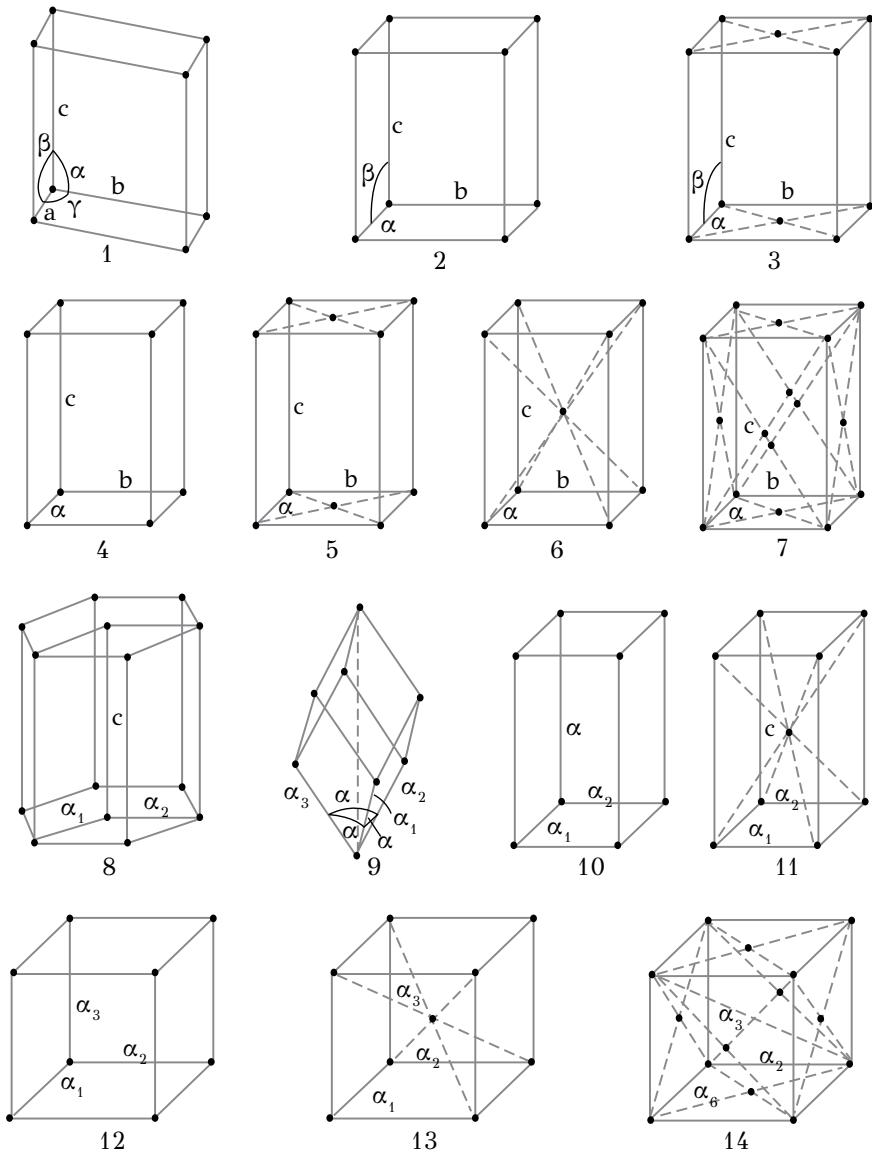


Figure 2.2: The unit cells of the 14 Bravais lattices. (1) triclinic, (2) primitive monoclinic, (3) base-centered monoclinic, (4) primitive orthorhombic, (5) base-centered orthorhombic, (6) body-centered orthorhombic, (7) face-centered orthorhombic, (8) hexagonal, (9) rhombohedral, (10) primitive tetragonal, (11) body-centered tetragonal, (12) primitive cubic, (13) body-centered cubic, (14) face-centered cubic.

Reprinted from: C.S. Barrett, T.B. Massalski, (1966). Structure of Metals, with permission from T.B. Massalski.

these systems will be discussed. The direction of a line connecting the origin with a point, whose coordinates are u , v and w , is described by the notation $[uvw]$. In this notation we do not use fractions but the smallest integers corresponding to the coordinates of the point. In Figure 2.3, since the line connecting the origin 0,0,0 and point 2,2,0 passes also through point 1,1,0, the notation [110] is used to describe this crystallographic direction and any other direction parallel to it.

Table 2.1
The seven crystal systems

Crystal System	Axes and angles	Examples
Triclinic	$a \neq b \neq c$ $\hat{\alpha} \neq \hat{\beta} \neq \hat{\gamma} \neq 90^\circ$	K_2CrO_7
Monoclinic	$a \neq b \neq c$ $\hat{\alpha} = \hat{\gamma} = 90^\circ \neq \hat{\beta}$	$\beta - S$ $CaSO_4 \cdot 2H_2O$ (Plaster)
Orthorhombic	$a \neq b \neq c$ $\hat{\alpha} = \hat{\beta} = \hat{\gamma} = 90^\circ$	$\alpha - S$ Ga Fe_3C (Cementite)
Tetragonal	$a = b \neq c$ $\hat{\alpha} = \hat{\beta} = \hat{\gamma} = 90^\circ$	$\beta - Sn$ TiO_2
Cubic	$a = b = c$ $\hat{\alpha} = \hat{\beta} = \hat{\gamma} = 90^\circ$	Cu, Ag, Au Fe $NaCl$
Hexagonal	$a_1 = a_2 = a_3 \neq c$ $\hat{\alpha} = \hat{\beta} = 90^\circ, \hat{\gamma} = 120^\circ$ Three equal coplanar axes at 120° angle. The fourth axis is perpendicular to them	Zn, Cd, Mg
Rhombohedral	$a_1 = a_2 = a_3 \neq c$ $\hat{\alpha} = \hat{\beta} = 90^\circ, \hat{\gamma} = 120^\circ$ Three equal coplanar axes at 120° angle. The fourth axis is perpendicular to them	As, Sb, Bi

Note: The angle \hat{a} is the angle between axes b and c .

Reprinted from: C.S. Barrett, T.B. Massalski, (1966). Structure of Metals, with permission from T.B. Massalski.

Negative indices are noted with a bar on top of the respective number. So $[0\bar{1}0]$ describes the direction of a line passing through 0,0,0 and 0,-1,0. The four diagonal directions of the cube have indices [111], [1 $\bar{1}$ 1], [11 $\bar{1}$] and [$\bar{1}$ 11]. The group of these

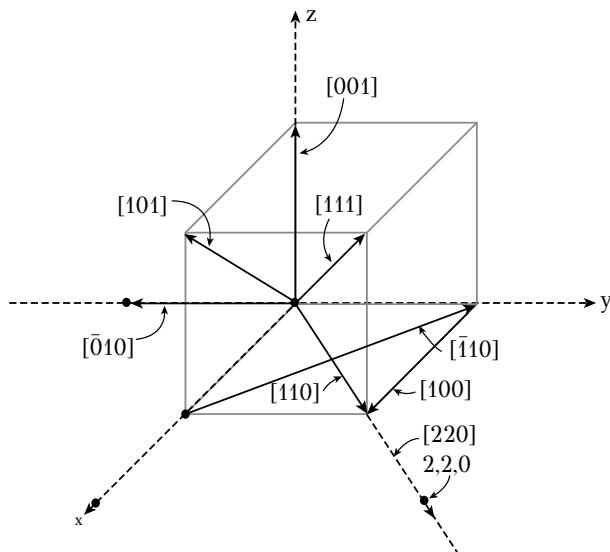


Figure 2.3: Crystallographic directions in the cubic system.

equivalent directions is noted by $\langle 111 \rangle$. In general, in order to describe equivalent directions we use the notation $\langle uvw \rangle$. For the description of crystal planes we use the so called *Miller indices*. These indices describe the orientation of crystal planes with respect to the reference axes, without providing the position of the planes relative to the origin. The indices are derived from the intersection of a plane with the crystallographic axes, using the following procedure: (a) find the intersection of the plane with the three axes, (b) take the inverse of these numbers, (c) find the smallest integers with the same ratio, h , k , l and (d) enclose these integers in parentheses (hkl) . The indexing of some planes is depicted in Figure 2.4.

The plane intersecting the axes in $1/2$, 1 and ∞ has Miller indices (210) . All parallel planes have the same Miller indices. The notation (hkl) refers to a single plane and all the planes parallel to it. A group of equivalent crystal planes is noted by $\{hkl\}$. For example the six faces of the unit cell of a cubic crystal have Miller indices: (100) , (010) , (001) , $(\bar{1}00)$, $(0\bar{1}0)$ and $(00\bar{1})$. The group of the above equivalent crystal planes is noted by $\{100\}$. All equivalent crystal planes of the group $\{hkl\}$ exhibit the same arrangement of atoms.

In the hexagonal system, four indices $(hkil)$ are used. These indices correspond to the intersection of a plane with axes a_1 , a_2 , a_3 and c , as depicted in Figure 2.5. Equivalent crystal planes in the hexagonal system are described by the interchange of the first three indices, since axes a_1 , a_2 and a_3 are equivalent. This means that planes $(\bar{1}\bar{1}00)$, $(10\bar{1}0)$, $(01\bar{1}0)$, $(\bar{1}100)$, $(\bar{1}010)$ and $(0\bar{1}10)$ are equivalent. The first three indices are related by the expression $i = -(h+k)$.

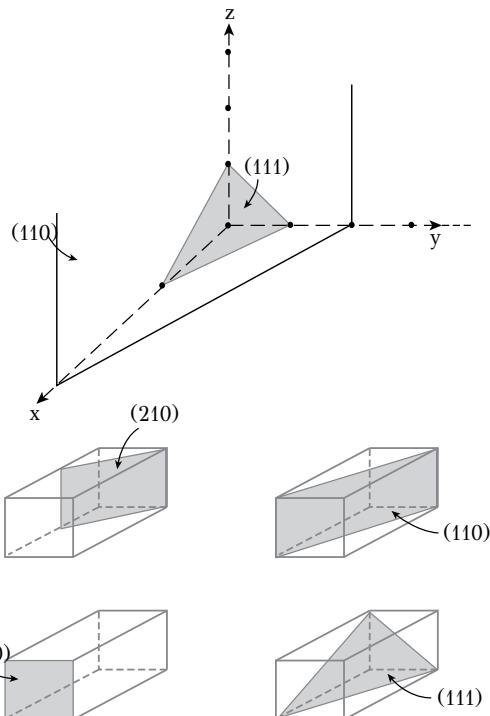


Figure 2.4: Miller indices for the description of crystal planes.

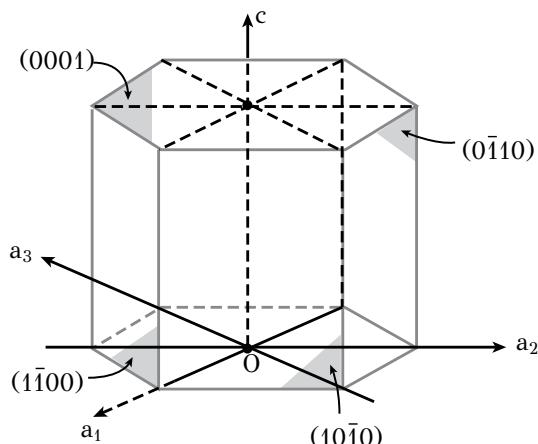


Figure 2.5: Crystal planes in the hexagonal system.

Reprinted from: C.S. Barrett, T.B. Massalski, (1966). Structure of Metals, with permission from T.B. Massalski.

2.4 THE CRYSTAL STRUCTURE OF METALS

The metal atoms, as electron donors, have one to three valence electrons. These electrons are donated to the whole crystal, since there are no other atoms, to act as electron acceptors and form other bonds, as ionic or covalent bonds. In this way the valence electrons are shared between the atoms of the crystal, forming a cloud of free electrons. Metallic crystals consist of positive metal ions surrounded by the electron cloud. In metallic bonding, the attraction between the array of positive metal ions and the free electron cloud balances the repulsion between the positive metal ions and is responsible for the structural stability of the crystal. Free electrons migrate between the immobile positive metal ions and provide the crystal with high thermal and electrical conductivity. The crystal structure can then be described by the lattice of positive metal ions. To continue our discussion of crystal structure, it is convenient to use the *hard-sphere model*, i.e., to describe the metal ions as hard and equally-sized billiard balls. Crystal structure is then derived by the different ways, space is filled with these balls. Goldschmidt and Laves described the basic principles of space filling in metals as:

1. atoms in a crystal structure are stacked in close packed arrangements in order to efficiently fill space
2. atoms occupy positions in a close packed arrangement resulting in the highest symmetry possible
3. atoms are stacked as to have the largest number of nearest neighbors

The number of nearest neighbors, in equal distances from an atom, is called the *coordination number*, *CN*. According to the Goldschmidt-Laves principles, metal atoms form close-packed structures. The basic crystal structures in metals are

1. body-centered cubic (*BCC*)
2. face-centered cubic (*FCC*)
3. hexagonal close-packed (*HCP*)

The unit cells of these structures are shown in Figure 2.6 in two ways. On the left, the position of each atom is represented by a dot. In this way it is easier to describe atomic planes and directions in the cell. On the right, the unit cell is shown using the hard-sphere model, highlighting the size of the atoms relative to the interatomic distances. In *BCC* and *FCC*, the size of the unit cell is defined by the cube edge, called the *lattice parameter* *a*. In *HCP* the unit cell size is defined by two lattice parameters, the edge of the hexagonal base *a* and the height of the hexagonal prism *c*. As depicted in Figure 2.6a, in *BCC*, each cube corner is occupied by an atom, while an additional atom is located at the cube center, in position $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$. On plane (110) the atoms touch along the $<111>$ directions, as depicted in Figure 2.7.

The $<111>$ directions are the *close-packed directions* in *BCC*. As shown in Figure 2.8, the atoms at the cube corners do not belong exclusively in one unit cell but they are shared between the eight neighboring unit cells meeting at the particular corner. The number of atoms corresponding to the *BCC* cell is therefore $8 \times (1/8) + 1 = 2$.

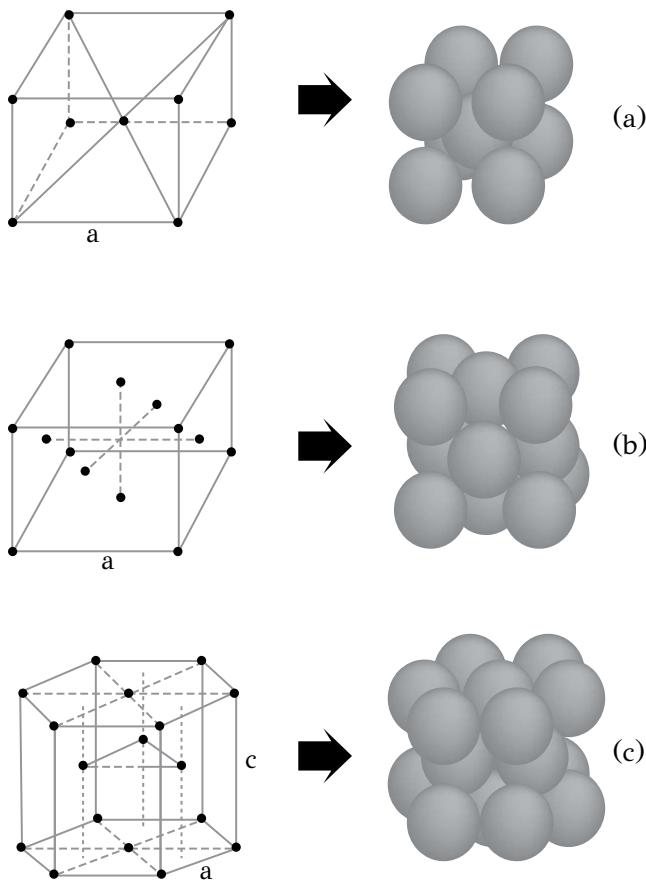


Figure 2.6: The unit cells of the three basic crystal structures: (a) BCC, (b) FCC and (c) HCP.

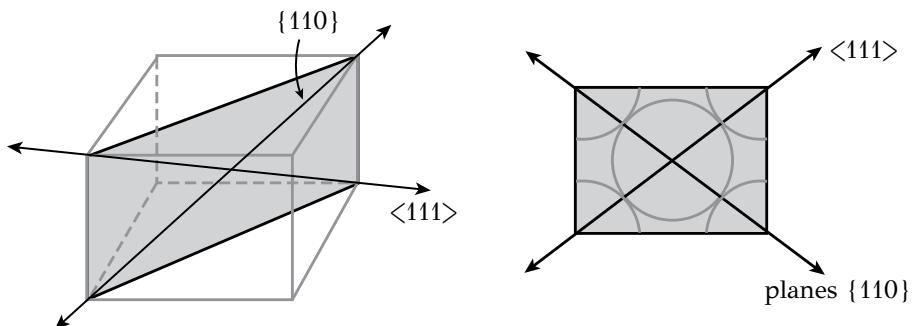


Figure 2.7: The 110 planes of BCC with the $<111>$ close-packed directions corresponding to the cube diagonal.

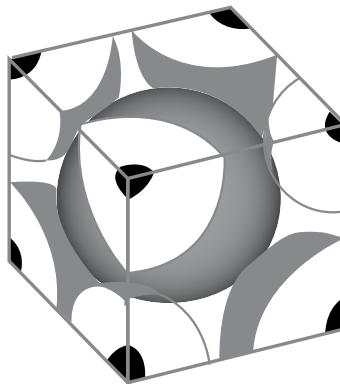


Figure 2.8: The unit cell of *BCC*. The atoms at the cube corners are shared with the neighboring unit cells.

As shown in Figure 2.6b, in *FCC*, the atoms occupy the cube corners and the face centers. There is no atom in the cube center. The *close-packed planes* in *FCC* are the diagonal planes of the cube denoted as the $\{111\}$ planes. On those planes the atoms touch along the $\langle 110 \rangle$ directions as depicted in Figure 2.9. Thus *FCC* has

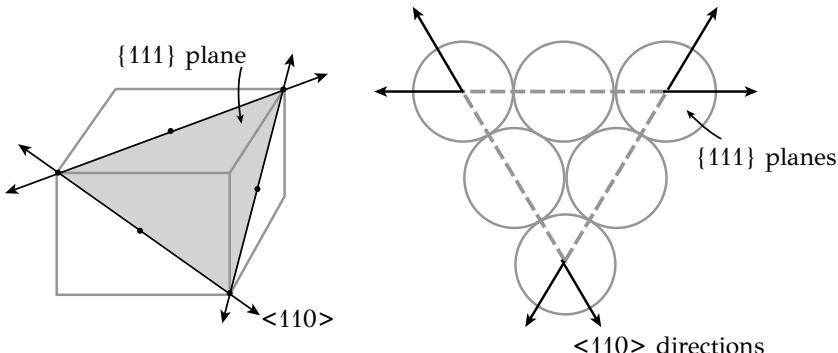


Figure 2.9: The close-packed planes $\{111\}$ and the $\langle 110 \rangle$ close-packed directions in *FCC*.

both close-packed planes $\{111\}$ and close-packed directions $\langle 110 \rangle$. As shown in Figure 2.10, the atoms at the cube corners contribute by $1/8$ to the unit cell, while the atoms at the face centers contribute by $1/2$. Then the number of atoms per unit cell is $8 \times (1/8) + 6 \times (1/2) = 4$, which is double than the *BCC*.

The unit cell of *HCP* is a hexagonal prism, as depicted in Figure 2.6c. The lattice parameters are defined by c and a . The ideal c/a ratio is the one that gives *HCP* an equivalent packing to that of *FCC*. The ideal c/a ratio is 1.63. In reality, the c/a ratio varies between 1.57 and 1.83 for most *HCP* metals. This variation is attributed

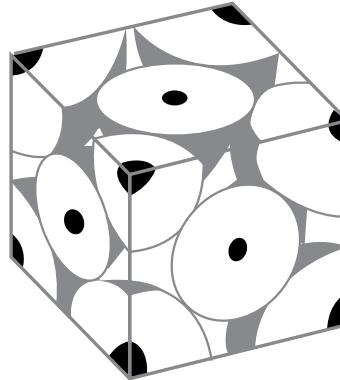


Figure 2.10: The unit cell of *FCC*. The atoms at the cube corners are shared with the neighboring unit cells.

to the electronic structure of atoms and has a significant influence on the mechanical properties of these metals, since it affects crystal plasticity, an issue to be discussed in Chapter 7. In *HCP* the close-packed plane is the basal plane (0001). The atomic packing of the {0001} planes is equivalent to the packing of the {111} planes of *FCC*. Instead of assembling unit cells, the crystal structures of *FCC* and *HCP* could be then constructed by stacking close-packed planes one on top of the other, where the atoms of the plane above are placed on the voids between the atoms of the plane below (like the eggs in an egg carton). The procedure is shown in Figure 2.11. Starting with the first close-packed plane A, the second plane B is formed by placing atoms in the voids of plane A. For the third plane there are two choices. In the first case, atoms can be placed in voids of plane B, which correspond also to voids in plane A, creating a new plane C. The stacking sequence is then ABCABC... and corresponds to the *FCC* structure. In the second case the atoms are placed in voids of plane B, which correspond to atomic positions in plane A. In this way plane A is repeated and the stacking sequence is then ABAB... corresponding to the *HCP* structure. So while the *FCC* and *HCP* structures are equivalent in their ability to fill space, they exhibit a different *stacking sequence* of close-packed planes: ABCABC... for the {111} planes of *FCC* and ABAB... for the {0001} planes of *HCP*. The different stacking sequence generates important differences in the properties of these structures.

The efficiency for filling space can be described by the *atomic packing factor*, *APF*, which is the ratio of the volume of the atoms belonging to the unit cell by the unit cell volume

$$APF = \frac{z(\frac{4}{3}\pi R^3)}{V_C} \quad (2.1)$$

where *z* is the number of atoms per unit cell, *R* the atomic radius and *V_c* the cell volume. *APF* expresses the fraction of cell volume occupied by atoms. For *FCC*, $4R = a(\sqrt{2})$ and *z* = 4. Then *APF* = 0.74, meaning that 74% of space is occupied by atoms and the other 26% is void space. For *HCP* the *APF* is also 0.74 as expected.

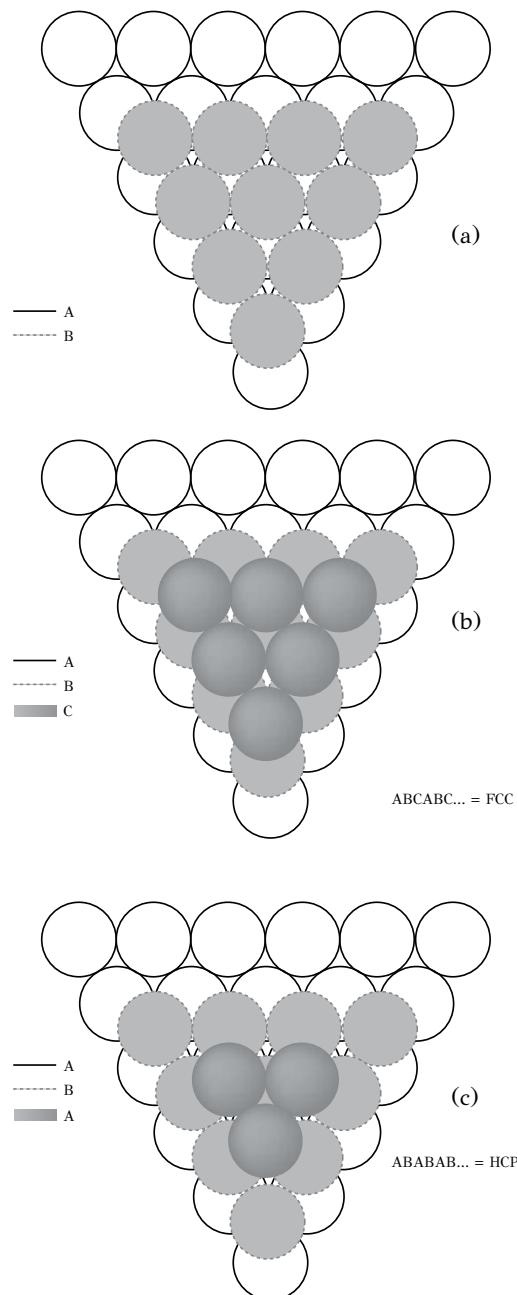


Figure 2.11: Stacking of hard spheres to generate *FCC* and *HCP* structures: (a) stacking of planes *A* and *B*, (b) stacking sequence *ABCABC...* corresponding to the *FCC* structure, (c) stacking *ABABAB...* corresponding to the *HCP* structure.

For *BCC* the $APF = 0.68$. Therefore, the *FCC* and *HCP* structures exhibit a denser packing than *BCC*. Another characteristic number of the crystal structure is the coordination number discussed above, which expresses the number of nearest neighbors in equal distances from an atom. $CN=12$ for *FCC* and *HCP*, while $CN=8$ for *BCC*. The *APF* and *CN* numbers are gathered together with other characteristic quantities in Table 2.2. The simple cubic system is also included for reference.

Table 2.2

Characteristic quantities of the basic crystal structures of metals. Numbers in parentheses correspond to the c/a ratio for the *HCP* metals

Structure	<i>FCC</i>	<i>HCP</i>	<i>BCC</i>	Simple Cubic
Coordination Number <i>CN</i>	12	12	8	6
Atomic Radius <i>R</i>	$\frac{1}{4}\alpha\sqrt{2}$	$\frac{1}{2}\alpha$	$\frac{1}{4}\alpha\sqrt{3}$	$\frac{1}{2}\alpha$
Number of atoms per unit cell	4	6	2	1
Atomic packing factor, <i>APF</i>	0.74	0.74	0.68	0.52
Examples	<i>Cu, Ag, Au, Ni, Al Pb, Pt γ-Fe</i>	<i>Mg (1.62) Ti (1.59) Zr (1.59) Be (1.56) Zn (1.86)</i>	<i>W, Mo, V Ba, Na Zr, α-Fe</i>	$\alpha - Po$

2.5 ALLOTROPY

Several metals change crystal structure with a change in temperature or pressure or even when subjected to a thermal or mechanical treatment. The different crystal structures of a metal are called *allotropic forms* and the phenomenon is termed *allotropy*. The same phenomenon in the case of compounds is called polymorphism. The reason for the existence of allotropic forms will be discussed in Chapter 4 in the framework of free energy and thermodynamic equilibrium. The discussion here will be limited to the basic examples of allotropic metals. The most important one is iron (*Fe*) which exhibits a *BCC* structure (termed α – *Fe*) up to $910^{\circ}C$, where it changes to an *FCC* structure (termed γ – *Fe*). Iron remains *FCC* up to $1400^{\circ}C$ where it changes back to the *BCC* structure (termed δ – *Fe*) and remains *BCC* up

to the melting point. The allotropy exhibited by iron is responsible for the wide diversity of microstructures and properties in steels. It is not an exaggeration to say that steel technology is based on the allotropy of iron. Another good example is Titanium (Ti), which exhibits two allotropic forms, α and $\beta - Ti$. The first, $\alpha - Ti$ has a *HCP* structure while $\beta - Ti$ has a *BCC* structure. The allotropic forms and the corresponding transition temperature of several metals are shown in Figure 2.12.

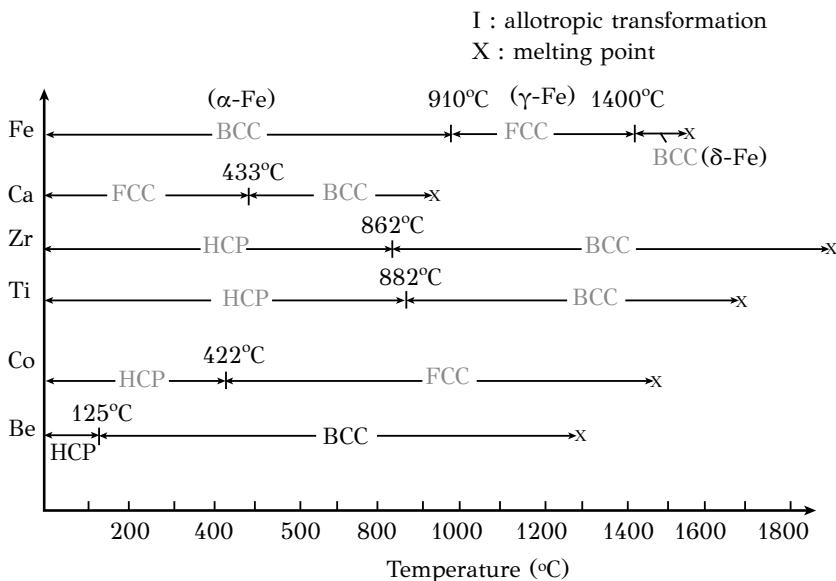


Figure 2.12: Allotropic forms and transition temperatures of some metals.

2.6 CRYSTAL STRUCTURE EFFECTS

Although the important effects of crystal structure on the development of microstructure and properties of metals will be discussed in detail in the next chapters, it is convenient at this stage to present shortly the basic effects. The mechanical behavior of metals, including the resistance of metals to plastic deformation, fracture, fatigue and creep, depend, in a large degree, on crystal structure. Diffusion, which controls several processes, including phase transformations, also depends on crystal structure. The major effects of crystal structure refer to the following issues:

Alloying. The potential of a metal to form alloys and solid solutions depends on its crystal structure. For example, the solid solubility of nickel in copper is higher than the solubility of zinc in copper, since copper and nickel have the same crystal structure (*FCC*) while zinc is *HCP*.

Plastic deformation. Close-packed planes and close-packed directions play an important role in the plastic deformation of metals. Plastic deformation in metals proceeds with the glide of crystal defects, called dislocations, on close-packed planes.

Such close-packed planes are available in the *FCC* structure, the {111} planes, and so *FCC* metals exhibit a better response to deformation processing. Steel is also processed at higher temperatures, where, due to allotropy, it exhibits an *FCC* structure.

Diffusion. The migration of atoms within the crystal (diffusion) is slower in dense crystal structures. For example, the diffusion of carbon in $\gamma - Fe$ (*FCC*) is slower than the diffusion of carbon in $\alpha - Fe$ (*BCC*). *FCC* metals, exhibit a higher creep resistance than *BCC* metals, since creep is assisted by diffusion. In fact most heat-resistant alloys, such as the *Ni*-based superalloys, exhibit an *FCC* crystal structure.

2.7 SOLID SOLUTIONS

The mechanical properties of pure metals are not sufficient for demanding structural applications. For this reason metallic alloys are used instead of pure metals. For example, steels are used for the construction of bridges instead of iron. Steel is an iron-carbon alloy and exhibits a much higher mechanical strength than pure iron. Brass, which is a copper-zinc alloy, is used instead of pure copper in water pipes and aluminum-copper alloys are used, instead of pure aluminum, in aeronautics. The first step towards the understanding of the formation and behavior of metallic alloys is to understand the structure of solid solutions. As depicted in Table 2.2, the atomic packing factor is well below unity even for the close-packed structures of *FCC* and *HCP*. This means that there is enough available space for a metal *A* to host other atoms of another component *B* in the lattice in order to form a solid solution. There are two types of solid solutions: *interstitial* and *substitutional* solid solutions. In the first case, atoms *B* are small enough to occupy the interstitial positions between atoms *A*, as depicted in Figure 2.13a.

Examples of interstitial solid solutions are austenite and ferrite in steels. In both cases carbon occupies interstitial positions in the *FCC* lattice of austenite or the *BCC* lattice of ferrite. In substitutional solid solutions, atoms *B* occupy regular positions in the lattice by substitution, as shown in Figure 2.13b. An example is a copper-nickel solid solution where copper atoms of the host lattice are substituted with *Ni* atoms. In both types of solid solutions there is a *limited solid solubility* due to differences in atomic size, amongst other differences, between the two metals, to be discussed below. Substitutional solid solutions can be either *random* (Figure 2.13b) or *ordered* (Figure 2.13c) exhibiting either long or short-range order. In random solid solutions, the *B* atoms are randomly distributed in the lattice. This is the most usual case. In ordered solutions, the *B* atoms form a periodic array, i.e., a lattice of *B* atoms within the lattice of *A* atoms, called *superlattices*. In certain cases the *B* atoms form *clusters*, as shown in Figure 2.13d. These clusters are either randomly or orderly distributed in the host lattice.

2.7.1 INTERSTITIAL SOLID SOLUTIONS

Atoms able to occupy interstitial positions in the lattice have a small atomic size. The atomic radius is shorter than 1 \AA . Most important interstitial solutes are *H*(0.46 \AA), *B*(0.97 \AA), *C*(0.77 \AA), *N*(0.71 \AA) and *O*(0.60 \AA). Solid solubility in interstitial solid

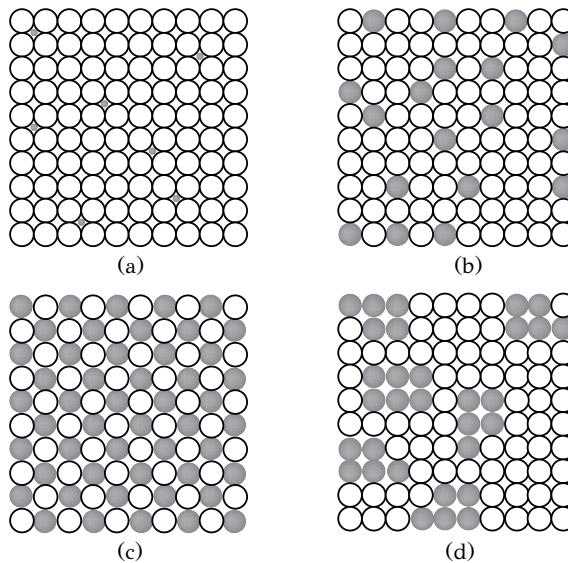


Figure 2.13: Solid solutions: (a) interstitial solid solution, (b) random substitutional solid solution, (c) ordered substitutional solid solution, (d) clustered substitutional solid solution.

solutions is controlled by the size of the interstitial sites, i.e., the void space between the host atoms. In the *FCC* structure there are two types of interstitial sites, the *octahedral* (Figure 2.14a) and *tetrahedral sites* (Figure 2.14b). Octahedral sites are larger and are surrounded by 6 atoms, which form an octahedron. Tetrahedral sites are surrounded by 4 atoms forming a tetrahedron. In *BCC* there are also octahedral (Figure 2.15a) and tetrahedral sites (Figure 2.15b) surrounded by 6 and 4 atoms respectively. The tetrahedral sites are larger. In *HCP* there are also octahedral (Figure 2.16a) and tetrahedral sites (Figure 2.16b) with the octahedral sites being larger. As expected, the size of the interstitial sites depends on the atomic size of the host metal. Using the hard-sphere model, suppose that the host atoms have a radius R , the question is what is the maximum size r of a smaller sphere that could fit in an interstitial position? The answer is shown in Table 2.3, where the size (radius) r of the interstitial positions and the respective coordination number is given as a function of atomic size R of the host atom. The importance of the size of intersitial sites will be elucidated considering ferrite and austenite in steels. In austenite, $\gamma - Fe$ (*FCC*), the octahedral site fits an atom of 0.52\AA radius, while the tetrahedral site fits an atom with 0.28\AA radius. This means that the octahedral site can fit a carbon atom (0.77\AA) with a relatively small lattice strain, while carbon cannot fit in the tetrahedral site. In ferrite, $\alpha - Fe$ (*BCC*), the larger tetrahedral site can fit an atom with 0.36\AA radius while the octahedral site can fit an atom with 0.19\AA radius. It is obvious that the formation of an interstitial solid solution of carbon in ferrite is more difficult than in austenite and in

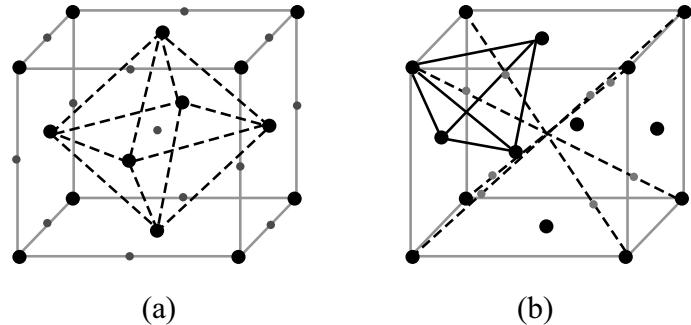


Figure 2.14: Interstitial sites in FCC: (a) octahedral, (b) tetrahedral.

Reprinted from: C.S. Barrett, T.B. Massalski, (1966). Structure of Metals, with permission from T.B. Massalski.

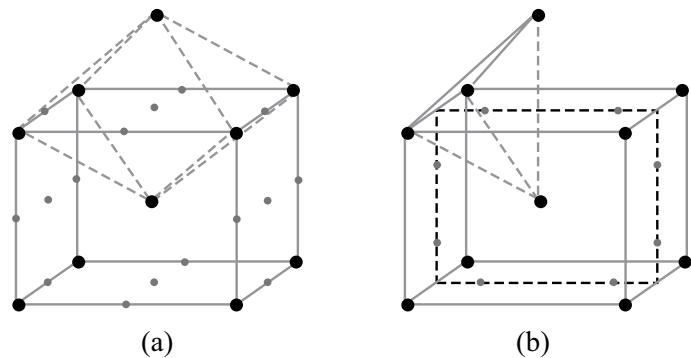


Figure 2.15: Interstitial sites in BCC: (a) octahedral, (b) tetrahedral.

Reprinted from: C.S. Barrett, T.B. Massalski, (1966). Structure of Metals, with permission from T.B. Massalski.

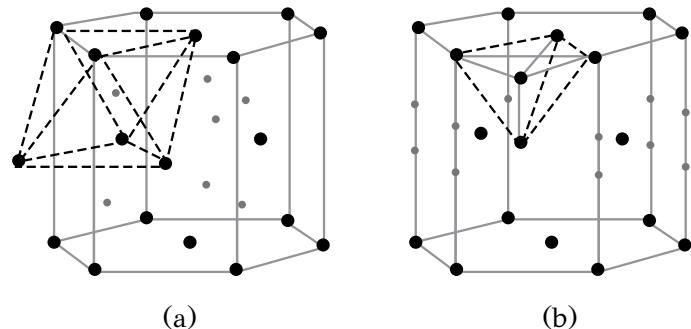


Figure 2.16: Interstitial sites in *HCP*: (a) octahedral, (b) tetrahedral.

Reprinted from: C.S. Barrett, T.B. Massalski, (1966). Structure of Metals, with permission from Mc Graw Hill.

any case the formation of such a solid solution is accompanied by large lattice strains. On the other hand, ferrite with an *APF* of 0.68 has more available free space, relative to austenite, with an *APF* of 0.74, to host interstitial atoms. However this space is distributed in smaller interstitial sites leading to a higher solid solubility of carbon in austenite than in ferrite. The maximum solid solubility of carbon in austenite is 2.1wt%, while it is only 0.021wt%, about 100 times smaller, in ferrite. Returning to the interstitial sites in ferrite, the carbon atoms finally occupy the octahedral sites despite the fact they are smaller. This is because the occupation of an octahedral site by carbon, e.g., at the center of the top face of the unit cell (Figure 2.15a), causes a displacement of two adjacent iron atoms, which are at a distance $\alpha/2$ from the carbon atom, while the occupation of a tetrahedral site causes a displacement of four iron atoms, located at the corners of the tetrahedron (Figure 2.15b). As a result the lattice strain is smaller in the case of the octahedral site. At the same time this strain is anisotropic. A consequence of this strain anisotropy is that interstitial atoms can change positions with the application of an external stress. If a tensile stress is applied in the [001] direction of the unit cell of Figure 2.17a then the carbon atoms located at the transverse [010] and [100] edges will jump to the octahedral sites located on the [001] edges, since these sites become more spacious under the application of stress (Figure 2.17b). When the stress is removed, the unit cell will resume its initial

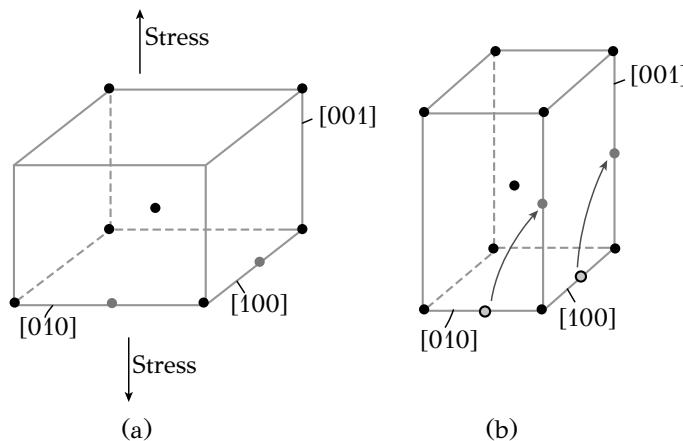


Figure 2.17: Snoek effect: (a) application of stress in the [001] direction, (b) migration of carbon atoms from the [010] and [100] edges to the octahedral sites in the [001] edge.

shape and the carbon atoms will return to their initial positions. The migration of the carbon atoms takes time and this is expressed by a time delay for the recovery of the initial elastic strain. The phenomenon is called the *Snoek effect* and is one of the basic mechanisms of *anelasticity* and *internal friction* in metals.

The size of interstitial sites is not the only factor influencing the formation and the solid solubility of interstitial solid solutions. The interstitial solutes are dissolved

Table 2.3
Size and coordination number of interstitial positions in FCC, HCP and BCC

Structure	Interstitial positions	Maximum radius r of Interstitial positions	Coordination Number
<i>FCC / HCP</i>	Octahedral	$0.414R$	6
	Tetrahedral	$0.225R$	4
<i>BCC</i>	Octahedral	$0.154R$	6
	Tetrahedral	$0.291R$	4

easier in the lattice of transition metals. Transition metals include *Fe, Co, Ni, Mn, Cr, Mo, Ti, W* and *V*. The ability of transition metals to form interstitial solid solutions is attributed to their electronic structure. On the other hand interstitial solutes do not dissolve easily in non-transition metals. For example, graphite crucibles are used for melting of these metals, since carbon is practically insoluble in most non-transition metals.

2.7.2 SUBSTITUTIONAL SOLID SOLUTIONS

Substitutional solid solutions between two metals *A* and *B* are formed by substitution of host atoms *A* by atoms *B* in the lattice. Several factors influence solid solubility, amongst them differences in atomic size and electronic structure between the two metals. General rules regarding the limits of solid solubility have been established by Hume-Rothery¹. These rules describe the necessary conditions for *complete solid solubility* between two metals, *A* and *B*. In short, the *Hume-Rothery rules* are:

1. the difference in atomic size should be less than 15%
2. small difference in electronegativity
3. same crystal structure
4. same valence

A difference in atomic size above 15%, leads to *limited solid solubility*. The size difference expresses the lattice strains developed when an atom *B* is hosted in the lattice of metal *A*. An illustration of the rule is shown in Figure 2.18, which depicts the calculated atomic radii vs atomic number for several metals. The range ($\pm 15\%$) of favorable sizes for the formation of solid solutions in Copper is indicated. Elements falling within this range satisfy the first Hume-Rothery rule. However in order

¹W. Hume-Rothery, Elements of structural metallurgy, The Institute of Metals, London, 1961

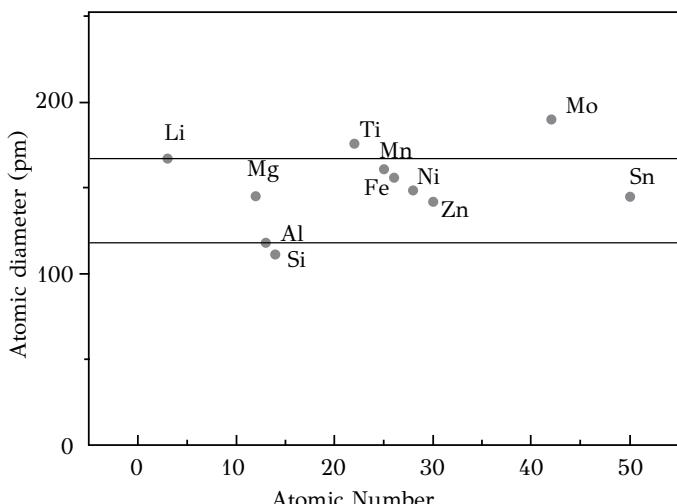


Figure 2.18: Range of favorable atomic size for the formation of substitutional solid solutions in copper according to the first Hume-Rothery rule ($\pm 15\%$).

to achieve complete solid solubility they should satisfy the other three rules as well. For example Zn falls within the range of favorable sizes of Cu. However Zn does not exhibit complete solid solubility in copper, because it has a *HCP* structure relative to the *FCC* structure of copper.

2.8 INTERMETALLIC COMPOUNDS AND INTERMEDIATE PHASES

When the difference in electronegativity between two metals A and B is large, then, instead of solid solutions, the metals form *intermetallic compounds (IMCs)*, either stoichiometric or exhibiting a homogeneity range. Intermetallic compounds also form at distinct alloy compositions or when the solid solubility is exceeded. As the stability of *IMCs* increases, the solid solubility decreases. The factors governing the formation, composition and crystal structure of intermetallic compounds include electronic structure, electronegativity, atomic radii of components and chemical bonding. It is not the scope of this chapter to cover the vast field of *IMCs*. The discussion here is restricted to examples of *IMCs*, which form in several alloy systems and influence their properties. In carbon steels, the *IMC* of interest is the iron carbide (Fe_3C) or cementite, which exhibits an orthorombic crystal structure. In $Al - Cu$ alloys, the *IMC* Al_2Cu is responsible for precipitation strengthening, achieved during the aging treatment of the alloy. In $Al - Mg - Si$ alloys, precipitation strengthening comes from another *IMC*, Mg_2Si . High-temperature nickel alloys are characterized by the formation of Ni_3Al , which is stable at high temperatures and provides creep resistance. The structure of Ni_3Al is *FCC* and forms as fine particles within the *FCC* Ni alloy matrix. Several *IMCs* are stable in a range of compositions of B in A and,

therefore, can be considered as solid solutions, called *intermediate phases*. An example is phase *CuZn*, which forms in the binary *Cu – Zn* alloy. The *CuZn* phase has a *BCC* structure while *Cu* is *FCC* and *Zn* is *HCP*. *IMCs* and intermediate phases are important constituents of alloys and appear in the phase diagrams, which will be discussed in Chapter 4.

2.9 A FIRST LOOK AT THE MICROSTRUCTURE OF ALLOYS

The microstructure of an alloy can be described using the structural elements discussed previously. Crystal structure begins to evolve during the solidification of a liquid metal (Figure 2.19a).

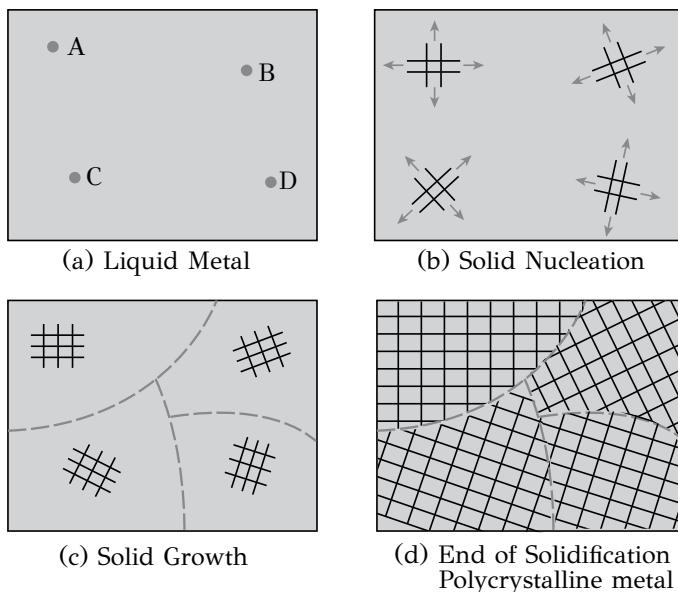


Figure 2.19: Grain formation: nucleation and growth of crystals with random orientations during the solidification of a metal.

The liquid phase is amorphous, i.e., it does not exhibit any crystal structure. With the drop in temperature, when the temperature reaches the freezing point, crystals begin to emerge at several locations in the liquid. The first solid nuclei form at, say, locations *A*, *B*, *C* and *D* (Figure 2.19b). The crystals then grow at the expense of the liquid phase. The solid at locations *A*, *B*, *C* and *D* exhibits the same crystal structure but a different lattice orientation, which is random (Figure 2.19c). At the end of solidification several crystals have formed, called *grains*, which have the same crystal structure with different orientation. Grains are separated by *grain boundaries*. Additional elements of microstructure are described in Figure 2.20a. The first observation is that the alloy consists of several grains, i.e., it is a *polycrystalline material*. Grain boundaries are regions of lattice disregistry, arising from the missorientation

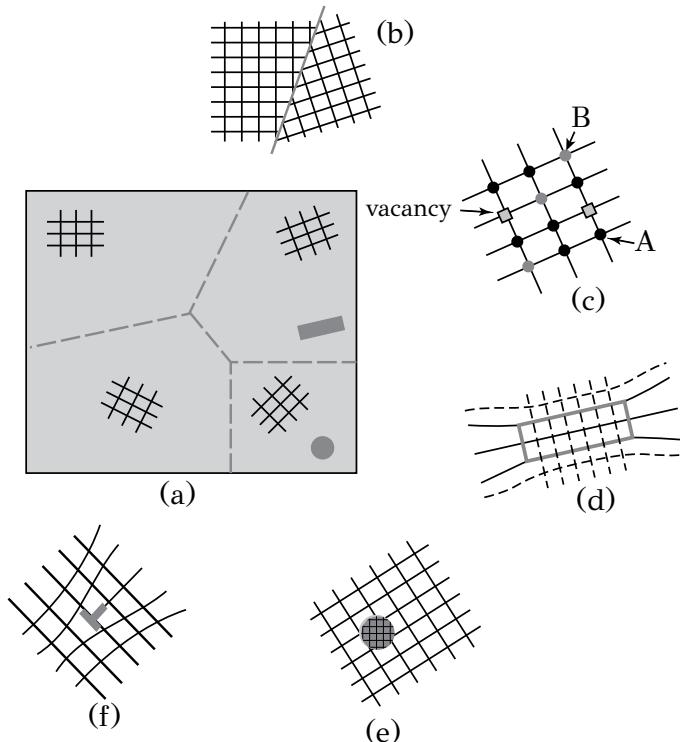


Figure 2.20: Basic elements of the microstructure: (a) schematic representation of microstructure, (b) grain boundary, (c) solid solution and vacancies, (d) plate-like coherent intermetallic compound, (e) spherical incoherent intermetallic compound, (f) edge dislocation.

between grains. In this sense, grain boundaries constitute a structural imperfection, to be discussed in detail in Chapter 3. It is however important to note that *grain size* plays a key role in mechanical properties of metallic alloys. A fine grain size improves both strength and fracture toughness of the alloy. It should also be noted that the anisotropy associated with crystallinity refers to single crystals. Due to multiple grain orientations, polycrystals appear as macroscopically isotropic. However, as will be discussed in Chapter 7, during plastic deformation, the grains in polycrystalline alloys evolve with preferred orientations leading to anisotropy and texture effects, a matter also to be discussed in Chapter 7. The formation of a solid solution, by the occupation of lattice sites by A or B atoms is depicted in Fig. 2.20c. Several lattice sites are not occupied by atoms. They are called *vacancies*. Vacancies are point defects, which aid the transport of atoms through the lattice, i.e., *diffusion*. They will be discussed in Chapter 3. Suppose that in this alloy system the conditions favor the formation of intermetallic compounds in the form of thin plates (Figure

2.20d) or spherical particles (Figure 2.20e). In the first case the plate exhibits the same crystal structure with the matrix except from a small difference in the lattice parameter between the two phases. Continuity of crystal planes from the matrix to the plate can be maintained and in this case the plate is *coherent* with the matrix. Coherency is accompanied by coherency strains arising from the distortion of crystal planes. Coherency strains are an important source of *strain energy* in the crystal. In the second case the spherical particle has a completely different crystal structure than the matrix and coherency cannot be sustained (Figure 2.20e). The particle is then *incoherent*. In this case there are no coherency strains, but due to the large number of broken bonds at the interface of the particle with the matrix, the formation of incoherent particles is accompanied by a large *interfacial energy*. The spherical particle of Figure 2.20e exhibits a higher interfacial energy than the plate of Figure 2.20d. The shape of intermetallic compounds is governed by the equilibrium between interfacial energy and strain energy. The shape minimizing the total energy is finally selected. These structural details are discussed in Chapter 6. One important structural imperfection, in the form of an extra half crystal plane is shown in Figure 2.20f. This imperfection is called *dislocation* and is the most important imperfection in metals, since it is responsible for the plastic deformation of metallic alloys. It is the glide of dislocations in certain crystal planes, called *slip planes*, and in certain crystallographic directions, called *slip directions*, which causes the plastic deformation of crystals. On the other hand, strengthening is achieved by forming obstacles to dislocation glide. Strengthening mechanisms are simply mechanisms by which several microstructural elements, such as grain boundaries, solute atoms and intermetallics impede dislocation motion. The structural characteristics of dislocations will be discussed in Chapter 3. Strengthening mechanisms is the subject of Chapter 8. As stated in the introduction of this book, properties depend on structure. Consequently, structure can be developed in such a way as to achieve desirable properties. The development of microstructure is perhaps the most interesting issue in the study of Physical Metallurgy. The structure can be tailored through processing. The structure of a metal can change from one with coarse grains to one with fine grains, exhibiting higher strength. The scientific field of obtaining required properties through the control of processing in order to obtain the right microstructure is called *alloy design*. The principles of alloy design will be discussed in Chapter 12. However in order to understand the mechanisms of microstructural evolution, the fundamentals of thermodynamics and kinetics of structure should be discussed first.

2.10 THERMODYNAMICS AND KINETICS OF STRUCTURE

The structure of metals is not in a stationary state. It responds to external stimuli, such as a rise in temperature or plastic deformation. The structure responds with various processes ranging from the generation of new lattice defects to the change of crystal structure through phase transformations. Considering that a metallic alloy, with its current microstructure, constitutes a “system”, two questions are raised: *Given the initial conditions, towards which direction will the system move, if an external stimulus is applied? How fast will the system move to that direction?* The answer in the

first question is given by *thermodynamics* while the answer to the second question is provided by *kinetics*. Thermodynamics defines the potential for a reaction or a transformation to take place while kinetics defines the rate of the reaction or transformation. The term “system” refers to an alloy, the structure of which consists of one or several phases. A *phase* is a part of the system with homogeneous structure, composition and properties. Each phase is physically distinct from any other phase in the system. In metallic alloys, the solid solutions and intermetallic compounds are the phases of the system. The components of the system are the elements (A, B, \dots), which constitute the alloy. The chemical composition of the alloy or a phase is defined by the *atomic or weight fractions* of the components. As an example consider a steel, at room temperature, containing 0.4wt% carbon. The components of the system are *Fe* and *C*. The phases present are ferrite and cementite. Ferrite is an interstitial solid solution of carbon in iron with *BCC* structure. Cementite is an intermetallic compound, which corresponds to the iron carbide Fe_3C . The composition of ferrite is 0.01wt% carbon while the composition of cementite is 6.67wt% carbon. Now assume that the temperature is raised to $850^\circ C$. The alloy will respond with a change in crystal structure. Phase constitution is different at $850^\circ C$. The phases present are only one: austenite. Austenite is an interstitial solid solution of carbon in iron with *FCC* structure. The composition of austenite will be 0.4wt% carbon. The potential (or feasibility) for this transformation is provided by thermodynamics. The rate of the transformation is provided by kinetics. A key concept in this analysis is thermodynamic equilibrium. The system responds to external stimuli and is moving towards the state of thermodynamic equilibrium. This concept is discussed in the next section.

2.10.1 THERMODYNAMIC EQUILIBRIUM

Thermodynamic equilibrium is a very important concept not only in physical metallurgy but in other scientific fields as well. There are several forms of equilibrium. *Mechanical equilibrium* is defined as the state of minimum potential energy. In the mechanical analog of Figure 2.21, the sphere in position (b) is in a state of *unstable mechanical equilibrium* since any minor displacement to either direction (left or right) will cause a permanent change in position to either position (a) or (c).

In (c) the sphere is at a state of *metastable equilibrium*, since there is another position, (a), with lower potential energy. In position (a) the sphere is at a state of *stable equilibrium*, since this state corresponds to the minimum potential energy in the system. The state of stable mechanical equilibrium is expressed by the minimization of potential energy

$$dU = 0$$

where $U = mgh$ is the potential energy, m is the mass and h the height of the sphere from the reference plane. *Thermal equilibrium* corresponds to a state where no temperature gradients exist and, therefore, there is no heat flow in the system. A system is at a state of *chemical equilibrium* when the chemical reaction ceases or the rate of the forward reaction (reactants \rightarrow products) and the inverse reaction are equal. The system is said to be at a state of *thermodynamic equilibrium* when it is at mechani-

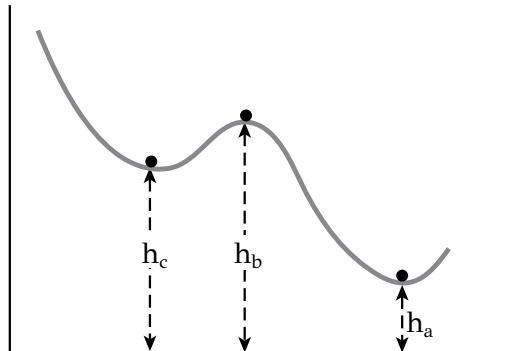


Figure 2.21: Mechanical analog for the definition of mechanical equilibrium: (a) stable equilibrium, (b) unstable equilibrium, (c) metastable equilibrium.

cal, thermal and chemical equilibrium simultaneously. The properties of the system (temperature, pressure, chemical composition,...) do not change with time. Thermodynamic equilibrium can be expressed with the aid of *Gibbs free energy* G or the Helmholtz free energy F , where

$$G = H - TS \quad (2.2)$$

$$F = E - TS \quad (2.3)$$

In the above relations H is the enthalpy, E is the internal energy, S the entropy and T the absolute temperature. As mechanical equilibrium is expressed by the minimization of potential energy, thermodynamic equilibrium can be expressed by the minimization of the Gibbs or Helmholtz free energies

$$dG = d(H - TS)_{T,P} = 0 \quad (2.4)$$

$$dF = d(E - TS)_{T,V} = 0 \quad (2.5)$$

In physical metallurgy most processes take place under constant pressure. For this reason thermodynamic equilibrium is expressed by the minimization of the Gibbs free energy $G = H - TS$.

2.10.2 INTERNAL ENERGY

Internal energy is the sum of potential and kinetic energy of the atoms in the structure of a metal. Internal energy arises from the atomic vibrations around lattice sites while potential energy comes from the bond energy. In a closed system, the internal energy does not change. In an open system, exchanging heat with the surroundings, internal energy changes according to the first law of thermodynamics. Consider a system, which received an amount of heat dQ from the surroundings and delivered an amount of work dW . During this process the system moved from state A to state

B , with a change in internal energy $dE = E_B - E_A$. According to the first law of thermodynamics

$$dE = dQ - dW \quad (2.6)$$

In the above equation the value of dE depends only on the initial and final states A and B of the system. Internal energy is a state variable. On the other hand dQ and dW depend on the process by which the system moved from state A to state B . Heat Q and work W are process variables. During a structural transformation in a metallic system, the heat exchanged causes a change of the internal energy of the system. Changes in volume are also important and they are taken into account by adding a PV term to obtain enthalpy, discussed in the next section.

2.10.3 ENTHALPY

The enthalpy of a system is

$$H = E + PV$$

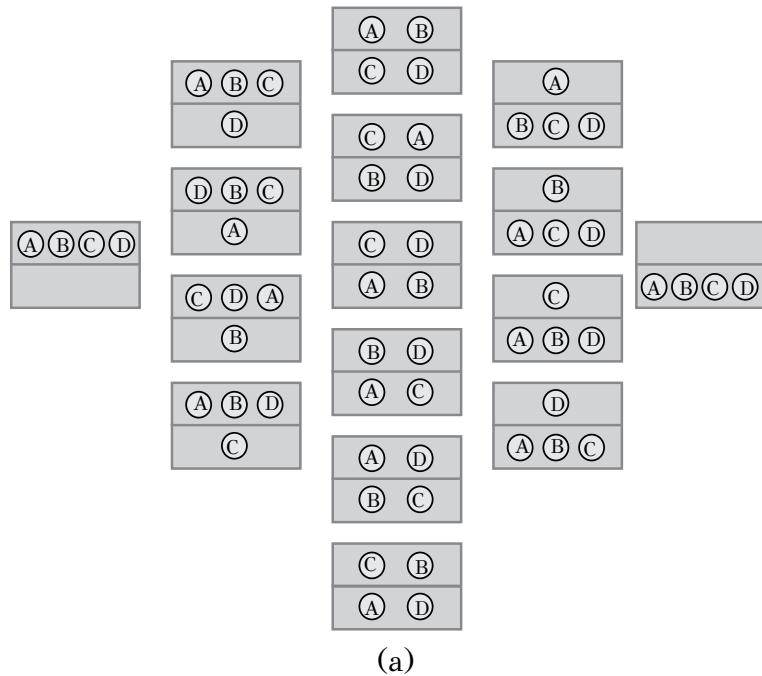
where P is the pressure and V the volume of the system. Enthalpy is also a state variable. With the use of internal energy and enthalpy, the specific heat can be defined as the amount of heat necessary to raise the temperature by one degree. The specific heats are defined by the following relations:

$$\begin{aligned} C_V &= \left(\frac{\partial Q}{\partial T} \right)_V = \left(\frac{\partial E}{\partial T} \right)_V \\ C_P &= \left(\frac{\partial Q}{\partial T} \right)_P = \left(\frac{\partial H}{\partial T} \right)_P \end{aligned} \quad (2.7)$$

where C_V is the specific heat under constant volume and C_P is the specific heat under constant pressure. In physical metallurgy we deal with systems that are under constant pressure. As a result, C_P is used for the calculation of the Gibbs free energy of the system.

2.10.4 ENTROPY

In statistical mechanics, according to Boltzmann, the entropy expresses the number of combinations or number of ways with which the atoms can be arranged in the crystal. Consider the *distribution* of four spheres A , B , C and D in two boxes. The number of available *combinations* is $2^4 = 16$, depicted in Figure 2.22a. All four spheres can be placed in the first box (4/0), or three spheres in the first and one sphere in the second box (3/1), or two spheres in the first and two in the second box (2/2), or one sphere in the first and three spheres in the second (1/3) and finally all four spheres in the second box. Each of the above distributions can be obtained by different combinations. For example, the distribution 3/1 can be obtained by four combinations, ABC/D , BCD/A , CDA/B , and DAB/C . Accordingly the distribution 2/2 can be obtained by six combinations. The number of available combinations



(a)

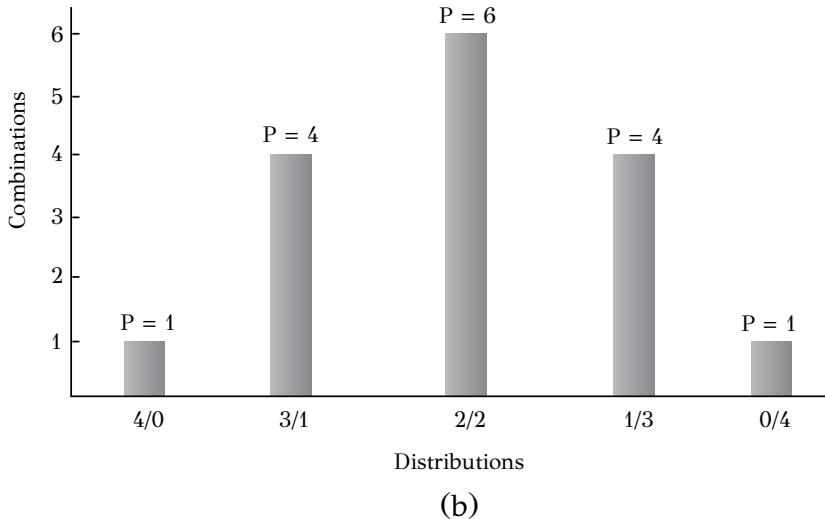


Figure 2.22: The 16 possible combinations for the distribution of four spheres A, B, C and D in two boxes and histogram indicating the number of combinations for each distribution.

for each distribution is depicted in Figure 2.22b. Assuming that the spheres are distributed randomly in the boxes, each individual combination has exactly the same probability as any other combination. For example the combination of placing all spheres in the first box ($ABCD/0$) has the same probability with the combination of placing two spheres in the first box and two spheres in the second box (e.g., AB/CD). However, as depicted in Figure 2.22 the distribution 2/2 can be obtained with the largest number of combinations, which is 6. The probability of obtaining an even distribution (two spheres in each box) is $6/16$ and is larger than the probability of any other distribution. Entropy can then be used as a measure of the probability of an individual distribution of atoms in a crystal. The distribution with the highest probability is the one that can be obtained with the highest number of combinations and exhibits the maximum entropy. In other words the most probable distribution is the one exhibiting the maximum entropy. The definition of entropy, according to the Boltzmann equation, is

$$S = k \ln p \quad (2.8)$$

where k is the Boltzmann constant ($k = 1.38 \times 10^{-23} J/K$) and p is the number of combinations for each distribution. Entropy has dimensions of energy/temperature (J/K). Applying the Boltzmann equation to the distributions of spheres of Figure 2.22 gives $S = k \ln 6 = 1.79k$ for the even distribution 2/2, $S = 1.38k$ for the 3/1 and 1/3 distributions, while the distributions 4/0 and 0/4 exhibit zero entropy. Entropy can then be considered a measure of randomness and disorder in the system. The formation of a solid solution with mixing of atoms A and B introduces randomness and disorder in the system and increases its entropy. The entropy increment is called *entropy of mixing*, ΔS_m . Consider the distribution of two A atoms and two B atoms in four lattice sites. There are six combinations (Figure 2.23): $AABB$, $ABAB$, $BBAA$, $BABA$, $ABBA$ and $BAAB$. Since $p = 6$, entropy is increased from a zero value before mixing to $1.79k$ after mixing. Thus $\Delta S_m = 1.79k$.

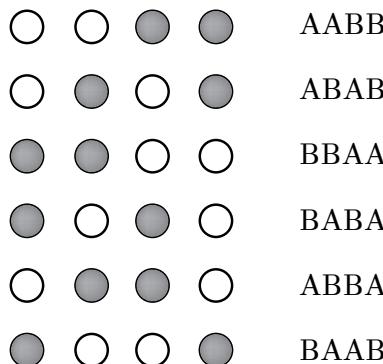


Figure 2.23: The six possible combinations for the distribution of two atoms A and two atoms B in four lattice sites.

Generalizing the above example, consider a lattice with N lattice sites, to be oc-

cupied by n atoms A and $N - n$ atoms B (Figure 2.24). The number of combinations p_{AB} for the distribution of atoms A and B in N lattice sites is

$$p_{AB} = \frac{N!}{n!(N-n)!} \quad (2.9)$$

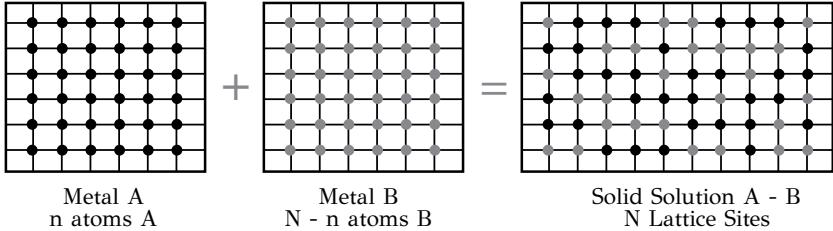


Figure 2.24: Formation of a solid solution from mixing atoms A and B.

The entropy of mixing arising from the formation of the solid solution is then $\Delta S_m = S_{AB} - S_A - S_B$ where S_{AB} is the entropy of the solid solution and S_A, S_B are the entropies of pure metals A and B before mixing. Using Equation (2.8) the entropy of mixing becomes

$$\Delta S_m = k(\ln p_{AB} - \ln p_A - \ln p_B)$$

and since $p_A = p_B = 1$, then $S_A = S_B = 0$ and

$$\Delta S_m = k \ln p_{AB}$$

Using the Relation (2.9)

$$\Delta S_m = k \ln \frac{N!}{n!(N-n)!}$$

Now use the Stirling theorem

$$\ln(N!) = N \ln(N) - N$$

and the above relation becomes

$$\Delta S_m = -Nk \left[\frac{n}{N} \ln \frac{n}{N} + \frac{N-n}{N} \ln \frac{N-n}{N} \right] \quad (2.10)$$

The mole fractions of atoms A and B in the solid solution are defined as

$$X_A = n/N \quad \text{and} \quad X_B = (N-n)/N$$

where $X_A + X_B = 1$. Relation (2.10) becomes

$$\Delta S_m = -Nk(X_A \ln X_A + X_B \ln X_B)$$

For one mole of solution, N is equal to the Avogadro number and $Nk = R$, the gas constant. The above relation becomes

$$\Delta S_m = -R(X_A \ln X_A + X_B \ln X_B) \quad (2.11)$$

which defines the entropy of mixing, which corresponds to the increase of entropy arising from the mixing of the two components A and B with mole fractions X_A and X_B respectively. Since both X_A and X_B are less than unity, the logarithms are negative and the entropy of mixing is positive. The value of ΔS_m depends on the composition of the solid solution (Figure 2.25). The curve is symmetrical with a maximum value

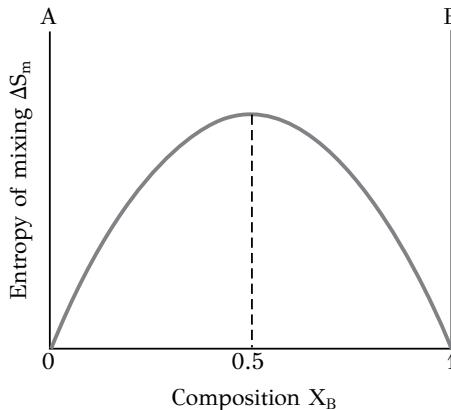


Figure 2.25: Entropy of mixing as a function of atomic fraction X_B . The entropy of mixing is maximized at $X_B = 0.5$.

at $X_B = 0.5$. The maximum entropy then appears for a solution of 50% A and 50% B , i.e., with an even distribution of A and B , just like the even distribution of spheres in the two boxes, discussed above. In summary, the formation of a solid solution always increases the entropy. The maximum entropy appears at the composition $X_B = 0.5$. The fact that $\Delta S_m > 0$ does not mean that the solid solution will finally form. As will be discussed in Chapter 4, the solid solution will be at thermodynamic equilibrium only if its formation decreases the Gibbs free energy. Since $\Delta G = \Delta H - T\Delta S$, the sign of ΔG depends, in addition to ΔS , on ΔH and temperature. In generalizing the importance of the Boltzmann equation for entropy $S = k \ln p$, p could be the number of microstates corresponding to a given macrostate of the system. Returning to the distribution of spheres in the two boxes, one macrostate could be the distribution 3/1, of three spheres in the first and one sphere in the second box. The microstates are the four possible combinations of obtaining the distribution 3/1. In this specific example, the number of microstates is $p = 4$. If the number of microstates for a given macrostate is large, the system exhibits randomness and disorder. On the other hand, if the number of microstates is small, the system exhibits order. In metals a macrostate corresponds to the observable macroscopic state of the system. Then the microstates could be the number of ways, atoms fill the quantum states of energy

levels or the number of ways, carbon atoms occupy interstitial sites in the crystal of $\alpha - Fe$. The first corresponds to an energy distribution while the second to a configurational distribution. The two distributions then correspond to two contributions in entropy. The first is the *vibrational entropy*, S_{vib} , and the second is the *configurational entropy*, S_{conf} . These entropic contributions will be involved in calculating the density of point defects (vacancies and interstitials) in Chapter 3.

2.10.5 GIBBS FREE ENERGY AND THERMODYNAMIC EQUILIBRIUM

Thermodynamic equilibrium can be expressed by the minimization of the Gibbs free energy, as stated in Relation (2.4). The definition of entropy from classical thermodynamics is

$$dS = \frac{dQ_{rev}}{T} \quad (2.12)$$

where dQ_{rev} is the amount of heat exchanged by the system during a reversible process. According to the second law of thermodynamics

$$dQ \leq dQ_{rev} \quad (2.13)$$

The definition of Gibbs free energy is

$$G = H - TS = E + PV - TS$$

The differential of the above expression is

$$dG = dE + pdV + VdP - TdS - SdT$$

With the use of (2.12) the above relation becomes

$$dG = dE + pdV + VdP - dQ_{rev} - SdT$$

Using the first law of thermodynamics (2.6) it becomes

$$dG = dQ - dW + pdV + VdP - dQ_{rev} - SdT$$

Considering that the only work performed by the system is the one arising from a change in volume, i.e., $dW = pdV$, the above relation becomes

$$dG = dQ - dQ_{rev} + VdP - SdT$$

which for the case of constant temperature and pressure becomes

$$(dG)_{p,T} = dQ - dQ_{rev}$$

Using the second law of thermodynamics (2.13) the above relation becomes

$$(dG)_{p,T} \leq 0 \quad \text{or} \quad d(H - TS)_{p,T} \leq 0 \quad (2.14)$$

which expresses that the Gibbs free energy, under constant temperature and pressure, is minimized at thermodynamic equilibrium. Some important remarks regarding thermodynamic equilibrium are given below:

1. Most systems in physical metallurgy are studied under constant pressure. In this case the criterion for thermodynamic equilibrium is the minimization of the Gibbs free energy.
2. While the equality in Relation (2.14) is valid at thermodynamic equilibrium, the inequality is valid for a spontaneous change. This simply means that all irreversible processes lead to a reduction in free energy, which is then minimized when the system reaches a state of thermodynamic equilibrium. The reduction of free energy during a process is the *driving force* of that process.
3. In a reversible process, such as the allotropic transformation of a metal, the free energy change is zero. The two allotropic forms coexist at equilibrium and exhibit the same free energy.
4. According to (2.14), thermodynamic equilibrium is favored by low enthalpy and high entropy. At low temperatures solid phases are more stable, since the strong bonding reduces the internal energy and thus the enthalpy of the material. At high temperatures, the term $-TS$ dominates and liquid phases are stabilized, which allow a greater atomic mobility with higher disorder and higher entropy.
5. The entropic term $-TS$ favors a process, if this process increases the entropy of the system. However the strength of this term depends on temperature. A process that is not thermodynamically feasible at low temperature can take place at high temperatures. On the other hand, processes, which introduce order, decrease the entropy. In this case, the term $-TS$ acts against the process.

2.10.6 KINETICS OF STRUCTURE

Thermodynamic equilibrium can be expressed by the minimization of the Gibbs free energy. Thermodynamics provides a valuable tool for the calculation of the driving force of a reaction or a transformation. The rate of the transformation is provided by kinetics. Consider the free energy of a system, composed by a number of atoms, undergoing a structural change from an initial condition with free energy G_1 to a final condition with free energy G_2 (Figure 2.26). The *driving force* for this transformation is the free energy difference $\Delta G = G_2 - G_1$. Since $\Delta G < 0$, the transformation is thermodynamically feasible. The transformation is, however, associated with an energy barrier ΔG^* . The system should be activated, i.e., its energy should increase by the amount ΔG^* above G_1 , before the transformation leads the system to the lower energy level G_2 . The additional energy required to surmount the *energy barrier* ΔG^* is called *activation energy*. The concept of the activation energy can be explained with the *Maxwell-Boltzmann energy distribution*, depicted in Figure 2.27, where the number of atoms per unit energy $n(E)$ is plotted against the energy of atoms. The distribution is nonsymmetrical and exhibits a “tail” at high energies. The total area under the entire curve is equal to the total number of atoms in the system. The area above the value ΔG^* , corresponds to the number of atoms exhibiting energies above the barrier ΔG^* . This number is proportional to $\exp(-\Delta G^*/kT)$. A higher energy barrier ΔG^* will reduce the number of atoms with energy above the barrier. At the same time a higher temperature (T_2 in Figure 2.27) shifts the energy distribution curve to the right and the number of atoms exhibiting energies higher than the bar-

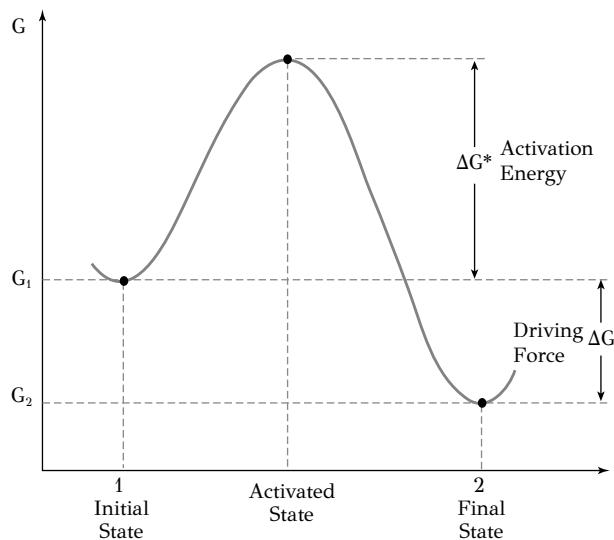


Figure 2.26: Variation of free energy during the transformation from an initial state (1) to a final state (2). ΔG is the thermodynamic driving force and ΔG^* is the activation energy of the transformation.

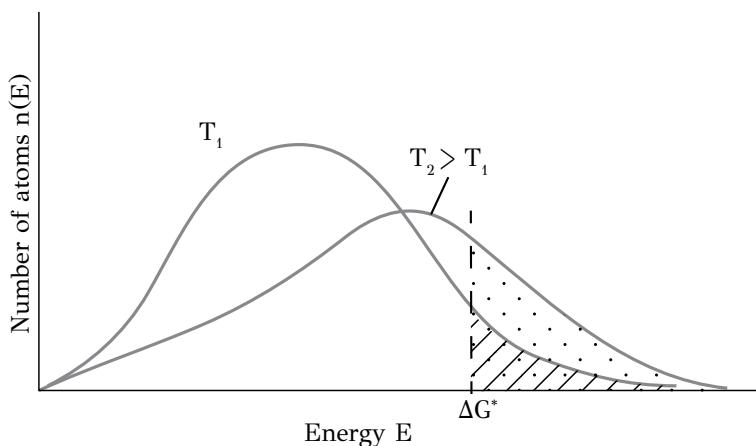


Figure 2.27: Maxwell-Boltzmann energy distribution indicating the activation energy ΔG^* .

rier ΔG^* increases. Since the rate of the transformation or reaction depends on the number of atoms with energies above the barrier, it is concluded that the reaction rate depends on both, the energy barrier (activation energy) and temperature. This is described by the *Arrhenius law* of rates of reactions

$$r = r_o \exp\left(-\frac{\Delta G^*}{kT}\right) \quad (2.15)$$

where r_o is a constant and T is the absolute temperature. In general, most processes in metallurgy follow the Arrhenius kinetics. For example the rate of diffusion of atoms, i.e., the transport of atoms through the lattice, follows Arrhenius kinetics and the diffusion rate increases exponentially with temperature. Due to the Maxwell-Boltzmann distribution, the change of reaction rate with increasing temperature is much greater than expected from the corresponding increase of atomic velocity. Consider an increase in temperature by $10K$ from 295 to $305K$ corresponding to a velocity increase by $(305/295)^{1/2}$ or about 1.6% whereas the rate of reaction increases by about 100% . This means that the rate of reaction depends significantly on the activation energy.

Taking the logarithms of both terms of (2.15)

$$\ln r = \ln r_o - \frac{\Delta G^*}{k} \left(\frac{1}{T}\right) \quad (2.16)$$

indicates a linear relationship between $\ln r$ and $1/T$, depicted in Figure 2.28.

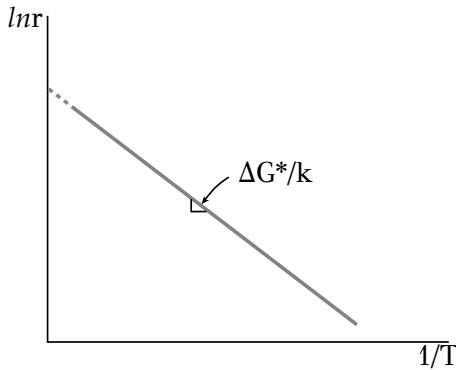


Figure 2.28: Arrhenius kinetics: the temperature dependence of the rate of a reaction. The activation energy can be determined from the slope.

The slope is $\Delta G^*/k$, which means that the activation energy can be determined by experimental measurement of the reaction rate as a function of temperature.

In several cases a reaction or transformation consists of a series of reactions or processing steps. Consider a simple example of metal oxidation (Figure 2.29) where the steps are:

1. oxygen adsorption at the oxide surface

2. diffusion of oxygen to the oxide/metal interface
3. reaction of oxygen with the metal to form oxide

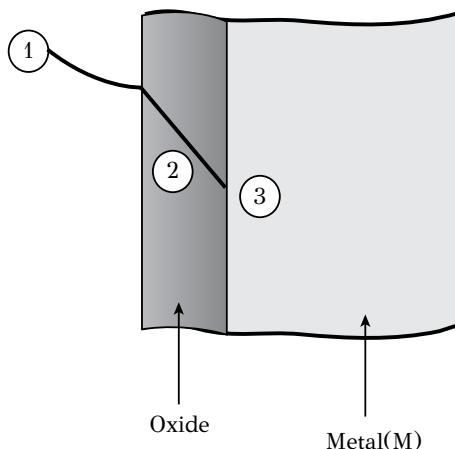


Figure 2.29: Reaction steps during the oxidation of a metal: (1) oxygen adsorption at the metal surface, (2) oxygen diffusion, (3) chemical reaction for the formation of metal oxide M_xO_y .

Each of the above reaction steps is characterized by an individual activation energy and, thus, reaction rate. The overall reaction rate is however controlled by the reaction step exhibiting the lowest rate or highest activation energy. This step is then called the *rate-limiting step*. This allows us to consider that the rate of a complex reaction, taking place in a series of steps, is equal to the rate of the slowest reaction. A good example is the formation of pearlite in steels during the slow cooling of austenite. If the steel is alloyed with substitutional elements, such as *Cr*, then the rate of pearlite formation depends on the diffusion of *Cr* and not of carbon, since *Cr* diffuses much slower than carbon in steels.

2.11 SYNOPSIS

1. Metals are crystalline materials. Atoms in a crystal structure occupy positions in an arrangement, which is characterized by periodicity. The atomic arrangement is systematically repeated in three-dimensional space.
2. Atomic arrangement can be described with the aid of the unit cell. Its repetition in three directions generates the crystal structure.
3. Atoms in metals are arranged in close-packed structures. The most important crystal structures of metals are base-centered cubic (*BCC*), face-centered cubic (*FCC*) and hexagonal close-packed (*HCP*).

4. The crystal structure influences several phenomena in physical metallurgy such as the potential of a metal to form alloys, the plastic deformation of metals and diffusion, which is the transport of atoms through the lattice.
5. Metals form two types of solid solutions: interstitial and substitutional solid solutions. The size of interstitial sites determines the maximum solid solubility in interstitial solid solutions. The formation of substitutional solid solutions is controlled by the Hume-Rothery rules. Most alloy systems do not fulfill these rules and, therefore, solid solutions exhibit a limited solid solubility.
6. Intermetallic compounds form in most alloy systems. These compounds are either stoichiometric or exhibit a homogeneity composition range. The factors governing the formation, composition and crystal structure of intermetallic compounds include electronic structure, electronegativity, atomic radii of components and chemical bonding.
7. The microstructure of metals consists of the different phases (solid solutions and intermetallic compounds) and crystal imperfections including point defects (vacancies and interstitials), line defects (dislocations), grain boundaries and interfaces. The microstructure of a metal influences its properties. Microstructure can be altered through processing with the activation of phase transformations.
8. Thermodynamics defines the potential for a reaction or a transformation to take place while kinetics defines the rate of the reaction or transformation.
9. A system is said to be at a state of thermodynamic equilibrium when it is at mechanical, thermal and chemical equilibrium simultaneously. The criterion for thermodynamic equilibrium is the minimization of the Gibbs free energy. Any spontaneous transformation decreases the free energy of the system.
10. The reduction in free energy accompanying a transformation is the thermodynamic driving force of the transformation. The energy barrier, that should be surmounted in order for the transformation to take place, is the activation energy of the transformation.
11. The activation energy is defined by the Maxwell-Boltzmann energy distribution. The rate of a transformation is then exponentially dependent on activation energy and temperature via the Arrhenius law.
12. In a reaction or transformation consisting of several steps, the rate of the overall reaction is controlled by the slowest step, which exhibits the highest activation energy. This step is called the rate-limiting step.

2.12 REVIEW QUESTIONS

- Calculate the atomic density, in atoms/mm^2 , of the (100) and (111) planes of lead (Pb). The atomic radius for lead is 1.7495\AA .
- The distance d_{hkl} between parallel (hkl) planes in cubic crystals is $d_{hkl} = a/\sqrt{(h^2 + k^2 + l^2)}$ where a is the lattice parameter and h,k,l the Miller indices. Calculate the interplanar spacing of the (100) plane and compare it with that of (111) planes for lead (Pb) with atomic radius 1.7495\AA .
- Calculate the atomic density, in atoms/m , of the close-packed direction of vanadium (V). Given: atomic mass 50.94g/mol , density $= 5.8\text{gr/cm}^3$.
- Show that in the *FCC* structure the (111) plane has a higher atomic density than the (100) plane. Hint: express the atomic density for each plane as a function of atomic radius.
- Show that the ideal c/a ratio in *HCP* is equal to 1.63.
- Show that the atomic packing factor (*APF*) is 0.74 for *HCP* and 0.68 for *BCC*. Discuss the effect of this difference on the diffusion in these metals.
- The atomic radius of copper is 1.276\AA and its atomic mass 63.57gr/mol . Calculate its density in gr/cm^3 .
- The atomic volume Ω is the volume corresponding to each atom of the crystal. The molar volume V_m is the volume corresponding to one mole of metal. Calculate the atomic and the molar volumes of gold given the atomic radius 1.44\AA .
- Calculate the volume change $\Delta V/V$ (%) arising from the transformation of iron from *FCC* to *BCC* during cooling from 1000°C . What is the respective length change $\Delta L/L$?

BIBLIOGRAPHY–FURTHER READING

- Buerger, M. J. (1978). *Elementary Crystallography*, MIT Press.
- Phillips, F. C. (1963). *An Introduction to Crystallography*, Longman, London.
- Bloss, F. D. (1971). *Crystallography and Crystal Chemistry*, Holt, Rinehart and Winston Inc, USA.
- Moffatt, W. G., Pearsall, G. W. and Wulff, J. (1964). *The Structure and Properties of Materials. Vol.1 : Structure*, John Wiley, New York.
- Smith, C. S. (1981). *A Search for Structure*, MIT Press.
- Barrett, C. S., Massalski, T.B. (1966). *The Structure of Metals*, Mc Graw - Hill, New York.
- Ragone, D. V. (1995). *Thermodynamics of Materials*, Vol. 1,2, John Wiley.
- Lee, H-G. (1999). *Chemical Thermodynamics for Metals and Materials*, Imperial College Press & World Scientific Publishing.
- Allen, S.M., Thomas, E.L. (1999). *The Structure of Materials*, John Wiley & Sons.

3 Structural imperfections

3.1 INTRODUCTION

The main characteristic of crystal structure, discussed in the previous chapter, is the periodic arrangement of atoms. However, in a real metal, the crystal structure is not perfect and contains several types of imperfections, which disrupt the periodicity of the crystal. Structural imperfections have a strong influence on physical and mechanical properties of metals as well as on the evolution of phase transformations in metallic materials. Imperfections can be classified according to dimension:

- Point defects, such as vacancies and interstitials
- Linear imperfections, such as edge and screw dislocations
- Surface imperfections, such as grain boundaries and interfaces
- Three-dimensional defects, such as voids and inclusions

All the above defects play a key role in the mechanical behavior of metals and the development of microstructure through phase transformations. For example, several phase transformations take place by atomic diffusion, where point defects, such as vacancies are essential. In addition, diffusion can be accelerated by the presence of dislocations or grain boundaries, which act as high diffusivity paths. The plastic deformation of metals takes place by dislocation glide, while strengthening is accomplished by impeding dislocation motion. The most common obstacles to dislocation glide are other dislocations, second phase particles or precipitates and grain boundaries. The three-dimensional imperfections, such as voids and inclusions, play a key role in the ductile fracture of metals. The examples discussed above, highlight the importance of imperfections in the behavior and properties of metals and alloys. The specific structural imperfections will be discussed in the following sections.

3.2 POINT DEFECTS

3.2.1 VACANCIES AND INTERSTITIALS

In a perfect crystal all lattice sites are occupied by atoms. The two most important point imperfections or defects are, vacancies and interstitials, depicted in Figure 3.1. A vacancy is an empty lattice site while the interstitial is an atom, which is displaced from its normal lattice site to a position between its neighboring atoms. In pure metals these atoms are called self-interstitials.

Consider now the concentration of point defects in a metal. Both vacancies and interstitials introduce randomness in the system and increase the entropy. On the other hand, the formation of vacancies and interstitials requires a certain amount of energy. Thus point defects will be at thermodynamic equilibrium in a metal provided that their formation leads to a reduction of the Gibbs free energy. Then at each

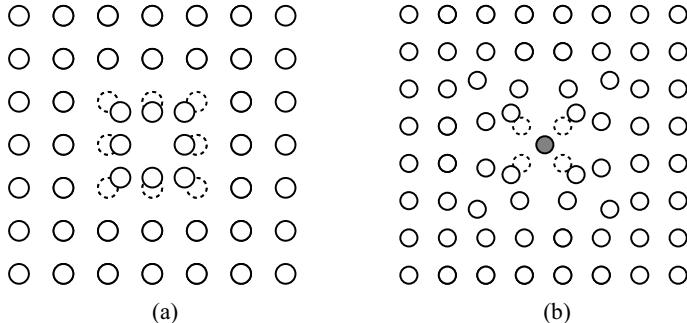


Figure 3.1: Point defects in simple cubic structure: (a) vacancy, (b) self-interstitial. Dotted circles correspond to original atomic positions.

temperature there will be a certain concentration of point defects at thermodynamic equilibrium in the crystal. In order to evaluate the vacancy concentration, consider the formation of a single vacancy by removing one atom from the bulk and placing it at the surface of the crystal, as depicted in Figure 3.2. The free energy change associated with this process will be calculated. If n vacancies are formed in a crystal of N lattice sites, the free energy change will be

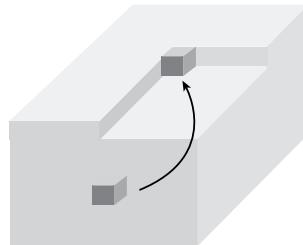


Figure 3.2: Removal of an atom from the bulk to form a vacancy.

$$\Delta G = \Delta H - T\Delta S \cong \Delta E - T\Delta S$$

assuming that the term $p\Delta V$ is very small. The change of the internal energy due to the formation of n vacancies is

$$\Delta E = n\Delta E_v$$

where ΔE_v is the vacancy energy of formation. The respective entropy change, after forming n vacancies, is

$$\Delta S = n\Delta S_{vib} + \Delta S_{conf}$$

where ΔS_{vib} is the change of vibrational entropy per vacancy and ΔS_{conf} is the change of configurational entropy of the lattice. The free energy change then becomes

$$\Delta G = n\Delta E_v - T(n\Delta S_{vib} + \Delta S_{conf})$$

The configurational entropy is related to the number p of the combinations with which n vacancies are distributed in a lattice of N sites. According to the treatment in the previous chapter, the configurational entropy is

$$\Delta S_{conf} = k \ln p$$

The number of combinations is

$$p = \frac{N!}{n!(N-n)!}$$

and the configurational entropy becomes

$$\Delta S_{conf} = k \ln \frac{N!}{n!(N-n)!}$$

Using the Stirling theorem, $\ln(N!) = N \ln N - N$, we find that

$$\Delta S_{conf} = -kN \left[\frac{n}{N} \ln \frac{n}{N} + \frac{N-n}{N} \ln \frac{N-n}{N} \right]$$

Because $n/N < 1$ and $(N-n)/N < 1$, we get $\Delta S_{conf} > 0$, i.e., the configurational entropy increases with the introduction of vacancies. The vibrational entropy is related to the distribution of energy states in the crystal and specifically with the change of the vibrational frequency of atoms due to the presence of vacancies. An estimation of ΔS_{vib} , from statistical mechanics, is

$$\Delta S_{vib} \approx 3k \ln \left(\frac{v}{v'} \right)$$

where v is the vibration frequency of atoms around a vacancy and v' is the respective vibration frequency in a perfect crystal. A vacancy raises the vibration amplitude of the atoms located around the vacancy and decreases the vibration frequency, then $v/v' > 1$ and $\Delta S_{vib} > 0$. Therefore, the total entropy change associated with the introduction of vacancies in a crystal is positive. Regarding the sign of ΔE_v , the removal of one atom from the lattice, in order to create a vacancy, requires energy (break of atomic bonds). Thus $\Delta E_v > 0$. The free energy change ΔG arising from the introduction of n vacancies in the crystal is then

$$\Delta G = n\Delta E_v - T(n\Delta S_{vib}) + TNk \left[\frac{n}{N} \ln \frac{n}{N} + \frac{N-n}{N} \ln \frac{N-n}{N} \right]$$

The above relation is depicted in Figure 3.3. Despite the positive contribution of the vacancy energy of formation ΔE_v , the free energy initially decreases due to the increase of entropy and then increases. Thermodynamic equilibrium corresponds to the point of minimum free energy and the number of vacancies at thermodynamic equilibrium is then n_v^e , given by the relation

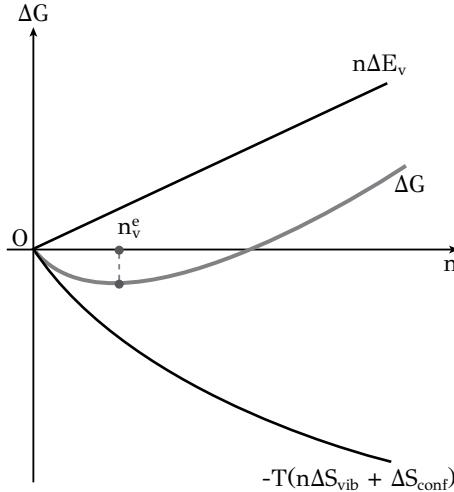


Figure 3.3: Free energy change with the introduction of vacancies in the lattice.

$$\frac{\partial \Delta G}{\partial n} = 0$$

which becomes

$$\frac{n_v^e}{N} = X_v^e = \exp \frac{\Delta S_{vib}}{k} \exp \left(-\frac{\Delta E_v}{kT} \right) \quad (3.1)$$

where X_v^e is the fraction of vacancies. The value of ΔE_v is between 80 and 160 kJ/Mol , while $\Delta S_{vib}/k$ is between 1 and 2. Relation (3.1) indicates that the concentration of vacancies increases exponentially with temperature, following an Arrhenius-type temperature dependence. The graph of $\ln X_v^e$ vs $1/T$ is linear with slope $-\Delta E_v/k$. Following similar arguments for interstitials we get

$$X_i^e = \exp \frac{\Delta S_{vib}}{k} \exp \left(-\frac{\Delta E_i}{kT} \right) \quad (3.2)$$

where X_i^e is the fraction of interstitials and ΔE_i is the energy of formation of interstitials. Because this is much larger than the energy of formation of vacancies $\Delta E_i = 7\Delta E_v$ the concentration of interstitials is several orders of magnitude lower than the concentration of vacancies at the same temperature. Relations (3.1) and (3.2) express the concentration of vacancies and interstitials at thermodynamic equilibrium in the lattice. However it is possible to generate concentrations of vacancies and interstitials above the equilibrium concentrations, through processing. For example, heating of a metal at a high temperature followed by quenching, generates the so-called quenched-in vacancies. Vacancies and interstitials, in excess of the equilibrium concentrations, are also produced during the cold working of metals, due to plastic deformation, a matter to be discussed in later sections of this chapter.

3.2.2 VACANCIES AND DIFFUSION

As discussed above, vacancies are empty lattice sites. As such they aid the movement of atoms in the lattice, since an atom can jump from its current position to the vacant site. As depicted in Figure 3.4, atom A occupies the vacant site and the vacancy now occupies the prior A site. In the same way, atom B can move to the position of the vacancy and the vacancy to the prior B site. In this way the movement of atoms in the lattice is accomplished, i.e., diffusion. This vacancy mechanism of diffusion is the basis of substitutional diffusion in metals. It is through diffusion that several processes in metals and alloys are accomplished, such as phase transformations, the nucleation and growth of a new phase, annealing, creep or even the enrichment of a metal surface with an alloying element, as in the carburizing and nitriding of steels. For diffusion to take place in substitutional solid solutions, vacancies are required and the higher the concentration of vacancies the higher is the diffusion rate. Diffusion will be discussed in detail in Chapter 5.

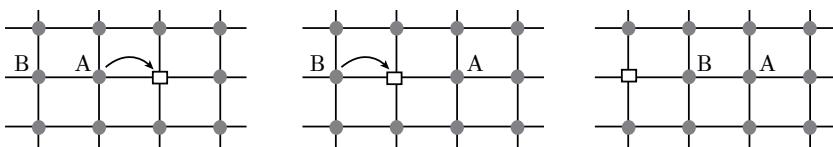


Figure 3.4: The role of vacancies in atomic diffusion.

3.3 LINEAR IMPERFECTIONS – DISLOCATIONS

3.3.1 THE PLASTIC DEFORMATION OF CRYSTALS

Elastic deformation in simple crystals is homogeneous in the atomic scale, since all atoms are subjected to the same displacement in the entire crystal volume. On the other hand, as it will be shown below, plastic deformation is extremely inhomogeneous, since it is not produced simultaneously in the entire crystal volume but only in certain regions of the crystal. Consider a metal specimen, e.g., aluminum, with a polished surface, subjected to loading as to be plastically deformed. As depicted in Figure 3.5, the original smooth surface, after deformation, exhibits striations called *slip lines*. These slip lines change their orientation from grain to grain. If we consider that grains differ in orientation of the lattice, then it is concluded that slip lines are related to crystal structure. As depicted in Figure 3.6, it appears that slip lines correspond to the edges of crystal planes, which have slipped under the action of a shear stress, just like a stack of playing cards. Slip lines indicate that plastic deformation takes place by the slip of crystal planes. It will be shown below that these crystal planes are specific crystal planes of the crystal. Slip takes place on close-packed planes and in close-packed directions. In FCC metals slip planes correspond to the {111} close-packed planes, while slip directions correspond to the $\langle 110 \rangle$ close-packed directions. In HCP metals, slip takes place on the basal planes {0001} in the

$\langle < \bar{1}2\bar{1}0 > \rangle$ directions. Finally, in *BCC* metals, where there are no closed packed planes but only close packed directions, slip takes place in the close-packed directions $\langle 111 \rangle$ on the $\{112\}$ or $\{110\}$ planes. The slip plane and slip direction constitute the *slip system* of the metal. *FCC* metals have four close-packed planes $\{111\}$ and three close-packed directions $\langle 110 \rangle$ in each plane. They exhibit, therefore, twelve slip systems. Plastic deformation requires the activation of at least one slip system. A slip system becomes operational by the application of a certain stress on the slip plane, called the critical resolved shear stress (*CRSS*), which is directly related to the mechanical strength of the metal. In the next section, the *CRSS* of perfect crystals will be calculated.

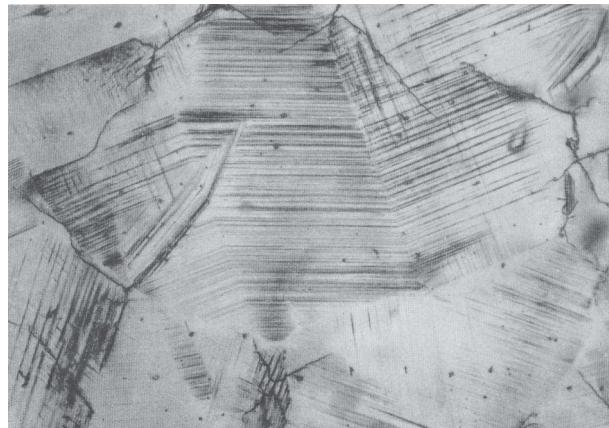


Figure 3.5: Slip lines in a stainless steel.

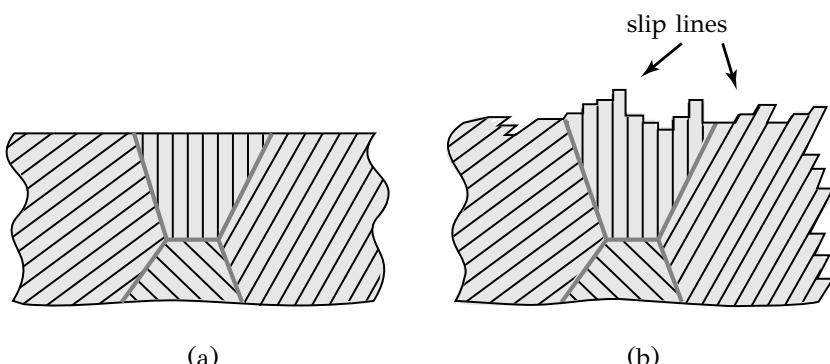


Figure 3.6: Slip lines: (a) originally polished surface before deformation, (b) slip lines corresponding to slipped crystal planes.

3.3.2 THE STRENGTH OF PERFECT CRYSTALS

A perfect crystal, which is free of point defects and dislocations, can be plastically deformed by shear only by the homogeneous slip of crystal planes. In this process all atoms, which belong to these planes participate. The shear resistance of a perfect crystal is called, *ideal shear strength*, τ_i . In order to calculate it, consider two crystal planes of an *FCC* crystal with interplanar spacing d and atomic spacing b on the plane, as depicted in Figure 3.7. Under the action of the shear stress τ , the upper plane is displaced by u relative to the lower plane. The periodicity of the crystal structure imposes a corresponding periodicity to the shear resistance. In order to simplify calculations, consider that the resistance is a sinusoidal function of displacement u , and vanishes at $u = 0$, $u = b$ and at $u = b/2$, the position of unstable equilibrium. Then the shear strength is

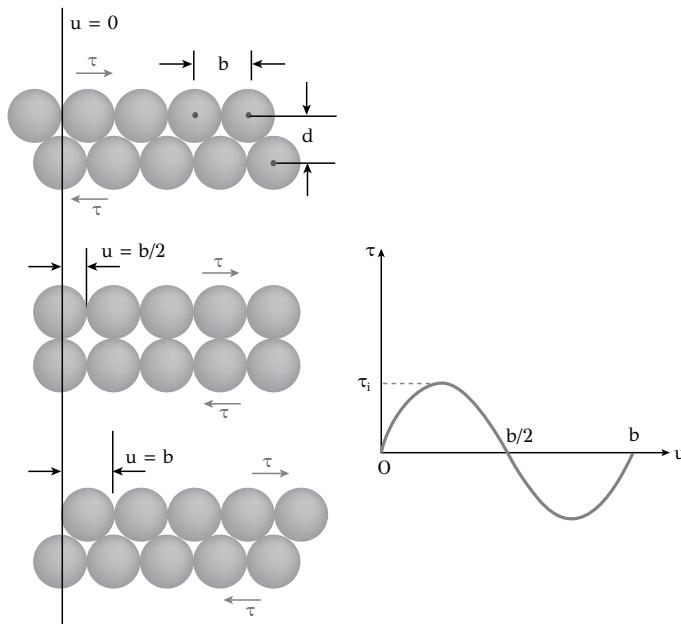


Figure 3.7: Ideal shear strength: the resistance of a perfect crystal.

$$\tau = \tau_i \sin\left(\frac{2\pi u}{b}\right)$$

At the limit of very small displacements $\sin(2\pi u/b) \approx 2\pi u/b$ and the above relation becomes

$$\tau \approx \tau_i \frac{2\pi u}{b}$$

For small strains, the shear strain u/d can be considered elastic and related to the shear modulus via Hooke's law

$$\tau = \mu \frac{u}{d}$$

Equating the above two relations and solving for τ_i we get

$$\tau_i \approx \frac{\mu}{2\pi} \frac{b}{d}$$

and because $b \approx d$ the ideal shear strength becomes

$$\tau_i \approx \frac{\mu}{2\pi} \approx 0.16\mu \quad (3.3)$$

In aluminum, the yield strength in shear is between $1MPa$ for pure aluminum single crystals and $300MPa$ for hard aluminum alloys. The respective values for iron range between 40 and $1200MPa$. Dividing the above values with the shear moduli $\mu_{Al} = 27GPa$ and $\mu_{Fe} = 82GPa$ we get

$$\tau_y = 0.00004\mu - 0.012\mu \text{ for aluminum}$$

$$\tau_y = 0.005\mu - 0.016\mu \text{ for iron}$$

It is apparent that the real shear strength of metals is significantly lower than the ideal shear strength of 0.16μ . This difference led three researchers, G.I. Taylor ¹, E. Orowan ² and M. Polanyi ³ in 1934, who worked independently, to propose the following:

- there exist structural imperfections, called dislocations, in the crystal structure
- the glide of these dislocations produces plastic deformation in crystalline solids
- a significantly lower shear stress than the ideal shear strength is required for dislocation glide

The first observation of dislocations was reported by Hedges and Mitchel in $AgCl$ crystals using a chemical etching technique. Today dislocations are routinely observed with the transmission electron microscope (*TEM*), a technique established by Hirsch, Horne and Whelan in 1954. Summarizing the above discussion, the plastic deformation of metals takes place by the homogeneous slip of crystal planes. For this to take place, a large shear stress, equal to the ideal shear strength, is required. However the actual shear strength of metals is much lower than the ideal shear strength. This simply means that crystals are not perfect and, therefore, contain imperfections,

¹G.I. Taylor, Proc. Royal Soc., A145, p.362 , 1934

²E. Orowan, Z. Phys., 89, p.605, 1934

³M. Polanyi, Z. Phys., 89, p.660, 1934

which are responsible for the significantly lower strength. These imperfections are linear and are called dislocations. Their glide causes the slip of crystal planes at a much lower stress than the ideal shear strength of perfect crystals.

3.3.3 DISLOCATIONS AND PLASTIC DEFORMATION IN METAL CRYSTALS

Long before the proposal of Taylor, Orowan and Polanyi for the existence of dislocations in metals, Volterra, a mathematician back in 1907, was studying the behavior of elastic cylinders, which he cut, deformed and rejoined. In this way he introduced linear imperfections in the otherwise continuous medium, long before their discovery in real materials. It is worth looking at the work of Volterra because it provides insight to the geometrical characteristics of dislocations and the stress fields associated with their presence in the crystal. Three Volterra cylinders are depicted in Figure 3.8. In all three cases the cylinders are sectioned, displacements are imposed at the two sides of the section and then the two sections are joined to the new position. In case a_1 the cut is across plane zx and the displacement b of the two sides is imposed in the x direction. After joining the two sides, the condition a_2 is obtained, with the displacement b causing elastic deformation of the cylinder. In case b_1 the cut is performed across plane zy and the displacement along the x direction, so that the sides of the section are moving apart. At condition b_2 this displacement causes elastic deformation of the cylinder. In order to close the gap, a cylindrical sector should be inserted. In case c_1 the cut is performed across the zx plane (as in case a_1) and the displacement is performed along the z direction, corresponding to a shear on the zx plane parallel to the z -axis. The condition c_2 is obtained, where the displacement b causes deformation of the cylinder around the z -axis. In this case the bases of the cylinder do not remain flat but form a helical surface. With the process described above, the following remarks can be made:

- In all three cases, displacements are introduced at the plane of the cut
- Cases a_2 and b_2 , where the displacement b is perpendicular to the z -axis, are equivalent to inserting extra material along the z -axis of the cylinder.
- In case c_2 the displacement is parallel to the z -axis of the cylinder
- In all cases the displacement causes the development of a stress field round the z -axis of the cylinder.

In all three cases, linear imperfections have been introduced along the z -axis of the cylinder. Now the Volterra process will be repeated in a perfect crystal. In Figure 3.9a the cut is performed across the zx plane and the displacement along the x direction, just like the Volterra case a_2 , described above. The displacement b of the upper part relative to the lower part of the crystal (Figure 3.9b) introduces an extra half plane in the crystal (Figure 3.9c). The edge of this extra half plane is parallel to the z -axis. In Figure 3.10a the cut is performed across the zy plane along the x direction, forming a gap (Figure 3.10b). In order to close the gap, just like the Volterra case b_2 , an extra half plane should be added (Figure 3.10c). The edge of this extra half plane is parallel to the z -axis. The introduction of the extra half plane in the cases of Figures 3.9

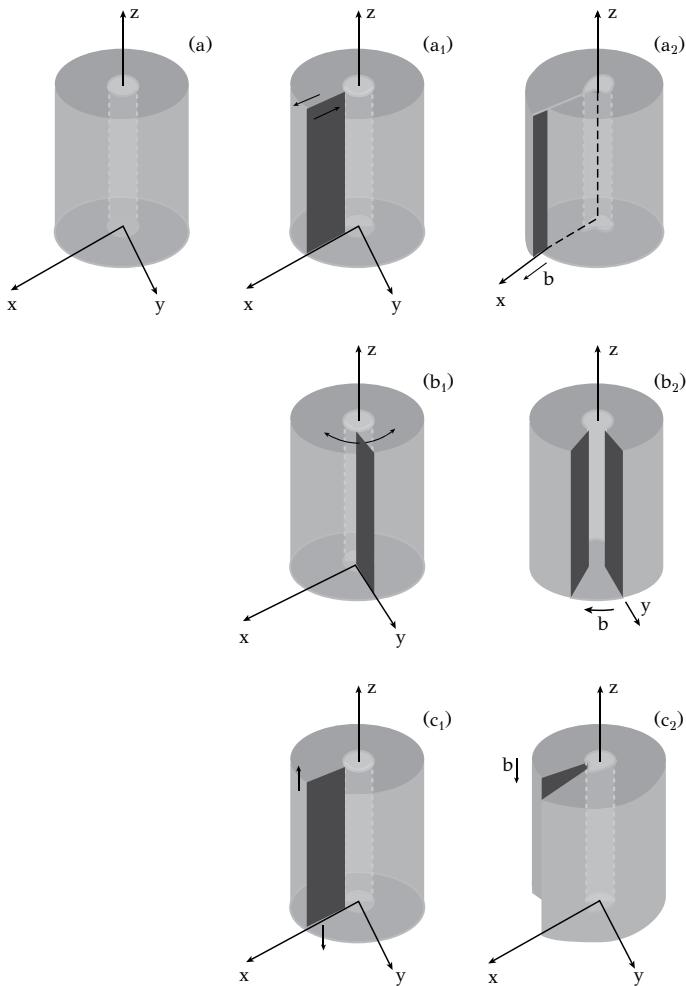


Figure 3.8: Linear imperfections in Volterra cylinders: (a) cylinder before deformation, ($a_1 - a_2$) section zx , displacement parallel to x , ($b_1 - b_2$) section zy , displacement parallel to x , ($c_1 - c_2$) section zx , displacement parallel to z . Cases a_2 and b_2 correspond to edge dislocations while case c_2 to a screw dislocation.

and 3.10 disrupts the atomic order and develops stress fields around the edge of the extra half plane. Essentially linear imperfections are introduced, which correspond to the Volterra imperfections. The linear imperfection is called *edge dislocation*. In both cases the displacement b is perpendicular to the dislocation line. Now consider Figure 3.11a where the cut is performed across plane zx and the displacement along the z -axis (Figure 3.11b), just like the Volterra case c_2 . Atomic order is disrupted and a stress field is developed along the z -axis, while a helical surface is formed on the front xy plane. The linear imperfection along the z -axis is called *screw dislocation*. In this case the displacement b is parallel to the dislocation line.

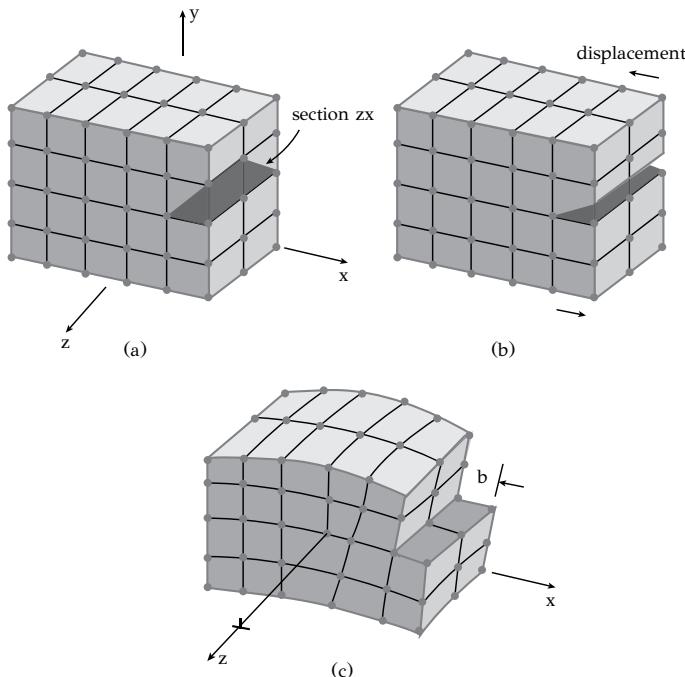


Figure 3.9: Sectioning in (a) and displacement in (b) of the upper part relative to the lower part of the crystal for the formation of an edge dislocation in (c).

At this point, having introduced the edge and screw dislocations, it will be shown that their glide causes plastic deformation in the crystal. In Figure 3.12a the crystal containing an edge dislocation is loaded with a shear stress τ . The assumption is made here that the stress τ is lower than the ideal shear strength, but high enough to move the dislocation from position A to position B (Figure 3.12b). In order for this move to be completed, a rearrangement of atoms 1, 2 and 3 is required, involving the dissociation of the 2-3 bond and the formation of the new 1-3 bond. Plane A is now perfect while plane B becomes the extra half plane, bringing the dislocation at position B , which is the new equilibrium position of the dislocation. The continuous application of stress brings the dislocation to position C (Figure 3.12c) and finally to

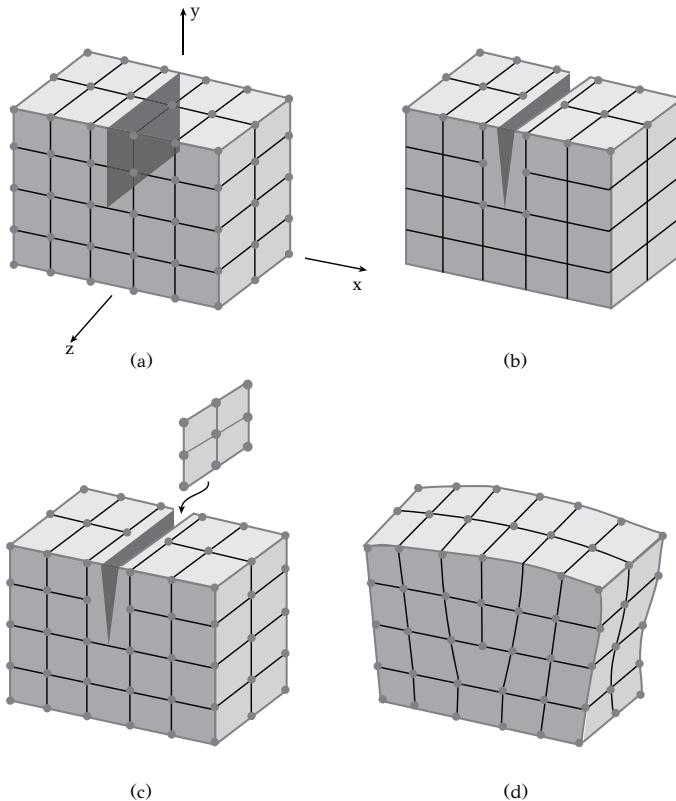


Figure 3.10: Sectioning in (a) and insertion of an extra half plane in (b) for the formation of an edge dislocation in (c).

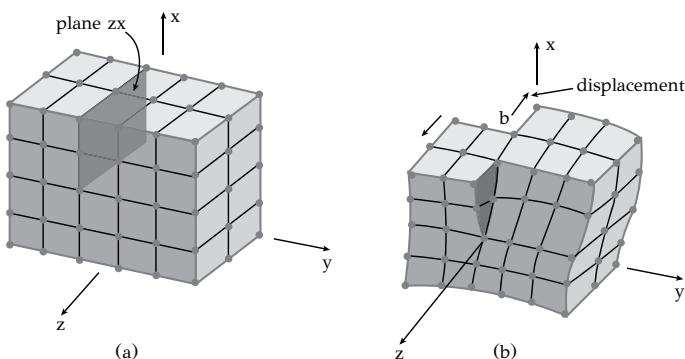


Figure 3.11: Sectioning in (a) and displacement in (b) for the formation of a screw dislocation.

position *D* (Figure 3.12d). The stress causes a step-wise movement of the dislocation from one position to the next, till the dislocation reaches the surface of the crystal. This step-wise movement is called *dislocation glide* while the plane where the glide takes place is called *slip plane*. The net result of dislocation glide is the displacement of the top section of the crystal, by *b*, relative to the bottom part, corresponding to plastic deformation by shear. The same result could have been produced by homogeneous shear of the upper relative to the bottom part, except that this would require a much higher stress and specifically a stress above the ideal strength of the crystal. This is because, in homogeneous shear, all bonds on the slip plane should dissociate, while in dislocation glide, the dissociation of only one bond at a time is required. This explains the difference between the actual and the ideal strength of crystals. The symbol for the edge dislocation is \perp , where the horizontal line symbolizes the slip plane and the vertical line symbolizes the extra half plane.

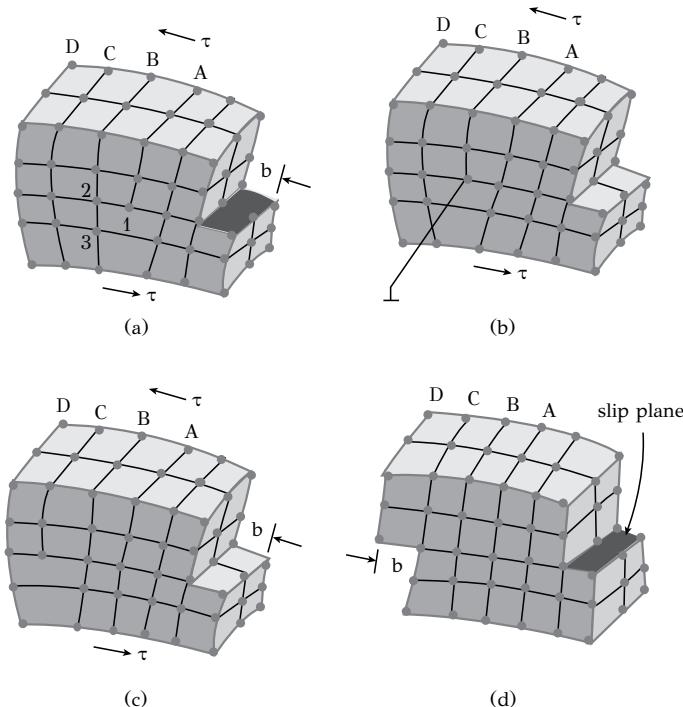


Figure 3.12: The glide of the edge dislocation in (a,b,c) causes shear of the crystal in (d).

It will be shown now that the glide of a screw dislocation causes plastic deformation. A crystal containing a screw dislocation is loaded in shear, as depicted in Figure 3.13a. Under the action of the applied shear stress τ , the dislocation moves from position *A* to the new equilibrium position *B* (Figure 3.13b) and then to position *C* (Figure 3.13c). The glide of the dislocation causes the displacement of the right

part of the crystal relative to the left part by b , i.e., it causes plastic deformation by shear. As in the case of the edge dislocation discussed above, the required stress for dislocation glide is much lower than the ideal strength. The discussion above can be summarized with the following remarks:

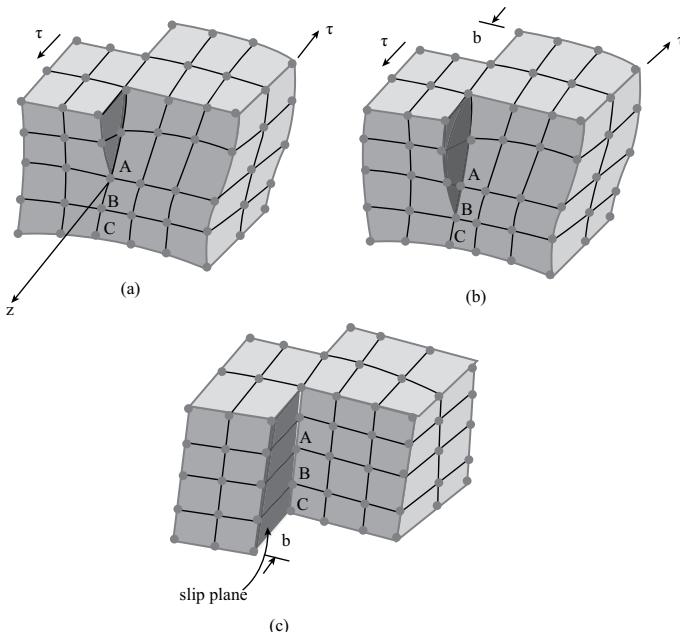


Figure 3.13: The glide of a screw dislocation in (a,b) causes shear of the crystal in (c).

- The shear stress causes a step-wise movement of dislocations from one equilibrium position to the next in the crystal. This movement is called glide of dislocations and takes place on the slip plane (hatched plane in Figure 3.12 and 3.13).
- At every position during glide, the dislocation line is the boundary between the part of the crystal that has slipped and the part which has not.
- The glide of an edge dislocation takes place in a direction parallel to the applied shear stress while the glide of a screw dislocation takes place in a direction perpendicular to the applied stress. In both cases the glide causes plastic deformation of the crystal.
- While elastic deformation is homogeneous, plastic deformation is inhomogeneous, since not all atoms of the crystal participate in the process. Only the atoms involved in the movement of the dislocation from one position to the next, participate.
- The stress required for dislocation glide is much lower than the ideal crystal

strength. This explains the difference between the actual and ideal strength of a metal.

As discussed in Chapter 2, metals are crystalline solids, exhibiting periodicity of atomic positions. It was shown that an imperfection of the crystal structure, the dislocation, is responsible for the plastic deformation of metals. The thorough study of dislocations is, therefore, important, because it allows the understanding of the mechanisms of plastic deformation as applied to the deformation processing of metals and also to the strengthening mechanisms for the development of high strength in metallic alloys.

3.3.4 GEOMETRICAL CHARACTERISTICS OF DISLOCATIONS

Dislocations are not straight lines in the crystal. In most cases they are curved. The direction of the dislocation line is expressed by the tangent unit vector \vec{s} . In curved dislocations, \vec{s} changes its direction along the dislocation line. It was shown, in the previous section, that the glide of one dislocation causes a displacement b of one part of the crystal relative to the other. The direction and magnitude of slip, caused by a dislocation, is expressed by the *Burgers vector* \vec{b} of the dislocation line. A dislocation then is fully characterized by vectors \vec{s} and \vec{b} . The direction of \vec{s} can be chosen arbitrarily, but once chosen, then the direction of \vec{b} is automatically fixed. The Burgers vector is determined by constructing the Burgers circuit, which is a path, atom by atom, around the dislocation line. Examples of how to determine the Burgers vector in edge and screw dislocations are depicted in Fig. 3.14. We will follow the *SF* convention, i.e., we will connect the start (*S*) with the finish (*F*) of the circuit. The Burgers circuit is usually orthogonal with equal number of lattice vectors on opposite sides. In order to construct the Burgers circuit we (a) select the direction of \vec{s} and (b) place the thumb parallel to \vec{s} and point to the direction of \vec{s} , then the rest of the fingers indicate the direction of rotation of the Burgers circuit. The procedure is applied to the edge dislocation in Figure 3.14a. The Burgers circuit is constructed with four lattice vectors on each side. The vector connecting the start (*S*) and finish (*F*) of the circuit is the Burgers vector of the edge dislocation. In Figure 3.14b the direction of \vec{s} has been reversed. Following the same procedure as before, a reversal in the direction of the Burgers vector is obtained. The procedure for screw dislocations is depicted in Figure 3.14c and 3.14d. Summarizing, the following remarks can be made:

- The Burgers vector of an edge dislocation is perpendicular to \vec{s} . When this is valid for the entire dislocation line, then this is a pure edge dislocation.
- The Burgers vector of a screw dislocation is parallel to \vec{s} . When this is valid for the entire dislocation line, then it is a pure screw dislocation.

It will be shown in the next section that most dislocations are not pure but they exhibit a mixed character of edge and screw simultaneously.

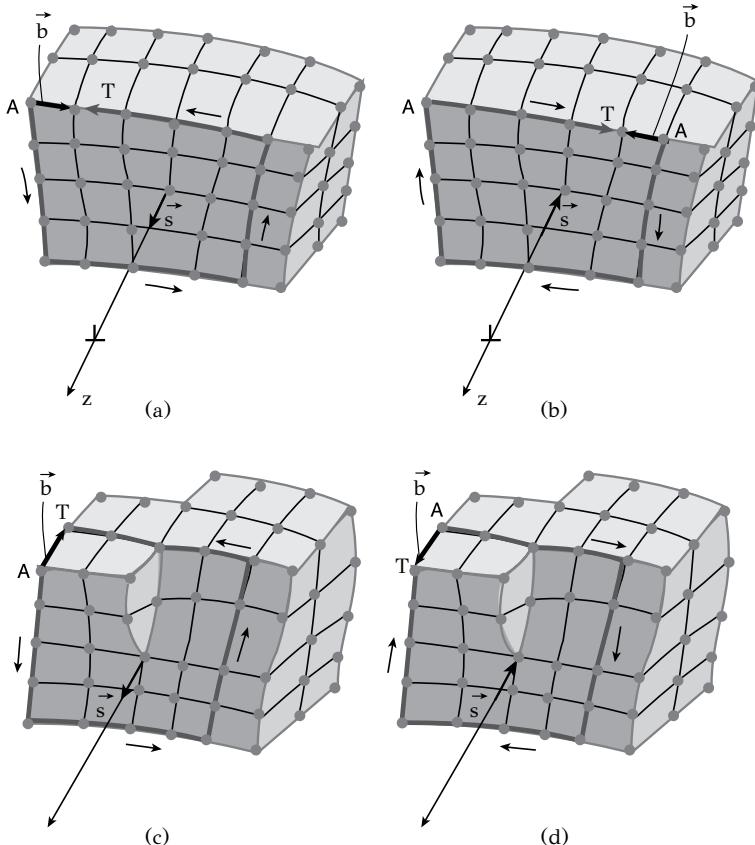


Figure 3.14: Determination of the Burgers vector in edge dislocations in (a,b) and screw dislocations in (c,d).

3.3.5 MIXED DISLOCATIONS AND DISLOCATION LOOPS

When a dislocation is not a pure edge neither a pure screw then it is a *mixed dislocation*. In this case the dislocation line forms a random angle with the Burgers vector (Figure 3.15) as in curved dislocations. The Burgers vector can be analyzed in two components, \vec{b}_e perpendicular to \vec{s} and \vec{b}_s parallel to \vec{s} . The vectors \vec{b}_e and \vec{b}_s are the edge and screw components of the mixed dislocation respectively. The extra half plane of an edge dislocation can expand inside the crystal, forming pure edge dislocations in the front and upper sides of the crystal, which are part of the mixed dislocation AB (Figure 3.16a). In addition, an edge dislocation in one direction can become a screw dislocation in another direction, as in the mixed dislocation depicted in Figure 3.16b.

An important characteristic of dislocations is that they cannot terminate inside the crystal. They can terminate either at the crystal surface, at grain boundaries or on

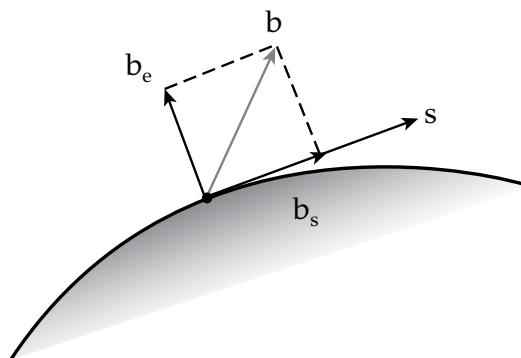


Figure 3.15: The edge and screw components of a mixed dislocation.

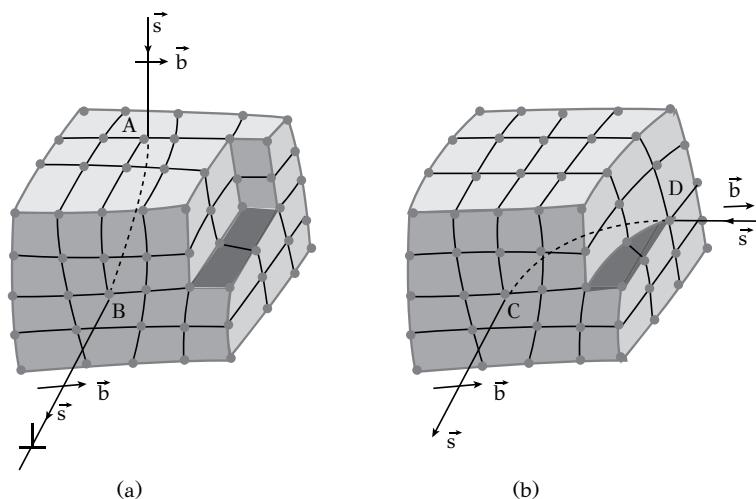


Figure 3.16: Examples of mixed dislocations: (a) dislocation AB with an edge character in A and B, (b) dislocation CD with edge character in C and screw character in D.

themselves forming *dislocation loops*. It will be shown that the glide of a dislocation loop, under the action of an applied shear stress, causes plastic deformation of the crystal. Consider a shear stress τ acting on the slip plane $Z_1Z_2Z_3Z_4$ of the crystal at the center of Figure 3.17. A circular dislocation loop with Burgers vector \vec{b} is on the slip plane. According to the discussion above, the circular loop is a mixed dislocation since the unit vector \vec{s} varies in direction along the dislocation line (the Burgers vector is constant). It will be shown that the stress causes the dislocation loop to expand and deform the crystal. The conditions at points A , B , C and D will be examined in detail. At point A , \vec{s} is perpendicular to \vec{b} . At this point the loop exhibits an edge dislocation character. The detailed structure of this edge dislocation is depicted in the section Z_1Z_2 of Figure 3.17a. Under the action of the stress τ , this dislocation glides from position A to position A' , causing the displacement of the upper part of the crystal relative to the bottom part by \vec{b} . At the opposite point C , the vector \vec{s} is perpendicular to \vec{b} and the dislocation loop has again an edge character, as in point A . The only difference is that in this case the extra half plane of the dislocation is below the slip plane, as depicted in Figure 3.17c. This dislocation will glide in the direction opposite to \vec{b} , from position C to position C' , causing a displacement of the upper part of the crystal relative to the bottom part by \vec{b} . At this point, it should be noted, that by convention, when the extra half plane of an edge dislocation is above the slip plane, then it is a *positive edge dislocation*. In the opposite case it is a *negative edge dislocation*. Positive edge dislocations glide to the direction of \vec{b} while negative edge dislocations glide to a direction opposite to \vec{b} . At point A of the loop the dislocation is a positive edge, while at point C it is a negative edge dislocation. Now consider points B and D . At point B of the loop (Figure 3.17b), the Burgers vector \vec{b} is opposite to \vec{s} . This is a *left handed screw dislocation*, which under the action of the shear stress τ glides in a direction perpendicular to \vec{b} , from position B to position B' , causing the displacement of the upper part of the crystal relative to the bottom part by \vec{b} . At the opposite point D , the Burgers vector \vec{b} points to the same direction with \vec{s} and the dislocation is a *right-handed screw dislocation*. Under the action of the shear stress τ , this dislocation will glide perpendicular to \vec{b} from position D to position D' , causing the displacement of the upper part of the crystal relative to the lower part by \vec{b} . It is then clear that under the shear stress τ , points A , B , C and D of the loop, glide to positions A' , B' , C' and D' respectively. This simply means that the loop expands by gliding on the slip plane. When the loop terminates at the crystal surface (at points A' , B' , C' and D') it causes the displacement of the upper part of the crystal relative to the bottom part by \vec{b} , the magnitude of the Burgers vector. Thus the expansion of a dislocation loop under the action of a shear stress causes plastic deformation of the crystal. The consecutive positions of the loop (1, 2, 3,...) as it expands are depicted in Figure 3.18, together with the final result of dislocation glide, which is the plastic deformation of the crystal. At every position, the dislocation line marks the boundary of the crystal that has slipped and the part which has not.

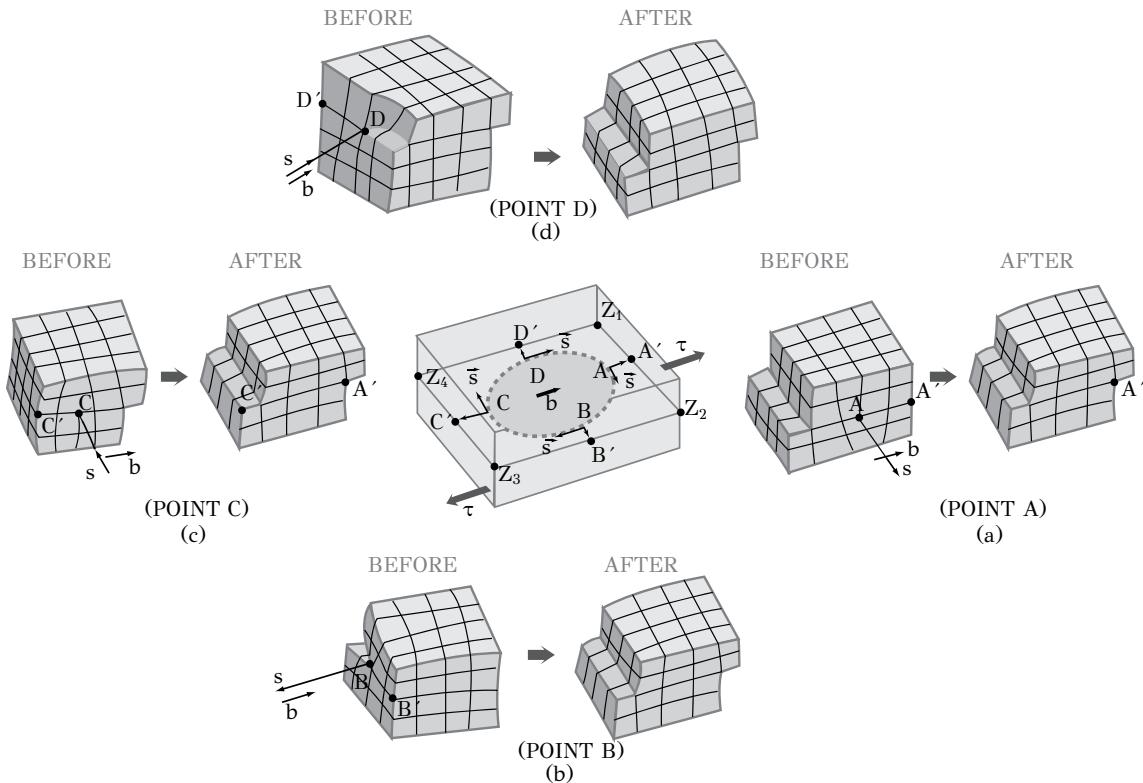


Figure 3.17: Detailed structure of a dislocation loop: (a) and (c) points A and C of the loop with edge character, (b) and (d) points B and D of the loop with screw character. The glide of the dislocation loop at points A, B, C and D causes shear of the crystal.

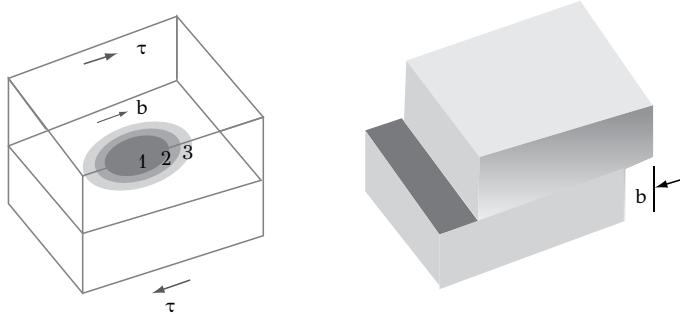


Figure 3.18: Dislocation loop which glides, expands and causes shear of the crystal.

3.3.6 ENERGY AND DENSITY OF DISLOCATIONS

All metals and alloys contain dislocations. While in the case of point defects, there is an equilibrium concentration at every temperature, dislocations always raise the free energy of the crystal and their concentration is not at equilibrium in the crystal. The increase in free energy is mainly attributed to the elastic strains around the core of the dislocations. These strains are caused by the displacement of atoms away from their equilibrium positions and their magnitude is proportional to the magnitude of the Burgers vector. While these issues will be discussed in more detail in the following section, it is important to say that the *elastic energy* of a dislocation, per unit length, is

$$E = \frac{1}{2} \mu b^2 \quad (3.4)$$

where μ is the shear modulus of the crystal. The dislocation density ρ is defined as the total dislocation line per unit volume of the crystal (in units of mm^{-2}). An alternative definition is the number of dislocations crossing a unit surface inside a crystal (again in units of mm^{-2}). The mean distance between dislocations is then $1/\sqrt{\rho}$. In well-annealed metals, the dislocation density ranges between 10^4 and $10^6 mm^{-2}$. With plastic deformation, the dislocation density increases rapidly and for metals subjected to cold working, the dislocation density ranges between 10^9 and $10^{12} mm^{-2}$. This simply means that dislocation multiplication takes place during plastic deformation. More specifically, dislocation sources, such as the Frank-Reed source, are activated and emit dislocation loops. These sources will be discussed in detail in Chapter 7, when referring to the plastic deformation of metals.

3.3.7 PERFECT AND PARTIAL DISLOCATIONS, STACKING FAULTS

When the Burgers vector of a dislocation connects two adjacent lattice sites, i.e., it is a lattice vector, then it characterizes a *perfect dislocation*. Consider various examples of dislocations in *FCC* and *BCC* structures, keeping in mind, that glide takes place

on close-packed planes in close-packed directions. The close-packed plane (111) of *FCC* containing three close-packed directions is depicted in Figure 3.19a. The Burgers vectors of perfect dislocations connect two adjacent lattice sites on the close-packed directions and their magnitude is equal to $\frac{1}{2}$ the face diagonal of the unit cell. So the three Burgers vectors for the (111) slip plane are $\vec{b} = \frac{1}{2}[011]$, $\frac{1}{2}[\bar{1}01]$, $\frac{1}{2}[\bar{1}\bar{1}0]$. The magnitude of these three vectors is $\alpha\sqrt{2}/2$. When one of the above dislocations glides on the (111) slip plane and ends up at the crystal surface, it will produce a slip step equal to $\alpha\sqrt{2}/2$. The close-packed directions [1̄11] and [1̄̄11] on the plane (110) of *BCC* are depicted in Figure 3.19b. The corresponding Burgers vectors of perfect dislocations are $b = \frac{1}{2}[\bar{1}11]$, $\frac{1}{2}[1\bar{1}1]$ with magnitude $a\sqrt{3}/2$.

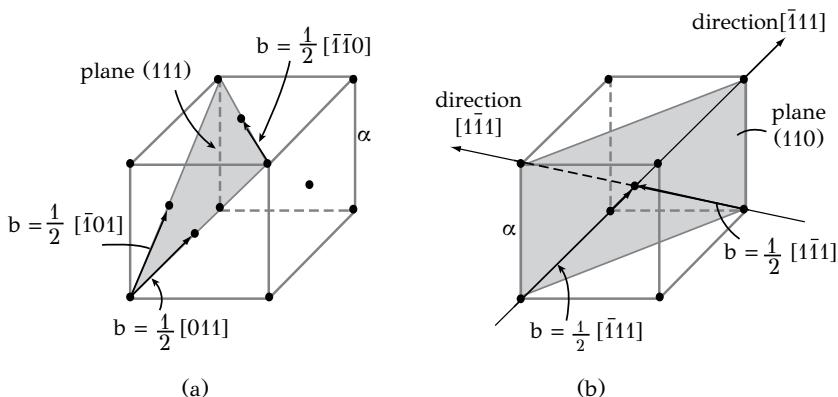


Figure 3.19: Burgers vectors of perfect dislocations in (a) FCC and (b) BCC crystals.

Partial dislocations are those whose Burgers vector is shorter than a lattice vector connecting two adjacent lattice sites. Several times the dissociation of a perfect dislocation into partial dislocations is energetically favored. The glide of these partials has the same effect as the glide of the perfect dislocation. Consider the dissociation of a perfect dislocation in *FCC* structure, as depicted in Figure 3.20a. The (111) slip plane is indicated. Three Burgers vectors are defined between the atomic positions *K*, *L* and *M*, corresponding to perfect dislocations $\frac{1}{2}[\bar{1}10]$, $\frac{1}{2}[0\bar{1}1]$ and $\frac{1}{2}[\bar{1}01]$. Now consider Figure 3.20b where the glide of the dislocation with Burgers vector $\vec{b} = \frac{1}{2}[\bar{1}10]$ brings the dislocation from position *K* to position *L*, causing the glide of plane (111) by one atomic spacing in the direction [1̄10]. The same glide can be accomplished in two smaller steps. The first step is glide from position *K* to position *N* while the second step is glide from position *N* to position *L*, in a zigzag path. Glide *KN* corresponds to the glide of a dislocation $\vec{b}_1 = \frac{1}{6}[\bar{2}11]$ while glide *NL* corresponds to the glide of a dislocation $\vec{b}_2 = \frac{1}{6}[1\bar{2}1]$. The dislocations with Burgers vectors b_1 and b_2 are partial dislocations, since their Burgers vectors are not full lattice vectors. The result of the glide of the two partials is the same as the glide of the perfect dislocation $\vec{b} = \frac{1}{2}[\bar{1}10]$. The dissociation of a perfect dislocation into two partials can be described by the reaction

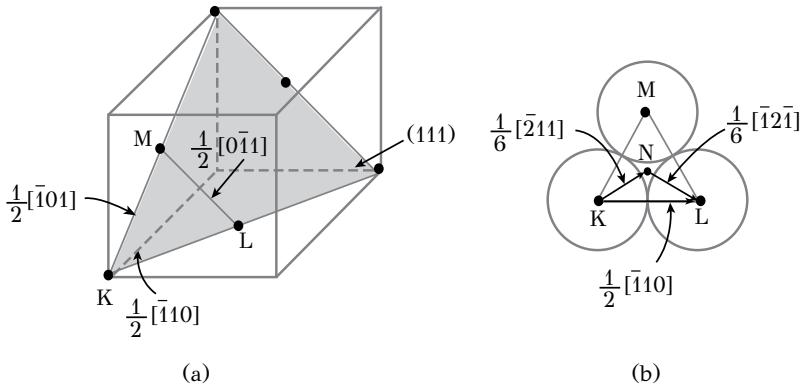


Figure 3.20: (a) Perfect dislocations on the (111) close-packed plane of FCC, (b) the glide of the perfect dislocation KL is accomplished in two steps KN and NL , corresponding to partial dislocations.

$$\vec{b} \rightarrow \vec{b}_1 + \vec{b}_2 \quad (3.5)$$

which can be written as

$$\frac{1}{2}[\bar{1}10] \rightarrow \frac{1}{6}[\bar{2}11] + \frac{1}{6}[\bar{1}2\bar{1}]$$

Taking into account that the elastic strain energy of a dislocation is proportional to b^2 , it can be shown that the dissociation of a perfect dislocation is energetically favored when

$$b_1^2 + b_2^2 < b^2 \quad (3.6)$$

i.e., when the dissociation leads to a decrease of the elastic energy of the crystal. The Relation (3.6) is known as the *Frank's rule*. The partial dislocations in the FCC structure, described above, are called *Shockley partial dislocations*. According to the above discussion, slip on the close-packed planes and in the close-packed directions of FCC, should always take place by dissociation of perfect dislocations into Shockley partials. However the dissociation leads to the formation of another crystal imperfection, in the form of an interface, which actually raises the free energy of the crystal. It was shown in Chapter 2, that the FCC crystal could be constructed by stacking (111) close-packed planes one on top of the other, with the stacking sequence $ABCABC\dots$. The atomic arrangements in two layers of this stack is depicted in Figure 3.21a, together with a perfect edge dislocation with $\vec{b} = \frac{1}{2}[\bar{1}10]$. The dissociation of this dislocation into Shockley partials, according to reaction (3.20), is depicted in Figure 3.21b. The perfect dislocation has been substituted with two partials, which, according to Frank's rule, exhibit a lower elastic energy. A closer look at Figure 3.21b indicates that the atomic positions between the two partials are not *B* positions but they correspond to *C* positions. In other words the stacking of crystal

planes between the two partials is not $ABCABC\dots$ but $ABCACABC$, meaning that a layer of *HCP* crystal, with stacking $ACAC$, intervenes in the normal *FCC* stacking sequence. This *HCP* layer, in the otherwise *FCC* crystal structure, is called a *stacking fault*. It is an interfacial imperfection of the crystal, which is accompanied by an interfacial energy, called the *stacking fault energy (SFE)*, in J/m^2 . The stacking fault energies (*SFE*) range between 10 and $300mJ/m^2$ depending on the alloy composition. For example, pure copper has an *SFE* of about $100mJ/m^2$, while this value is reduced to only $10\ mJ/m^2$ for a $Cu-Al$ alloy. The lower the *SFE*, the larger the number of stacking faults in an alloy.

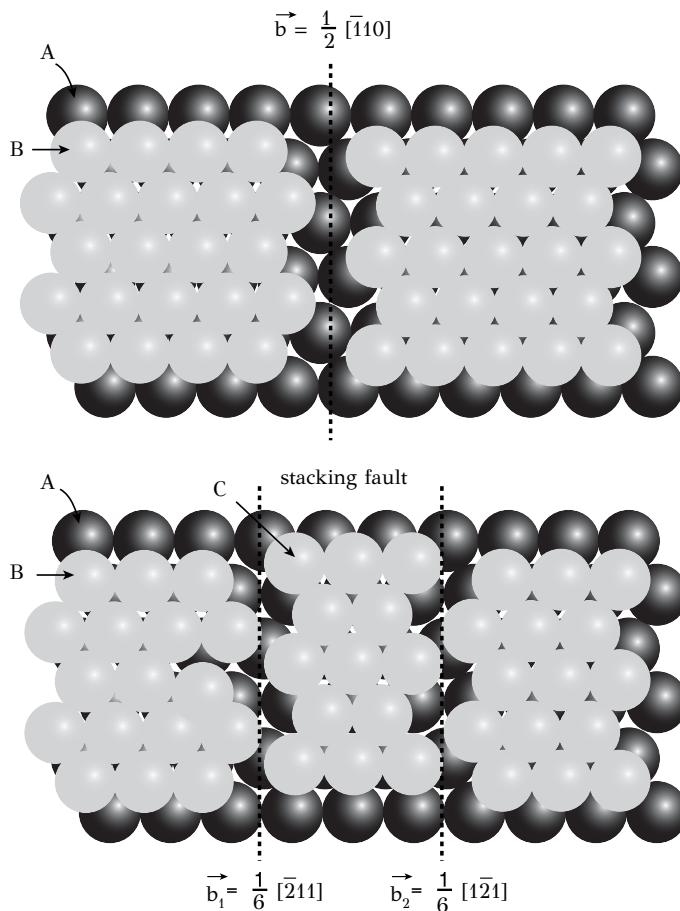


Figure 3.21: (a) Perfect dislocation on the (111) close-packed plane of *FCC*, (b) dissociation of the perfect dislocation into partials and formation of a stacking fault.

Returning to the above discussion, as to when the dissociation into partials will take place, it should be noted that the dissociation on one hand reduces the elastic

energy, but on the other hand it increases the energy, by the *SFE* amount, due to the formation of the stacking fault. The dissociation will take place only if it leads to a reduction of the total energy. So in the case of a low *SFE* (as in brass, for example), the dissociation is energetically favored and is accompanied by the formation of stacking faults. In this case slip in the (111) planes takes place by the glide of partial dislocations. If on the other hand the *SFE* is large (as in aluminum), then the dissociation into partials is not energetically favored and, consequently, stacking faults do not form. In this case slip in (111) planes takes place by the glide of perfect dislocations.

3.3.8 DISLOCATION MOVEMENTS: GLIDE, CROSS-SLIP, AND CLIMB

It was shown in the previous section that the movement of an edge dislocation on the slip plane leads to the shear of the crystal. This movement is called *dislocation glide*. A critical shear stress is required for dislocation glide, termed the *critical resolved shear stress (CRSS)*, which will be examined in detail in Chapter 7. The CRSS is substantially lower than the ideal strength of the crystal. The CRSS expresses the resistance exhibited by the crystal on dislocation glide. It is composed by the resistance exhibited by the crystal lattice itself, called lattice resistance, and additional resistances, such as those exhibited by grain boundaries, solute atoms and precipitates. These resistances form the basis of strengthening mechanisms of metals and will be discussed in detail in Chapter 8. The slip plane of an edge dislocation contains both the dislocation line and the Burgers vector. This simply means that the glide of the edge dislocation is confined to a specific slip plane. In the case of a screw dislocation, however, the dislocation line and the Burgers vector are parallel and do not define a specific slip plane. Therefore, the glide of a screw dislocation is not confined to a single slip plane. The direction of glide of a dislocation on the slip plane depends on the sign of the dislocation, i.e., positive or negative edge and right handed or left handed screw dislocation. In any case, under the application of the same shear stress, dislocations with opposite signs glide in opposite directions. In order for a dislocation to glide on the slip plane, a shear stress should be applied on the slip plane and in the direction of the Burgers vector, irrespectively of the direction of the dislocation line. It was discussed above that the glide of a screw dislocation is not confined to a single slip plane. As an example consider α -iron with *BCC* crystal structure (Figure 3.22a). The close-packed direction in *BCC* is [111], corresponding to the cube diagonal. Potential slip planes, which contain the [111] direction are $(\bar{1}10)$ and $(\bar{1}\bar{1}2)$. A dislocation loop with Burgers vector $\frac{1}{2}[111]$ glides on plane $(\bar{1}10)$, as depicted in Figure 3.22b. At point A of the loop the dislocation exhibits a screw character. If for some reason dislocation glide is impeded on plane $(\bar{1}10)$, then the dislocation can glide on the $(\bar{1}\bar{1}2)$ plane, which also contains the [111] direction. The change of slip plane by a dislocation is called *cross slip*. Cross slip of screw dislocations is important for the plastic deformation of metals, because it allows the dislocations to continue their glide in the crystal and produce plastic deformation. Very frequent cross slip results in a wavy slip surface. On the other hand, edge dislocations cannot cross slip. The only way for an edge dislocation to change slip plane is by *climb* to

another parallel slip plane. The climb process is depicted in Figure 3.23. If the last atom of the extra half plane A_1 jumps into the vacancy O_1 , then the dislocation climbs to the slip plane above. Climb can continue if atom A_2 jumps into vacancy O_2 , and so on. The upward movement of the dislocation is called positive climb (Figure 3.23a). The downward movement is accomplished by the addition of interstitial atoms in the extra half plane (e.g., I_1 and I_2 in Figure 3.23b) and is called negative climb. Both positive and negative climb require the diffusion of point defects, vacancies and interstitials, towards the dislocation. Climb is, therefore, thermally activated. It will be shown in Chapter 7 that climb is one of the major mechanisms by which recovery takes place, during the annealing of metals. Recovery decreases the hardness of a metal, which has been subjected to cold working. Dislocation climb is also a major mechanism during the creep of metals, i.e., the plastic deformation at high temperatures. As a final remark, it should be noted that dislocation glide and cross slip do not change the atomic arrangement of the crystal. They are conservative processes. On the other hand, dislocation climb requires the diffusion of point defects and is, therefore, a non-conservative process.

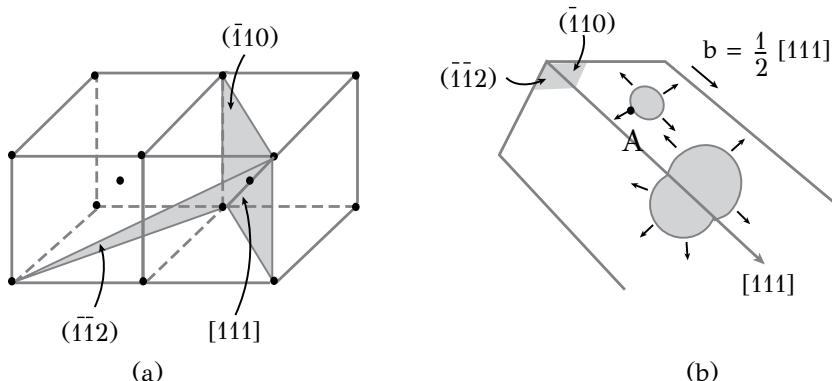


Figure 3.22: Cross-slip: (a) close-packed direction $[111]$ in BCC, (b) switch of slip plane by a dislocation loop.

3.3.9 JOGS AND KINKS , DISLOCATION REACTIONS

If the dislocation climb, discussed above, is limited only to a segment of the dislocation line, then two steps are formed, termed dislocation *jogs*. Jogs transfer a segment of the dislocation line to a different slip plane, as depicted in Figure 3.24a. Jogs should not be confused with dislocation *kinks*, which are steps on the same slip plane with the rest of the dislocation line (Figure 3.24b,c). Jogs and kinks are dislocation segments, with the same Burgers vector as the main dislocation line to which they belong. Jogs and kinks follow the same rules regarding conservative and non-conservative motion. Therefore, kinks sharing the same slip plane with the rest of the line, do not impede the glide of the dislocation, whether it is an edge or a

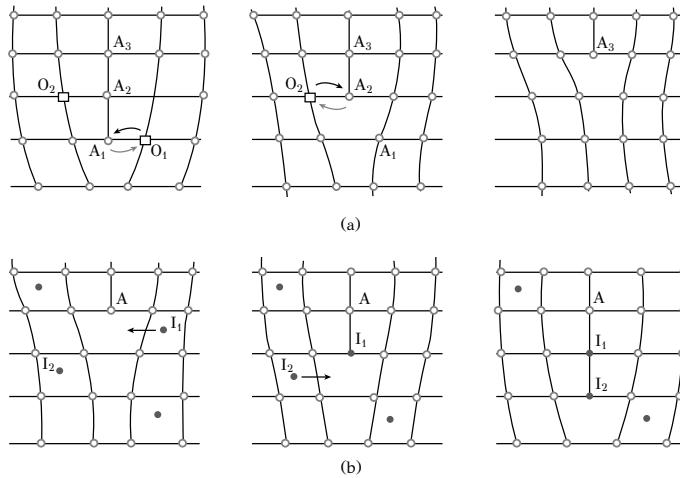


Figure 3.23: (a) Positive climb of an edge dislocation with vacancy diffusion towards the dislocation, (b) negative climb with diffusion of self-interstitials towards the dislocation.

screw dislocation. Jogs on a screw dislocation, however, exhibit an edge character and can move only along the dislocation line, as depicted in Figure 3.24d. However a motion, perpendicular to the Burgers vector, requires climb, which is thermally activated. This impedes the glide of the screw dislocation and leads to the production of point defects during dislocation glide, in the case where glide takes place under a high stress. It is important to note that jogs on screw dislocations can form when two screw dislocations intersect, as shown in Figure 3.25. Screw dislocation *ab* with Burgers vector \vec{b}_1 glides on the slip plane P_1 . Screw dislocation *cd* with Burgers vector \vec{b}_2 intersects the slip plane at point *O*. After intersection a jog is created in each dislocation. In dislocation *ab* the jog has an edge character with Burgers vector \vec{b}_2 . This jog brings the dislocation segment *bb'* in the slip plane P_2 , which is below P_1 . The glide of dislocation *ab* is impeded since it requires the climb of the jog, which is thermally activated. At the same time a jog with Burgers vector \vec{b}_1 has formed on dislocation *cd*, which also has an edge character. It is obvious that dislocation intersections impede further dislocation glide and constitute one of the basic mechanisms of strain hardening, discussed in detail in Chapter 7.

3.3.10 PLASTIC DEFORMATION DUE TO DISLOCATION GLIDE

It has been shown that dislocation glide, under the action of applied stresses, causes plastic deformation. The plastic strain will be calculated as a function of dislocation density, based on the observation that two atoms, located on the two opposing sides of the slip plane, will be displaced relative to each other by b , when the dislocation passes from their position. In Figure 3.26 consider a crystal with dimensions L, H

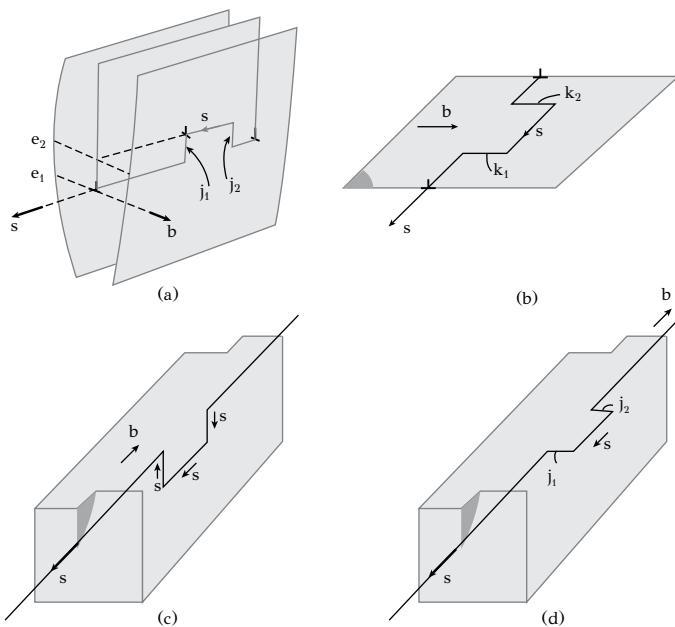


Figure 3.24: (a) Jogs on an edge dislocation, (b) kinks on an edge dislocation, (c) kinks on a screw dislocation, (d) jogs on a screw dislocation.

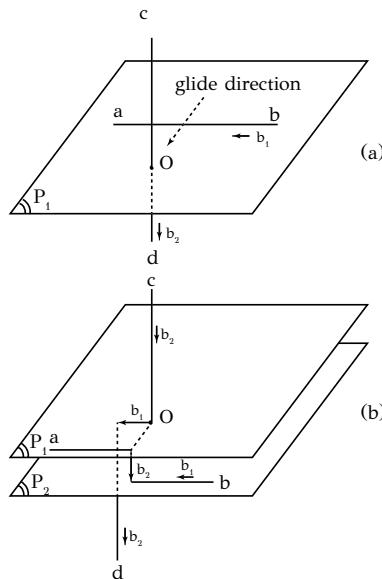


Figure 3.25: (a) Screw dislocation ab glides on slip plane P_1 and intersects screw dislocation cd . (b) The intersection results in jog formation on both screw dislocations.

and W , containing N mobile edge dislocations. When these dislocations glide for a distance W across the crystal, they produce a shear displacement S . Two of these dislocations A and B are shown in Figure 3.26a, gliding under the action of the shear stress τ . In Figure 3.26b, dislocation A has completed its glide by a distance W , while dislocation B has covered a distance x_B and has, therefore, contributed by $(x_B/W)b$ to the displacement S . In general, any dislocation (i) , which has undergone glide by a distance x_i , contributes by $(x_i/W)b$ to the displacement S . If the number of gliding dislocations is N then the total displacement S is

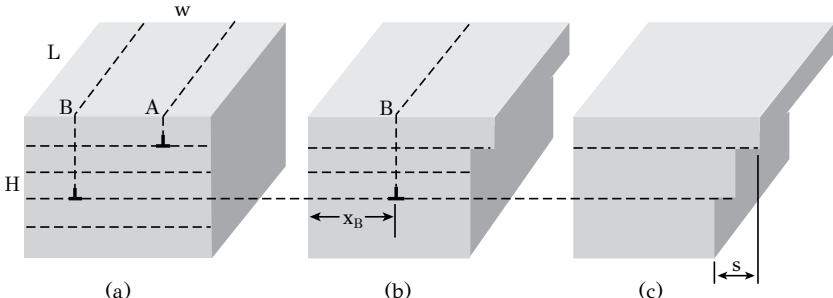


Figure 3.26: Plastic deformation by shear of the crystal caused by dislocation glide.

$$S = \frac{b}{W} \sum_{i=1}^N x_i$$

and the macroscopic plastic shear strain γ will be

$$\gamma = \frac{S}{H} = \frac{b}{WH} \sum_{i=1}^N x_i$$

Considering that the mean distance of dislocation glide is

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

and that the mobile dislocation density ρ_m is the total dislocation line divided by the crystal volume,

$$\rho_m = \frac{NL}{LWH}$$

the plastic strain becomes

$$\gamma = \rho_m b \bar{x} \quad (3.7)$$

and the strain rate is

$$\dot{\gamma} = \rho_m b \bar{v} \quad (3.8)$$

where \bar{v} is the average dislocation velocity. The dislocation velocity is a function of the applied shear stress acting on the slip plane

$$\bar{v} = A\tau^m \quad (3.9)$$

The dislocation velocity increases with stress, however it is limited by the velocity of the shear waves in the material. In the low velocity range ($10^{-7} - 10^{-1} m/s$), the exponent m in Relation (3.9) takes values $m = 25 - 35$, while at higher velocities ($1 - 10^4 m/s$) it decreases to $m = 1 - 5$.

3.3.11 ENERGY AND STRESS FIELDS OF DISLOCATIONS

Dislocations, as line defects, disrupt the periodicity of the crystal and cause lattice distortions, which raise the energy of the crystal. While the deformation associated with a vacancy extends only a short distance from the vacancy, the deformation associated with a dislocation extends at much longer distances, up to $100b$ from the dislocation core. This raises the elastic energy around the dislocation line considerably, while it favors long-range dislocation interactions. These interactions lead, in general, to a decrease of the free energy of the crystal. In order to understand these interactions and how they affect processes, such as the annealing of metals, the dislocation energy should be discussed. There are two contributions to the dislocation energy E :

$$E = E_s + E_o$$

where E_s is the elastic energy due to lattice distortions caused by the dislocation and E_o is the energy of the dislocation core. It is considered here that the size of the dislocation core is $r_o = 2 - 5b$. Inside this region the distortions are so large that it is impossible to compute stresses by applying the linear elasticity theory. Consider first the screw dislocation. The elastic strain around the screw dislocation corresponds to case c_2 of the Volterra cylinder in Figure 3.8c, which is reproduced in Figure 3.27. The shear strain of the stripe at position r is

$$\gamma_{z\theta} = \frac{b}{2\pi r}$$

According to Hooke's law, the corresponding shear stress is

$$\sigma_{z\theta} = \mu \gamma_{z\theta} = \frac{\mu b}{2\pi r} \quad (3.10)$$

The stress field around a screw dislocation is pure shear. The elastic strain energy per unit length of the stripe, with thickness dr , is

$$dE = \frac{1}{2} \sigma_{z\theta} \gamma_{z\theta} dV = \frac{\mu b^2}{4\pi} \frac{dr}{r}$$

and integrating from r_o to a distance R around the dislocation line

$$E = \int_{r_0}^R dE = \frac{\mu b^2}{4\pi} \ln \frac{R}{r_0}$$



Figure 3.27: Deformation of a thin stripe of material around a screw dislocation.

The total dislocation energy per unit length of dislocation line is then

$$E = \frac{\mu b^2}{4\pi} \ln \frac{R}{r_o} + E_o \quad (3.11)$$

The stress field around an edge dislocation is more complicated than that of screws. It involves shear on the slip plane as well as tension and compression below and above the slip plane respectively (Figure 3.28). The stresses are

$$\sigma_{rr} = \sigma_{\theta\theta} = -\frac{\mu b}{2\pi(1-\nu)} \frac{\sin \theta}{r} \quad (3.12)$$

$$\sigma_{r\theta} = \frac{\mu b}{2\pi(1-\nu)} \frac{\cos \theta}{r} \quad (3.13)$$

The stresses decay with $1/r$ and the energy of the edge dislocation is

$$E = \frac{\mu b^2}{4\pi(1-\nu)} \ln \frac{R}{r_0} + E_o \quad (3.14)$$

The elastic strain energy of the dislocations can be approximated by

$$E = \frac{1}{2} \mu b^2 \quad (3.15)$$

The magnitude of the dislocation energy is considerable. If $r_o = 10\text{\AA} = 10^{-7}\text{cm}$, $R = 10^{-4}\text{cm}$ and $b = 3\text{\AA} = 3 \times 10^{-8}\text{cm}$, then $E = 10^8\text{eV/cm}$. This corresponds to

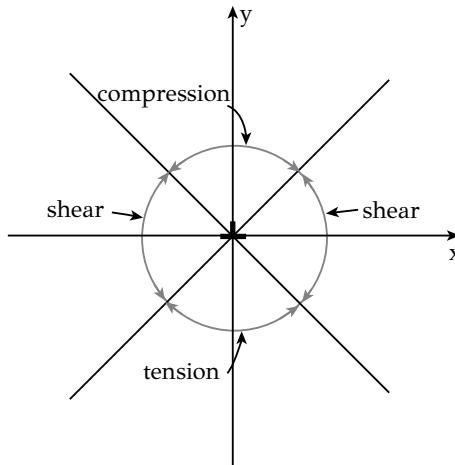


Figure 3.28: Stress field around an edge dislocation.

6eV per crystal plane crossed by the dislocation line. For comparison, the energy of a vacancy is 1eV , the energy of fusion is approximately 0.1eV per atom and the energy of vaporization is 1eV per atom. Relations (3.11) and (3.14) provide the total energy of screw and edge dislocations respectively, per unit length of dislocation line. Two important remarks can be made:

- The energy E varies with $\ln R$ and the long-range stress field allows the dislocation to interact with point defects, solute atoms and other dislocations at long distances, up to $100b$, from the dislocation core.
- The energy of the dislocation is proportional to b^2 . This means that dislocation reactions, leading to a reduction of strain energy, are favored. For example, a dislocation can dissociate into partials, if the Franks rule is obeyed.

3.3.12 FORCES ON DISLOCATIONS

Three forces acting on a dislocation line will be discussed in this section: the glide force, the line tension of a dislocation and the force required to bend a dislocation between obstacles.

Glide force. Plastic deformation of metals takes place by the glide of dislocations on the close-packed planes in the close-packed directions of the lattice. Glide takes place under the action of a shear stress, which acts on the slip plane, parallel to the slip direction. This stress is called resolved shear stress, since the macroscopic applied stress is resolved on the slip plane and in the slip direction. As discussed in section 3.3.8, for glide to be activated, the resolved shear stress must acquire the critical value $CRSS$. Consider now that a shear stress greater than the $CRSS$ acts on a dislocation. This stress corresponds to an applied force F , which is the *glide force*, acting perpendicular to the dislocation line. Considering Figure 3.26, the work dW

per unit length of dislocation, performed by F during glide by distance x_i is

$$\frac{dW}{L} = Fx_i$$

On the other hand the shear stress performs work given by

$$dW = (\tau A) \left(\frac{x_i}{W} b \right) = (\tau WL) \left(\frac{x_i}{W} b \right)$$

where the first term corresponds to the shear force acting on the slip plane with area $A = WL$ and the second term to the shear displacement of the crystal when the dislocation glides for a distance x_i . The above relation becomes

$$\frac{dW}{L} = \tau bx_i$$

From the above relations the glide force of a dislocation is

$$F = \tau b \quad (3.16)$$

The glide force is perpendicular to the dislocation line.

Line tension. The energy of a dislocation is proportional to its length and decreases as its length decreases. Another way to express this observation is to say that the dislocation exhibits a *line tension*, which is a force acting to reduce its length, like in a rubber band. The line tension T has units of energy per unit length and according to the Relation (3.15) it is given by

$$T = \frac{1}{2} \mu b^2 \quad (3.17)$$

Bending force. Consider now the force required to bend a dislocation between obstacles. In Figure 3.29a a dislocation with length dL bends with a radius R . The line tension T acts to reduce the dislocation length and imposes a force F_1 in the O_x direction

$$F_1 = 2T \sin \frac{d\theta}{2}$$

For small angles θ , $\sin \frac{d\theta}{2} \approx \frac{d\theta}{2}$ and $d\theta = \frac{dL}{R}$, the above relation becomes

$$F_1 = \frac{T dL}{R}$$

The force required to maintain the dislocation in the curved position is, according to Relation (3.16),

$$F_2 = \tau b dL$$

At equilibrium $F_1 = F_2$ and the stress τ is

$$\tau = \frac{T}{bR}$$

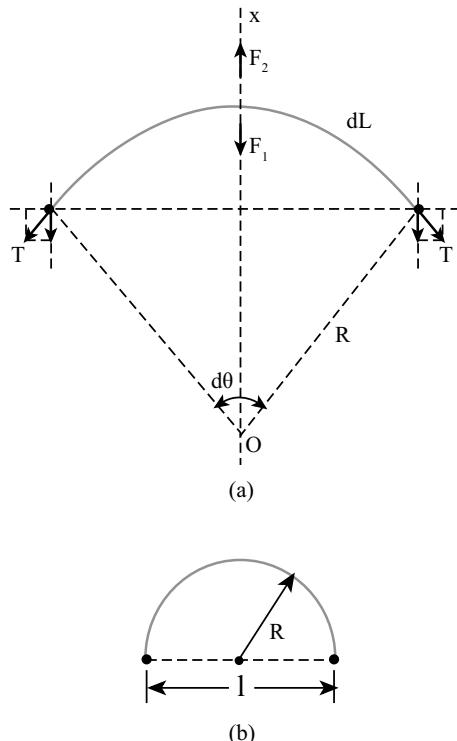


Figure 3.29: (a) Bowing of a dislocation between two obstacles, under the action of a shear stress, (b) the critical condition for obstacle by-passing.

Using the Relation (3.17) the stress becomes

$$\tau = \frac{\mu b}{2R}$$

As discussed above, when the dislocation encounters obstacles on the slip plane, it bends between them. In order to bypass these obstacles, the dislocation bends to a semicircle (Figure 3.29b). If the distance between obstacles is $l = 2R$, then the above relation gives

$$\tau = \frac{\mu b}{l} \quad (3.18)$$

Relation (3.18) expresses an important result: the stress required by a dislocation to bypass obstacles on the slip plane, in other words the resistance to dislocation glide, is inversely proportional to the distance between obstacles. If a fine and dense dispersion of obstacles is formed in the material, the strength of the material will be increased, since high resistance to dislocation glide will be exhibited. This observation forms the basis of alloy design, including chemical composition and processing, for precipitation strengthening in many metallic systems.

3.3.13 OBSERVATION OF DISLOCATIONS

As discussed in section 3.2.2, dislocations were first observed in the mid 50s with the development of the transmission electron microscope (*TEM*). According to this technique, an electron beam is transmitted through a very thin metal foil. The beam is diffracted from the crystal planes. However the lattice distortions associated with dislocations cause the rotation of crystal planes. The specimen is tilted relative to the electron beam to a position that the diffraction from the distorted planes is strong. In this way a black line-type image is taken, which corresponds to the projection of a curved dislocation on the microscope screen. A bright-field *TEM* image in a $Fe - Mo - C$ steel is depicted in Figure 3.30a. The thin black lines correspond to projection of dislocations, as indicated in Figure 3.30b. The black needle-shaped particles are Mo_2C carbides, which have precipitated during the tempering of the steel. From images like Figure 3.30a, the interaction of dislocations with obstacles, leading to alloy strengthening, can be studied. Dislocations and dislocation loops are also shown in Figure 3.30c in an $Al - Mg - Si$ alloy.

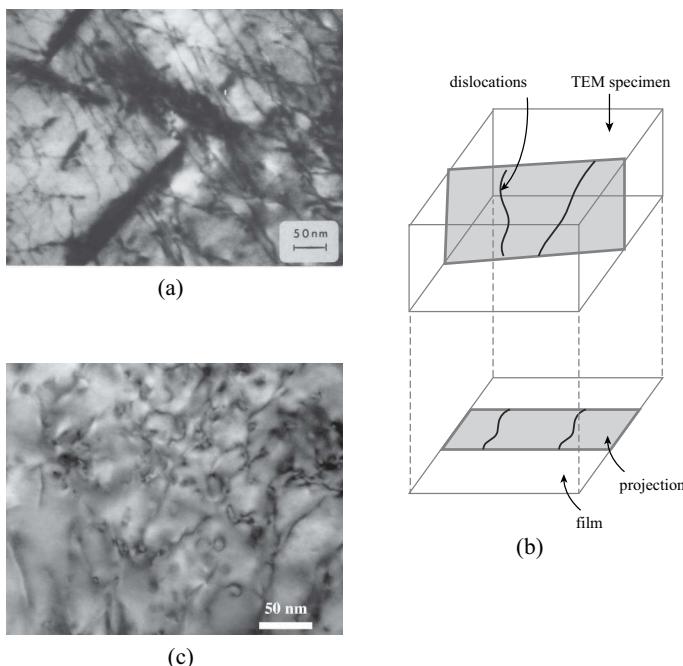


Figure 3.30: (a) Dislocations in a $Fe - Mo - C$ steel, (b) schematic depicting the projection of dislocations on the image plane, (c) dislocations and dislocation loops in a $Al - Mg - Si$ alloy.

3.3.14 DISLOCATION EFFECTS

The most important effect of dislocations is their role in the plastic deformation of metals. Dislocation glide causes plastic deformation, since by the glide of a single dislocation, the crystal is sheared by the magnitude of the Burgers vector. Dislocation glide takes place at specific slip systems consisting of the slip plane and slip directions. Slip takes place at the close-packed planes in the close-packed directions. It follows that, since the *FCC* structure exhibits the highest number of slip systems from all other crystal structures, *FCC* metals are easier to deform than other metals. Plastic deformation and more specifically crystal plasticity, will be discussed in detail in Chapter 7. A second, equally important, effect of dislocations is related to the strengthening of metals. Strengthening mechanisms are simply mechanisms by which various microstructural obstacles impede dislocation glide. Examples of strengthening mechanisms are solid solution hardening, strain hardening and precipitation hardening. In the first case, in a *Cu – Zn* solid solution, dislocation glide is impeded by the strain fields, generated by the *Zn* atoms in substitutional sites. In the second case, strain hardening, dislocations are impeded by other dislocations, which cross the slip plane. Finally in the third case, precipitation hardening, dislocation glide is impeded by a fine dispersion of intermetallic compounds formed through precipitation. A good example is the strengthening of *Al – Cu* alloys by a fine dispersion of the Al_2Cu intermetallic phase, which forms during the aging treatment of the alloy. The strengthening mechanisms will be discussed in detail in Chapter 8. A third effect of dislocations is related with phase transformations in metals and alloys. Dislocations assist the formation of a new phase, since they act as sites for the heterogeneous nucleation of the new phase. If, for example, in the case of the *Al – Cu* alloy discussed above, the alloy is subjected to cold rolling before the aging heat treatment, then the dislocation density will be increased. This simply means that there will be more sites available for the heterogeneous nucleation of the Al_2Cu phase during aging, forming a finer particle dispersion, which leads to higher strength. Finally, another effect of dislocations is related to their role as high-diffusivity paths. In the case of edge dislocations, for example, there is enough free space below the extra half plane. This space forms essentially a channel for the diffusion of atoms. In this way, diffusion is accelerated and consequently the rate of diffusional phase transformations is increased.

3.4 INTERFACES

3.4.1 INTRODUCTION – INTERFACIAL ENERGY

Interfaces constitute an integral part of the microstructure of metals and, therefore, have a direct effect on the mechanical properties of metals and alloys. The external surfaces as well as the internal interfaces influence a series of phenomena, such as oxidation and corrosion, friction and wear, intergranular fracture, brittle fracture, fatigue and creep, nucleation and growth in phase transformations, recrystallization and grain growth during annealing. In addition the interfaces play an important role

in materials processing such as casting, welding, surface treatments and sintering. The discussion of interfaces in this chapter is only an introduction to the subject of interfaces. The interested reader should consult the bibliography at the end of the chapter for a more detailed treatment. The primary interfaces in the microstructure of metals and alloys are:

- free surfaces of crystals
- grain boundaries
- interphase boundaries (boundaries between phases)
- stacking faults

Metals are surrounded by *free surfaces*, where the metal is in contact with the working environment. *Grain boundaries* separate grains, i.e., regions of crystal with the same crystal structure but with different orientation. *Interphase boundaries* separate two phases, which can have the same or different crystal structure and chemistry. *Stacking faults* represent a disruption of the normal stacking of crystal planes. An example of grain boundaries and interphase boundaries is shown in Figure 3.31 in a ferritic-pearlitic steel. Most phase transformations in metals take place by nucleation and growth of a new phase β in the matrix of phase α . The kinetics of this phase transformation depends on the fine structure of the α/β interface. Interphase boundaries are therefore the most important, but at the same time the most complicated, type of interface. Grain boundaries play a key role in recrystallization, the generation of new undeformed grains in a microstructure that has been subjected to cold working. It is clear, from the above discussion, that interfaces play a key role in the development of microstructure in metals and alloys.



Figure 3.31: Microstructure of a carbon steel ($Fe - 0.4C$): Grain boundaries and interphase boundaries between ferrite (white) and pearlite (black).

Interfaces in metals increase the free energy. In a pure metal, the change of free energy by the generation of a new surface dA is

$$dG = Vdp - SdT + \gamma dA$$

where γ is the interfacial energy, in units of J/m^2 . It represents the work that should be consumed in order to generate a unit area of interface under constant pressure and temperature, i.e.,

$$\gamma = \left(\frac{\partial G}{\partial A} \right)_{p,T} \quad (3.19)$$

It will be shown, in the next chapters, that the interfacial energy γ influences several important phenomena in metals.

3.4.2 FREE SURFACES

The surface energy of a free surface arises from the energy of “broken” bonds of surface atoms. While the coordination number in the bulk of an *FCC* crystal is 12, the same number at the surface is less than 12, meaning that the surface atoms form fewer bonds than the atoms in the bulk. The energy of the free surface of a crystal depends on the number of broken bonds. However this number varies from one crystal plane to the other, since the atomic density of crystal planes varies. This makes the surface energy highly anisotropic. Low index planes, such as {111} in *FCC* exhibit higher atomic densities (fewer broken bonds) and have lower surface energies. The anisotropy of the energy of free surfaces determines the *equilibrium shape* of a single crystal. In most cases a faceted crystal is obtained. For a single crystal surrounded by facets A_1, A_2, \dots with corresponding energies $\gamma_1, \gamma_2, \dots$ the equilibrium shape is defined by the equation

$$\sum_i A_i \gamma_i = \min, i = 1, 2, \dots \quad (3.20)$$

which expresses the minimization of the total surface energy for a given crystal volume. Relation (3.20) defines that the equilibrium shape of a single crystal, termed the *Wulff shape*, is a polyhedron, where the largest faces exhibit the lowest energy. Values of the surface energies for various metals appear in Table 3.1 and indicate that the surface energy of a free surface scales with the melting point. In general metals exhibit high surface energies. The surface energy of the water-air interface is about $70mJ/m^2$. Polymers have surface energies in the range $10 - 30mJ/m^2$.

3.4.3 GRAIN BOUNDARIES

A grain boundary separates two regions of a crystal with the same crystal structure but with different lattice orientation. The difference in orientation is expressed with the angle θ (misorientation angle), depicted in Figure 3.32a. If the grain in the right is rotated by an angle θ anticlockwise around an axis parallel to the grain boundary, then the two orientations will coincide. The boundary is then called a *tilt boundary*.

Table 3.1
Surface energy of several metals

Metal	Melting Point $^{\circ}\text{C}$	Surface Energy (mJ/m^2)
<i>Sn</i>	232	680
<i>Al</i>	660	1080
<i>Ag</i>	961	1120
<i>Au</i>	1063	1390
<i>Cu</i>	1084	1720
$\delta - \text{Fe}$	1536	2080
<i>Pt</i>	1769	2280
<i>W</i>	3407	2650

Reprinted from: H. Jones, *The surface energy of solid metals*, Metal Science Journal, 5 (1971) 15, with permission from Elsevier.

There is another type of grain boundary, the *twist boundary*, where the rotation by θ is performed around an axis perpendicular to the boundary, as shown in Figure 3.31b. Grain boundaries can be also distinguished by the magnitude of the misorientation angle θ . For $\theta < 10 - 15^{\circ}$ the grain boundaries are called *low-angle boundaries*. It can be considered that low-angle boundaries can be represented by dislocation arrays, edge dislocations in tilt boundaries and screw dislocations in twist boundaries, as depicted in Figure 3.33 for a tilt boundary. The interfacial energy of low-angle boundaries consists of the total energy of the dislocations at the boundary. The dislocation energy depends on the distance between dislocations, which is approximately

$$s = \frac{b}{\sin \theta} \cong \frac{b}{\theta}$$

For very low values of $\theta (2 - 4^{\circ})$ the distance s between dislocations is large and the interfacial energy is small. As θ increases, the number of dislocations and, therefore, the grain boundary energy increases, as depicted in Figure 3.34. Further increase of θ leads to saturation of dislocations at the boundary. Further increase of the grain boundary energy is limited by the self-accommodation of the stress fields of dislocations. For $\theta > 15^{\circ}$, the boundaries are called *high-angle boundaries*. They are associated with a high degree of disregistry of atoms at the grain boundary. Their structure is more “open” as they contain a larger number of broken bonds (Figure 3.35a). There is also more space for the segregation of foreign atoms and impurities, phenomena that will be discussed in later chapters of the book. Grain boundaries in a Ni-Cu Monel alloy are depicted in Figure 3.35b. Due to atomic disregistry at the boundary, high-angle grain boundaries arrest dislocations and impede dislocation glide from grain to grain. This leads to significant strengthening, and to the associated grain size dependence of the yield strength of metals, an issue to be discussed in Chapter 8.

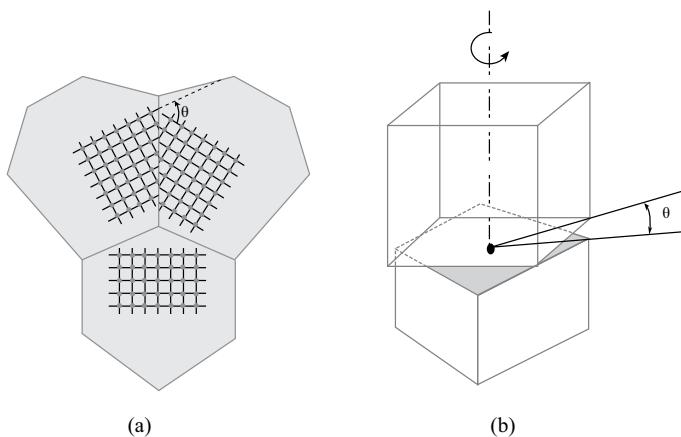


Figure 3.32: (a) Tilt boundary and (b) twist boundary. θ is the misorientation angle.

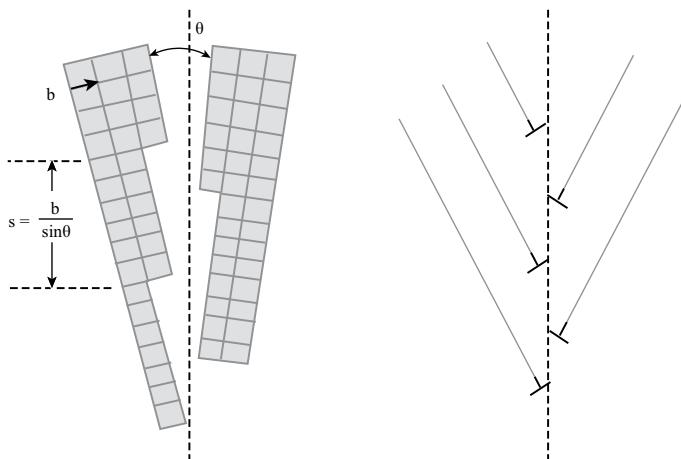


Figure 3.33: Low-angle tilt boundary formed by dislocation arrays.

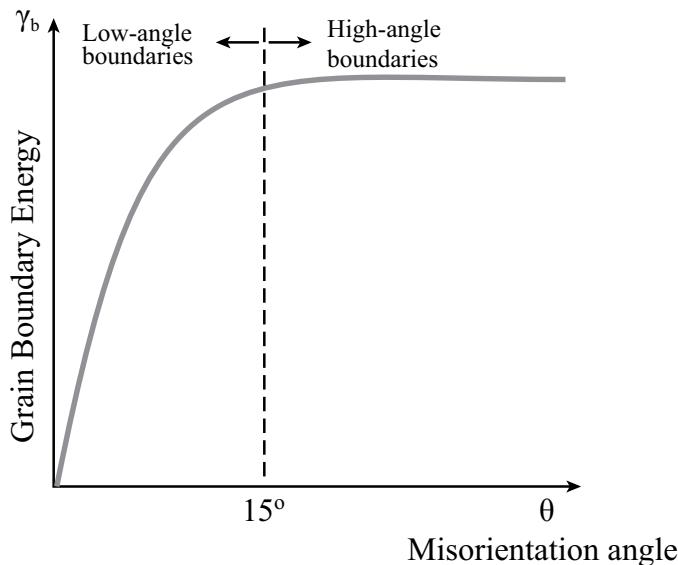


Figure 3.34: Change of grain boundary energy with the misorientation angle θ between two grains

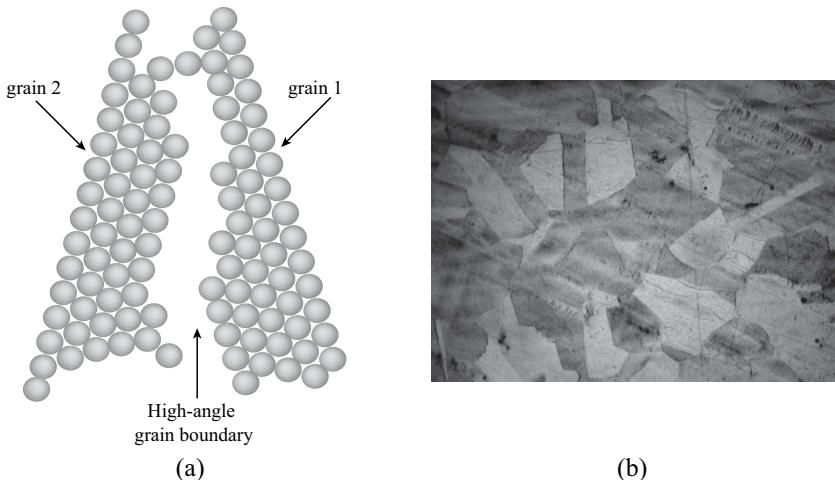


Figure 3.35: (a) Structure of a high-angle grain boundary, (b) grains and grain boundaries in a $Ni - Cu$ alloy (Monel).

3.4.4 INTERPHASE BOUNDARIES

While grain boundaries separate two grains of the same phase, *interphase boundaries* separate two different phases. These phases can exhibit a different crystal structure or chemical composition. Depending on whether there is full, partial or no coincidence of the two crystal lattices at the boundary, interphase boundaries are distinguished to *coherent*, *semicoherent* and *incoherent boundaries*. Coherent boundaries separate two phases, α and β , and the two lattices exhibit full coincidence at the boundary. In other words, continuity of crystal planes from one phase to the other is maintained, as depicted in Figure 3.36. Due to the change of phase and chemical composition, the atoms at the boundary exhibit a higher energy than the atoms in the bulk. This is the source of the *coherent interfacial energy*, γ_{coh} . The lattice parameters of the two phases are generally different. In this case the two lattices are distorted at the boundary in order to maintain coherency. The strains that develop are called *coherency strains* and raise the total free energy of the crystal. When the difference between the lattice parameters of the two phases, or lattice misfit, is large, the coherency strains increase and it might be energetically more favorable not to maintain coherency in the whole boundary. In this case the boundary is called semicoherent boundary, and the coherency strains are relaxed through the formation of *misfit dislocations*, as depicted in Figure 3.37. The lattice misfit δ is given by

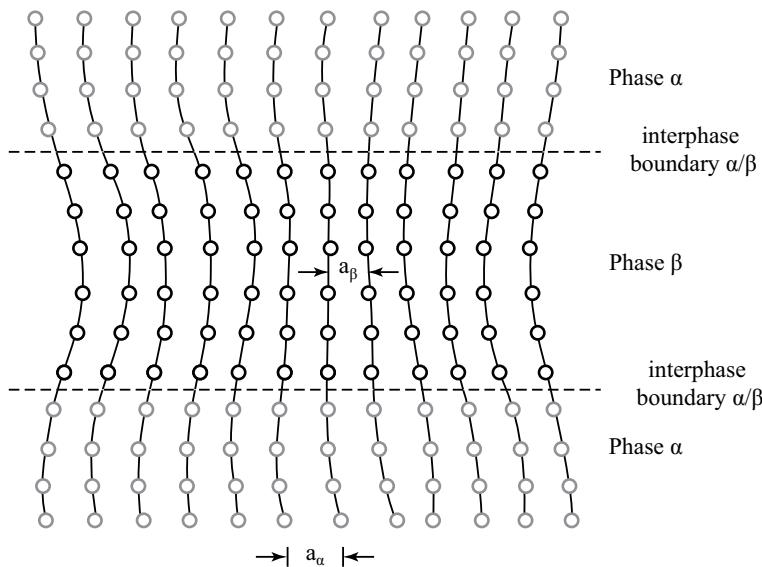


Figure 3.36: Coherent interphase boundary between two phases α and β with small lattice misfit, causing coherency strains.

$$\delta = \frac{\alpha_\alpha - \alpha_\beta}{\alpha_\alpha}$$

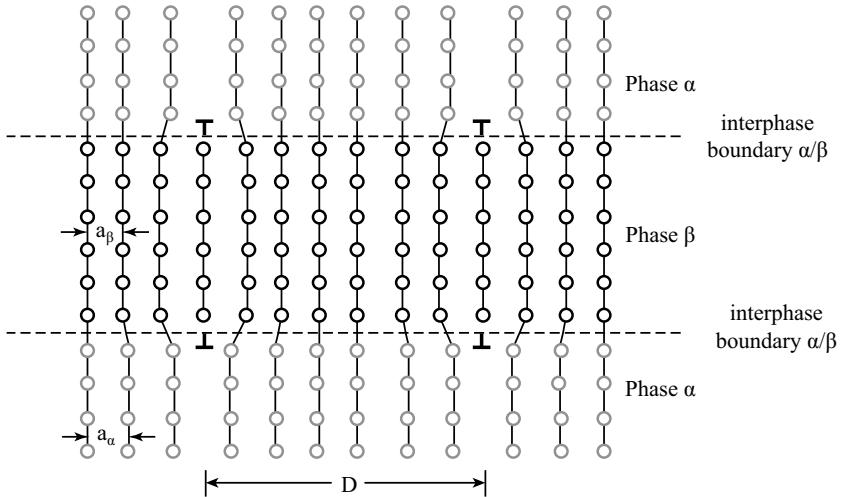


Figure 3.37: Semicoherent interphase boundary between two phases α and β with a large lattice misfit, which is accommodated by the formation of misfit dislocations.

The spacing D between misfit dislocations is

$$D = \frac{b}{\delta}$$

where b is the Burgers vector of misfit dislocations,

$$b = \frac{\alpha_\alpha + \alpha_\beta}{2}$$

With the introduction of misfit dislocations, coherency between the two phases is maintained partially and the coherency strains are relaxed. However the interfacial energy increases, due to the local strains generated by the misfit dislocations. In this case the *semicoherent interfacial energy* γ_{sem} is composed of two parts, the coherent part γ_{coh} and the part γ_D arising from the misfit dislocations. The last part γ_D depends on the number of dislocations at the boundary, which in turn depends on the lattice misfit δ . For very large values of the lattice misfit ($\delta > 0.25$), even partial coherency cannot be maintained. In this case the boundaries are called incoherent boundaries, depicted in Figure 3.38. Their structure is similar to high-angle boundaries or grain boundaries, discussed in the previous section. The *incoherent interfacial energy* γ_{inc} is very large and is associated to the large number of broken bonds at the boundary, related to the different crystal structure and chemical composition of the two phases. The range of values of the interfacial energy for the three types of interphase boundaries is given in Table 3.2.

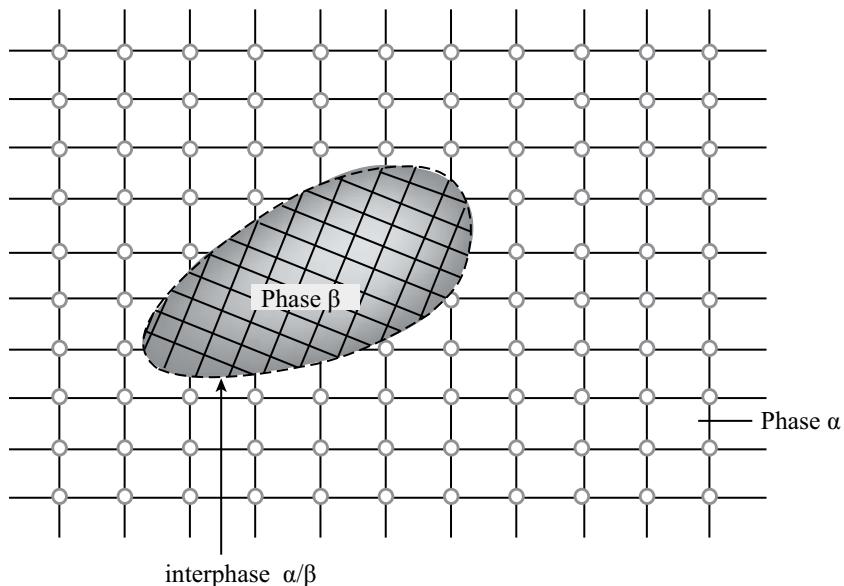


Figure 3.38: Incoherent interphase boundary.

Table 3.2
Interfacial energy of interphase boundaries

Interphase boundary	Interfacial Energy (mJ/m^2)
Coherent	1-200
Semicohesive	200-500
Incoherent	500-1000

3.4.5 INTERFACE EFFECTS

The interfaces influence in many ways the development of microstructure and the mechanical behavior of metals and alloys. Only the major effects will be discussed here. Grain boundaries impede dislocation glide and contribute to strengthening. The finer the grain size, the higher the strength. Grain boundary strengthening is discussed in detail in Chapter 8, however a relation between the yield strength σ_y and the mean grain size d , the Hall-Petch relation, is given here

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

where σ_i and k_y are constants. Grain boundaries influence the kinetics of phase transformations in two ways: (a) they provide sites for the heterogeneous nucleation of the new phase and (b) they act as high diffusivity paths, enhancing the growth of the new phase. Grain boundaries play an important role in creep, the plastic deformation of metals at high temperatures. Fine-grained alloys exhibit low creep strength, especially under conditions of diffusional flow, a subject that is discussed in detail in Chapter 9. Finally, grain boundaries attract tramp elements, such as P and S . The segregation of these elements weakens the grain boundaries and causes intergranular fracture. The nature of interphase boundaries (coherent, semicoherent, incoherent) plays an important role: (a) on the kinetics of phase transformations and (b) in the development of high strength in an alloy. As discussed in Chapter 6, the precipitation of a new phase from a supersaturated solid solution, follows a precipitation sequence. Precipitation starts with the formation of metastable phases which are initially coherent with the matrix, in order to reduce the interfacial energy. As these phases grow, the associated coherency strains increase. Precipitation then continues with the formation of semicoherent phases, which exhibit lower coherency strains. The precipitation ends with the formation of the equilibrium phase, which is incoherent with the matrix. During this precipitation sequence, the hardness of the alloy changes. This change is associated with the different ways with which the dislocations interact with the interphase boundaries. These interactions, which form the basis of precipitation strengthening, will be discussed in detail in Chapter 8.

3.5 SYNOPSIS

1. Structural imperfections have a significant influence on physical and mechanical properties of metals. They can be classified as (a) point defects, such as vacancies and interstitials, (b) linear imperfections, such as edge and screw dislocations, (c) surface imperfections, such as grain boundaries and interfaces and (d) three-dimensional defects, such as voids and inclusions.
2. At each temperature there exists a certain concentration of point defects at thermodynamic equilibrium in the crystal. The concentration of vacancies and interstitials increases exponentially with temperature, following Arrhenius-type temperature dependence. The concentration of interstitials is several orders of magnitude

lower than the concentration of vacancies at the same temperature.

3. Vacancies play a key role in the diffusion of atoms, especially the substitutional diffusion.
4. Plastic deformation takes place by dislocation glide in specific slip planes (close-packed planes) and slip directions (close-packed directions), which constitute the slip systems of a metal.
5. A slip system becomes operational by the application of a certain stress on the slip plane, called the critical resolved shear stress (CRSS), which is directly related to the mechanical strength of the metal.
6. The glide of an edge dislocation takes place in a direction parallel to the applied shear stress while the glide of a screw dislocation takes place in a direction perpendicular to the applied stress. In both cases the glide causes plastic deformation of the crystal.
7. At every position during glide, the dislocation line is the boundary between the part of the crystal that has slipped and the part which has not.
8. The stress required for dislocation glide is much lower than the ideal crystal strength. This explains the difference between the actual and ideal strength of a metal.
9. The direction and magnitude of slip, caused by a dislocation, is expressed by the Burgers vector \vec{b} of the dislocation line. The Burgers vector is determined by the Burgers circuit.
10. The Burgers vector of an edge dislocation is perpendicular to the dislocation line. When this is valid for the entire dislocation line, then this is a pure edge dislocation. The Burgers vector of a screw dislocation is parallel to the dislocation line. When this is valid for the entire dislocation line, then it is a pure screw dislocation.
11. When a dislocation is not a pure edge neither a pure screw, it is a mixed dislocation. In this case the dislocation line forms a random angle with the Burgers vector.
12. Dislocations cannot terminate inside the crystal. They can terminate either at the crystal surface, at grain boundaries or on themselves, forming dislocation loops. The expansion of a dislocation loop under the action of a shear stress causes plastic deformation of the crystal.
13. The elastic strain energy of a dislocation is associated with strains caused by the

displacement of atoms away from their equilibrium positions around the core of the dislocation line. The elastic strain energy is proportional to the square of the Burgers vector.

14. In several cases, the dissociation of a perfect dislocation into partial dislocations is energetically favored. The glide of these partials has the same effect as the glide of the perfect dislocation.
15. The dissociation of a perfect dislocation in FCC crystals into partial dislocations is accompanied by the formation of a stacking fault, characterized by the stacking fault energy (SFE). Despite the fact that the dissociation decreases the elastic strain energy of the crystal, if the Frank's rule is obeyed, the dissociation will take place only if it decreases the total energy of the crystal.
16. In FCC metals with low SFE, slip takes place by the glide of partial dislocations, while in FCC metals with high SFE, slip takes place by the glide of perfect dislocations.
17. The change of slip plane by a screw dislocation is called cross slip. Cross slip of screw dislocations is important for the plastic deformation of metals, because it allows the dislocations to continue their glide in the crystal and produce plastic deformation.
18. The change of slip plane by an edge dislocation is called climb and requires the diffusion of vacancies or interstitials to the dislocation.
19. Jogs and kinks are steps on the dislocation line. Jogs transfer a segment of the dislocation line to a different slip plane while kinks are steps on the same slip plane with the rest of the dislocation line. Kinks and jogs on an edge dislocation do not impede the glide of the dislocation. However jogs on a screw dislocation, impede its glide, since climb is required in order for the jogs to move with the dislocation.
20. The plastic strain rate resulting from dislocation glide is proportional to the mobile dislocation density and the average dislocation velocity.
21. The stress field around a screw dislocation is pure shear. The stress field around an edge dislocation involves shear on the slip plane as well as tension and compression below and above the slip plane respectively.
22. The energy of a dislocation is proportional to b^2 . The long-range stress field allows the dislocation to interact with point defects, solute atoms and other dislocations at long distances, up to $100b$, from the dislocation core.
23. Regarding forces acting on dislocations: the glide force acts perpendicular to the

dislocation line. The line tension acts to reduce the dislocation length. The bending force acts to bend the dislocation between obstacles. The stress required in order for a dislocation to bypass obstacles on the slip plane, in other words the resistance of the obstacles to dislocation glide, is inversely proportional to the distance between obstacles.

24. The major effects of dislocations are related to plastic deformation, strengthening and phase transformations. In the last case, dislocations enhance both the nucleation and growth of a new phase, by providing sites for the heterogeneous nucleation and by serving as high-diffusivity paths for the diffusional growth of the new phase.
25. The primary interfaces in metals are the free surfaces, grain boundaries, interphase boundaries and stacking faults. All interfaces are characterized by an interfacial energy γ .
26. The surface energy of free surfaces is highly anisotropic. The equilibrium shape of a single crystal is the one that minimizes the total surface energy.
27. A grain boundary separates two regions of a crystal with the same crystal structure but with different lattice orientation. Grain boundaries are distinguished in tilt and twist boundaries as well as low-angle and high-angle boundaries.
28. Interphase boundaries separate two different phases, which may have different crystal structure or chemical composition. Depending on whether there is full, partial or no coincidence of the two crystal lattices at the boundary, interphase boundaries are distinguished in coherent, semicoherent and incoherent boundaries.

3.6 REVIEW QUESTIONS

1. A divacancy is a lattice defect consisting of two vacancies side by side. Consider the formation of divacancies in the *FCC* structure. The connection between the two vacancies in the pair is characterized by a binding energy and an associated binding entropy defined by:

$$E_{2v}^b = 2E_{1v} - E_{2v}$$

$$S_{2v}^b = 2S_{1v} - S_{2v}$$

where the notation $1v$ and $2v$ refers to single vacancies and divacancies respectively. Show that the fraction of divacancies, at thermodynamic equilibrium, is given by

$$X_{2v}^e = 6(X_{1v}^e)^2 \exp(-S_{2v}^b/k) \exp(E_{2v}^b/kT)$$

where X_{1v} is the fraction of single vacancies in the crystal.

2. Calculate and compare the equilibrium fraction of single vacancies, divacancies and self-interstitials in copper at $27^\circ C$ ($300K$) and at $1000^\circ C$ ($1273K$). The energies of formation of single vacancies and interstitials are 85 and $385 kJ/mol$ respectively while the binding energy of divacancies (defined in problem 1) is $30 kJ/mol$. Take the entropic term $S/k = 1.5$ in all cases ($k = 1.38 \times 10^{-23} J/k$ and $R = 8.314 J/molK$).
3. Show that the application of hydrostatic pressure reduces the equilibrium vacancy concentration.
4. The thermal expansion of metals arises mainly from the change of the lattice parameter with temperature. Considering that the formation of vacancies contributes to thermal expansion, the equilibrium vacancy concentration in a metal can be experimentally determined by dilatometry and X-ray diffraction (XRD). Show that, to a good approximation, the vacancy concentration is

$$X_v^e \cong 3 \left(\frac{\Delta L}{L_o} - \frac{\Delta \alpha}{\alpha_o} \right)$$

where $\Delta L/L_o$ is the length change at temperature T relative to a reference length L_o at a reference temperature T_o . In addition $\Delta \alpha/\alpha_o$ is the change of the lattice parameter at temperature T with respect to the reference lattice parameter at temperature T_o .

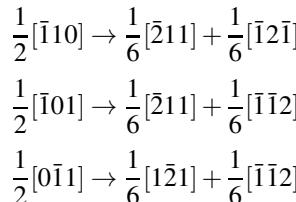
5. The equilibrium vacancy concentration is given by

$$X_v^e = \exp(S_{vib}/k) \exp(-E_v/kT)$$

The energy of vacancy formation can be determined from the slope of the Arrhenius-type diagram of $\ln X_v$ vs $1/T$, which is usually a straight line. However, in principle, the vibrational entropy as well as the vacancy formation energy are both temperature dependent due to thermal expansion. Considering these effects show that: (a) the $\ln X_v$ vs $1/T$ plot exhibits curvature and (b) the vacancy formation energy at any temperature T can be determined from the equation

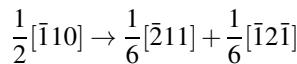
$$\left[\frac{\partial \ln X_v^e}{\partial (1/T)} \right]_p = -\frac{E_v(T)}{k}$$

6. Consider the dissociation of perfect dislocations into partials according to the reactions (see Figure 3.20):



- (a) Show that the Burgers vector is conserved after dissociation
- (b) Show that Frank's rule is obeyed for each reaction

7. Consider the dissociation of a perfect dislocation into partials according to the reaction

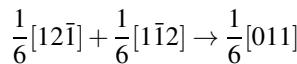


Show that the separation distance between the two partials is

$$d = \frac{\mu b^2}{4\pi(1-\nu)(SFE)}$$

where SFE is the energy of the stacking fault formed between the two partials and b is the magnitude of the Burgers vector of the two partial dislocations.

8. Show that the reaction between partial dislocations in an *FCC* metal (known as the Lomer-Cottrell lock)



is energetically feasible. Show that the new dislocation formed by this reaction cannot glide, i.e., it is a sessile dislocation. What are the consequences of the formation of Lomer-Cottrell locks on the plastic deformation of metals?

BIBLIOGRAPHY–FURTHER READING

1. Seeger, A. (1955). *Theory of Lattice Imperfections*, Handbuch der Physik, Vol.7, part 1, Springer-Verlag.
2. Wollenberger, H.J. (1983). Point Defects, in *Physical Metallurgy*, R.W. Cahn and P. Haasen (eds.) North-Holland Physics Publishing, Amsterdam.
3. Weertman, J., & Weertman, J.R. (1964). *Elementary Dislocation Theory*, The Macmillan Company, New York.
4. Cottrell, A. H. (1963). *Dislocations and Plastic Flow in Crystals*, Oxford University Press.
5. Hull, D., & Bacon, D. J. (1984). *Introduction to Dislocations*, Pergamon Press.
6. Hirth, J. P., & Lothe, J. (1982). *Theory of Dislocations*, John Wiley.
7. Nabarro, F.R.N. (ed), (1979). *Dislocations in Solids*, Vol.1-5, North-Holland.
8. Mc Clintock, F. A., & Argon, A. S. (1966). *Mechanical Behavior of Materials*, Addison-Wesley.
9. Baluffi, R.W. (ed), (1980). *Grain Boundary Structure and Kinetics*, American Society of Metals.
10. Murr, L. E. (1975). *Interfacial Phenomena in Metals and Alloys*, Addison-Wesley, Massachusetts USA.



Taylor & Francis
Taylor & Francis Group
<http://taylorandfrancis.com>

4 Alloy thermodynamics and phase diagrams

4.1 INTRODUCTION

The most important tool for the study of alloys will be presented in this chapter: *the phase diagram*. Phase diagrams are maps, in the temperature-composition space, of the alloy constitution under conditions of thermodynamic equilibrium. The term constitution includes the phases and the composition of phases present in the system. With the aid of the phase diagram and for each alloy composition it is possible:

- to determine the fraction of phases and the composition of phases as a function of temperature
- to predict the changes in the alloy constitution for any change in temperature
- to determine parameters, which are useful for the processing of the alloy, such as liquidus, solidus and solvus temperatures, solidification ranges, eutectic, peritectic and eutectoid temperatures and invariant reactions, temperatures for the dissolution or precipitation of phases and intermetallic compounds, solubility limits and more.

Since phase diagrams are representations of thermodynamic equilibrium, they should be obtained directly from the minimization of the Gibbs free energy of the system. The total free energy is the sum of the free energies of all phases in the system. If the free energies of phases are expressed as functions of temperature and composition, it is then a straightforward matter to determine the conditions, under which the total free energy is minimized, i.e., the conditions of thermodynamic equilibrium. This procedure will be applied first in one-component systems, i.e., pure metals, where the free energy is only a function of temperature. Then the composition will be introduced as a variable and two-component systems, or binary systems, will be discussed. The steps in this process are the following:

- determine the free energy $G(T, X)$ of a binary solid solution
- introduce the concept of chemical potential, through which the equilibrium between two phases can be described
- introduce the Gibbs phase rule, which correlates the number of phases with the number of components and the degrees of freedom in a system
- construct binary phase diagrams from the corresponding free energy functions $G(T, X)$

Before proceeding, it is necessary to distinguish between the composition of the alloy and the composition of phases.

Alloy composition. The components of the system are the elements making up the alloy. For example, steel is an iron-carbon alloy. The components are *Fe* and *C*. The alloy composition refers to the relative amounts of the components in the alloy. The alloy composition can be expressed by either by the *atomic fraction* (or mole fraction) or the *weight fraction* (or weight%). Consider a binary alloy with components *A* and *B*. The composition of the alloy can be expressed by the atomic fraction of *B*, denoted by X_B . Between the two atomic fractions the following relation holds

$$X_A + X_B = 1$$

So if in a *Cu – Zn* alloy, the atomic fraction of zinc is $X_{Zn} = 0.25$, then $X_{Cu} = 0.75$, meaning that the alloy is composed of 25% *Zn* atoms and 75% *Cu* atoms. The atomic fraction coincides with the mole fraction. So if the alloy contains n_A moles of component *A* and n_B moles of component *B*, then

$$X_A = \frac{n_A}{n_A + n_B}$$

and

$$X_B = \frac{n_B}{n_A + n_B}$$

The other way to express the alloy composition is by the weight fraction of *B*, denoted by W_B . A *Cu – Zn* alloy with $W_{Zn} = 0.18$ is composed by 82wt% *Cu* and 18wt% *Zn*. The two ways of expressing the composition, atomic or weight fraction, are equivalent. If the atomic fraction X_B is known, then the weight fraction can be derived by the following relation

$$W_B = \frac{X_B(AW)_B}{X_A(AW)_A + X_B(AW)_B}$$

where $(AW)_A$, $(AW)_B$ are the atomic weights of the components *A* and *B* respectively.

Phase composition. A phase is a part of the system with homogeneous chemical composition, structure and properties, distinct from the other phases of the system. Phases are the solid solutions and intermetallic compounds, which form in a specific alloy. An alloy may be composed of one or more phases and be a single-phase, a two-phase or a multiphase alloy. The number of phases is not arbitrary but depends on the number of components and the number of variables (temperature, pressure) as described by the Gibbs phase rule, to be discussed in a following section. Phases are denoted by small greek symbols $\alpha, \beta, \gamma, \dots$. In a binary system, the composition of a phase is expressed by the atomic fraction X_B or the weight fraction W_B of the second component *B*. If a *Cu – Zn* alloy with composition X_{Zn} , consists of two phases α and β , then the compositions of the two phases are expressed by X_{Zn}^α and X_{Zn}^β . In any case, the alloy composition is distinct from the phase composition.

4.2 FREE ENERGY OF ONE-COMPONENT SYSTEMS (PURE METALS)

Pure metals are one-component systems and the free energy is a function of temperature, $G = G(T)$. The specific heat at constant pressure was defined in Chapter 2

as

$$C_p = \left(\frac{dH}{dT} \right)_p$$

Provided the temperature dependence of C_p is known, the enthalpy can be calculated by integrating the above relation

$$H = \int_{298}^T C_p dT$$

The lower integration limit is considered as the reference state and is defined as $H = 0$ at $T = 298K$ ($25^\circ C$), called the *stable element reference (SER)*. The variation of C_p and H with temperature is shown in Figure 4.1a and b respectively.

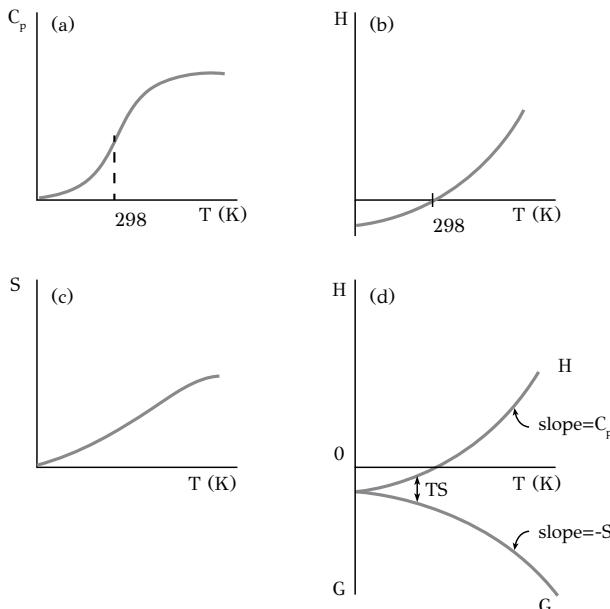


Figure 4.1: Variation of C_p , H , S and G with temperature for a pure metal.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). Phase Transformations in Metals and Alloys, with permission from CRC press.

The variation of entropy with temperature can also be obtained from the temperature dependence of the specific heat, since

$$\frac{C_p}{T} = \left(\frac{dS}{dT} \right)_p$$

Taking $S = 0$ at $T = 0$ and integrating

$$S = \int_0^T \frac{C_p}{T} dT$$

The variation of entropy with temperature is depicted in Figure 4.1c. Knowing the temperature dependence of H and S , the free energy can be derived from the relation $G = H - TS$. The free energy is shown in Figure 4.1d. The free energy is reduced with temperature with a rate

$$\left(\frac{dG}{dT} \right)_P = -S$$

The melting of a pure metal under constant pressure is a phase transformation and can be studied with the change of the free energy with temperature. The system is composed of two phases, solid (S) and liquid (L). The enthalpies and free energies of the solid and liquid phases are shown in Figure 4.2 as a function of temperature.

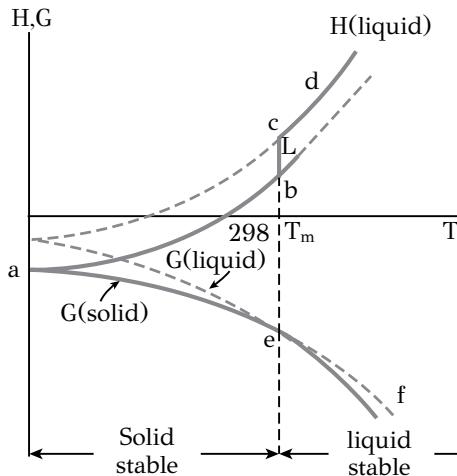


Figure 4.2: Variation of enthalpy (H) and free energy (G) for the solid and liquid phases of a pure metal with temperature. L is the latent heat of fusion and T_m the melting point.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). Phase Transformations in Metals and Alloys, with permission from CRC press.

At all temperatures the enthalpy of the liquid is higher than the enthalpy of the solid ($H^L > H^S$) due to the higher internal energy of the liquid. Heating the solid along line ab , at the melting point T_m the temperature remains stable during the solid \rightarrow liquid transformation (melting), while the enthalpy rises from point b on the H^S curve to point c on the H^L curve. The difference L corresponds to the *latent heat of fusion*. It is the heat supplied to the solid to complete the transformation (fusion).

The corresponding free energy curves of the solid and liquid phases cross at point e , the melting point T_m . At this temperature the free energies of the solid and liquid are equal and the solid and liquid phases coexist. Below T_m , $G^S < G^L$ and the solid phase is stable. Above T_m , $G^L < G^S$ and the liquid is the stable phase. In studying solidification, we are interested in the difference of free energy between the liquid and the solid phase at a temperature away from the melting point. A magnification around point e , where the free energy curves cross, is depicted in Figure 4.3. At an undercooling ΔT , the difference in free energies is $\Delta G = G^S - G^L$. This corresponds to the *driving force for solidification* of the metal.

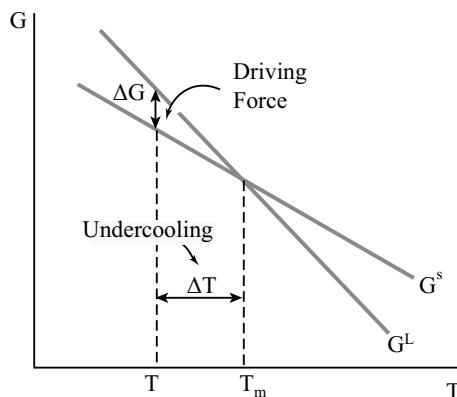


Figure 4.3: Variation of free energy of solid and liquid with temperature around the melting point. At an undercooling ΔT , the driving force for solidification is ΔG .

Allotropic transformations of pure metals can also be described with the variation of free energy with temperature. Consider the allotropic forms of iron, one is *BCC* and the other is *FCC*. The free energies of the two allotropic forms and the liquid are shown in Figure 4.4 as a function of temperature. Below 910°C *BCC* $\alpha - \text{Fe}$ has the

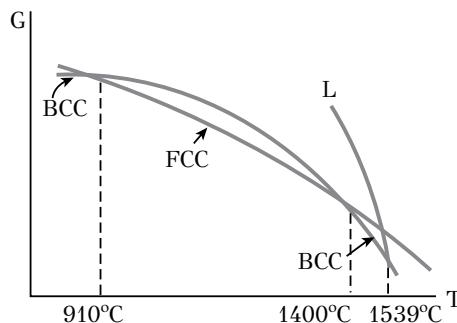


Figure 4.4: Variation of the free energy of $\alpha - \text{Fe}$ (BCC), $\gamma - \text{Fe}$ (FCC) and liquid (L) of pure iron with temperature.

lowest free energy and is, therefore, the stable phase. Between 910 and 1400°C the *FCC* γ -Fe exhibits the lowest free energy and is, therefore, the stable phase. The *BCC* δ -Fe becomes again the stable phase between 1400 and 1539°C, the melting point of iron. Above the melting point, liquid Fe is the stable phase, since it exhibits the lowest free energy between all phases. At 910°C α -Fe coexists with γ -Fe and at 1400°C, γ -Fe coexists with δ -Fe at thermodynamic equilibrium.

4.3 FREE ENERGY OF SOLID SOLUTIONS

While the free energy of a pure metal is only a function of temperature, $G = G(T)$, the free energy of a binary solid solution is a function of temperature and composition, $G = G(T, X)$. As discussed above, once the free energies of phases are expressed as functions of temperature and composition, it is then a straightforward matter to determine the conditions, under which the total free energy is minimized, i.e., the conditions of thermodynamic equilibrium. In this chapter, the free energy of a substitutional solid solution will be discussed. The model to be presented takes into account only first nearest neighbor interactions and is called the *regular solution model*. Consider a binary solid solution of metals A and B. The composition is expressed by the atomic fraction X_B . The free energy of the system before mixing or, in other words, the free energy of the pure components is

$$G_o = X_A G_A + X_B G_B \quad (4.1)$$

where G_A and G_B are the free energies of the pure metals. After mixing, the free energy of the solid solution is

$$G = G_o + \Delta G_m \quad (4.2)$$

where ΔG_m is the *free energy of mixing*. The solid solution will form only if mixing leads to a reduction in the free energy of the system, i.e., $G < G_o$, or $\Delta G_m < 0$. The free energy of mixing can be written as

$$\Delta G_m = \Delta H_m - T \Delta S_m \quad (4.3)$$

where ΔH_m and ΔS_m are the *enthalpy of mixing* and the *entropy of mixing* respectively. The entropy of mixing was discussed in Chapter 2. For a binary solid solution it is given by

$$\Delta S_m = -R(X_A \ln X_A + X_B \ln X_B) \quad (4.4)$$

From the above relation it is clear that $\Delta S_m > 0$ for all compositions ($0 < X_B < 1$). This means that the second term ($-T \Delta S_m$) of (4.3) is always negative. The sign of ΔG_m is then determined by the sign and magnitude of the enthalpy of mixing ΔH_m . The enthalpy of mixing can be written as

$$\Delta H_m = H - H_o \quad (4.5)$$

where H_o is the enthalpy of the pure metals (before mixing), which is given by

$$H_o = X_A H_A + X_B H_B$$

where H_A and H_B are the enthalpies of pure components A and B . The enthalpy of the solid solution can be determined under the assumption that it is composed of the bond energies of the first nearest neighbors. Three kinds of bonds can be distinguished in the solid solution with corresponding energies (enthalpies):

- N_{AA} bonds $A - A$ with bond energy H_{AA}
- N_{BB} bonds $B - B$ with bond energy H_{BB}
- N_{AB} bonds $A - B$ with bond energy H_{AB}

The enthalpy of the solid solution is then

$$H = N_{AA}H_{AA} + N_{BB}H_{BB} + N_{AB}H_{AB} \quad (4.6)$$

The number of bonds N_{AA} , N_{BB} and N_{AB} can be determined from the following argument. Consider that the atoms A and B are randomly distributed in the lattice sites and exhibit a specific coordination number Z (number of first nearest neighbors). If the atomic fraction of A is X_A then the probability that a lattice site is occupied by A is also X_A . The probability a lattice site to be occupied by B is X_B . The probability two atoms A and B to be first neighbors is X_AX_B . The probability two atoms A to be first neighbors is X_A^2 , while for two atoms B is X_B^2 . If the coordination number in the solid solution is Z ($Z = 8$ for *BCC* and $Z = 12$ for *FCC*, *HCP*) and there are N lattice sites in the solid solution, then the total number of bonds is $\frac{1}{2}ZN$. The number of $A - A$ bonds equals the total number of bonds times the probability for a bond to be an $A - A$ bond

$$N_{AA} = \frac{1}{2}ZNX_A^2$$

In the same way, the number of $B - B$ and $A - B$ bonds is

$$N_{BB} = \frac{1}{2}ZNX_B^2$$

$$N_{AB} = ZNX_AX_B$$

where in the last equation both $A - B$ and $B - A$ bonds have been counted. Substituting the number of bonds in (4.6), the enthalpy of the solid solution becomes

$$H = \frac{1}{2}ZN \left[X_AH_{AA} + X_BH_{BB} + 2X_AX_B \left(H_{AB} - \frac{H_{AA} + H_{BB}}{2} \right) \right] \quad (4.7)$$

The enthalpies of the pure metals A and B is derived from the above relation for $X_A = 1$ and $X_B = 1$ respectively

$$H_A = \frac{1}{2}ZNH_{AA}$$

$$H_B = \frac{1}{2}ZNH_{BB}$$

The enthalpy of the pure metals before mixing can be considered as the enthalpy of a mixture between those metals without taking into account the interaction between atoms (mechanical mixture), then

$$H_o = X_A H_A + X_B H_B = \frac{1}{2} ZN (X_A H_{AA} + X_B H_{BB})$$

which is equal with the first two terms of (4.7). The third term of (4.7) is the *excess term*, arising from the interaction between atoms A and B in the solid solution. If the mechanical mixture is considered as the *ideal solution*, then the excess term expresses the deviation of the solution from the ideal behavior. According to (4.5), the excess term is the enthalpy of mixing ΔH_m

$$\Delta H_m = ZN X_A X_B \left[H_{AB} - \frac{H_{AA} + H_{BB}}{2} \right]$$

Setting

$$ZN \left[H_{AB} - \frac{H_{AA} + H_{BB}}{2} \right] = \Omega \quad (4.8)$$

where Ω is the interaction parameter, the enthalpy of mixing can be written as

$$\Delta H_m = \Omega X_A X_B \quad (4.9)$$

As discussed above the term ΔH_m expresses the deviation from the ideal behavior ($\Delta H_m = 0$). The sign of ΔH_m depends on the sign of $H_{AB} - \frac{H_{AA} + H_{BB}}{2}$.

Combining (4.3), (4.4) and (4.9), the free energy of mixing becomes

$$\Delta G_m = \Omega X_A X_B + RT (X_A \ln X_A + X_B \ln X_B) \quad (4.10)$$

The sign of ΔG_m depends on the sign of $H_{AB} - \frac{H_{AA} + H_{BB}}{2}$ and temperature. Three cases can be distinguished:

- Case 1: $H_{AB} < \frac{H_{AA} + H_{BB}}{2}$ or $\Omega < 0$. In this case the A – B bonds are more preferable than the A – A or B – B bonds, since the energy of the A – B bond is lower than the mean energy of the A – A and B – B bonds, favoring the formation of a solid solution. The enthalpy of mixing (excess term) is negative. The ΔH_m curve is parabolic due to the $X_A X_B$ term. The ΔH_m , ΔS_m and ΔG_m curves are shown in Figure 4.5a for a certain temperature. The free energy of mixing ΔG_m is negative for all compositions and mixing is therefore thermodynamically feasible. The solid solution exhibits complete solid solubility at this temperature.
- Case 2: $H_{AB} = \frac{H_{AA} + H_{BB}}{2}$ or $\Omega = 0$. There is no special preference for A – A, B – B or A – B bonds. An ideal solid solution forms with $\Delta H_m = 0$. The free energy of mixing is $\Delta G_m = -T\Delta S_m$ and the entropy of mixing makes $\Delta G_m < 0$ for all compositions at the specific temperature (Figure 4.5b). Mixing is thermodynamically feasible and the solid solution exhibits complete solid solubility at the specific temperature. The term $-T\Delta S_m$ is also known as the *ideal or entropic term*.

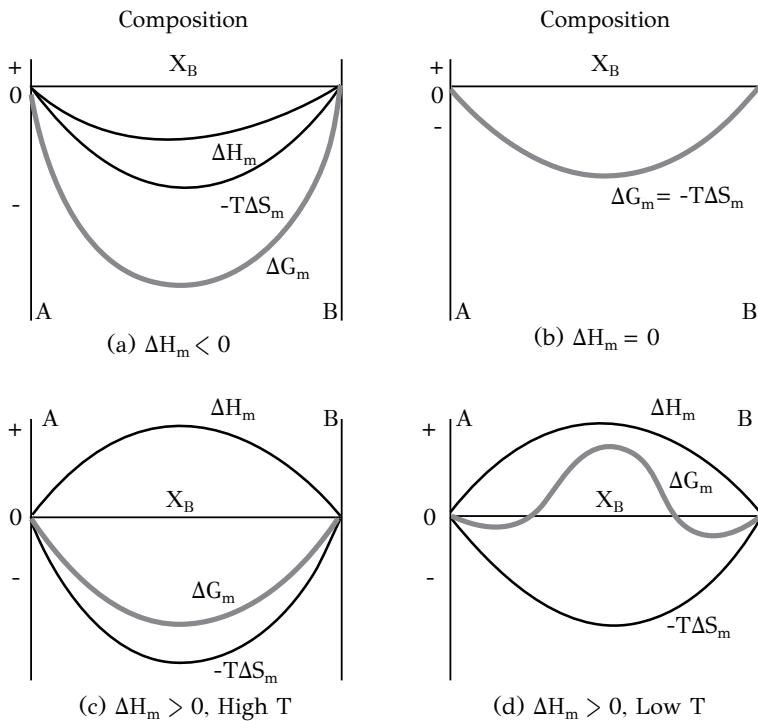


Figure 4.5: Effect of enthalpy of mixing ΔH_m and temperature on the free energy of mixing ΔG_m for the formation of a solid solution.

- Case 3: $H_{AB} > \frac{H_{AA} + H_{BB}}{2}$ or $\Omega > 0$. $A - A$ and $B - B$ bonds are more preferable than the $A - B$ bonds. This means that there is a trend for the solid solution to decompose in two phases, one rich in A and one rich in B , since in this configuration the number of $A - A$ and $B - B$ bonds is maximized. The enthalpy of mixing (excess term) ΔH_m is positive and the sign of ΔG_m depends on temperature due to the entropic term $-T\Delta S_m$. At low temperatures the term ΔH_m dominates the term $-T\Delta S_m$ and ΔG_m becomes positive, leading to decomposition in two phases. As the temperature increases, the term $-T\Delta S_m$ increases and above a specific temperature the term $-T\Delta S_m$ dominates, making ΔG_m negative, favoring the formation of a single solid solution. The curves of ΔH_m , ΔS_m and ΔG_m are shown in Figure 4.5c and d for a high and low temperature respectively.

After the derivation of ΔG_m , the free energy of a solid solution (Relation 4.2) can be written as

$$G = X_A G_A + X_B G_B + \Omega X_A X_B + RT (X_A \ln X_A + X_B \ln X_B) \quad (4.11)$$

where the free energy of the solid solution is expressed as a function of temperature

and composition. Free energy curves are shown in Figure 4.6, before mixing in Figure 4.6a, after mixing and for a high temperature in Figure 4.6b and finally for a low temperature in Figure 4.6c. In the last case the free energy curve exhibits a double

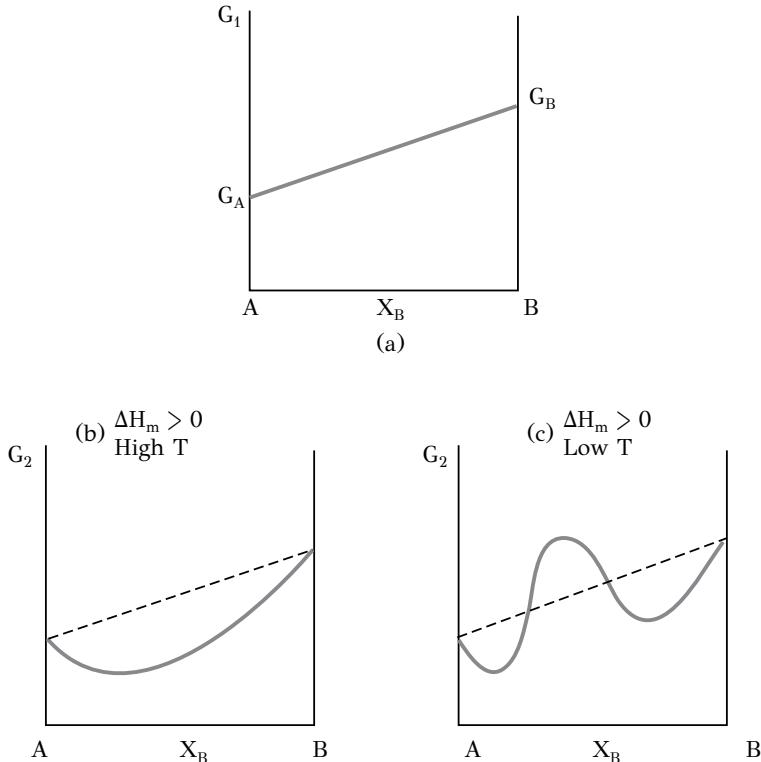


Figure 4.6: Free energy of a solid solution with positive enthalpy of mixing:
(a) before mixing, (b) High T, (c) Low T.

well with two minima, indicating the potential of the system to decompose in two phases. Equation (4.11) is usually written in the form

$$G = G^o + G^{id} + G^{xs} \quad (4.12)$$

where G^o is the free energy of pure components, G^{id} is the ideal or entropic term ($-T\Delta S_m$) and G^{xs} is the excess term (ΔH_m). The regular solution can be generalized for many components as

$$G = \sum_i X_i G_i + RT \sum_i X_i \ln X_i + \sum_i \sum_{j>i} X_i X_j \Omega_{ij} \quad (4.13)$$

In the regular solution model, the interaction coefficients Ω are functions of temperature. However, in general, the interaction coefficients of the excess term are also

a function of composition. Models describing the compositional dependence of the interaction parameters as well as models describing the free energy of interstitial solid solutions and intermetallic compounds will be discussed in Chapter 12.

4.4 CHEMICAL POTENTIAL AND THERMODYNAMIC EQUILIBRIUM

The chemical potential is another significant concept, developed by Gibbs, which plays a key role in the thermodynamics of solutions and in the thermodynamic equilibrium between phases. Apart from equilibrium, the chemical potential gradients express the principal driving forces for diffusion in alloys, as will be discussed in Chapter 5. The chemical potential of a component i is defined by

$$\mu_i = \left(\frac{\partial G}{\partial n_i} \right)_{T,P,n_j \neq n_i} \quad (4.14)$$

and expresses the free energy change, which results from the addition or removal of one mole of component from the solution. In a binary solution the free energy change of a phase is then

$$dG = -SdT + VdP + \mu_A dn_A + \mu_B dn_B$$

which for constant temperature and pressure becomes

$$dG = \mu_A dn_A + \mu_B dn_B \quad (4.15)$$

where μ_A and μ_B are the chemical potentials of components A and B in the phase. Consider the free energy of a phase, e.g., a solid solution, in Figure 4.7.

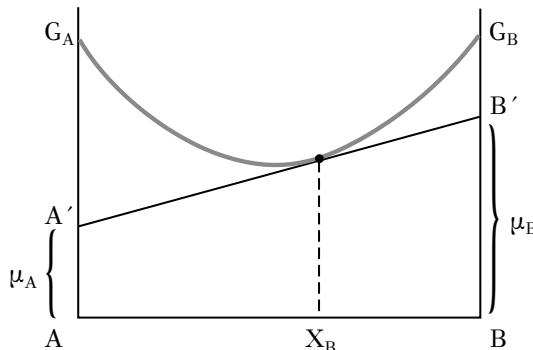


Figure 4.7: The tangent to the free energy curve of a solid solution defines the chemical potentials of the components at a specific composition.

For a solution with composition X_B , the tangent to the free energy curve is drawn. The tangent intersects the vertical axes of the pure components at A' and B' . Segments AA' and BB' correspond to the chemical potentials μ_A and μ_B of the components A and B respectively. At each composition a similar tangent can be drawn and the

chemical potentials can then be determined as a function of composition as $\mu_B = \mu_B^o + RT \ln X_B$. In an ideal solution ($\Delta H_m = 0$), the chemical potential of a component, e.g., component B , is

$$\mu_B = \mu_B^o + RT \ln X_B \quad (4.16)$$

where X_B is the mole fraction of component B and μ_B^o is the chemical potential in the standard state of B . In real solutions, with either $\Delta H_m < 0$ or $\Delta H_m > 0$, the chemical potential is

$$\mu_B = \mu_B^o + RT \ln a_B \quad (4.17)$$

where a_B is the *activity* of component B in the solution. The activity is related to the mole fraction with the relation

$$\alpha_B = \gamma_B X_B \quad (4.18)$$

where γ_B is the *activity coefficient*, which is also a function of composition. The activity coefficient indicates the departure of the solution from ideality. If $\gamma_B = 1$, the solution is ideal. If $\gamma_B < 1$ or $\gamma_B > 1$ the solution exhibits a negative or positive departure from ideality. Taking the standard state to correspond to the free energy of the pure components B , Relation (4.16) for an ideal solution becomes

$$\mu_B = G_B + RT \ln X_B \quad (4.19)$$

while Relation (4.17) for real solutions becomes

$$\mu_B = G_B + RT \ln a_B \quad (4.20)$$

and with the aid of (4.18) it becomes

$$\mu_B = G_B + RT \ln(\gamma_B X_B) \quad (4.21)$$

Now consider a binary alloy $A - B$, which is composed of two phases α and β . It will be shown that when the two phases are at thermodynamic equilibrium at constant temperature and pressure, the chemical potentials of the components at each phase are equal, i.e., $\mu_A^\alpha = \mu_B^\alpha$ and $\mu_A^\beta = \mu_B^\beta$. For this purpose consider that a small quantity of component A , dn_A , is transferred from phase β to phase α , as depicted in Figure 4.8.

The free energy change in the two phases is

$$dG^\alpha = \mu_A^\alpha dn_A^\alpha$$

$$dG^\beta = \mu_A^\beta dn_A^\beta$$

The total free energy change for the system is

$$dG = dG^\alpha + dG^\beta = \mu_A^\alpha dn_A^\alpha + \mu_A^\beta dn_A^\beta$$

The quantity of A removed from phase α , dn_A^α , is equal to the quantity of A added to phase β , dn_A^β , so $dn_A^\alpha = -dn_A^\beta$, then the above equation becomes

$$dG = (\mu_A^\alpha - \mu_A^\beta) dn_A^\alpha \quad (4.22)$$

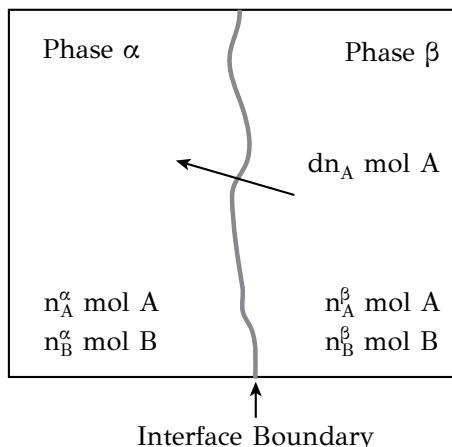


Figure 4.8: Transfer of dn_A mol of component A from the β phase to the α phase in a binary alloy $A - B$.

For thermodynamic equilibrium under constant temperature and pressure, $dG = 0$ and from Equation (4.22) we get

$$\mu_A^\alpha = \mu_A^\beta \quad (4.23)$$

When two phases are in equilibrium, the chemical potentials of all components in the two phases are equal. This is equivalent to the minimization of the free energy in the system. Consider that $dG < 0$. Then the system is not at equilibrium and from Equation (4.22) we get

$$dG = (\mu_A^\alpha - \mu_A^\beta) dn_A^\alpha < 0$$

In this case the quantity dn_A^α has the opposite sign from the quantity $(\mu_A^\alpha - \mu_A^\beta)$ so that $dG < 0$. If the chemical potential of A in phase β is larger than the chemical potential of A in phase α , i.e., $\mu_A^\beta > \mu_A^\alpha$, then A has the tendency to move from phase β to phase α and then $dn_A^\alpha > 0$. The transport of a component from one phase to another under the action of a chemical potential gradient is called diffusion. When enough component A is transferred from phase β to phase α , so that $\mu_A^\alpha = \mu_A^\beta$, then thermodynamic equilibrium is established.

The conditions of thermodynamic equilibrium between two phases in a binary system are shown in Figure 4.9. Consider that the two components have different crystal structures, for instance A is FCC and B is BCC. In the free energy-composition diagram of Figure 4.9, which is drawn for a constant temperature T_o , there are two free energy curves, one for phase α and one for phase β , corresponding to the two different crystal structures. Phase α is a solid solution, rich in A, with an FCC structure. Phase β is a solid solution, rich in B, with a BCC crystal structure. The free energy curves intersect the vertical axes at G_A , G'_A , G_B , and G'_B . G_A and G_B

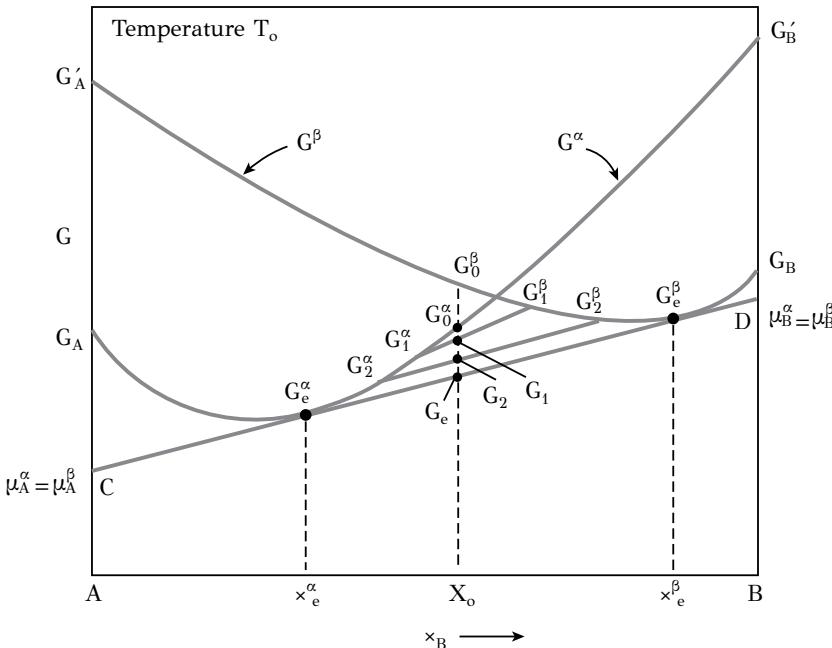


Figure 4.9: Free energy curves of phases α and β of a binary alloy $A - B$ at temperature T_o . The equilibrium between phases is described by the common tangent.

are the free energies of pure components A and B at T_o . G'_A is the free energy that A could have if it exhibited a *BCC* structure. Similarly G'_B is the free energy that B could have if it exhibited the *FCC* structure. It appears, from Figure 4.9, that alloys rich in A have a lower free energy when they form a solid solution of phase α . Similarly alloys rich in B have a lower free energy when they form a solid solution of phase β . The picture is, however, not so clear close to the intersection of the two free energy curves. Consider an alloy with composition X_o . If this alloy could exist entirely as phase β , its free energy would be G_o^β . If the same alloy could exist entirely as phase α , its free energy would be smaller, G_o^α . The free energy of the alloy could be further reduced to the value G_1 if a mixture of phases α and β could form, with free energies G_1^α and G_1^β respectively. The free energy could be further reduced to the value G_2 with the formation of a mixture of phases α and β with free energies G_2^α and G_2^β respectively. The minimum free energy G_e of the alloy is obtained for a mixture of phases α and β with energies G_e^α and G_e^β corresponding to the contact points of the common tangent to the free energy curves of phases α and β . With the formation of this phase mixture thermodynamic equilibrium is established, since at points C and D $\mu_A^\alpha = \mu_B^\beta$ and $\mu_A^\beta = \mu_B^\alpha$ respectively. At equilibrium the alloy with composition X_o at temperature T_o is composed of two phases α and β with compositions X_e^α and

X_e^β . These compositions are constant for any alloy composition between the contact points, $X_e^\alpha < X_o < X_e^\beta$. It should be noted that the free energies G_e^α and G_e^β are not the minimum free energies of phases α and β . What matters is the minimization of the total free energy of the system and the equality of chemical potentials of the components in the two phases. This is accomplished with the common tangent to the two free energy curves. The fractions of phases that coexist at equilibrium, at temperature T_o , can be derived from the *lever rule* as follows

$$f_\alpha = \frac{X_e^\beta - X_0}{X_e^\beta - X_e^\alpha} \quad (4.24)$$

$$f_\beta = \frac{X_0 - X_e^\alpha}{X_e^\beta - X_e^\alpha} \quad (4.25)$$

Summarizing the above discussion, for a certain temperature T_o and depending on the alloy composition X_o , the binary alloy $A - B$ could take the following forms at thermodynamic equilibrium:

- For $0 < X_0 < X_e^\alpha$: single-phase α with fraction $f_\alpha = 1$
- For $X_e^\alpha < X_0 < X_e^\beta$: two-phase mixture $\alpha + \beta$, with fractions f_α and f_β given by the lever rule
- For $X_e^\beta < X_0 < 1$: single phase β with $f_\beta = 1$

The next step is to study the effect of temperature on the equilibrium between the two phases. This will lead to the construction of the phase diagrams. However an important rule will be discussed first, the *Gibbs phase rule*.

4.5 THE GIBBS PHASE RULE

Consider an alloy of composition X_o at a temperature T_o . When two phases α and β are in thermodynamic equilibrium, then their compositions are determined by the contact points of the common tangent to the free energy curves of phases α and β . These compositions, X_e^α and X_e^β , are fixed for the temperature T_o and are independent of the alloy composition X_o . This means that there are limitations on the number of independent variables of the system, when there is thermodynamic equilibrium between phases. These limitations are expressed by the *Gibbs phase rule*. The rule correlates the degrees of freedom F , i.e., the number of variables (e.g., temperature, pressure, alloy composition), which can vary independently without disturbing the equilibrium, with the number of components C and the number of phases P , participating in the equilibrium. The Gibbs phase rule takes the form

$$F = C - P + 2 \quad (4.26)$$

where the number 2 refers to the variables of pressure and temperature. In order to derive the phase rule, consider a system consisting of C components and P phases at

thermodynamic equilibrium. The composition of each phase is determined by $(C - 1)$ terms, since in a quaternary system $A - B - C - D$, as an example, the compositions are related with $X_A + X_B + X_C + X_D = 1$. The compositions of P phases are determined by $P(C - 1)$ terms. Considering in addition the variation of temperature and pressure, the total number of variables is $P(C - 1) + 2$. Now consider the number of variables, which are fixed by the equilibrium. We already know that the chemical potential of each component is the same in all phases. For a binary system with two components A and B and three phases α , β and γ , the chemical potentials are related by

$$\mu_A^\alpha = \mu_A^\beta = \mu_A^\gamma$$

$$\mu_B^\alpha = \mu_B^\beta = \mu_B^\gamma$$

For each component there are two independent equations, which determine the equilibrium between phases. For example, for component A , the equations are $\mu_A^\alpha = \mu_A^\beta$ and $\mu_A^\beta = \mu_A^\gamma$. The third equation

$$\mu_A^\alpha = \mu_A^\gamma$$

is derived from the first two. For each component, therefore, there are $(P - 1)$ independent equations describing the equality of chemical potentials at thermodynamic equilibrium between the phases. Generalizing for a system with C components and P phases, at thermodynamic equilibrium the following relations between chemical potentials hold:

$$\mu_A^\alpha = \mu_A^\beta = \mu_A^\gamma = \dots = \mu_A^p$$

$$\mu_B^\alpha = \mu_B^\beta = \mu_B^\gamma = \dots = \mu_B^p$$

$$\mu_C^\alpha = \mu_C^\beta = \mu_C^\gamma = \dots = \mu_C^p$$

.

.

$$\mu_K^\alpha = \mu_K^\beta = \mu_K^\gamma = \dots = \mu_K^p$$

The number of independent equations is $C(P - 1)$. This means that $C(P - 1)$ variables are determined by the fact that the system is at thermodynamic equilibrium. The degrees of freedom can be derived if the number of the variables fixed by equilibrium is subtracted from the total number of variables

$$F = [P(C - 1) + 2] - [C(P - 1)] = C - P + 2$$

which is the same as (4.26). In metallic systems, variations in pressure cause only negligible effects in the equilibrium between phases, and the phase rule becomes

$$F = C - P + 1 \tag{4.27}$$

Applying the Gibbs phase rule (4.27) for the case of a binary alloy for the equilibrium between two phases, $C = 2$ and $P = 2$, gives $F = 1$. This means that the only

variable, which can vary independently, without disturbing the equilibrium, is the alloy composition. If the alloy composition changes from X_o to X'_o the phase compositions at equilibrium will not change. However if the temperature changes from T_o to T'_o then the system will acquire a new state of thermodynamic equilibrium. The Gibbs phase rule, as expressed by the Relation (4.26), requires that pressure, temperature and alloy composition are the only variables, which affect the equilibrium between phases. The effect of interfacial energy or other fields, e.g., magnetic fields, has been neglected. In the case, however, that such an effect becomes important, the constant in (4.27) should be increased by one unit, for each variable taken into account. The Gibbs phase rule can be applied in order to derive the possible equilibrium conditions in an alloy system. Using Equation (4.27), the equilibrium conditions for a binary system ($C = 2$) appear in Table 4.1

Table 4.1
Equilibrium conditions for a binary system

Components	Phases	Degrees of Freedom	Equilibrium
$C = 2$	$P = 1$	$F = 2$	Bivariant
$C = 2$	$P = 2$	$F = 1$	Univariant
$C = 2$	$P = 3$	$F = 0$	Invariant

Depending on the degrees of freedom the system may exhibit the following equilibrium conditions:

Bivariant equilibrium. The binary system is a single-phase system ($P = 1$) and, therefore, exhibits two degrees of freedom ($F = 2$). The temperature and the composition of the phase (or alloy) can vary independently without disturbing the equilibrium.

Univariant equilibrium. When two phases ($P = 2$) coexist in a binary system, then the system exhibits one degree of freedom ($F = 1$). This means that if one variable is selected, the rest of the variables are automatically fixed. If the temperature is selected, then the compositions of the phases participating in the equilibrium are fixed for that temperature. These compositions are determined by the contact points of the common tangent to the free energy curves, as discussed above.

Invariant equilibrium. When three phases ($P = 3$) coexist in a binary system, then the system exhibits zero degrees of freedom ($F = 0$). This means that both the temperature and the compositions of the phases participating in the equilibrium are fixed. When heat is supplied or removed from the alloy, the temperature and the

composition of the phases remain constant. Only the phase fractions change up to the point where one phase is consumed and the equilibrium becomes univariant, or two of the phases are consumed and the equilibrium becomes bivariant. Both cases correspond to *invariant reactions*, since during the reaction the three phases coexist with $F = 0$ at a constant temperature and constant phase compositions. The invariant reactions encountered in binary systems are depicted in Table 4.2.

Table 4.2
Invariant reactions in binary alloy systems

Reaction	Name	Example
$L = \alpha + \beta$	Eutectic reaction	$Ag - Cu$ ($780^\circ C$)
$\gamma = \alpha + \beta$	Eutectoid reaction	$Fe - C$ ($723^\circ C$)
$L_1 = \alpha + L_2$	Monotectic reaction	$Cu - Pb$ ($995^\circ C$)
$\alpha = \beta + L$	Metatetic reaction	$Ag - Li$ ($317^\circ C$)
$L = \alpha + \beta$	Peritectic reaction	$Fe - Ni$ ($1512^\circ C$)
$\alpha + \beta = \gamma$	Peritectoid reaction	$Al - Cu$ ($686^\circ C$)
$L_1 + L_2 = \alpha$	Syntectic reaction	$K - Zn$ ($592^\circ C$)

Liquid phases are denoted by L , L_1 and L_2 , while solid phases are denoted by α , β and γ . The *invariant reactions*, which include at least one liquid phase, end up with the syllable *-tectic*, while those concerning only solid phases end up with the syllable *-tectoid* in their name.

4.6 EQUILIBRIUM PHASE DIAGRAMS IN BINARY SYSTEMS

The common tangent construction, discussed in section 4.4 and Figure 4.9, indicated that at a certain temperature the compositions of the two phases coexisting at equilibrium are given by the contact points to the $G - X$ curves. If a series of such $G - X$ diagrams, corresponding to a series of temperatures, is considered, then the phase compositions at equilibrium can be determined as a function of temperature. This set of points or the locus of contact points of the common tangent to the $G - X$ curves of the two phases constitute the phase boundary curves in the equilibrium phase diagram. In other words the binary phase diagram is constructed from lines of univariant equilibria and cross points of invariant equilibria. The *phase diagram* is then a map, in the temperature-composition space, which provides the alloy constitution, i.e., the phases present and the composition of these phases at thermodynamic equilibrium. The common tangent construction will be used in the following sections to construct

various types of phase diagrams in binary alloy systems.

4.6.1 BINARY SYSTEM WITH COMPLETE SOLID SOLUBILITY

The simplest case of phase diagram arises when metals A and B exhibit complete solid solubility, i.e., all Hume-Rothery rules, discussed in Chapter 2, are satisfied. In such a system there is only one solid phase (*S*) and one liquid phase (*L*). The free energy of the pure components A and B varies with temperature, as depicted in Figure 4.1d. The melting points of the pure components correspond to the condition $G_A^L = G_A^S$ at temperatures $T_m(A)$ and $T_m(B)$. The variation of free energy of the liquid (*L*) and the solid (*S*) with temperature is depicted in Figure 4.10.

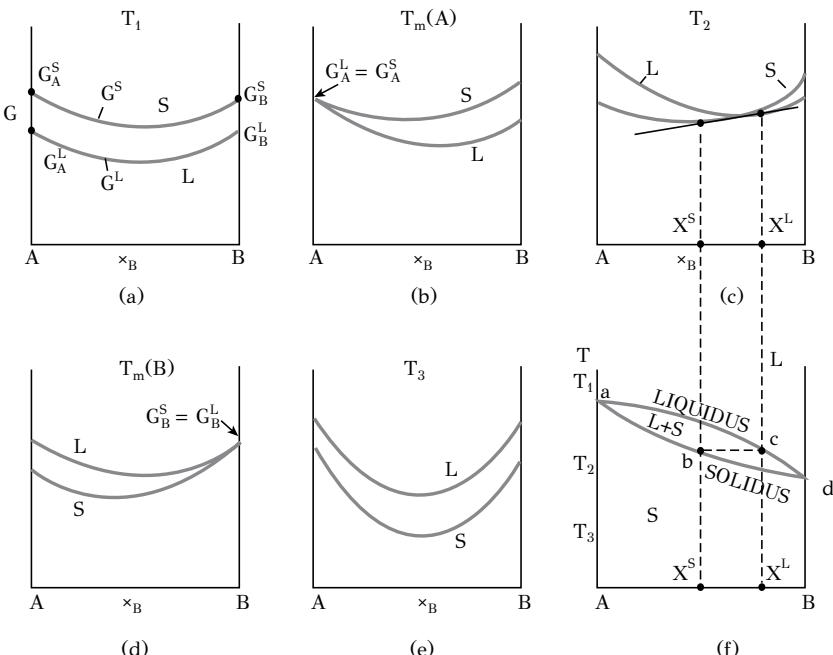


Figure 4.10: Construction of a phase diagram in a binary system exhibiting complete solid solubility, from the respective free energy curves of the solid and liquid phases.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). Phase Transformations in Metals and Alloys, with permission from CRC press.

Starting at a high temperature T_1 , with $T_1 > T_m(A) > T_m(B)$, the liquid exhibits a lower free energy than the solid for all alloy compositions. The liquid is then the stable phase at T_1 for all alloy compositions, from pure A ($X = 0$) to pure B ($X = 1$), as depicted in Figure 4.10a. A decrease in temperature has two effects. The first is that the free energies G_A^L and G_B^L increase faster than G_A^S and G_B^S . The second is that the curvature of the $G - X$ curves decreases due to the smaller contribution of the $-T\Delta S_m$ term to the free energy. At $T_m(A)$ the condition $G_A^S = G_A^L$ is satisfied (Figure

4.10b). This condition corresponds to point a in the phase diagram of Figure 4.10f. At a lower temperature T_2 , the free energy curves of the solid and liquid phase cross (Figure 4.10c) and the common tangent defines the compositions X^S and X^L . Alloys with composition X_B between 0 and X^S are solid at equilibrium, between X^L and 1 are liquid, while between X^S and X^L the alloy consists of two phases, solid and liquid, with compositions X^S and X^L respectively. These compositions appear as points b and c in the phase diagram (Figure 4.10f). Between T_2 and $T_m(B)$ the free energy of the liquid G^L continues to increase faster than G^S while the locus of the contact points X^S and X^L correspond to the *solidus* and *liquidus* curves in the phase diagram. At $T_m(B)$ the condition $G_B^S = G_B^L$ is satisfied. Below the melting point of B, the free energy of the solid G^S is lower than the free energy of the liquid G^L for all alloy compositions and all alloys consist of a single solid phase (solid solution) and the equilibrium is bivariant. Above the liquidus the alloy is liquid and the equilibrium is again bivariant. Between the solidus and the liquidus curves the alloy consists of a mixture of two phases, solid and liquid, and the equilibrium is univariant. Below the solidus curve, the solid phase exhibits complete solid solubility. This condition corresponds to a negative enthalpy of mixing ($\Delta H_m < 0$), which leads to $\Delta G_m < 0$ for all alloy compositions.

4.6.2 BINARY SYSTEM WITH A MISCELLIBILITY GAP

Consider now a binary alloy system $A - B$ where the solid phase α is a solid solution with a positive enthalpy of mixing ($\Delta H_m > 0$). The components A and B have the tendency to form two separate solid phases. The free energy curves of the liquid and the solid phase are shown for different temperatures in Figure 4.11.

At a high temperature T_1 the liquid exhibits a lower free energy than the solid for all alloy compositions. At temperature T_2 the free energy curves cross at two points. The common tangent construction defines the phase compositions at equilibrium. For alloy compositions X_B between 0 and X_2^α as well as between $(X_2^\alpha)'$ and 1, the alloy is stable as a single α phase¹. Between X_2^L and $(X_2^L)'$ the alloy is stable as single-phase liquid. Between X_2^α and X_2^L as well as between $(X_2^L)'$ and $(X_2^\alpha)'$ the alloy is stable as a mixture of solid and liquid phases. The points X_2^α , X_2^L , $(X_2^L)'$ and $(X_2^\alpha)'$ are depicted also in the phase diagram of Figure 4.11d. At both temperatures T_1 and T_2 , the free energy curve of the solid exhibits a positive curvature because the term $-T\Delta S_m$ dominates at high temperatures. However at a lower temperature T_3 , the term ΔH_m dominates the term $-T\Delta S_m$ and the free energy curve for the solid obtains a negative curvature for a range of alloy compositions. As depicted in Figure 4.11c the solid phase has the tendency to separate in two solid phases α_1 and α_2 between the alloy compositions $X_3^{\alpha_1}$ and $X_3^{\alpha_2}$. The region $\alpha_1 + \alpha_2$ of the phase diagram is termed the *miscibility gap*. Outside the miscibility gap the alloy is stable as a homogeneous solid of single α phase. However if an alloy is cooled inside the miscibility gap, phase α will decompose in two solid phases, α_1 and α_2 , with the same crystal structure but

¹The subscript in the compositions, e.g. X_1 , corresponds to the temperature, e.g., T_1 .

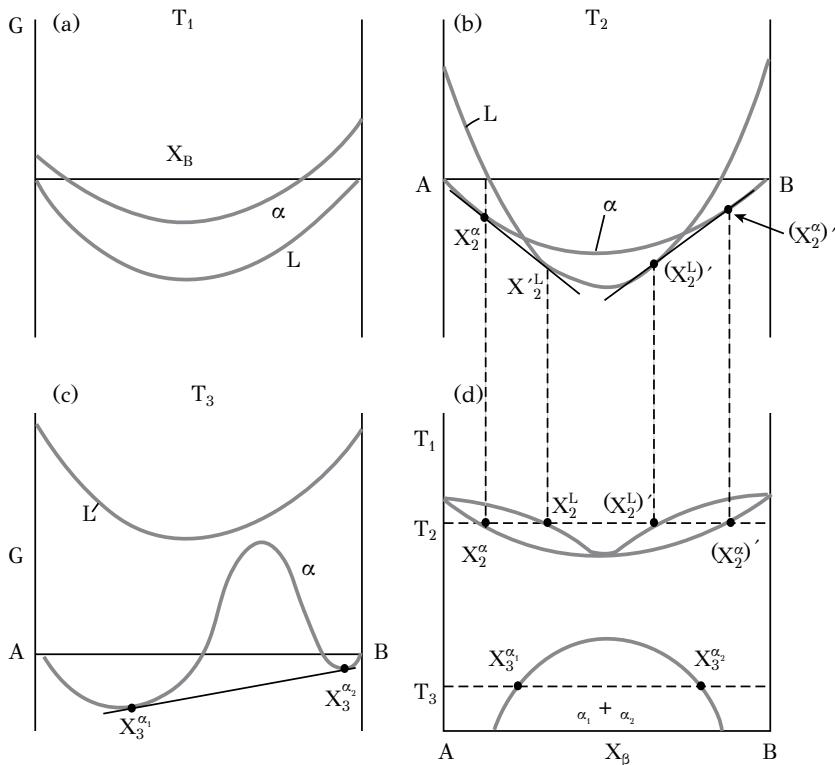


Figure 4.11: Free energy curves (a,b,c) and (d) binary phase diagram with a miscibility gap.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). *Phase Transformations in Metals and Alloys*, with permission from CRC press.

different composition. This transformation is termed *spinodal decomposition* and will be discussed in detail in Chapter 6.

4.6.3 BINARY SYSTEM WITH EUTECTIC POINT

In cases where the enthalpy of mixing ΔH_m takes high positive values, the miscibility gap can extend inside the region of the liquid phase. In this case a phase diagram with a eutectic point forms as shown in Figure 4.12e.

Consider two components A and B with different crystal structures. In this system one liquid phase L and two solid phases α and β appear. The free energy curves of these phases at temperatures T_1 , T_2 , T_3 and T_4 are depicted in Figure 4.12a,b,c and d respectively. The common tangent construction defines the phase compositions at equilibrium for temperatures T_1 and T_2 . At temperature T_3 the three curves L, α and β have a common tangent. This temperature is termed the *eutectic temperature* T_E .

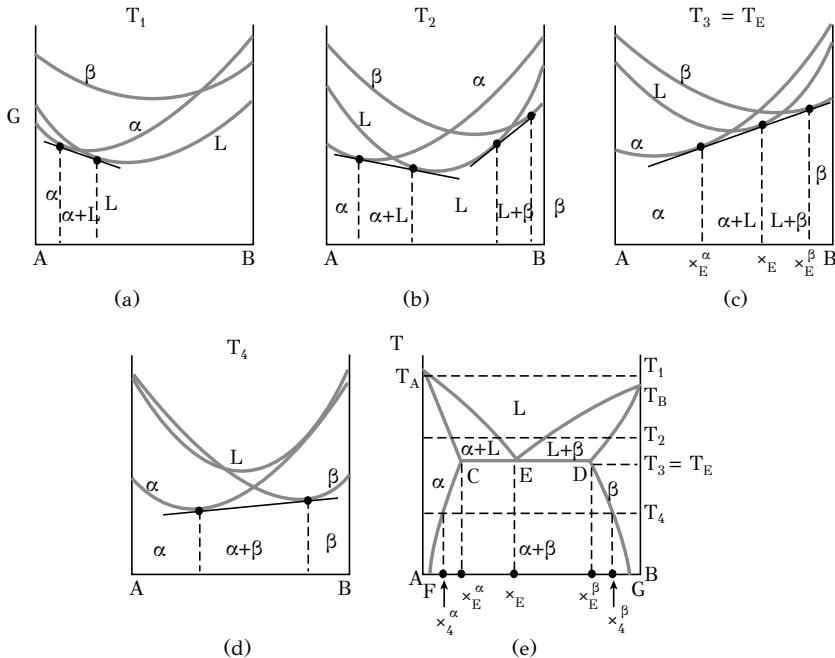


Figure 4.12: Construction of a eutectic phase diagram from the respective free energy curves.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). Phase Transformations in Metals and Alloys, with permission from CRC press.

Application of the Gibbs phase rule for $P = 3$ provides $F = C - P + 1 = 0$ and the equilibrium is invariant. The three phases coexist and at the eutectic temperature the invariant reaction, called the *eutectic reaction* takes place

$$L(X_E) = \alpha(X_E^\alpha) + \beta(X_E^\beta) \quad (4.28)$$

where the compositions of phases are included in parentheses. During the eutectic reaction the liquid with composition X_E is consumed to form a mixture of phase α and β with compositions X_E^α and X_E^β respectively. Below the eutectic temperature the liquid disappears and a univariant equilibrium between phases α and β is established. The *liquidus* and *solidus lines* for the α phase are $T_A E$ and $T_A C$ respectively. For the β phase the respective lines are $T_B E$ and $T_B D$. For the binary system the liquidus line is $T_A E T_B$ and the solidus line is $T_A C E D T_B$, which includes the horizontal eutectic line CED . The lines CF and DG are the limits of solid solubility or *solvus lines*. The solvus CF corresponds to the limit of solid solubility of component B in the solid solution of the α phase as a function of temperature. The solvus DG corresponds to the limit of solid solubility of component A in the solid solution of the β phase as a function of temperature. For example, at temperature T_4 , the solubility limit of B in

the α phase is X_4^α .

The equilibrium solidification of a binary *eutectic alloy*, i.e., one with the eutectic composition X_E , can be studied with the aid of the phase diagram of Figure 4.13a.

During cooling from the liquid phase, the alloy starts to solidify at the eutectic temperature. It should be noted that the term eutectic stems from the greek “e‘uthktiko”, which means “easy melting”, since the eutectic temperature is well below the melting points of both components A and B. At the eutectic temperature, the eutectic reaction (4.28) takes place. According to the reaction, the liquid is consumed and a mixture of two solid phases α and β forms. In most cases the mixture is formed in a lamellar arrangement of phases α and β , depicted in Figure 4.13b, called eutectic structure. The eutectic reaction continues at the eutectic temperature T_E up

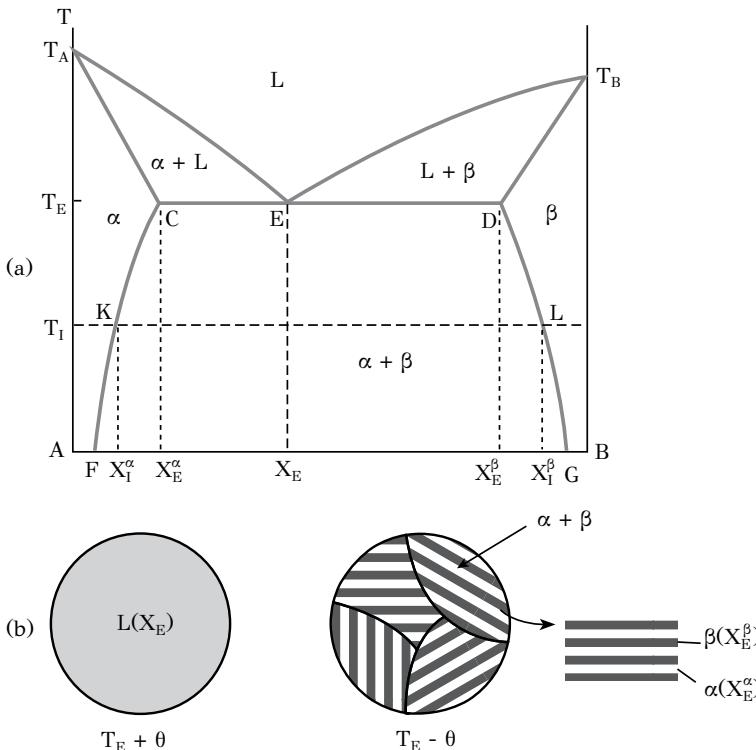


Figure 4.13: (a) Binary phase diagram with eutectic and (b) the development of the eutectic lamellar structure in the eutectic mixture ($\alpha + \beta$).

to the point of consumption of all liquid. During the eutectic reaction three phases L , α and β coexist in an invariant equilibrium. The relative fractions of phases α and β

can be derived from the lever rule at the eutectic temperature T_E :

$$f_\alpha = \frac{X_E^\beta - X_E}{X_E^\beta - X_E^\alpha}$$

$$f_\beta = \frac{X_E - X_E^\alpha}{X_E^\beta - X_E^\alpha}$$

Below the eutectic temperature the phases α and β are in univariant equilibrium and their composition follows the solvus lines CF and DG . The phase compositions can be determined from the phase diagram at any temperature by drawing the tie line at that temperature, connecting the two solvus lines. For example for temperature T_1 the tie line is KL and the compositions of the α and β phases are X_1^α and X_1^β respectively. It is also possible to determine the fractions of the α and β phases at T_1 using the lever rule:

$$f_\alpha = \frac{X_1^\beta - X_E}{X_1^\beta - X_1^\alpha}$$

$$f_\beta = \frac{X_E - X_1^\alpha}{X_1^\beta - X_1^\alpha}$$

Now consider the solidification of *hypoeutectic alloys*, i.e., alloys with composition lower than the eutectic composition X_E . Three alloys L , M and N will be considered (Figure 4.14).

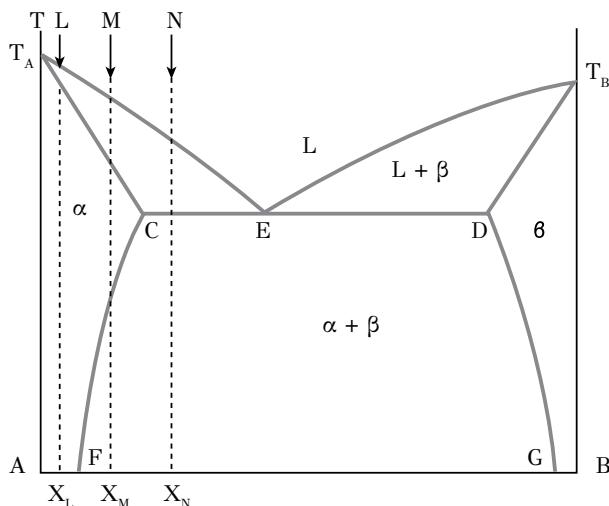


Figure 4.14: Binary phase diagram with three hypoeutectic alloys L , M and N .

The solidification of the alloy L , with composition X_L , starts at the liquidus temperature T_1 (Figure 4.15). The solid forms as α phase with composition X_1^α . Thus the

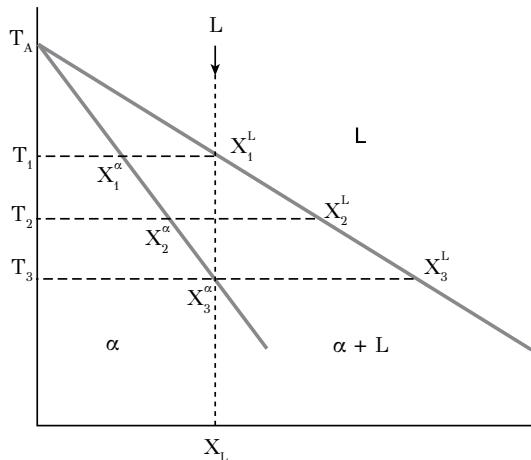


Figure 4.15: Magnification of upper left corner of Figure 4.14 with the cooling of the hypoeutectic alloy L .

solid phase α forms with a leaner composition, meaning that component B is rejected into the liquid from the growing solid. The liquid is then enriched in B and its composition rises following the liquidus line. At a lower temperature T_2 , the composition of the α phase is X_2^α and the liquid phase is X_2^L . At T_2 the α phase is in univariant equilibrium with the liquid phase and the relative phase fractions can be determined by the lever rule:

$$f_\alpha = \frac{X_2^L - X_L}{X_2^L - X_2^\alpha}$$

$$f_L = \frac{X_L - X_2^L}{X_2^L - X_2^\alpha}$$

As the temperature drops further, the fraction of the solid phase increases. At T_3 the liquid phase disappears and the solidification is complete. The alloy is now solid, as α phase with composition equal to the alloy composition X_L . During cooling from T_1 to T_3 the composition of the solid varied from X_1^α to $X_3^\alpha = X_L$ and that of the liquid from $X_1^L = X_L$ to X_3^L in order to maintain the equilibrium between the phases. The change of the composition of the phases is performed by atomic transport termed *diffusion*, a matter to be discussed in Chapter 5. During cooling from T_3 down to room temperature, the alloy remains single α phase with composition X_L .

For alloy M (Figure 4.16a), with composition X_M , the solidification from temperature T_1 up to T_3 takes place in exactly the same way as for alloy L , discussed above. At temperature T_3 the solidification is complete and the alloy exists as single-phase α with composition X_M . At lower temperatures, below the solvus line, a solid-state reaction takes place, the *precipitation* of β phase from the α phase solid solution. At temperature T_4 the β phase forms with composition X_4^β and the β phase is in

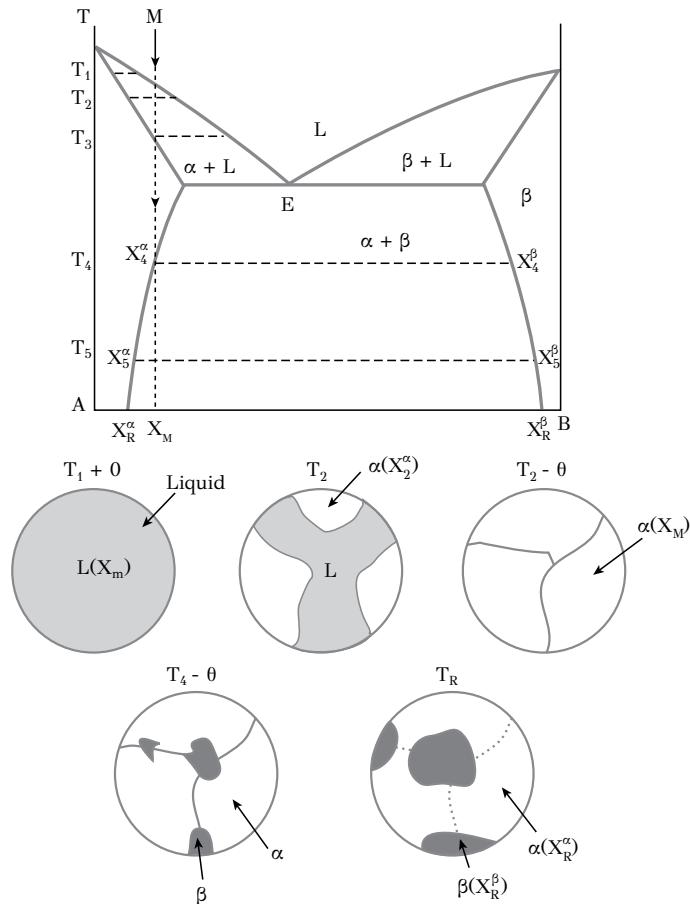


Figure 4.16: Binary eutectic phase diagram with the solidification path of alloy M, depicting changes in the structure of the alloy during cooling.

equilibrium with the α phase with composition X_4^α . At a lower temperature T_5 , as the solubility of B in the α phase decreases, the fraction of the β phase increases. The fractions of the α and β phases are determined from the lever rule. At room temperature T_R the alloy with composition X_M consists of two solid phases with compositions X_R^α and X_R^β respectively. During cooling from T_4 to T_R , the α phase rejects component B to the β phase and its composition drops from X_4^α to X_R^α . On the other hand the β phase is enriched in B and its composition increases from X_4^β to X_R^β . This atomic exchange is termed *partitioning* of component B and is accomplished by atomic diffusion. The structure of alloy M is depicted schematically in Figure 4.16b for various temperatures. The precipitation of the β phase is accomplished by the processes of nucleation and growth, to be discussed in detail in Chapter 6.

The solidification of alloy N (Figure 4.17a), with composition X , starts at temperature T_1 with the formation of the α phase with composition X_1^α . During cooling

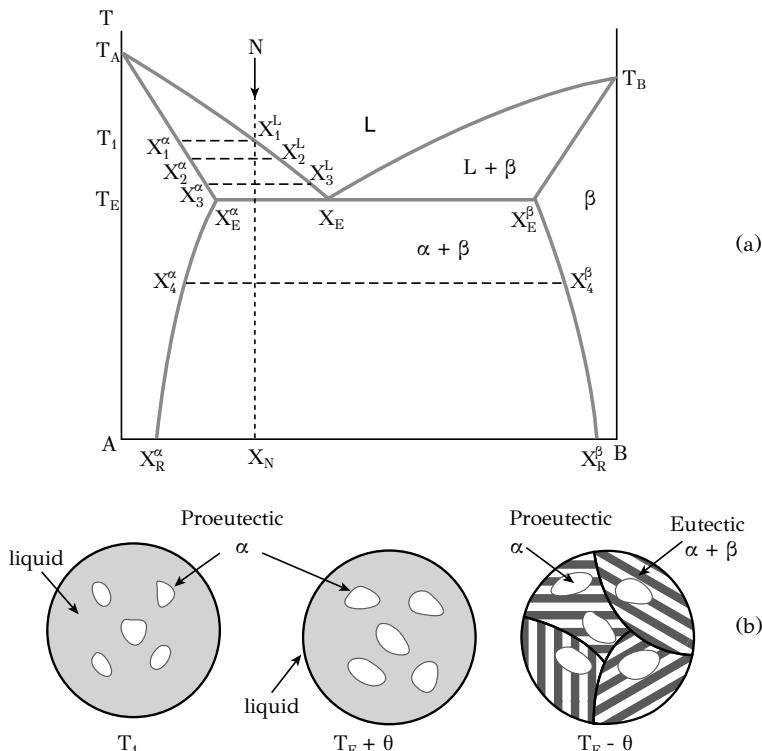


Figure 4.17: Binary eutectic phase diagram with the solidification path of alloy N , depicting changes in the structure of the alloy during cooling.

from T_1 to T_3 , the liquid is enriched in B and its composition rises from X_1^L to X_3^L . At the eutectic temperature T_E the α phase is in equilibrium with the liquid, which has obtained the eutectic composition. Solidification of the remaining liquid continues with the eutectic reaction, during which the liquid is consumed and two solid phases α and β form, with compositions X_E^α and X_E^β respectively. The alloy remains at the eutectic temperature till all liquid is consumed. The solidification of alloy N takes place in two stages. The first stage is from temperature T_1 to temperature T_E . In this stage the alloy solidifies with the formation of α phase, called *proeutectic* α phase, since it forms before the eutectic reaction. At the same time the liquid is enriched in B and reaches the eutectic composition at the eutectic temperature. In the second stage the alloy solidifies by the eutectic reaction, where the remaining liquid is consumed and a mixture of phases α and β form. Cooling below the eutectic changes the relative fractions of the α and β phases according to the lever rule. The α phase should be distinguished between proeutectic α formed between T_1 and T_E

and eutectic α formed by the eutectic reaction. The structure of the alloy is depicted schematically for temperatures T_1 , $T_E - \theta$ and $T_E + \theta$ in Figure 4.17b with the formation of proeutectic α phase and the eutectic mixture with the lamellar structure of α and β phases.

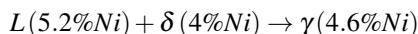
Hypereutectic alloys, with composition higher than the eutectic composition, solidify in a manner similar to that of hypoeutectic alloys, except that the first phase forming during solidification is the β phase (proeutectic β). Summarizing the discussion on the equilibrium solidification of binary alloys involving a eutectic reaction (Figure 4.14), the following remarks can be made:

- Alloys with composition between 0 and X_F solidify forming single α phase.
- Alloys with composition between X_G and 1 solidify forming a single β phase.
- Alloys with composition between X_F and X_C solidify forming α phase while at lower temperature β phase forms by precipitation from the α phase.
- Alloys with composition between X_D and X_G solidify forming β phase while at lower temperature α phase forms by precipitation from the β phase.
- Alloys with composition between X_C and X_E solidify with the formation of proeutectic α phase and the formation of a eutectic mixture $\alpha + \beta$.
- Alloys with composition between X_E and X_D solidify with the formation of proeutectic β phase and the formation of a eutectic mixture $\alpha + \beta$.
- Alloys with composition X_E solidify forming only a eutectic mixture $\alpha + \beta$.
- The eutectic reaction takes place in alloys with composition between X_C and X_D .

Several alloys form eutectic systems. A few examples are $Ag - Cu$, $Pb - Sn$, $Pb - Sb$, $Al - Si$ and $Cd - Zn$.

4.6.4 BINARY SYSTEM WITH PERITECTIC POINT

A peritectic point is present in the upper left corner of the binary $Fe - Ni$ phase diagram (Figure 4.18a). The phase diagram is characterized by a large field of γ phase with *FCC* structure, while the *BCC* structure appears as the small fields of the α and δ phases at low and high temperatures respectively. The peritectic point is shown in magnification in Figure 4.18b. The peritectic composition is 4.6%Ni (wt%) while the peritectic temperature is $1518^{\circ}C$. Alloys with composition lower than 4%Ni form δ phase, while alloys with composition greater than 5.2%Ni form γ phase on solidification. Alloys with composition between 4 and 5.2%Ni involve the peritectic reaction during solidification. Now consider the solidification of the peritectic alloy with composition 4.6%Ni along the line *abcd*. The solidification starts at point *b* with the formation of δ phase, which is lean in Ni. From point *b* to point *c* more δ phase forms while the liquid is enriched in Ni. At point *c* the liquid composition is 5.2%Ni, while the composition of the δ phase is 4%. The fractions of the δ phase and the liquid are 61 and 39% respectively. At point *c* ($1518^{\circ}C$) the *peritectic reaction* takes place:



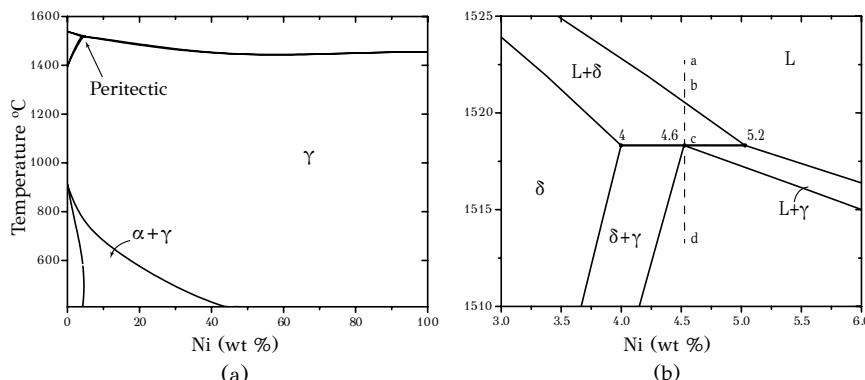


Figure 4.18: (a) The binary $Fe - Ni$ phase diagram, (b) the region with the peritectic point. (Calculated with Thermo-Calc)

The above reaction results in the formation of homogeneous γ phase with composition 4.6%Ni. Alloys with composition between 4 and 4.6%Ni (to the left of point c) at $1518^{\circ}C$ contain more than 61% δ phase and, consequently, after the peritectic reaction, enter the $\delta + \gamma$ region of the phase diagram. On the other hand alloys with composition between 4.6 and 5.2%Ni (to the right of point c) contain more than 39% liquid and after the peritectic reaction enter the $\gamma + L$ region. The solidification of these alloys is not completed at the peritectic temperature with the peritectic reaction but at a lower temperature with the solidification of the remaining liquid and the formation of γ phase.

4.7 EXAMPLES OF PHASE DIAGRAMS

The phase diagrams of two important binary systems will be discussed, the $Cu - Zn$ and the $Fe - C$ system.

4.7.1 THE BINARY SYSTEM CU-ZN

Brass is an important copper alloy based on the $Cu - Zn$ system. The binary phase diagram is shown in Figure 4.19.

The solid solutions between Cu and Zn exhibit limited solid solubility since the two metals have different crystal structures. Cu is *FCC* while Zn is *HCP*. The diagram contains seven solid phases: α (*FCC*), β (*disordered BCC*), β' (*ordered BCC*), γ (*cubic*), δ (*BCC*), ε (*HCP*) and η (*HCP*). The phases α and η are the *terminal solid solutions* $Cu - Zn$ and $Zn - Cu$ respectively while the remaining phases are *intermediate phases*. The diagram contains five peritectic points. The respective invariant reactions are:



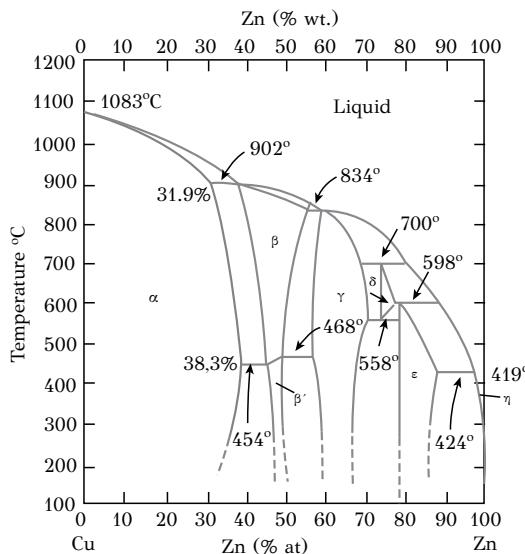
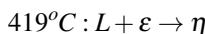
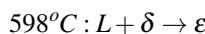
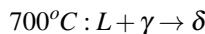
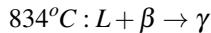
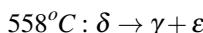


Figure 4.19: (a) The binary $Cu - Zn$ phase diagram.



The diagram contains also a eutectoid reaction:



The β phase (BCC) on cooling below $454 - 468^{\circ}C$ forms the β' phase, which is an ordered BCC phase. The unit cell of the β' phase is a cube where every corner is occupied by a copper atom while the cube center is occupied by a zinc atom, corresponding to a stoichiometry of $CuZn$. The ε phase has HCP structure with $c/a = 1.58$, smaller than the ideal c/a ratio, 1.63, while the η phase is also HCP with $c/a = 1.86$, larger than the ideal c/a ratio. Another phase, the ζ phase, does not appear in the phase diagram, since it is a non-equilibrium phase. It has an HCP structure with a c/a ratio close to the ideal c/a ratio. The ζ phase forms by strain-induced martensitic transformation at low temperatures.

4.7.2 THE BINARY SYSTEM Fe-C

The $Fe - C$ phase diagram, shown in Figure 4.20, is probably the most important phase diagram, since steel technology is based on the $Fe - C$ system.

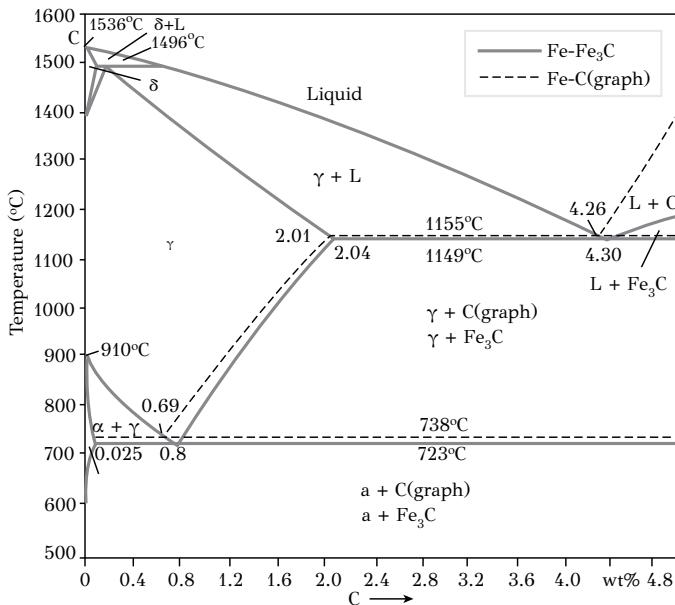


Figure 4.20: The binary Fe-C phase diagram. Full lines: equilibrium with cementite, dotted lines: equilibrium with graphite.

Dotted lines correspond to the equilibrium of iron with graphite, which is the stable form of carbon, while the full lines correspond to the equilibrium of iron with cementite (Fe_3C), which is a metastable form of carbon, forming instead of graphite. The difference between the two diagrams is small, except in the region of the eutectic where the difference is appreciable. The discussion here is limited to the metastable $Fe - C$ phase diagram involving the cementite phase. The stable phases appearing in the diagram are the following:

- α , ferrite (BCC)
- γ , austenite (FCC)
- δ , ferrite (BCC)
- Fe_3C , cementite (*orthorhombic*)

The phase of δ ferrite is the continuation of α ferrite at high temperatures. Another phase, martensite, does not appear in the phase diagram since it is a non-equilibrium phase, forming from the martensitic transformation of austenite. As has been discussed in Chapter 2, iron exhibits the phenomenon of allotropy. A simplified explanation of the allotropy of iron can be given considering the magnetic transitions of iron in Figure 4.21. Below the Curie temperature $T_c = 768^\circ C$, $\alpha - Fe$ is ferromagnetic, while it is paramagnetic above T_c . Then the stability of $\alpha - Fe$ at low temperatures can be explained by the lower internal energy of the ferromagnetic structure, characterized by aligned spins. On the other hand $\gamma - Fe$ is paramagnetic all the way above $-193^\circ C$. This affects the free energy of $\gamma - Fe$ at high temperatures, due to

the $-TS$ term, and the $\alpha \rightarrow \gamma$ transformation takes place at $910^\circ C$. The same effect ($-TS$ term) affects the free energy of $\alpha - Fe$ at higher temperatures and the iron returns to the *BCC* state ($\delta - Fe$) at a higher temperature, $1400^\circ C$. It maintains this structure up to the melting point of $1539^\circ C$.

The *Fe – C* diagram contains three invariant reactions:

<i>Peritectic</i>	$1496^\circ C : L + \delta \rightarrow \gamma$
<i>Eutectic</i>	$1149^\circ C : L \rightarrow \gamma + Fe_3C$
<i>Eutectoid</i>	$723^\circ C : \gamma \rightarrow \alpha + Fe_3C$

The phase mixture $\gamma + Fe_3C$, forming from the eutectic reaction, is called *ledeburite*. The phase mixture $\alpha + Fe_3C$, forming from the eutectoid reaction, is called *pearlite*. Carbon occupies interstitial sites in the crystal structure of iron. The maximum solid

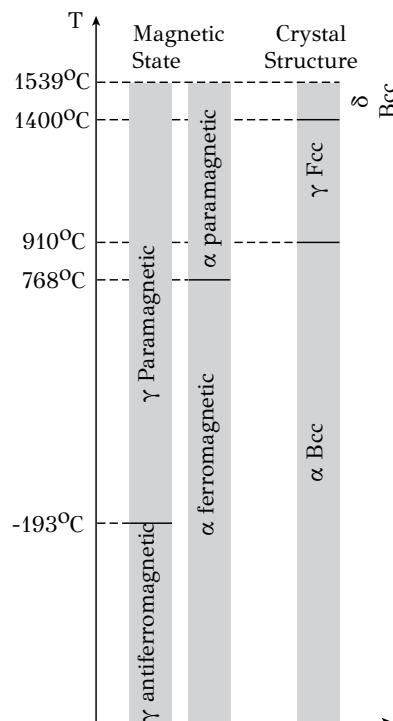


Figure 4.21: Correlation between the magnetic states and crystal structure of iron.

solubility in austenite is about $2\text{wt}\%$ while it is only $0.025\text{wt}\%$ in ferrite, a difference that has been discussed in Chapter 2. Besides carbon, other alloying elements occupy substitutional sites. Alloying elements can be distinguished in (a) elements which extend the γ field in the phase diagram (austenite stabilizers), such as *Ni* and *Mn* and (b) elements that restrict the γ field and form carbides or other intermetallic compounds, such as *Al*, *Si*, *Cr*, *Mo*, *V*, *Nb* and *Ti*. The *Fe – Ni* system is depicted

in Figure 4.22, while parts of the $Fe - Mo$ and $Fe - Cr$ systems are shown in Figure 4.23. The $Fe - C$ system will be presented in more detail in Chapter 10, in the

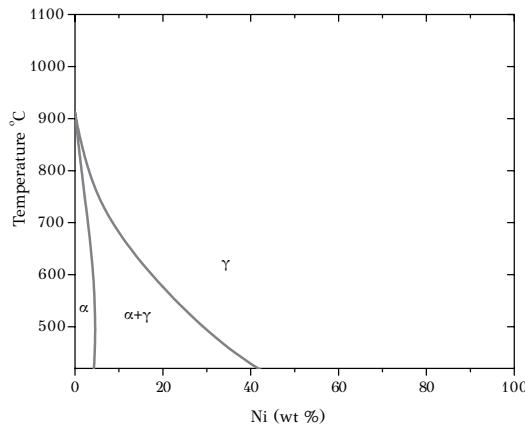


Figure 4.22: The $Fe - Ni$ binary phase diagram. (Calculated with Thermo-Calc)

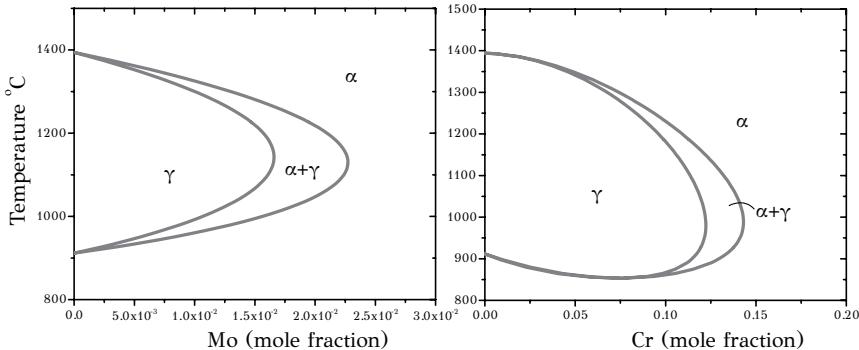


Figure 4.23: Sections of the $Fe - Mo$ and $Fe - Cr$ phase diagrams. (Calculated with Thermo-Calc)

discussion of phase transformations in steels.

4.8 CASE STUDY: SOLDER ALLOYS – THE Pb-Sn PHASE DIAGRAM

Soldering is a joining method where two metals are joined together by melting a filler metal, which has a lower melting point than the adjoining metals. In soldering the melting temperature is lower than $450^{\circ}C$. The most important characteristic of filler metals for soldering is their melting point (or eutectic temperature), if it is a eutectic alloy or their melting range, if it is a non-eutectic alloy. Other important characteristics are the mechanical and physical properties. In this case study, the most common

solder alloy based on the $Pb - Sn$ system will be discussed. The $Pb - Sn$ binary system has wide applications in soldering because of the low eutectic temperature of just $183^\circ C$. The Pb-Sn phase diagram is shown in Figure 4.24. The diagram has

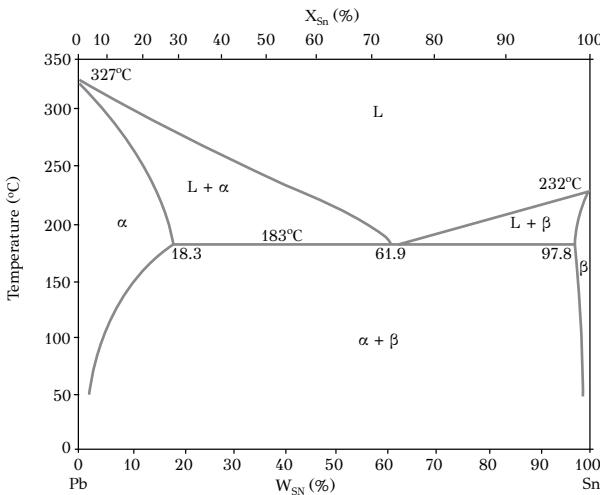
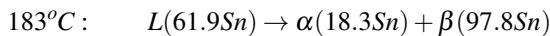


Figure 4.24: The $Pb - Sn$ phase diagram.

two composition axes. In the upper axis the composition is displayed as the mole or atomic fraction $X_{Sn}(\%)$ while in the lower axis the composition is displayed in weight $W_{Sn}(\%)$. In the following, the $W_{Sn}(\%)$ composition will be used. There are three phases in the diagram: Liquid (L), the α phase and the β phase. The α phase is the solid solution $Pb - Sn$, rich in Pb , while the β phase is the $Sn - Pb$ solid solution, rich in Sn . Both phases are substitutional solid solutions. There are three two-phase regions in the diagram: $L + \alpha$, $L + \beta$, and $\alpha + \beta$. There is also a eutectic point with a eutectic temperature $183^\circ C$ and eutectic composition $61.9Sn$. The eutectic reaction is:



where the phase compositions are included in parentheses. The melting points of the pure components are $327^\circ C$ for Pb and $232^\circ C$ for Sn . The solubility of Sn in Pb is higher than the solubility of Pb in Sn . More specifically the maximum solid solubility of Sn in Pb is 18.3% while the maximum solubility of Pb in Sn is just 2.2% at $183^\circ C$. Consider now a $Pb - 30Sn$ alloy at $250^\circ C$ in Figure 4.25. The alloy consists of a mixture of $\alpha + L$ phases. At the coordinates $250^\circ C, 30Sn$ of the phase diagram, a tie line is drawn, which is a horizontal line crossing the solidus and the liquidus lines at the points corresponding to the compositions $13Sn$ and $34Sn$ respectively, as depicted in Figure 4.25. The composition of the α phase is $13Sn$ while the composition of the

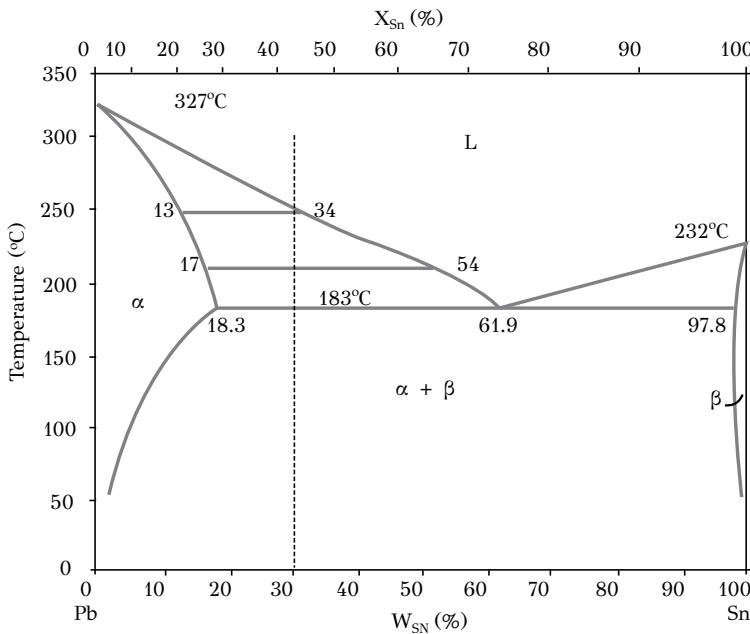


Figure 4.25: Tie lines at several temperatures for the alloy Pb-30Sn.

liquid phase is 34Sn. The fractions of the two phases are given by the lever rule as:

$$f_{\alpha} = \frac{34 - 30}{34 - 13} = \frac{4}{21} = 0.19$$

$$f_L = \frac{30 - 13}{34 - 13} = \frac{17}{21} = 0.81$$

In other words the alloy consists of 19% α phase and 81% Liquid phase. As the temperature drops, it is expected that the liquid will solidify and more α phase will form. If the above calculation is repeated for the temperature of 210°C the phase fractions are:

$$f_{\alpha} = \frac{54 - 30}{54 - 17} = \frac{24}{37} = 0.648$$

$$f_L = \frac{30 - 17}{54 - 17} = \frac{13}{37} = 0.352$$

i.e., the α phase increased to 64.8% while the amount of liquid phase decreased to 35.2%. In addition, the phase compositions changed from 250 to 210°C. The composition of the α phase became 17Sn (from 13Sn) and the composition of the liquid became 54Sn (from 34Sn). The alloy composition has remained 30Sn, however the phase fractions and phase compositions changed between 250 and 210°C. The change in phase compositions was accomplished by the transfer of atoms between the phases, a process called diffusion. If the temperature is decreased further in the

$\alpha + L$ region, more α phase will form. Consider that the temperature is just above the eutectic temperature, say $183^\circ C + \theta$. The phase fractions are:

$$f_\alpha = \frac{61.9 - 30}{61.9 - 18.3} = \frac{31.9}{43.6} = 0.731$$

$$f_L = \frac{30 - 18.3}{61.9 - 18.3} = \frac{11.7}{43.6} = 0.269$$

The alloy consists now of 73.1% α phase while there is 26.9% liquid remaining. The composition of this liquid is 61.9Sn, i.e., the liquid has obtained the eutectic composition and can now solidify through the eutectic reaction $L \rightarrow \alpha + \beta$. The temperature will remain constant at $183^\circ C$ till the reaction is completed and all liquid is consumed. At temperature $183^\circ C - \theta$ the alloy consists of α and β phases with compositions 18.3Sn and 97.8Sn and phase fractions:

$$f_\alpha = \frac{97.8 - 30}{97.8 - 18.3} = \frac{67.8}{79.5} = 0.852$$

$$f_\beta = \frac{30 - 18.3}{97.8 - 18.3} = \frac{11.7}{79.5} = 0.148$$

The amount of the α phase has increased to 85.2%. From that amount, the 73.1% is α phase formed before the eutectic reaction and, therefore, is called *proeutectic* α phase. The remaining $85.2 - 73.1 = 12.1\%$ is α phase through the eutectic reaction, together with the formation of β phase. The $Pb - 30Sn$ alloy, below the eutectic temperature consists of 73.1% proeutectic α phase and eutectic mixture $\alpha + \beta$ consisting of 12.1% α phase and 14.8% β phase. A further decrease of temperature inside the $\alpha + \beta$ region of the phase diagram results in the formation of more β phase, since the solid solubility of Sn in Pb is reduced. For example, at $90^\circ C$, the lever rule gives

$$f_\alpha = \frac{98 - 30}{98 - 5} = \frac{68}{93} = 0.731$$

$$f_\beta = \frac{30 - 5}{98 - 5} = \frac{25}{93} = 0.269$$

which means that the β phase increased from 14.8 at $183^\circ C$ to 26.9% at $90^\circ C$. The cooling curve of the $Pb - 30Sn$ alloy is shown in Figure 4.26. Starting from the temperature of $350^\circ C$ where the alloy is liquid, the alloy is cooled and solidification starts at the liquidus temperature $265^\circ C$, with the formation of α phase. The slope of the cooling curve decreases in the two-phase $\alpha + L$ region. Cooling from $265^\circ C$ down to the eutectic temperature results in the formation of proeutectic α phase, while the remaining liquid is enriched in Sn. At $183^\circ C$ the remaining liquid solidifies through the eutectic reaction, forming a mixture of $\alpha + \beta$ phases. The temperature remains constant until the eutectic reaction is completed. Then the alloy continues to cool down to room temperature. The development of the microstructure of the alloy at several stages of cooling is shown schematically in Figure 4.27, while a metallographic image of the $Pb - 30Sn$ alloy is compared against the $Pb - 40Sn$ and $Pb - 50Sn$ alloys in Figure 4.28. It is evident that the amount of eutectic mixture

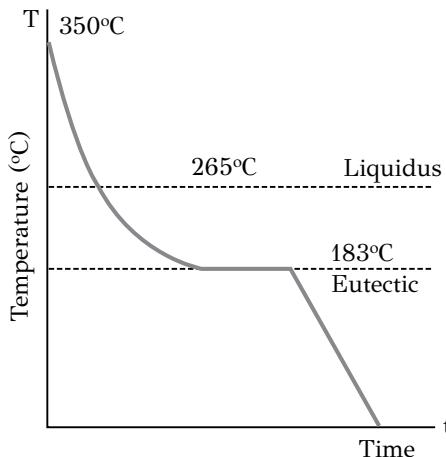


Figure 4.26: Cooling curve of alloy Pb-30Sn.

$(\alpha + \beta)$ increases with the Sn content of the alloy. The amount of the eutectic mixture can be derived using the lever rule, at $183^\circ C + \theta$, since the amount of the eutectic mixture equals the amount of liquid just before the eutectic reaction:

$$Pb - 30Sn : \quad f_L = f_{(\alpha+\beta)} = \frac{30 - 18.3}{61.9 - 18.3} = \frac{11.7}{43.6} = 0.26$$

$$Pb - 40Sn : \quad f_L = f_{(\alpha+\beta)} = \frac{40 - 18.3}{61.9 - 18.3} = \frac{21.7}{43.6} = 0.497$$

$$Pb - 50Sn : \quad f_L = f_{(\alpha+\beta)} = \frac{50 - 18.3}{61.9 - 18.3} = \frac{31.7}{43.6} = 0.727$$

The eutectic alloy $Pb - 61.9Sn$ consists of 100% eutectic $(\alpha + \beta)$ phase mixture.

Consider now a hypereutectic alloy, which is an alloy with composition higher than the eutectic composition, such as $Pb - 70Sn$. In this alloy solidification starts at $190^\circ C$ with the formation of proeutectic β phase. As the temperature decreases, within the two-phase $\beta + L$ region, more β phase forms while the remaining liquid is depleted from Sn . At $183^\circ C$ the liquid obtains the eutectic composition (61.9Sn) and solidifies through the eutectic reaction, forming a mixture of $(\alpha + \beta)$ phases. In this way the final microstructure, depicted in Figure 4.29, consists of proeutectic β phase and eutectic mixture of $(\alpha + \beta)$ phases. The fraction of the eutectic mixture is derived with the lever rule as above:

$$Pb - 70Sn : \quad f_L = f_{(\alpha+\beta)} = \frac{97.8 - 70}{97.8 - 61.9} = \frac{27.8}{35.9} = 0.774$$

so the alloy consists of 22.6% proeutectic β phase and 77.4% eutectic $(\alpha + \beta)$ mixture.

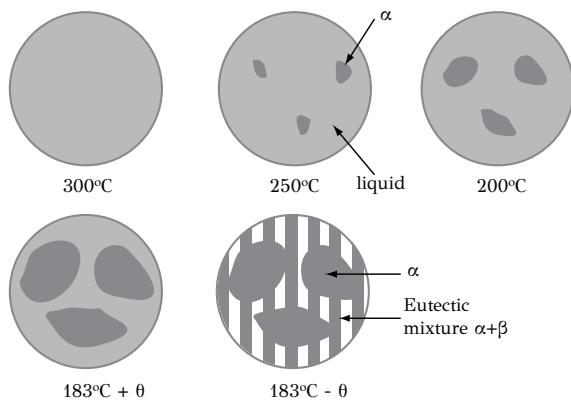


Figure 4.27: Development of the microstructure of the alloy $Pb - 30Sn$ at several stages of cooling.

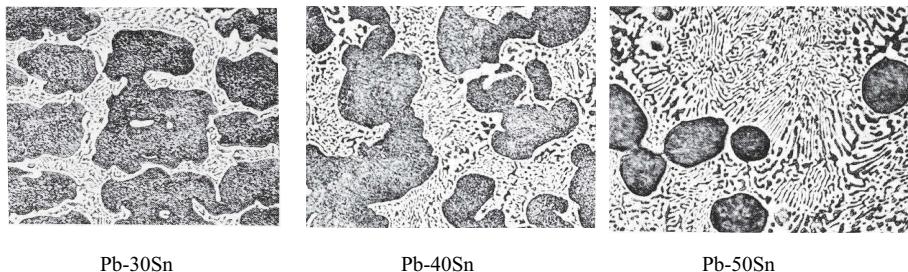


Figure 4.28: Microstructure of $Pb - 30Sn$, $Pb - 40Sn$ and $Pb - 50Sn$ alloys. The proeutectic α phase appears with dark color, surrounded by a eutectic mixture of $\alpha + \beta$. With the increase of Sn the percentage of the eutectic mixture increases.

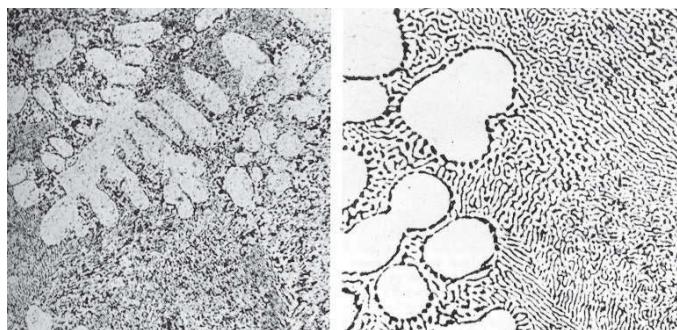


Figure 4.29: Microstructure of $Pb - 70Sn$ alloy consisting of the proeutectic β (white color) and the eutectic mixture of $\alpha + \beta$. The right picture is a higher magnification of the left.

Consider now the free energy-composition ($G - X$) diagrams of the $Pb - Sn$ system at various temperatures, shown in Figure 4.30. At $250^\circ C$, the common tangent

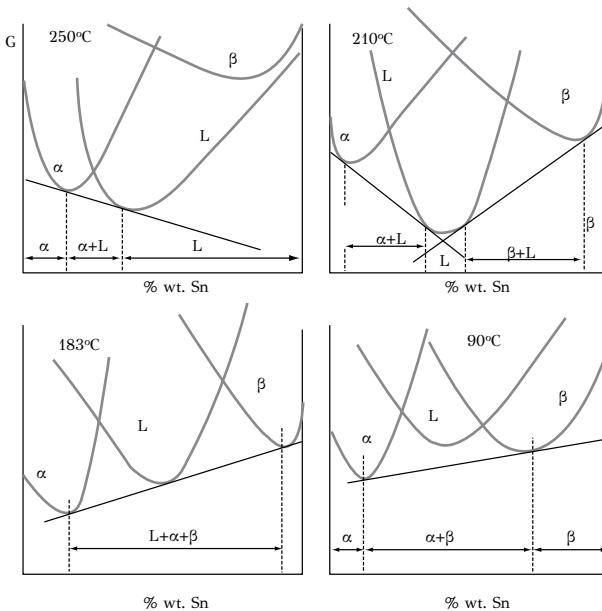


Figure 4.30: Free energy-composition curves of the $Pb - Sn$ system at various temperature.

to the α and L phases defines the compositions $13Sn$ and $34Sn$ for the α and L phases respectively. The curve for the β phase is higher and does not participate in the equilibrium. At $210^\circ C$, the curve of the L phase crosses both the curves of the α and β phases. Two common tangents are drawn corresponding to the equilibrium between the liquid (L) and the α and β phases respectively. Between the two-phase regions ($\alpha + L$) and ($\beta + L$) there is a single phase (L) region. At $183^\circ C$ (eutectic temperature), the three curves α , β and L have one common tangent. There is equilibrium between three phases (degrees of freedom $F = 0$) during the eutectic reaction. At $90^\circ C$ the curve of the liquid phase (L) moves up and the liquid does not participate at the equilibrium. The common tangent the curves of the α and β phases defines the compositions $5Sn$ and $98Sn$ of the α and β phases at equilibrium. A few words regarding the thermodynamic quantities. The free energy of mixing is $\Delta G_m = \Delta H_m - T\Delta S_m$. The sign of the entropy of mixing (ΔS_m) is always positive. Since there is no complete solid solubility displayed in the $Pb - Sn$ diagram and the components Pb and Sn do not form one solid solution but two solid solutions, α and β , the enthalpy of mixing (ΔH_m) is also positive. With $\Delta S_m > 0$ and $\Delta H_m > 0$, the sign of ΔG_m depends on temperature. For example, consider the alloy $Pb - 10Sn$. At $200^\circ C$ the alloy is 100% α phase, i.e., a single solid solution. For this case $\Delta G_m < 0$ due to the contribution of the $-T\Delta S_m$ term. However, at a lower temperature, e.g.,

50°C, the contribution of the $-T\Delta S_m$ term is small and $\Delta G_m > 0$. As a result the alloy decomposes in two phases α and β , as predicted by the phase diagram.

4.9 SYNOPSIS

1. Phase diagrams are maps, in the temperature - composition space, of the alloy constitution under conditions of thermodynamic equilibrium. The term constitution corresponds to the phases and the composition of phases present in the system. Since phase diagrams are representations of thermodynamic equilibrium, they can be obtained directly from the minimization of the Gibbs free energy of the system.
2. Pure metals are one-component systems and the free energy is only a function of temperature, $G = G(T)$. The free energy can be calculated at any temperature provided the temperature dependence of the specific heat C_p is known. Knowledge of the $G(T)$ curve can be used for the study of phase transformations, such as allotropic transformations or solidification, in pure metals.
3. The free energy of a binary solid solution is a function of temperature and composition, $G = G(T, X)$ and can be derived by a model, which takes into account only first nearest neighbor interactions and is called the regular solution model.
4. Central role in the regular solution model plays the excess term, which corresponds to the enthalpy of mixing. It arises from the interaction between atoms in the solid solution and expresses the departure of the solution from the ideal behavior, where there are no interactions.
5. When the enthalpy of mixing is negative ($\Delta H_m < 0$) the free energy of mixing becomes negative ($\Delta G_m < 0$). When $\Delta H_m = 0$, the solution is ideal and $\Delta G_m < 0$. In both cases the solid solution exhibits complete solid solubility. When the enthalpy of mixing is positive ($\Delta H_m > 0$) the sign of ΔG_m depends on temperature due to the entropic term $-T\Delta S_m$. At low temperatures the term ΔH_m dominates the term $-T\Delta S_m$ and ΔG_m becomes positive, leading to decomposition in two phases. Above a specific temperature the term $-T\Delta S_m$ dominates, making ΔG_m negative, favoring the formation of a single solid solution.
6. The chemical potential expresses the free energy change, which results from the addition or removal of one mole of component from the solution. The chemical potential is a function of temperature and composition. While in ideal solutions the chemical potential depends on the mole fraction of the component, in real solutions it is related to the activity of the component. The activity coefficient then indicates the departure of the solution from ideality.
7. When two phases are at thermodynamic equilibrium, at constant temperature and pressure, the chemical potentials of the components at each phase are equal and the total free energy is minimized. The condition of thermodynamic equilibrium

between two phases is expressed by the common tangent to the free energy curves of the two phases at a certain temperature. The two phases coexist at equilibrium and their relative fractions are given by the lever rule.

8. There are limitations on the number of independent variables of the system, when there is thermodynamic equilibrium between phases. These limitations are expressed by the Gibbs phase rule. The rule correlates the degrees of freedom F , i.e., the number of variables (e.g., temperature, pressure, alloy composition), which can vary independently without disturbing the equilibrium, with the number of components C and the number of phases P , participating in the equilibrium.
9. According to the Gibbs phase rule, a binary system can exhibit bivariant ($F = 2$), univariant ($F = 1$) and invariant ($F = 0$) equilibrium. In the case of bivariant equilibrium, the temperature and the composition of the phase (or alloy) can vary independently without disturbing the equilibrium. In the case of univariant equilibrium, if the temperature is selected, then the compositions of the phases participating in the equilibrium are fixed for that temperature. Finally in the case of invariant equilibrium, both the temperature and the compositions of the phases participating in the equilibrium are fixed.
10. During an invariant reaction in a binary system, when heat is supplied or removed from the alloy, the temperature and the composition of the phases remain constant. Only the phase fractions change up to the point where one phase is consumed and the equilibrium becomes monovariant, or two of the phases are consumed and the equilibrium becomes bivariant.
11. The locus of contact points of the common tangent to the $G - X$ curves of the two phases constitute the phase boundary curves in the equilibrium phase diagram. In other words the binary phase diagram is constructed from lines of univariant equilibria and cross points of invariant equilibria. The phase diagram is then a map, in the temperature-composition space, which provides the alloy constitution, i.e., the phases present and the composition of these phases at thermodynamic equilibrium.

4.10 REVIEW QUESTIONS

1. Consider the solidification of a pure metal. The liquid is undercooled by ΔT below the melting point T_m . The difference in free energies of the solid and liquid phases is the thermodynamic driving force for solidification. Show that the driving force can be approximated by

$$\Delta G = L \frac{\Delta T}{T_m}$$

where L is the latent heat of fusion.

2. (a) Construct the $Ag - Cu$ phase diagram using the following data:
 - Melting point of A and Cu : 960 and $1080^\circ C$ respectively

- Eutectic temperature: $780^{\circ}C$
 - Eutectic composition: 28 wt%Cu
 - Phase compositions at eutectic temperature: α phase: 9%Cu, β phase: 92%Cu
 - Max solid solubility at $400^{\circ}C$: α phase: 1%Cu, β phase: 0%Ag
- (b) Determine the liquidus temperature of the $Ag - 40Cu$ (wt%) alloy. Is there another alloy with the same liquidus temperature?
- (c) 26gr of $Ag - 7.5Cu$ alloy are melted with 376gr pure Cu forming a new alloy. Determine the composition, the liquidus and the solidus temperature of the new alloy.
- (d) How much copper (in gr) is contained in 100gr of eutectic alloy at $600^{\circ}C$?
- (e) 80gr of eutectic alloy solidify according to the eutectic reaction. What are the amounts (in gr) of the α and β phases forming from the above reaction?
- (f) Consider 120gr of $Ag - 15Cu$ alloy. How much copper (in gr) should be added in order to obtain a $Ag - 30Cu$ alloy?
3. Consider the $Ag - Cu$ phase diagram (Figure 4.19). It has a eutectic point at $Ag - 28Cu$ at the eutectic temperature of $780^{\circ}C$.
- (a) For the $Ag - 7Cu$, $Ag - 15Cu$, $Ag - 28Cu$ and $Ag - 70Cu$ alloys:
- describe the cooling path from $1000^{\circ}C$ to room temperature
 - sketch the expected final microstructure of the alloys
 - calculate the equilibrium phase fractions at $820^{\circ}C$
- (b) Draw the free energy - composition curves ($G - X$) of phases α , β and L at 820 , 780 and $760^{\circ}C$
- (c) What are the signs of enthalpy of mixing ΔH_m , entropy of mixing ΔS_m and free energy of mixing ΔG_m , of the $Ag - 5Cu$ alloy at 800 and $300^{\circ}C$?
- (d) Which alloy contains 30% proeutectic β phase at $800^{\circ}C$?
4. Consider the $Cu - Zn$ phase diagram and answer the following questions:
- (a) What are the solid solubility limits of Zn in Cu and Cu in Zn at $200^{\circ}C$?
- (b) Write the eutectoid reaction, indicating the eutectoid temperature and the phase compositions.
- (c) Draw the free energy - composition curves for the phases participating in the eutectoid reaction.
- (d) For the $Cu - 40Zn$ alloy at $200^{\circ}C$, what are the phase fractions and compositions?
5. The binary $Al - Cu$ system is the basis of the 2xxx alloys used for the construction of wings and fuselage of aircraft. Consider the $Al - Cu$ binary phase diagram to answer the following questions:
- (a) At room temperature the $Al - 4Cu$ alloy consists of α and θ (Al_2Cu) phases. At which temperature should the alloy be heated in order to dissolve the θ phase?
- (b) At a certain temperature T , the ratio $(X - X_T^{\alpha})/X$, where X is the alloy composition and X_T^{α} is the solid solubility limit of Cu in the α phase, is called supersaturation ratio. Calculate the supersaturation ratio of the $Al - 4Cu$ alloy at 450 , 400 , 350 and $300^{\circ}C$.

- (c) Consider that the $Al - 4Cu$ alloy is heated to $550^{\circ}C$ and then it is slowly cooled. Calculate the fraction of θ phase which forms by precipitation from the α phase at 450, 400, 350 and $300^{\circ}C$.
- (d) The artificial aging treatment is a heat treatment applied for the precipitation strengthening of $Al - Cu$ alloys. The alloy is heated to a high temperature to dissolve the θ phase (solution treatment). Then the alloy is cooled rapidly to form a supersaturated solid solution. Then the alloy is heated to an intermediate temperature (aging) to precipitate the θ phase. According to the above, propose the alloy composition, which allows for the maximum precipitation of θ phase, as well as the solution and aging temperatures for the heat treatment.

BIBLIOGRAPHY-FURTHER READING

1. Gaskell, D. R. (1981). *Metallurgical Thermodynamics*, McGraw Hill.
2. Ragone, D. V. (1995). *Thermodynamics of Materials*, Vol. 1,2, John Wiley.
3. Prince, A. (1996). *Alloy Phase Equilibria*, Elsevier, New York.
4. *ASM Handbook, Vol. 3: Alloy Phase Diagrams*, (1992). ASM, Metals Park, OH, USA.
5. Gordon, P. (1992). *Principles of Phase Diagrams in Materials Systems*, McGraw Hill, New York, USA.
6. Saunders, N., Miodownik, A. P. (1998). *CALPHAD: Calculation of Phase Diagrams*, A Comprehensive Guide, Pergamon.
7. Porter, D.A., Easterling, K.E., Sherif, M.Y. (2009). *Phase Transformations in Metals and Alloys*, CRC Press.
8. Hillert, M. (2008). *Phase Equilibria, Phase Diagrams and Phase Transformations*, Cambridge University Press, England.
9. Massalski, T.B., Okamoto, H., Subramanian, P.R., Kacprzak, L. (ed) (1990). *Binary Alloy Phase Diagrams*, 2nd edition, ASM International, Metals Park, OH, USA.



Taylor & Francis
Taylor & Francis Group
<http://taylorandfrancis.com>

5 Diffusion

5.1 INTRODUCTION

The development of microstructure in metals takes place by movements of atoms in the crystal lattice, a solid-state mass transfer, which is called *diffusion*. Most phase transformations are diffusive, i.e., diffusion is required in order for the transformation to take place. The transformation of austenite to pearlite in a carbon steel, through the eutectoid reaction, takes place by the diffusion of carbon atoms in the iron lattice. Basic materials processing in physical metallurgy, like the precipitation of a new phase, the homogenization of cast structures and the carburizing of steel are accomplished by the diffusion of atoms. The kinetics of these processes, i.e., the rate at which they are accomplished, depends entirely on the rate of diffusion. However it is not only phase transformations the reason why we are interested in diffusion. The deformation of metals at high temperatures, i.e., creep, takes place by mechanisms in which diffusion has an important contribution. In this way, creep resistant alloys can be developed, provided we understand ways to retard diffusion in these materials. Diffusion takes place always in the direction to reduce the free energy. It will be shown that the *driving force* for the diffusion of an alloying element is the *chemical potential gradient* $d\mu/dx$ of that element. It should be noted, however, that the laws describing diffusion (Fick's laws) correlate the diffusional flux of a component with the concentration gradient of that component, since in practice it is much easier to measure variations in concentration than differences in chemical potential. In the following paragraphs, the mechanisms of diffusion, the Fick's laws describing diffusion as well as the temperature dependence of diffusion will be discussed. The discussion on the relation between thermodynamics and diffusion, substitutional diffusion, the diffusion in multicomponent systems and the diffusion from non-chemical forces will follow. Finally, simple analytical solutions to common diffusion problems will be presented.

5.2 DIFFUSION MECHANISMS

The mechanisms by which atoms move in the lattice depends on the type of solid solution, either interstitial or substitutional. Interstitial atoms occupy the octahedral interstitial sites in *FCC* and *BCC* structures. In most cases, solid solutions are dilute and an interstitial atom is always surrounded by vacant interstitial sites. In this way the interstitial atom can move to the neighboring vacant interstitial site, provided there is enough energy for the atom to pass through iron atoms 1 and 2, as depicted in Figure 5.1a. The diffusion of interstitial atoms, like carbon in steels, is called *interstitial diffusion*. In steels carbon occupies interstitial sites in both ferrite and austenite. The diffusion of carbon takes place with consecutive jumps from one interstitial position to the next. In the case of substitutional solid solutions, like copper in nickel, the diffusion is called *substitutional diffusion*. Finally the diffusion of

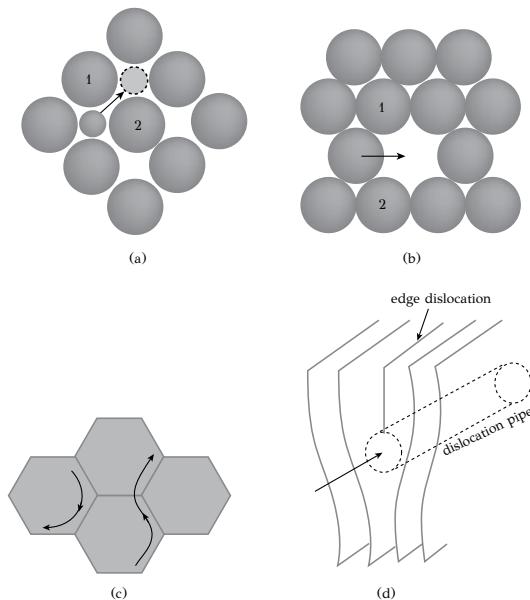


Figure 5.1: Diffusion mechanisms: (a) interstitial diffusion, (b) substitutional diffusion, (c) grain boundary diffusion, (d) dislocation diffusion.

copper atoms in copper is called *self-diffusion*. In these cases diffusion can take place only if the neighboring lattice site is occupied by a vacancy. In this way, if the atom acquires enough energy, it could jump to the vacancy, as depicted in Figure 5.1b. This jump is equivalent to the jump of the vacancy to the atom site and the mechanism has also been termed *vacancy diffusion*. In order for substitutional or self diffusion to take place two conditions should be fulfilled. There should be an appropriate vacancy concentration and the atoms must acquire enough thermal energy so that the vibration at their equilibrium position ends up to a successful jump to the vacancy. Both conditions depend strongly on temperature and this explains the temperature dependence of diffusion.

When the interstitial or substitutional diffusion takes place in the lattice, it is called *lattice diffusion* or *volume diffusion* or *bulk diffusion*. However diffusion can be assisted by structural imperfections. *Grain boundary diffusion* is the diffusion of atoms through grain boundaries, which exhibit a more “open structure” than the bulk of the crystal, as depicted in Figure 5.1c. In this case the rate of diffusion is higher than bulk diffusion. In fact diffusion is faster in metals with a finer grain size. This affects the deformation at high temperatures, i.e., creep, where diffusion plays an important role. As will be discussed in Chapter 9, alloys exhibiting high creep resistance, have either a coarse grain size or they are manufactured as single crystals. Dislocations assist diffusion, since the space below the extra half plane serves as a channel for the rapid diffusion of atoms, as depicted in Figure 5.1d. This type of

diffusion is called *dislocation pipe* or *core diffusion*. Both grain boundary and dislocation diffusion proceed at higher rates than bulk diffusion. Grain boundaries and dislocations operate, therefore, as *high-diffusivity paths* and influence the kinetics of diffusional phase transformations.

5.3 FICK'S FIRST LAW OF DIFFUSION – THE DIFFUSION COEFFICIENT

In this section the diffusive flux of a component will be correlated with its concentration gradient. Interstitial diffusion will be considered, where diffusion takes place with jumps of atoms in empty interstitial sites. During diffusion, the atoms of the host lattice remain fixed in their lattice sites. Consider the diffusion of carbon in austenite, which is an interstitial $Fe - C$ solid solution with *FCC* structure. Consider also that the solution is dilute, so that each carbon atom is surrounded by six empty interstitial sites. The carbon atom performs Γ successful jumps per unit time (jump frequency) to the empty interstitial sites. The jumps are completely random, i.e., carbon has the same probability to occupy any of the available six empty interstitial sites. In addition the jumps are uncorrelated, i.e., the direction of a jump is not related in any way with the direction of the previous jump. Consider, in addition, that the jump frequency Γ is independent from the carbon concentration in austenite. In Figure 5.2, a concentration gradient of carbon atoms in the solid solution is depicted. Let n_1 and

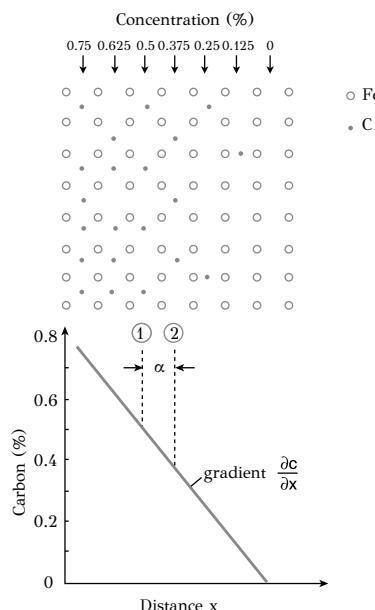


Figure 5.2: Interstitial diffusion of carbon in a $Fe - C$ solid solution. The diffusive flux is proportional to the concentration gradient.

n_1 be the concentrations (in atoms/m^2) in planes (1) and (2) respectively.

The spacing between these planes is equal to the lattice parameter α . The flux J_{12} (in $\text{atoms}/\text{m}^2\text{s}$) of carbon atoms from plane (1) to plane (2) is

$$J_{12} = \frac{1}{6}\Gamma n_1$$

while at the same time the flux of carbon atoms from plane (2) to plane (1) is

$$J_{21} = \frac{1}{6}\Gamma n_2$$

Because $n_1 > n_2$ the net flux of carbon atoms from plane (1) to plane (2) is

$$J = J_{12} - J_{21} = \frac{1}{6}\Gamma(n_1 - n_2)$$

The carbon concentration c (in atoms/m^3) for the two planes is $c_1 = n_1/\alpha$ and $c_2 = n_2/\alpha$, then

$$n_1 - n_2 = \alpha(c_1 - c_2) = -\alpha^2 \frac{\partial c}{\partial x}$$

and the net flux is

$$J = -\frac{1}{6}\Gamma\alpha^2 \frac{\partial c}{\partial x}$$

if we set

$$D = \frac{1}{6}\Gamma\alpha^2 \quad (5.1)$$

the net flux becomes

$$J = -D \frac{\partial c}{\partial x} \quad (5.2)$$

Relation (5.2) describes *Fick's first law of diffusion*. It is an empirical law, which correlates the diffusive flux of carbon atoms with the concentration gradient $\partial c/\partial x$. An analogous relation describes heat flow by conduction, where the heat flux q is proportional to the temperature gradient $\partial T/\partial x$. The (-) sign in Relation (5.2) arises from the fact that the concentration gradient $\partial c/\partial x$ is negative, while the flux is a positive quantity. The coefficient D in Relation (5.2) is called the *diffusion coefficient* and has units of m^2/s . Under the assumption, stated above, that the jump frequency is independent of concentration, the diffusion coefficient D should also be independent of concentration. In reality, however, the diffusion coefficient depends on the concentration of carbon in austenite. For example, at 1000°C , the diffusion coefficient of carbon in $\gamma-\text{Fe}$ is $D = 1.4 \times 10^{-11} \text{ m}^2/\text{s}$ for a concentration equal to $0.1\%\text{C}$ and $D = 8 \times 10^{-11} \text{ m}^2/\text{s}$ for a concentration $1\%\text{C}$. This means that the diffusion of carbon is faster in solid solutions with a higher carbon content, since carbon in interstitial sites stretches the lattice and generates a more "open" structure for the movement of carbon atoms. The jump frequency Γ can be calculated from Relation (5.1). In the same example as above, with $\alpha = 2.6\text{\AA}$ and $D = 1.4 \times 10^{-11} \text{ m}^2/\text{s}$ (1000°C), we get $\Gamma = 1.2 \times 10^9$ jumps/s. If the vibration frequency of the interstitial carbon atoms is

10^{13} Hz then it appears that only 1 out of 10^4 attempts is a successful jump leading to the occupation of the neighboring empty interstitial site. Another interesting issue is the net distance the carbon atom travels in the lattice during diffusion. This issue is related to the random walk problem, discussed in the next section.

5.4 RANDOM WALK AND DIFFUSION

A *random walk* consists of a series of consecutive and uncorrelated steps or jumps. By uncorrelated it is meant that the direction of a jump does not depend on the direction of the previous jump. The concept of random walk is important in understanding the diffusion process. In metallic materials the random walk is limited with respect to the length and direction of jumps. The length of the jump is limited by the lattice parameter, while the jump direction is limited by the available crystallographic directions and the symmetry of the lattice. For example when an atom jumps from a lattice site to a neighboring vacancy, the random walk is called a *lattice walk*. One important issue in random walk problems is to calculate the net distance travelled by an atom between an initial and a final position in the lattice. Consider an atom starting its random walk at position 1 and finishing in position 2, as depicted in Figure 5.3.

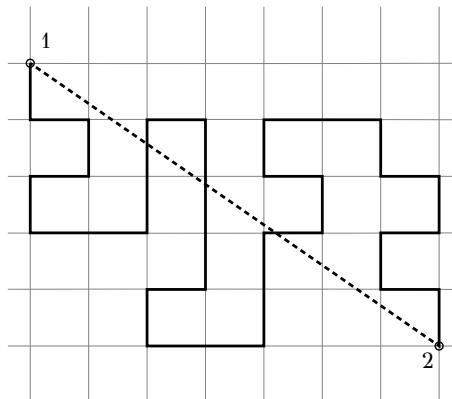


Figure 5.3: Random walk of an atom in a two-dimensional tetragonal lattice. The net diffusion distance is indicated by the dotted line.

It can be shown that if the atom performs n jumps of length α each, then the net displacement of the atom is not $r = n\alpha$ but

$$r = a\sqrt{n}$$

If the jump frequency is Γ then in time t , the net displacement or the net diffusion distance is

$$r = \alpha\sqrt{\Gamma t}$$

and with the help of Relation (5.1)

$$r = \sqrt{6Dt} \quad (5.3)$$

The quantity \sqrt{Dt} is very important in diffusion as it expresses the *net diffusion distance*. If the diffusion of carbon in austenite is considered, say at 1000°C , then the total distance covered by the carbon atom in 1.5 sec is 0.5 m while the net diffusion distance is about $10\mu\text{m}$.

5.5 FICK'S SECOND LAW OF DIFFUSION

Consider the carburizing process, which takes place by diffusion of carbon in a steel plate of thickness L at a constant temperature T (Figure 5.4a). Assuming that the

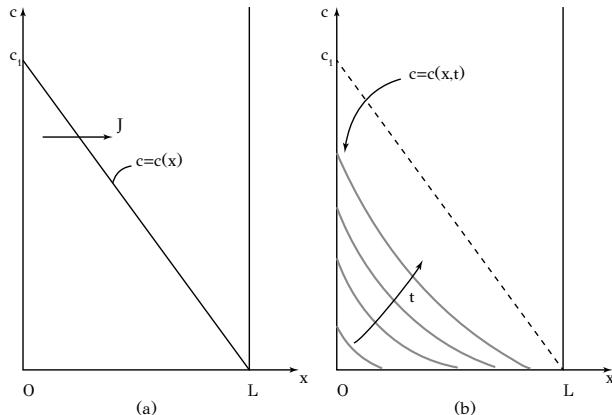


Figure 5.4: (a) Steady state conditions: constant flux, linear concentration profile, (b) non-steady state conditions with transient concentration profiles.

carburizing atmosphere is on the left side of the plate, a constant carbon concentration c_1 is established at the surface, while the concentration at the right side is $c = 0$. Assuming, in addition, that the diffusion coefficient is independent of concentration, Fick's first law gives

$$J = -D \frac{\partial c}{\partial x} = -D \frac{\Delta c}{\Delta x} = -D \frac{0 - c_1}{L} = \frac{Dc_1}{L}$$

The above relation provides the diffusive flux of carbon in the plate under steady state conditions. This means that as long as D , c_1 and L remain constant, then the flux J will remain constant. The question arises as to how the system approaches this steady state condition? Transient carbon concentration profiles are shown in Figure 5.4b from the beginning of the carburizing process ($t = 0$) up to the development of steady state. In the transient states the carbon concentration is a function of both position and time, i.e., $c = c(x, t)$. These transient profiles cannot be calculated with

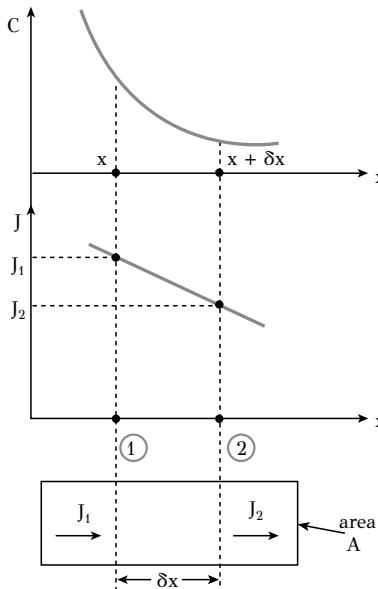


Figure 5.5: Concentration profile and associated flux during interstitial diffusion.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). *Phase Transformations in Metals and Alloys*, with permission from CRC press.

Fick's first law. We need Fick's second law. Consider a transient concentration profile in Figure 5.5. The flux at every position depends on the diffusion coefficient and the concentration gradient $\partial c / \partial x$. Now take a small segment δx . The flux of carbon atoms entering δx is

$$J_1 = -D \frac{\partial c}{\partial x}$$

and the flux of carbon atoms leaving δx is

$$J_2 = J_1 + \frac{\partial J}{\partial x} \delta x$$

where the second term on the *rhs* expresses the change of flux in segment δx . If the cross sectional area is A , then the number of carbon atoms entering δx in time δt is $J_1 A \delta t$. In the same time the number of atoms leaving δx is $J_2 A \delta t$. The increase of the carbon concentration in segment δx is then

$$\delta c = \frac{(J_1 - J_2) A \delta t}{A \delta x}$$

Applying the principle regarding the conservation of matter, the change in flux within segment δx is equal to the change in concentration and the above relations give

$$\frac{\partial c}{\partial t} = -\frac{\partial J}{\partial x}$$

Using Fick's first law for the flux we get

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial c}{\partial x} \right) \quad (5.4)$$

Relation (5.4) expresses *Fick's second law* and is a direct consequence of conservation of matter. It is a partial differential equation (*PDE*), the solution of which provides the transient concentration profiles $c(x, t)$ or in other words the evolution of the concentration field for a given flux. Methods for the solution of the *PDE* will be discussed in section 5.11. In the case where the diffusion coefficient is independent of concentration, Fick's second law becomes

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad (5.5)$$

It should be noted that the sign of $\partial c / \partial t$ depends on the sign of $\partial^2 c / \partial x^2$. If $\partial^2 c / \partial x^2 > 0$ then the carbon concentration increases with time. If, on the other hand, $\partial^2 c / \partial x^2 < 0$ the concentration decreases at the particular point of interest. The generalized Fick's first and second laws in three dimensions are:

$$\vec{J} = -D \nabla c \quad (5.6)$$

$$\frac{\partial c}{\partial t} = -\nabla \cdot \vec{J} \quad (5.7)$$

As stated above, Fick's first law is an empirical law, which correlates the diffusive flux of a component with the concentration gradient of that component. Fick's first law is, therefore, compatible with the framework of the Onsager principles of irreversible thermodynamics, where fluxes (of heat, mass, electric current, etc.) are caused by respective potential gradients. The basic principles of irreversible thermodynamics will be discussed in section 5.9. It is worth mentioning that the diffusive flux in Relation (5.6) is a vector, characterized by magnitude and direction.

Fick's second law is a consequence of conservation of mass. The *rhs* in Relation (5.7) is a scalar. Equation (5.7) is a *PDE*, which describes the development of the concentration field. The solution of this *PDE*, under the appropriate initial and boundary conditions, provides the spatial and temporal dependence of the concentration of a component in an alloy. In the case where the diffusion coefficient can be considered independent of concentration, simple analytical solutions of the *PDE* exist and are discussed in section 5.11. In most cases the diffusion coefficient depends on concentration. In these cases Equation (5.7) is a non-linear PDE, which can be solved by numerical methods, a matter to be discussed in section 5.12.

5.6 TEMPERATURE DEPENDENCE OF DIFFUSION

Consider now the effect of temperature on the diffusion rate, in other words the temperature dependence of the diffusion coefficient. Consider again carbon diffusion

in iron. The diffusion coefficient is associated with the jump frequency in empty interstitial sites, according to the Relation (5.1).

Consider first the energy changes during a jump (Figure 5.6). At equilibrium po-

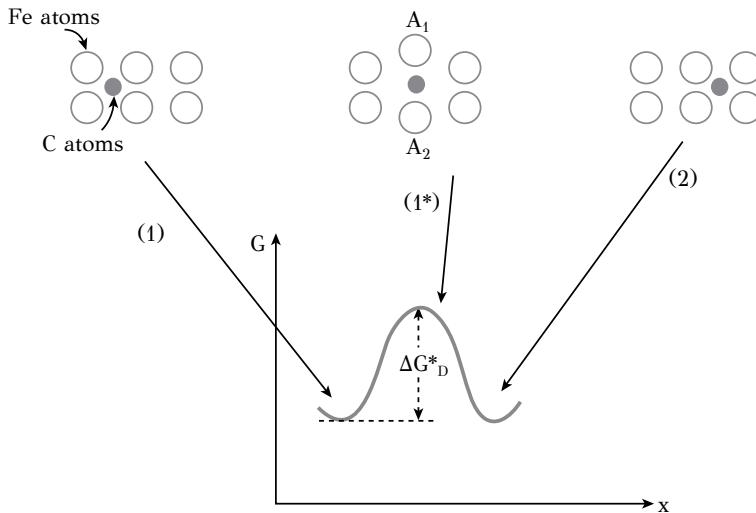


Figure 5.6: Interstitial diffusion: energy change during a diffusive jump of a carbon atom from its initial position (1), through the activated condition (1*) to the final position (2). ΔG_D^* is the activation energy for diffusion.

sition (1) the carbon atom is at a minimum free energy state G_1 . In order to perform a successful jump to the nearest neighboring interstitial site (2), it should pass from position 1^* , between the iron atoms A_1 and A_2 . Position 1^* corresponds to a higher energy, associated with the deformation required to expand the bond $A_1 - A_2$ in order for the carbon atom to pass through. This simply means that diffusion requires thermal activation. The probability of a carbon atom to acquire this extra energy ΔG_D^* and overcome this energy barrier is given by the Boltzmann factor $\exp(-\Delta G_D^*/RT)$. If v is the vibration frequency of interstitial atoms, then the jump frequency is

$$\Gamma = v z \exp(-\Delta G_D^*/RT)$$

where z is the coordination number of the interstitial sites. Then using Relation (5.1) we get

$$D = \left(\frac{1}{6} \alpha^2 z v \exp \frac{\Delta S_D^*}{R} \right) \exp \left(-\frac{\Delta H_D^*}{RT} \right)$$

Setting

$$D_o = \frac{1}{6} \alpha^2 z v \exp \frac{\Delta S_D^*}{R}$$

which is the temperature-independent term, we get

$$D = D_o \exp \left(-\frac{\Delta H_D^*}{RT} \right) \quad (5.8)$$

Relation (5.8) expresses the temperature dependence of the diffusion coefficient and is an Arrhenius-type expression. The quantity ΔH^* is the *activation enthalpy* for interstitial diffusion. The graphical representation of (5.8) is derived by taking the natural logarithms of the two sides of (5.8):

$$\ln D = \ln D_o - \frac{\Delta H_D^*}{RT}$$

and is depicted in Figure 5.7. The relation between $\ln D$ and $1/T$ is linear. The slope is equal to $-\Delta H_D^*/R$. This means that the activation enthalpy can be determined from experimental measurements of D as a function of temperature.

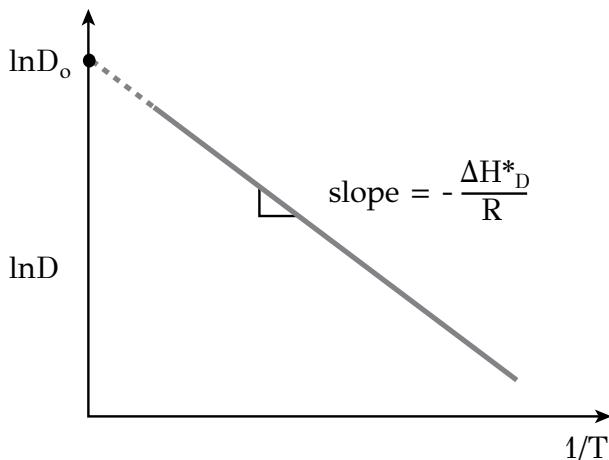


Figure 5.7: Arrhenius diagram showing the temperature dependence of the diffusion coefficient. The activation enthalpy can be determined from the slope.

According to the above discussion, the temperature dependence of diffusion is integrated in the diffusion coefficient. There is a direct relation between the diffusion coefficient and the atomic *mobility* in the crystal lattice. As a consequence the diffusion coefficient depends on the nature and strength of atomic bonding in the material. At the same time the diffusion coefficient depends on atomic packing, the way atoms arrange in space, i.e., crystal structure. In a close-packed structure with less free space the mobility is limited and the diffusion coefficient is lower than in a more “open” structure. We will try to quantify the above observations. Consider a metal where self-diffusion takes place by the jump of atoms in vacancy sites. The activation energy ΔH_D^* can be written as

$$\Delta H_D^* = \Delta H_v + \Delta H_m \quad (5.9)$$

The first term ΔH_v is the energy for vacancy formation, which was discussed in Chapter 3. ΔH_v depends on the nature and strength of atomic bonds. Since the bond strength influences directly the melting point T_m , it is a natural consequence that the

diffusion coefficient correlates through ΔH_V with the melting point. This correlation is depicted in Figure 5.8 for *FCC* metals. The relation between ΔH_D^* and T_m is linear.

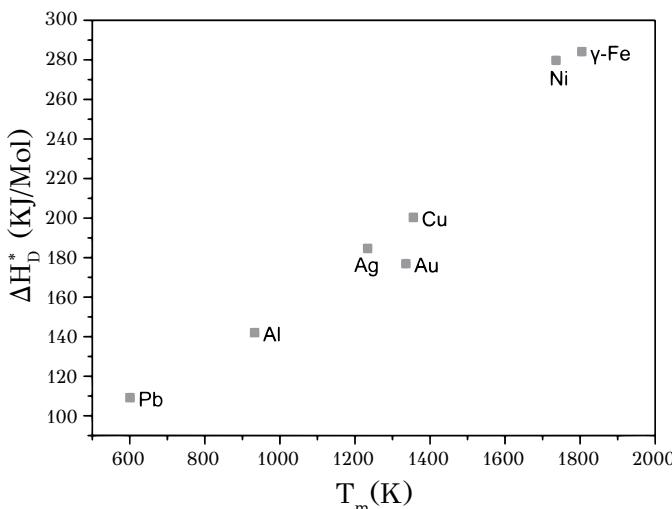


Figure 5.8: Correlation between the activation enthalpy for diffusion and the melting point for some *FCC* metals.

The quantity $\Delta H_D^*/RT_m$ is constant and approximately equal to 18 for most *FCC* metals. In addition, the diffusion coefficient for *FCC* metals at the melting point is equal to $D(T_m) = 10^{-12} m^2/s$. As a generalization, metals with the same crystal structure have the same diffusion coefficient at the same homologous temperature (T/T_m). The second term ΔH_m in Relation (5.9) is the energy required for the movement of atoms and depends on atomic packing. In interstitial diffusion, the formation of vacancies is not a prerequisite, then

$$\Delta H_D^* = \Delta H_m$$

A direct consequence is that interstitial diffusion is faster than substitutional diffusion. Several of the effects discussed above are depicted in Figure 5.9.

Some remarks are given below:

- a Diffusion of hydrogen in iron is faster than the diffusion of carbon in iron. This is attributed to the higher mobility of the hydrogen atom, since it is much smaller in size than carbon (lower ΔH_m).
- b Both hydrogen and carbon diffuse faster in *BCC* iron ($\alpha - Fe$) than in *FCC* ($\gamma - Fe$). This is attributed to the more open structure of *BCC* relative to *FCC*. Remember that the atomic packing factor of *FCC* is 0.74 while that of *BCC* is

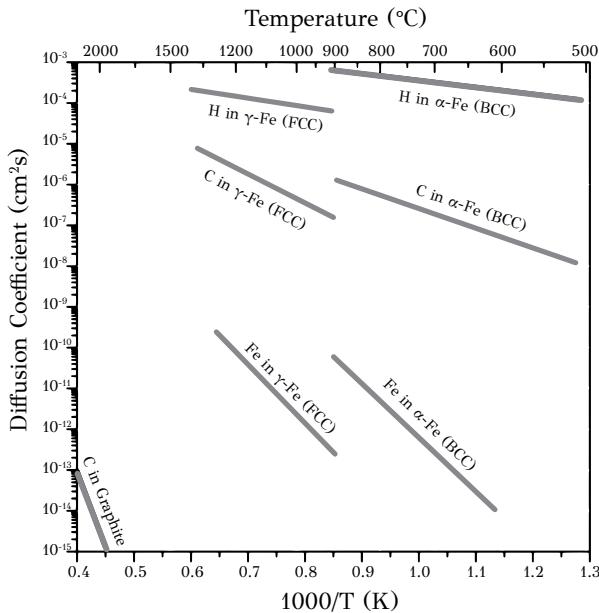


Figure 5.9: Arrhenius plot for several cases of diffusion: hydrogen in iron, carbon in iron, self diffusion of iron and self diffusion of carbon in graphite.

- 0.68, as discussed in Chapter 2.
- c The Arrhenius slope, i.e., the activation energy for the self-diffusion of iron is larger than the activation energy for the diffusion of carbon in iron. This is attributed to the fact the first case is substitutional diffusion while the second case is interstitial diffusion.
 - d The diffusion of carbon in graphite (self-diffusion of carbon) is slower than the self-diffusion of iron. Both cases concern substitutional diffusion, however graphite exhibits covalent bonding, which is much stronger than metallic bonding in iron. As a consequence the vacancy energy of formation in graphite is higher, leading to higher activation energy for diffusion.

5.7 THERMODYNAMICS AND DIFFUSION

5.7.1 DRIVING FORCE FOR DIFFUSION

In this section it will be shown that the *driving force* for diffusion of a component in a solid solution is the *gradient of the chemical potential* ($d\mu/dx$) of that component. In the course towards thermodynamic equilibrium, diffusion takes place always in a direction leading to reduction of free energy. It is then possible to observe diffusion

from regions with high concentration to regions with low concentration and vice versa. Consider two solid solutions between metals *A* and *B* with concentrations c_1 and c_2 (Figure 5.10). The first solution is richer in *B*, i.e., $c_1 > c_2$. The two solutions are brought in contact at a high temperature to activate diffusion. Now consider two cases.

In the first case (Figure 5.10a), the solid solution *A* – *B* has a negative enthalpy of mixing ($\Delta H_m < 0$) and, according to the discussion in Chapter 4, mixing of the two components and the formation of a homogenous solid solution with composition c_o is favored (Figure 5.10b). The formation of this solid solution takes place by diffusion of *B* from solution c_1 , which is rich in *B* to solution c_2 , which is lean in *B*. At the same time *A* diffuses from solution c_2 , which is rich in *A* to solution c_1 , which is leaner in *A*. The diffusion of both *A* and *B* takes place from regions with high to regions with low concentration. In this way the free energy is reduced to G_o (Figure 5.10c). Regarding the chemical potentials, as shown in Figure 5.10d, diffusion of both *A* and *B* takes place from regions with high to regions with low chemical potential since $\mu_A(c_2) > \mu_A(c_1)$ and $\mu_B(c_1) > \mu_B(c_2)$.

In the second case (Figure 5.10e), the solid solution *A* – *B* has a positive enthalpy of mixing ($\Delta H_m > 0$). According to the discussion in Chapter 4, this leads to the formation of a miscibility gap favoring the decomposition in two solid solutions, one rich in *B*, with composition c'_1 and one rich in *A* with composition c'_2 (Figure 5.10f). The free energies of the two solutions become G'_1 and G'_2 and the total free energy is reduced to G'_o (Figure 5.10g). This takes place by the diffusion of both *A* and *B* from regions with low concentration to regions with high concentration, i.e., diffusion against the concentration gradient dc/dx . This is called *uphill diffusion* and in this case the diffusion coefficient is negative. Uphill diffusion will be considered in the study of spinodal decomposition in Chapter 6. Regarding the chemical potentials, as shown in Figure 5.10h, diffusion of both *A* and *B* takes place from regions with high to regions with low chemical potential since $\mu_A(c_1) > \mu_A(c_2)$ and $\mu_B(c_2) > \mu_B(c_1)$.

In both cases discussed above, diffusion of *A* and *B* takes place from high to low chemical potential and causes reduction of the free energy. The driving force for diffusion is then the *chemical potential gradient* $d\mu/dx$. Diffusion stops when the chemical potentials of the components equalize in the two solutions, when thermodynamic equilibrium is established. However, in practice, it is easier to measure concentration differences than differences in the chemical potential and diffusion is correlated to a *concentration gradient* dc/dx . In this case the chemical potential gradient can be correlated to the concentration gradient by using a thermodynamic model for the description of the free energy of the solid solution. This coupling between thermodynamics and diffusion is accomplished through the thermodynamic factor, to be discussed in the next section.

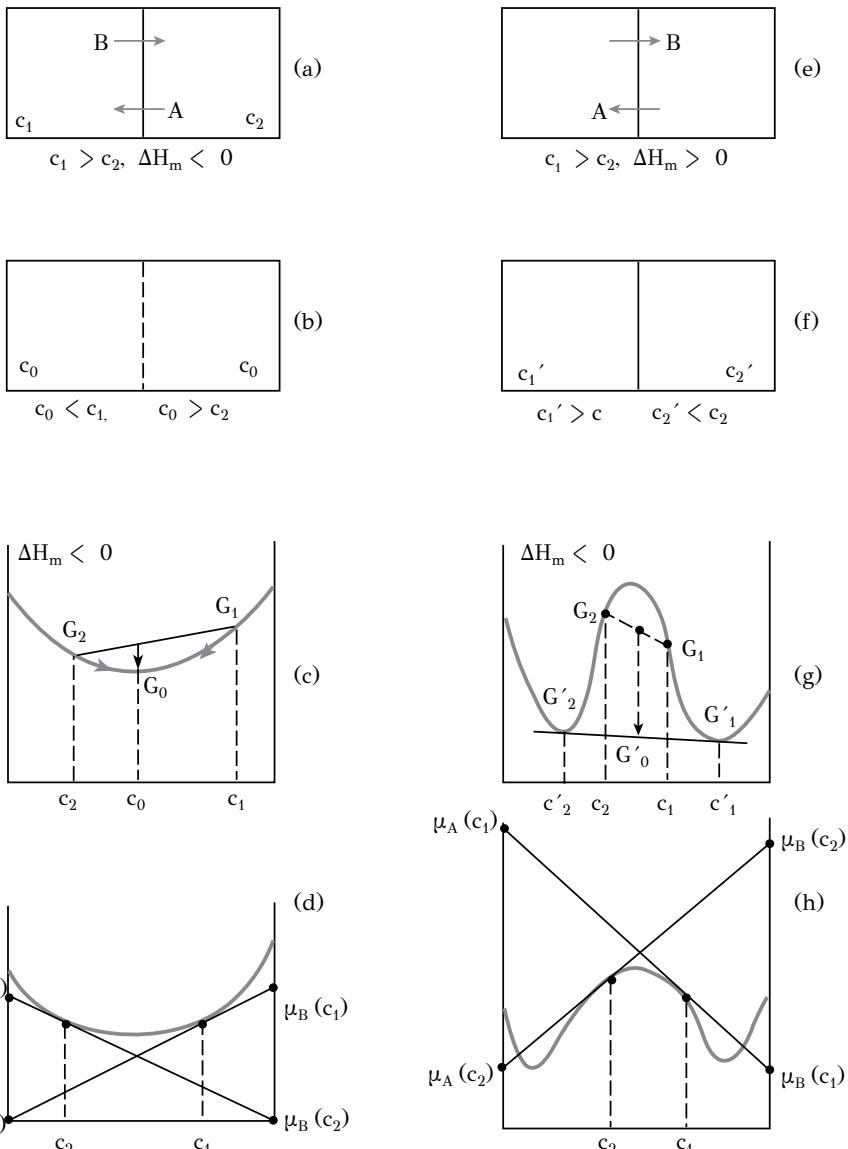


Figure 5.10: The driving force for diffusion: (a to d) diffusion from high to low concentration, (e to h) diffusion from low to high concentration (uphill diffusion). In both cases the diffusion takes place from high to low chemical potential (see discussion in text).

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). *Phase Transformations in Metals and Alloys*, with permission from CRC press.

5.7.2 CONCENTRATION DEPENDENCE OF THE DIFFUSION COEFFICIENT – THE THERMODYNAMIC FACTOR

As discussed above, the driving force for diffusion is the chemical potential gradient. Then the diffusive flux J is proportional to $d\mu/dx$

$$J_i = -M_i c_i \frac{\partial \mu_i}{\partial x} \quad (5.10)$$

where the quantity M_i is the mobility and c_i is the concentration of component i . The chemical potential gradient can be correlated to the concentration gradient with the relation

$$\frac{\partial \mu_i}{\partial x} = \frac{\partial \mu_i}{\partial c_i} \frac{\partial c_i}{\partial x}$$

and the Relation (5.10) becomes

$$J_i = -M_i c_i \frac{\partial \mu_i}{\partial c_i} \frac{\partial c_i}{\partial x}$$

Comparing the above relation with Fick's first law (Relation 5.2), the diffusion coefficient becomes

$$D_i = M_i c_i \frac{\partial \mu_i}{\partial c_i} \quad (5.11)$$

Relation (5.11) suggests that the diffusion coefficient integrates the composition dependence of the chemical potential. It is clear that a relation for $\partial \mu_i / \partial c_i$ is required from solution thermodynamics.

From Chapter 4, the chemical potential can be written as

$$\mu_i = G_i + RT \ln(\gamma_i X_i)$$

where G_i is the free energy of the pure component, γ_i the activity coefficient and X_i the atomic fraction of component i . Taking the derivative with respect to X_i we get

$$\frac{\partial \mu_i}{\partial X_i} = \frac{RT}{X_i} \left(1 + \frac{\partial \ln \gamma_i}{\partial \ln X_i} \right)$$

The concentration c_i is related to atomic fraction through the molar volume V_m

$$c_i = \frac{X_i}{V_m}$$

and the above relation becomes

$$\frac{\partial \mu_i}{\partial c_i} = \frac{RT V_m}{X_i} \left(1 + \frac{\partial \ln \gamma_i}{\partial \ln X_i} \right)$$

Finally Relation (5.11) becomes

$$D_i = M_i R T \left(1 + \frac{\partial \ln \gamma_i}{\partial \ln X_i} \right) \quad (5.12)$$

The quantity in parentheses is called the *thermodynamic factor*. Relation (5.12) is important, as it suggests that the composition dependence of the diffusion coefficient arises from the non-ideality of the solid solution, since for ideal solutions the thermodynamic factor is unity. In addition, in the case of very dilute solutions ($X_i \rightarrow 0$, $\gamma_i \rightarrow \text{const}$) the thermodynamic factor is unity. Thus in the cases of ideal or very dilute solutions the diffusion coefficient is simply

$$D_i = M_i RT \quad (5.13)$$

while it is independent from composition. In the case of non-ideal solutions the diffusion coefficient includes the thermodynamic factor as in (5.12). It is apparent that self-diffusion obeys Relation (5.13).

The above relations can be specified for a binary system $A - B$. Assuming that this system is a regular solution, then according to Chapter 4, the free energy of the solid solution is

$$G = X_A G_A + X_B G_B + \Omega X_A X_B + RT(X_A \ln X_A + X_B \ln X_B) \quad (5.14)$$

The first two terms express the free energies of the pure metals, the third term is the excess term expressing the deviation from ideality (enthalpy of mixing) and the fourth term is the entropic term (entropy of mixing). It can be shown that the thermodynamic factor can be written as

$$\left(1 + \frac{\partial \ln \gamma_B}{\partial \ln X_B}\right) = \frac{X_A X_B}{RT} \frac{\partial^2 G}{\partial X^2} \quad (5.15)$$

and

$$\left(1 + \frac{\partial \ln \gamma_B}{\partial \ln X_B}\right) = 1 - \frac{2\Omega}{RT} X_A X_B \quad (5.16)$$

Relation (5.15) correlates the thermodynamic factor with the curvature of the free energy ($G - X$ curve). This correlation is depicted in Figure 5.11. In regions where the free energy exhibits a positive curvature the thermodynamic factor is positive and, according to (5.12), the diffusion coefficient is positive. Then diffusion takes place according to Fick's first law from regions with high to regions with low concentration. However between the spinodal points K_1 and K_2 , the free energy exhibits a negative curvature and, as a consequence, the thermodynamic factor and the diffusion coefficient are negative. In this case uphill diffusion takes place. Uphill diffusion is the basic mechanism of spinodal decomposition, to be discussed in Chapter 6, which is a continuous phase transformation leading to the decomposition of a solid solution into two phases.

Relation (5.16) correlates the thermodynamic factor with the excess term in the free energy description given by Equation (5.14), i.e., with the enthalpy of mixing. According to the discussion in Chapter 4, the excess term expresses the deviation from the ideal solution. When the solution is ideal, then $\Omega = 0$ and the thermodynamic factor is unity. In this case the diffusion coefficient is given by Relation (5.13). In the cases where $\Omega < 0$ or $\Omega > 0$, according to Relation (5.16), the thermodynamic factor is either higher or lower than 1. In this case diffusion in a solid solution is either accelerated or retarded relative to self-diffusion.

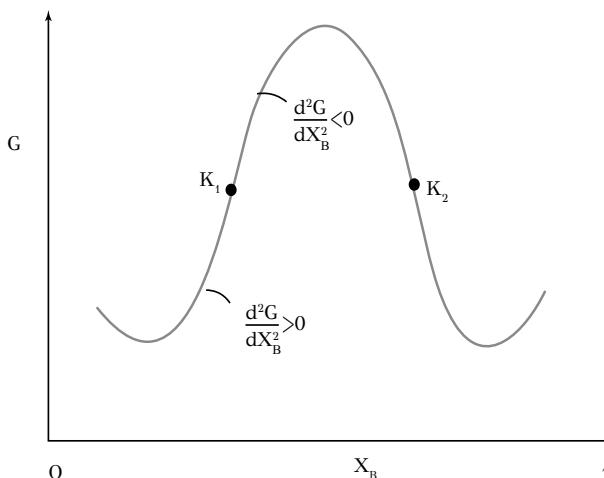


Figure 5.11: Free energy-composition diagram indicating the region between the spinodal points K_1 and K_2 , where the curvature of the free energy is negative. The associated thermodynamic factor and the diffusion coefficient are also negative (uphill diffusion).

5.8 SUBSTITUTIONAL DIFFUSION

Fick's first law and the diffusion coefficient were presented in section 5.3 based on interstitial diffusion. This was done for purpose. In interstitial diffusion, like the diffusion of carbon in γ -Fe, the carbon atom jumps in neighboring vacant interstitial sites. In addition, the jumps are uncorrelated, i.e., each jump is independent in direction from the previous jump. It can be considered that the carbon atoms move while the iron atoms are fixed. Then interstitial diffusion can be defined using the crystal lattice as a reference frame, called the C -frame¹. However, while in interstitial diffusion the carbon atoms are always surrounded by vacant interstitial sites, in substitutional diffusion, in a solid solution $A - B$, an atom of the lattice is not always surrounded by vacancies. Even more, the jump of atom A in a vacancy could be followed by the jump of a B atom in the prior A site. This means that the diffusion of A and B is not independent. As will be discussed below, concentration gradients of A and B in the solid solution cause displacement of the lattice. This generates the need to define diffusion in different reference frames or coordinate systems. The discussion in the following paragraphs concerns the various types of substitutional diffusion, the respective diffusion coefficients and the associated reference frames used for the definition of the diffusion coefficients. Three cases will be discussed: (a) self-diffusion, (b) diffusion in a homogeneous binary solid solution and (c) diffusion in a non-homogeneous binary solid solution exhibiting concentration gradients.

¹R.W. Balluffi, S.M. Allen & W.C. Carter, (2005). Kinetics of Materials, Wiley Interscience

5.8.1 SELF-DIFFUSION IN PURE METALS

The self-diffusion of a pure metal B can be studied by measuring the diffusive flux of a radioactive tracer isotope of B , denoted by B^* , which is chemically equivalent to B . Consider the diffusion couple in Figure 5.12.

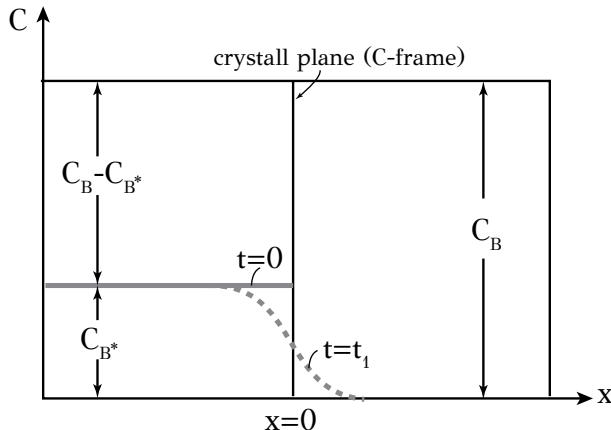


Figure 5.12: Diffusion couple for the study of self diffusion. B^* is a radioactive tracer isotope of B .

The right part of the couple is pure metal B while the left part is pure metal B containing a small quantity of tracer B^* in a homogeneous concentration. Self-diffusion takes place by the vacancy mechanism. B and B^* atoms jump in neighboring vacancies. During diffusion, B^* enters the right part of the couple. The crystal lattice remains stable during diffusion and consequently the crystal planes can serve as a reference frame for the measurement of the diffusive flux of B^* , i.e., the number of B^* atoms crossing the unit area of a crystal plane per unit time. This frame is called the C -frame. Fick's first law for tracer B in the C -frame is written as

$$J_{B^*} = -D_{B^*} \frac{\partial c_{B^*}}{\partial x}$$

where D_{B^*} is the self-diffusion coefficient of B . The diffusion coefficient is defined in the C -frame and is related to the mobility of B via the Relation (5.13)

$$D_B^* = M_B RT$$

The concentration profile of B^* after time t_1 corresponds to the dotted line in Figure 5.12.

5.8.2 SELF-DIFFUSION IN A HOMOGENEOUS SOLID SOLUTION

In a binary solid solution of metals A and B , which is homogeneous, there is no concentration gradient of either A or B . The chemical potentials of A and B are the

same in both sides of the couple. Since there is no chemical potential gradient, the driving force for diffusion is zero. Under these conditions only self-diffusion can take place. In this case the diffusion of B can be studied by the measurement of the diffusive flux of a radioactive tracer B^* , as in the previous case. The diffusion couple is depicted in Figure 5.13.

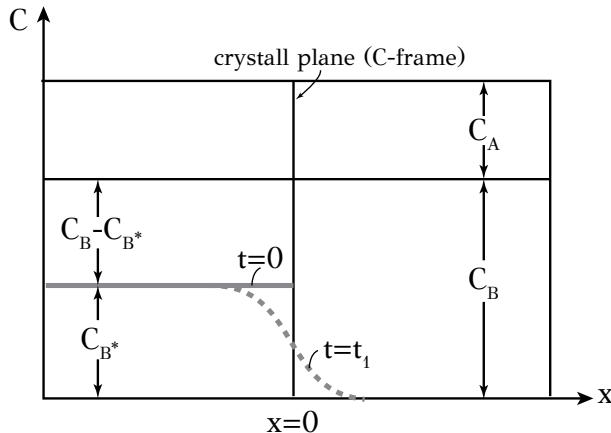


Figure 5.13: Diffusion couple for the study of substitutional diffusion in a homogeneous solid solution $A - B$. B^* is a radioactive tracer isotope of B .

In the left part there is a concentration c_{B^*} of B^* while in the right part there is initially no B^* . The concentration of B is homogeneous and equal to c_B in both parts of the couple. The diffusive flux of B^* in the C -frame is

$$J_{B^*} = -D_{B^*} \frac{\partial c_{B^*}}{\partial x}$$

where D_{B^*} is the diffusion coefficient of B in the homogeneous solid solution $A - B$. It is obvious that the diffusion coefficient of A is defined in a similar way. Both diffusion coefficients are related to the respective mobilities via the following equations

$$D_{A^*} = M_A R T$$

$$D_{B^*} = M_B R T$$

5.8.3 SUBSTITUTIONAL DIFFUSION IN A SOLID SOLUTION WITH CONCENTRATION GRADIENTS – INTERDIFFUSION

In a heterogeneous binary solid solution $A - B$ there are concentration gradients of both A and B . These concentration gradients establish respective chemical potential gradients, which drive the diffusion process. The diffusion in such a system is termed *interdiffusion*. If one of the components (say A) diffuses faster than the other (B), then regions initially rich in A will lose mass and shrink while regions initially rich in B

will gain this mass and expand. This mass transfer has been termed the *Kirkendall¹* effect and was firstly observed in the diffusion of Zn in brass ($Cu - Zn$). The diffusive fluxes of A and B in the C-frame are

$$J_A = -D_A \frac{\partial c_A}{\partial x}$$

$$J_B = -D_B \frac{\partial c_B}{\partial x}$$

The diffusion coefficients D_A and D_B are called *intrinsic diffusion coefficients* and are defined in the C-frame. The diffusive fluxes of A and B having opposite directions and magnitude result in a net flux of vacancies J_V given by the vector equation

$$\vec{J}_A + \vec{J}_B + \vec{J}_V = 0 \quad (5.17)$$

which is depicted in Figure 5.14. The diffusion couple consists of initially pure metals A and B. The concentration profile of the fast diffusing component (A) is shown after a certain diffusion time. Having a net diffusive flux of vacancies in the system means that vacancies are produced at some place and vacancies are vanished in another place. According to the Kirkendall effect, vacancies are vanished in the region initially rich in A (left side of the couple) while vacancies are produced in the region initially rich in B (right side of the couple). The mechanism of generation or destruction of vacancies is dislocation climb, which was discussed in Chapter 3. It is a non-conservative motion of dislocations perpendicular to the slip plane. This mechanism is depicted in Figure 5.14. In the case of vacancy destruction (left side of couple) the dislocation climbs upwards destroying the crystal plane. On the other hand, in the case of vacancy production (right side of couple) the dislocation climbs downwards forming a new crystal plane. In this way crystal planes are removed from the left side and added to the right side of the couple. Actually if one plane is marked with inert markers, which do not participate in the diffusion, the marked plane will move to the left. The velocity u of the marked plane (dotted line in Figure 5.14) is the velocity with which the crystal lattice is displaced by the diffusion process.

The velocity at which the lattice is displaced is required in order to express Fick's second law. Consider that the total number of atoms per unit volume is constant and equal to

$$c_o = c_A + c_B$$

and taking the gradients

$$\frac{\partial c_A}{\partial x} = -\frac{\partial c_B}{\partial x}$$

Using the Relation (5.17) the vacancy flux is

$$J_V = (D_A - D_B) \frac{\partial c_A}{\partial x}$$

¹A.D. Smigelskas and E.O. Kirkendall, Trans. AIME, 171 (1947) 130-142.

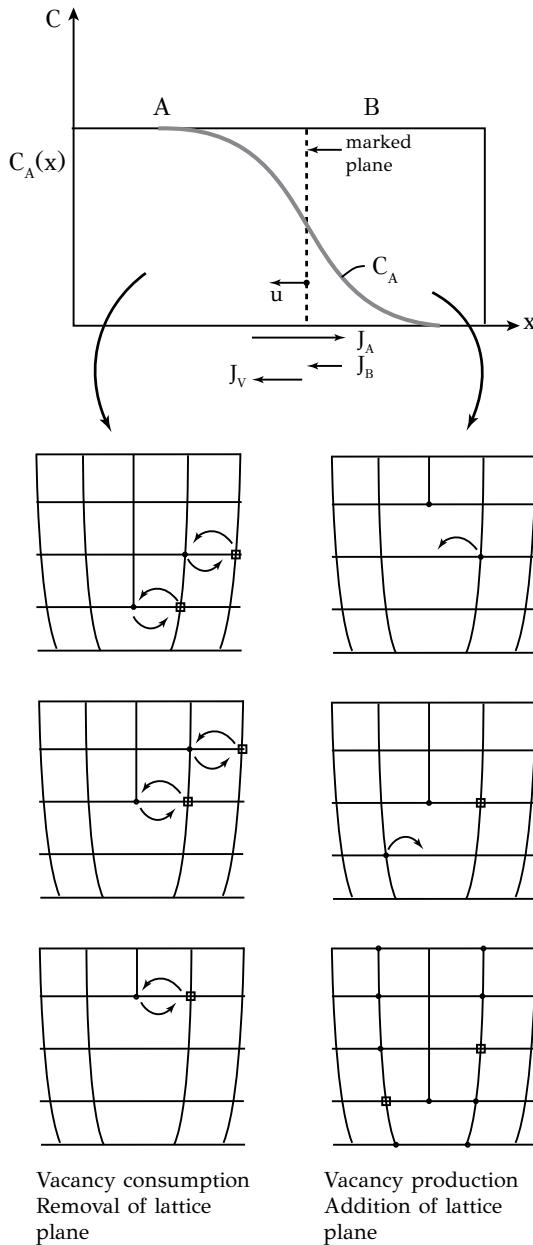


Figure 5.14: Diffusion couple for the study of substitutional diffusion in an inhomogeneous solid solution with concentration gradient. Removal and addition of crystal planes by dislocation climb on the left and right sides of the couple are indicated.

In a time δt the marked plane of area A is displaced by $u \cdot \delta t$, which corresponds to a volume $A \cdot u \cdot \delta t$. This volume contains $A \cdot u \cdot \delta t \cdot c_o$ atoms, which are replaced by $J_V \cdot A \cdot \delta t$ vacancies. Then

$$J_V = c_o u$$

and solving for the velocity u

$$u = (D_A - D_B) \frac{\partial X_A}{\partial x} \quad (5.18)$$

where it was considered that

$$X_A = \frac{c_A}{c_o}$$

One consequence of the displacement of crystal planes is that the C -frame cannot be used to determine the concentration profiles of A and B via Fick's second law. Since the volume of the couple remains constant, the Volume-fixed frame or V -frame¹ will be used, where the diffusive fluxes of A and B refer to a constant plane, such as the left end of the diffusion couple, which is fixed with respect to the laboratory. As in section 5.5, the principle for conservation of mass will guide us to Fick's second law. Consider a section δx in Figure 5.15.

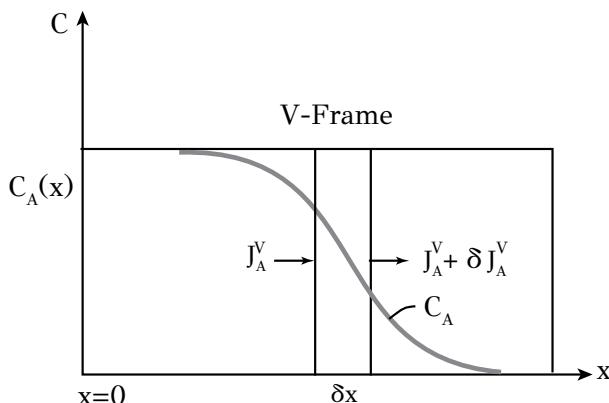


Figure 5.15: The V -frame used for the derivation of Fick's second law for interdiffusion in an inhomogeneous solid solution.

The flux of A entering the section δx consists of two contributions: the diffusive flux according to Fick's first law expressed in the C -frame and the flux $u \cdot c_A$ arising from the displacement of crystal planes with velocity u . This flux is

$$J_A^V = -D_A \frac{\partial c_A}{\partial x} + uc_A \quad (5.19)$$

¹R.W. Balluffi, S.M. Allen & W.C. Carter, (2005). Kinetics of Materials. Wiley Interscience

where the suffix V indicates that the flux refers to the V -frame. Using the Relation (5.18) the flux becomes

$$J_A^V = -(X_B D_A + X_A D_B) \frac{\partial c_A}{\partial x}$$

The above relation is Fick's first law in the V -frame. The diffusion coefficient

$$\tilde{D} = X_B D_A + X_A D_B \quad (5.20)$$

is termed *interdiffusion coefficient* and is defined in the V -frame. D_A and D_B are the intrinsic diffusion coefficients defined in the C -frame. Equation (5.20) is termed the *Darken equation*. Returning to Figure 5.15, the flux of A atoms entering section δx is provided by Equation (5.19). This influx causes a change in the concentration of A atoms equal to

$$\frac{\partial c_A}{\partial t} = -\frac{\partial J_A^V}{\partial x}$$

which becomes

$$\frac{\partial c_A}{\partial t} = \frac{\partial}{\partial x} \left(\tilde{D} \frac{\partial c_A}{\partial x} \right) \quad (5.21)$$

which is Fick's second law expressed in the V -frame. Interdiffusion can then be described in the V -frame with a single diffusion coefficient \tilde{D} , which is linked to the intrinsic diffusion coefficients of the components through the Darken equation.

Consider now the coupling to thermodynamics. In the case of diffusion in a homogeneous solid solution, the diffusion coefficients were defined as the self-diffusion coefficients D_A^* and D_B^* . In the case of diffusion in an inhomogeneous solid solution with concentration gradients of the two components A and B , the intrinsic diffusion coefficients D_A and D_B are related to the respective D_A^* and D_B^* via the thermodynamic factor

$$D_A = D_A^* \left(1 + \frac{\partial \ln \gamma_A}{\partial \ln X_A} \right)$$

$$D_B = D_B^* \left(1 + \frac{\partial \ln \gamma_B}{\partial \ln X_B} \right)$$

or using (5.16)

$$D_A = D_A^* \left(1 - \frac{2\Omega}{RT} X_A X_B \right) \quad (5.22)$$

$$D_B = D_B^* \left(1 - \frac{2\Omega}{RT} X_A X_B \right) \quad (5.23)$$

The intrinsic diffusion coefficients depend on composition and the excess term (non-ideality) of the solid solution. When $\Omega = 0$ the solution is ideal and the intrinsic diffusion coefficients are equal to the respective self-diffusion coefficients for a homogeneous solution and are independent of composition. When $\Omega < 0$ ($\Delta H_m < 0$) the Relations (5.22) and (5.23) indicate that $D_A > D_A^*$ and $D_B > D_B^*$ and diffusion

in an inhomogeneous solution is accelerated relative to the diffusion in a homogeneous solid solution. The opposite holds for the case $\Omega > 0$. The non-ideality of the inhomogeneous solid solution imposes, through the thermodynamic factor, the concentration dependence of the diffusion coefficients as well as an acceleration or deceleration of diffusion with respect to the homogeneous solid solution.

5.8.4 REVIEW OF DIFFUSION COEFFICIENTS

The diffusion coefficients for substitutional diffusion defined in the previous sections are summarized in Table 5.1.

- The coefficient D^* refers to self-diffusion in a pure metal, which takes place by the vacancy mechanism.
- The coefficients D_A^* and D_B^* refer to substitutional diffusion in a homogeneous solid solution, in the absence of any concentration gradients of A and B. These are actually self-diffusion coefficients independent of composition and are defined in the C-frame, which in this case coincides with the V-frame.
- The intrinsic diffusion coefficients D_A and D_B refer to the diffusion of A and B in an inhomogeneous solid solution in the presence of concentration gradients of A and B. They link the diffusive flux with the respective concentration gradients via Fick's first law. They are defined in the C-frame, which moves at a velocity u relative to the V-frame. They are linked to the respective self-diffusion coefficients D_A^* and D_B^* via the thermodynamic factor and depend on composition.
- The interdiffusion coefficient \bar{D} is a unique coefficient describing interdiffusion in an inhomogeneous solid solution in the presence of concentration gradients of A and B. It is defined in the V-frame. Fick's first and second laws are also defined in the V-frame. The interdiffusion coefficient depends on composition through the thermodynamic factor.

5.9 IRREVERSIBLE THERMODYNAMICS AND DIFFUSION

A system at thermodynamic equilibrium is characterized by zero gradients of intensive thermodynamic quantities or potentials, such as temperature, chemical or electrical potential. When the system is away from equilibrium, the potential gradients acquire non-zero values and operate as “forces”, which in turn cause “fluxes” of extensive quantities, such as heat, mass or electricity. As the system moves towards equilibrium, the free energy is dissipated and entropy is produced. *Irreversible thermodynamics* is based on the concept of entropy production during an irreversible process, which leads the system towards equilibrium. If the entropy production rate is denoted by $\dot{\sigma}$ (entropy per unit volume and the unit time), then the product $T\dot{\sigma}$, in units of $J/m^3 s$, expresses the free energy dissipation rate.

Table 5.1**Diffusion coefficients for substitutional diffusion**

Diffusion Coefficient	Symbol	Reference Frame	Relations
Self-diffusion in pure metals Self-diffusion Coefficient	D^*	C-Frame	$D^* = MRT$
Diffusion in homogeneous solid solution Self-diffusion Coefficients	D_A^*, D_B^*	C-Frame	$D_A^* = M_A RT$ $D_B^* = M_B RT$
Diffusion in inhomogeneous solid solution Intrinsic diffusion coefficients	D_A, D_B	C-Frame	$D_A = D_A^* \left(1 + \frac{\partial \ln \gamma_A}{\partial \ln X_A} \right)$ $D_B = D_B^* \left(1 + \frac{\partial \ln \gamma_B}{\partial \ln X_B} \right)$
Interdiffusion in inhomogeneous solid solution Interdiffusion coefficient	\tilde{D}	V-Frame	$\tilde{D} = X_B D_A + X_A D_B$ $J_A^V = -\tilde{D} \frac{\partial c_A}{\partial x}$ $\frac{\partial c_A}{\partial t} = \frac{\partial}{\partial x} \left(\tilde{D} \frac{\partial c_A}{\partial x} \right)$

The first postulate of irreversible thermodynamics is that the rate of entropy production is non-negative

$$\dot{\sigma} \geq 0$$

and that the product $T\dot{\sigma}$ can be written as

$$T\dot{\sigma} = - \sum_i J_i F_i \quad (5.24)$$

where the *rhs* is a sum of products between “fluxes” J_i and conjugate “forces” F_i . The forces can be expressed as potential gradients. If the potential gradients are denoted by ξ_i then equation (5.24) can be written as

$$T\dot{\sigma} = - \sum_i J_i \nabla \xi_i$$

As an example of the above equation consider the existence of a chemical potential gradient. With the help of Equation (5.10) the above relation can be written as

$$T\dot{\sigma} = \sum_i M_i c_i |\nabla \mu_i|^2$$

The second postulate of irreversible thermodynamics is that “near equilibrium” the fluxes are linearly related to the conjugate forces. This can be shown by considering the flux J as a function of force F and write the Taylor series expansion around the

equilibrium position ($F = 0$)

$$J(F) = J(0) + J'(0)\frac{F}{1!} + J''(0)\frac{F^2}{2!} + \dots$$

The flux J for $F = 0$ is zero, i.e., $J(0) = 0$ at equilibrium. Neglecting the 2nd order terms and above we get

$$J(F) = J'(0)F$$

from which it is apparent that the flux is linearly related to the conjugate force. The above relation is valid around the position $F = 0$, i.e., near equilibrium. Generalizing, it can be said that the flux J is linearly related to all conjugate forces (potential gradients) acting in the system, then

$$J_i = \sum_j L_{ij} F_j \quad (5.25)$$

The two postulates of irreversible thermodynamics expressed by (5.24) and (5.25) were developed by Onsager. The coefficients L_{ij} are referred to as the Onsager coefficients. The diagonal terms L_{ii} are called direct coefficients and they couple each flux to its conjugate force. The off-diagonal terms are called coupling coefficients and express additional coupling effects. The Onsager symmetry principle states that

$$L_{ij} = L_{ji} \quad (5.26)$$

Irreversible thermodynamics will be applied in the study of multicomponent diffusion, i.e., diffusion in alloy systems with many components. It will also be applied in the study of diffusion from non-chemical forces such as stress-induced diffusion and curvature-induced diffusion.

5.9.1 DIFFUSION IN MULTICOMPONENT SYSTEMS

Diffusion in multicomponent systems had been studied experimentally long before its theoretical foundation by Onsager. The most classic example is the Darken's experiment in 1949. Based on the Darken's experiment, the complex interactions between the alloying elements during multicomponent diffusion will be discussed. Consider the simultaneous diffusion of carbon and silicon in iron, in a $Fe - C - Si$ system. The diffusional flux of the two components C and Si is caused by two "forces", the chemical potential gradients of C and Si . According to (5.25) the diffusional fluxes in the x direction are

$$J_C = L_{CC} \frac{\partial \mu_C}{\partial x} + L_{CSi} \frac{\partial \mu_{Si}}{\partial x} \quad (5.27)$$

$$J_{Si} = L_{SiC} \frac{\partial \mu_C}{\partial x} + L_{SiSi} \frac{\partial \mu_{Si}}{\partial x} \quad (5.28)$$

In each of the above relations, the diffusive flux (e.g., carbon) has two contributions: the direct term, which is the chemical potential gradient of carbon with the

direct Onsager coefficient L_{CC} and the off-diagonal term, the chemical potential gradient of Si , with the coupling Onsager coefficient L_{CSi} . This means that even in the case where the chemical potential gradient of carbon vanishes, the diffusion of carbon will continue until the chemical potential gradient of Si also vanishes. According to the Onsager symmetry principle the coupling coefficients are equal

$$L_{CSi} = L_{SiC}$$

i.e., the influence of the chemical potential gradient of Si on carbon diffusion is the same with the influence of the chemical potential gradient of C on Silicon diffusion. Consider now a diffusion couple with the left side consisting of a $Fe - 0.4C - 4Si$ (wt%) alloy and the right side consisting of a $Fe - 0.4C$ (wt%) alloy. Both sides have equal length of 25mm. Initially both sides of the couple have the same carbon concentration. The left side has in addition 4% Si while the right side has no Si . The couple is now heated to 1050°C. The initial concentration profiles as well as the concentration and the chemical potential profiles for various diffusion times are shown in Figure 5.16. Consider two time periods. The first period is from $t = 0$ till $t = 10^8$ sec, where no Si diffusion is observed from the left to the right side. The second period is above 10^8 sec where diffusion of Si from the left to the right side takes place. In the first period, despite the fact that there is no carbon concentration gradient, diffusion of carbon takes place from left to right. This diffusion is actually an uphill diffusion and the driving force is primarily the gradient of the chemical potential of carbon, i.e., the direct term in (5.27). There is also a contribution of the coupling term, i.e., the second term in (5.27). It is worth noting that the presence of Si on the left side has raised the chemical potential of carbon on the left side. The difference of the carbon potential is shown in Figure 5.15c. The diffusion of carbon continues until the chemical potential gradient vanishes and carbon has the same chemical potential in both sides of the couple, at time $t = 10^8$ sec, the end of the first period. One should expect no further carbon diffusion, however carbon diffusion continues under the action of the coupling term, the gradient of the chemical potential of Si , $\partial\mu_{Si}/\partial x$, of Equation (5.27). Carbon diffusion starts exactly at the same time Si diffusion from left to right starts, as shown in Figure 5.16b. The Si influx in the right side causes the diffusion of carbon from the right to the left side. While in the first period (up to 10^8 sec) carbon diffuses from left to right under the action of $\partial\mu_C/\partial x$, in the second period (above 10^8 sec) the direction of carbon diffusion is reversed and carbon diffuses from right to left under the action of $\partial\mu_{Si}/\partial x$, which develops due to the diffusion of Si . The example discussed above leads to some important observations:

- Carbon diffusion takes place even when the initial concentration gradient is zero because the driving force for the diffusion of carbon is primarily the chemical potential gradient $\partial\mu_C/\partial x$ and also $\partial\mu_{Si}/\partial x$.
- When the chemical potential gradient of carbon vanishes, diffusion of carbon does not stop. It continues under the action of the coupling term $\partial\mu_{Si}/\partial x$, the chemical potential gradient of Si . This is a consequence of irreversible thermodynamics and the Onsager principles, Equation (5.25).

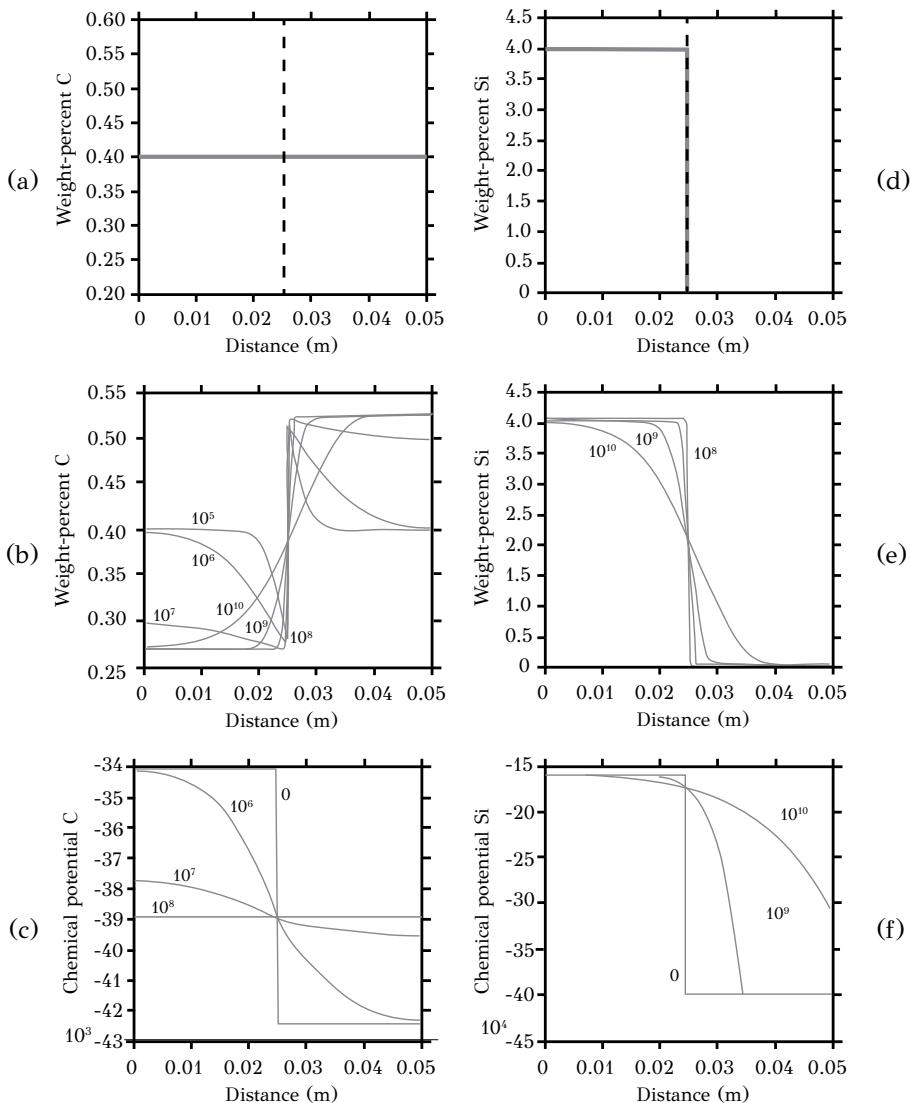


Figure 5.16: Diffusion couple ($Fe - 0.4C - 4Si/Fe - 0.4C$) for the study of multi-component diffusion: (a,d) initial concentration profiles of C and Si. (b,e) Transient concentration profiles of C and Si for various diffusion times. (c, f) Chemical potential profiles of carbon and Si for various diffusion times.

- The diffusion of carbon and silicon are influenced not only by the respective driving forces but by the respective mobilities as well. Carbon is an interstitial alloying element and its diffusion in the iron lattice is much faster than the diffusion of silicon, which is a substitutional element.
- Silicon influences the diffusion of carbon in two ways. The first is the effect on the direct term of (5.27) since it raises the chemical potential of carbon. The second is through the coupling term of (5.27).
- In addition to the effect on the driving forces, the alloying elements influence the interdiffusion coefficient through the thermodynamic factor, as discussed in section 5.8.

The example discussed above illustrates the complex interactions during multi-component diffusion. The treatment of such non-linear diffusion problems will be discussed in section 5.12.

5.9.2 DIFFUSION BY NON-CHEMICAL FORCES: CURVATURE AND STRESS

As discussed above, the diffusion of a component is caused by the gradient of the chemical potential of the component and additionally by the gradients of the chemical potentials of other components in the alloy. Besides gradients in chemical potential, diffusion can be driven by other “non-chemical forces”. The most important are interfacial curvature and mechanical stress.

Curvature effects. The microstructure of metallic alloys consists of several phases separated by interfaces. These interfaces are usually curved and are characterized by the interfacial energy γ . Curvature raises the free energy of a curved relative to a flat interface by the quantity

$$\Delta G = \frac{2\gamma V_m}{r} \quad (5.29)$$

which is known as the Gibbs-Thomson effect. This causes the development of an additional force, according to (5.25), which contributes to diffusion. A simple example is shown in Figure 5.17, which indicates an interface in a pure metal A exhibiting a convexity of radius r .

The local curvature increases the chemical potential of A directly below the convexity relative to the flat surface. The diffusion of A will take place by the vacancy mechanism from the convex region to the flat region and result in *surface smoothing* and a reduction of interfacial energy. Another important application of curvature-induced diffusion is in *coarsening* of a particle dispersion. When the particles of a dispersion have unequal sizes, the differences in curvature generates diffusive fluxes from the small to the large particles. The mean particle size increases leading to a degradation of mechanical strength. Coarsening will be considered in detail during the study of phase transformations in Chapter 6.

Stress effects. The influence of mechanical stresses on diffusion is more complicated. Consider the diffusion of carbon in iron. Under the absence of any stress, the diffusive flux is

$$J_i = -M_i c_i \nabla \mu_i$$

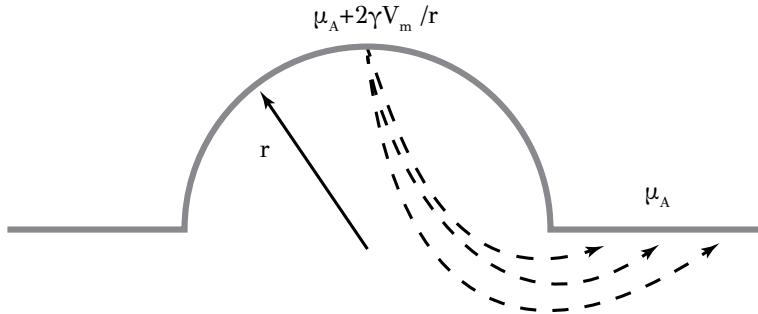


Figure 5.17: Diffusion induced by interfacial curvature.

Stresses affect diffusion in two ways. The first is the development of a coupling term in Equation (5.25). The second is the effect on mobility M . If the stress field is uniform, i.e., there are no stress gradients, the only driving force is the chemical potential gradient. In this case the effect of stress is limited on mobility and the diffusion coefficient. As discussed in section 5.6, carbon diffusion takes place by a series of jumps in vacant interstitial sites. The process is thermally activated and the temperature dependence of the diffusion coefficient is

$$D = D_o \exp\left(-\frac{\Delta H_D^*}{kT}\right) \quad (5.30)$$

The diffusive jump under the action of an applied uniform tensile or compressive stress in the y direction is depicted in Figure 5.18. The tensile stress stretches the bonds in the y direction and consequently reduces the energy barrier to diffusion. On the other hand the compressive stress increases the energy barrier. Therefore a tensile stress in the y direction, increases the diffusion coefficient in the x direction. The opposite is valid for the compressive stress. The effect of stress can be expressed mathematically by adding a term W in the activation energy of Equation (5.30)

$$D = D_o \exp\left(-\frac{\Delta H_D^* + W}{kT}\right)$$

where W is the work performed by the applied stress σ . In most cases the term W/kT is small and the above relation can be written as

$$D = D_o \left(1 - \frac{W}{kT}\right) \exp\left(-\frac{\Delta H_D^*}{kT}\right)$$

Since the work is a linear function of applied stress it follows that the diffusion coefficient will also be a linear function of stress. It is worth noting however that the application of a uniform stress changes the symmetry of the crystal. For example the application of stress in the y direction raises the diffusion coefficient in the x and z directions. The diffusion will be accelerated in these directions due to bond stretching in the y direction. Due to the Poisson effect ($\epsilon_y = -\epsilon_x$), diffusion in the

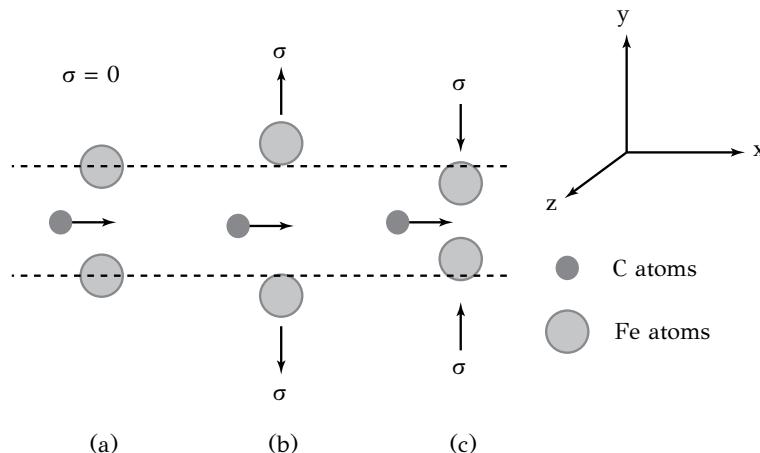


Figure 5.18: Diffusion induced by a homogeneous stress: (a) carbon atom passing through two iron atoms under no stress, (b) under tensile stress applied in the y direction, (c) under compressive stress in the y direction.

y direction will be retarded. As a result the application of a uniform stress makes diffusion anisotropic. Consider now the effect of applied stress on the driving force for diffusion. Take the example of carbon in iron with a uniform concentration c_o . Each carbon atom at an interstitial site operates as a dilatation center, since the lattice experiences a local dilatation by ΔV . This dilatation interacts with the hydrostatic pressure p of the stress field. When the stress field is uniform, then $\Delta p = 0$ and no excess driving force is developed. Consider now a non-uniform stress field with a finite gradient of hydrostatic pressure. A good example is the stress field around an edge dislocation (Figure 5.19), which was discussed in Chapter 3.

The hydrostatic component is

$$p = \frac{\mu b}{3\pi} \frac{1+\nu}{1-\nu} \frac{y}{x^2 + y^2}$$

where μ in this case is the shear modulus (not to be confused with the chemical potential) and b the Burgers vector. The hydrostatic pressure has a different sign above ($y > 0$) and below ($y < 0$) the slip plane of the dislocation. Above the slip plane hydrostatic compression dominates, while below the slip plane hydrostatic tension prevails due to the extra half plane of the dislocation. The gradient ∇p causes the development of a diffusive flux of carbon from the compressive to the tensile region of the dislocation, as depicted in Figure 5.19. In this way the compressive region is depleted of carbon with $c < c_o$ while the tensile region is enriched in carbon with $c > c_o$. The differences in carbon concentration between the two regions cause the development of a chemical potential gradient $\nabla \mu_C$, which acts against the pressure

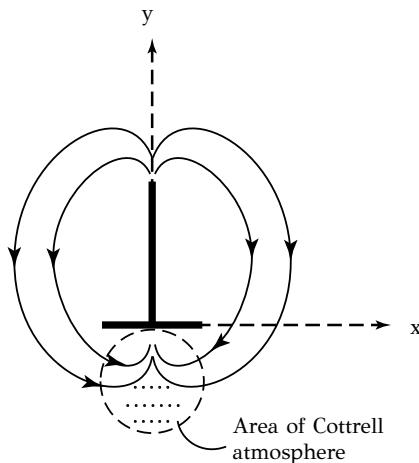


Figure 5.19: Diffusion of atoms induced by the inhomogeneous stress field around an edge dislocation for the formation of Cottrell atmosphere.

gradient ∇p until equilibrium is established according to the relation

$$J_C = L_{CC}(\nabla \mu_C + \nabla p \Delta V)$$

In this case an atmosphere of carbon atoms is established below the slip plane of the dislocation, termed the *Cottrell atmosphere*, which is responsible for the appearance of yield point effects and strain aging in steels. These effects will be discussed in detail in Chapter 8.

Another classic example of how stress affects diffusion is *diffusional flow*, which is a form of high-temperature creep of metals. In this case creep strains develop exclusively by diffusional flow of atoms driven by gradients in hydrostatic pressure at the scale of a grain. This effect leads to the grain size dependence of creep strain rate, a subject to be discussed in detail in Chapter 9.

5.10 EFFECTS OF DIFFUSION

Diffusion has important effects in metallurgical processes:

Diffusive phase transformations. These transformations take place by atomic diffusion, as will be discussed in Chapter 6. For example, the formation of pearlite in steels during the slow cooling of austenite takes place by the diffusion of carbon and other alloying elements. If the cooling of the austenite is fast enough, so that carbon is not given the required time to diffuse, then martensite forms. Martensitic transformation is then diffusionless. Another example of a diffusive transformation is the precipitation of the intermetallic compound Al_2Cu in $Al - Cu$ alloys, which takes place by the diffusion of Cu in Al . The nucleation, growth and coarsening of phases during precipitation take place by diffusion.

Segregation and homogenization. Diffusion plays an important role on the development of microstructure in cast alloys. During solidification alloying elements are rejected from the liquid to the solid phase. Due to limited diffusion in the solid, regions solidified first are poor in alloying elements while regions solidified last are rich in alloying elements, resulting in *segregation*. Segregation is undesirable since it degrades the mechanical properties and workability of the cast material. Segregation can be removed if a heat treatment is applied after casting for the *homogenization* of the cast structure. Homogenization takes place by the diffusion of alloying elements.

Creep. Diffusion plays an important role during the deformation of metals at high temperatures, i.e., during creep. It was shown (Chapter 3) that the plastic deformation of metals takes place by dislocation glide. At high temperatures dislocations acquire an extra degree of freedom and in addition to glide they can climb. Climb provides the dislocations the opportunity to change slip plane and continue their glide, if for some reason glide is impeded in the current slip plane. Climb takes place by diffusion of atoms towards or away from the dislocation. This creep mechanism is termed *dislocation creep*. At even higher temperatures, deformation takes place exclusively by diffusion. This creep deformation is termed *diffusional flow* and takes place by diffusion in the grain from compressive to tensile regions. The resulting mass transfer changes the shape of the grains and causes *grain boundary sliding*, a mechanism to be discussed in Chapter 9.

Heat treatment of steel. Diffusion is important in many heat and surface treatments of steels. For example carburizing and nitriding are accomplished by the diffusion of carbon or nitrogen for the surface enrichment of steels in order to increase surface hardness and wear resistance. Another heat treatment is the tempering of quenched steel. Martensite, which forms during quenching, is a supersaturated solid solution of carbon in iron. Martensite is hard but at the same time it is brittle. The diffusion of carbon out of solution during tempering leads to a reduction in hardness and an increase in ductility. The role of diffusion in the heat treatment of steel will be discussed in Chapter 10.

5.11 ANALYTICAL SOLUTIONS TO THE DIFFUSION EQUATION

Analytical solutions to the diffusion equation will be discussed in the following paragraphs. The diffusion equation (Fick's 2nd law) in one direction is

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial c}{\partial x} \right) \quad (5.31)$$

which for a composition-independent diffusion coefficient becomes

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad (5.32)$$

which in turn is a linear partial differential equation (*PDE*). The solution of the *PDE*, under certain initial and boundary conditions, provides the concentration profiles $c(x,t)$ of a component for one-dimensional diffusion in the x direction. In the

case where the composition dependence of the diffusion coefficient cannot be neglected, the diffusion Equation (5.31) is non-linear and can be solved by numerical methods, discussed in section 5.12. The solutions presented here are valid for interstitial diffusion as well as substitutional diffusion. In the last case the interdiffusion coefficient should be used.

5.11.1 TYPES OF DIFFUSION PROBLEMS

In most diffusion problems, the solution of the second order linear *PDE* (5.32) is sought under certain initial and boundary conditions. We distinguish between two types of solutions. The first type refers to solutions, which are valid at short times or to semi-infinite media, where the dimensions are much larger than the net diffusion distance. These solutions involve the error function. The second type refers to solutions, which are valid at long times or to finite media and involve Fourier series. In cylindrical coordinates the Fourier series are substituted by a series of Bessel functions. In order to solve the diffusion equation, several methods are applied, amongst them the method of separation of variables and the Laplace transform. Solutions to simple one-dimensional diffusion problems will be discussed in the next section. For a detailed treatment of the subject, the reader is directed to the suggested reading list at the end of this chapter, especially Krantz's book.

5.11.2 SHORT DIFFUSION TIMES OR SEMI-INFINITE MEDIA

Two cases will be considered: instantaneous sources and constant surface concentration.

Instantaneous source. Consider the concentration profile, which develops from an instantaneous source, i.e., the delivery of a specific quantity of component at time $t = 0$. Consider a planar source at $x = 0$ with quantity S (atoms/m^2) of radioactive tracer A^* as depicted in Figure 5.20a.

The general solution of (5.32) is

$$c(x,t) = \frac{B}{\sqrt{t}} \exp\left(-\frac{x^2}{4Dt}\right)$$

where B is a constant. The boundary conditions are

$$c(\infty, t) = 0$$

$$c(-\infty, t) = 0$$

An additional condition arises from the principle of mass conservation

$$\int_{-\infty}^{\infty} c dx = S$$

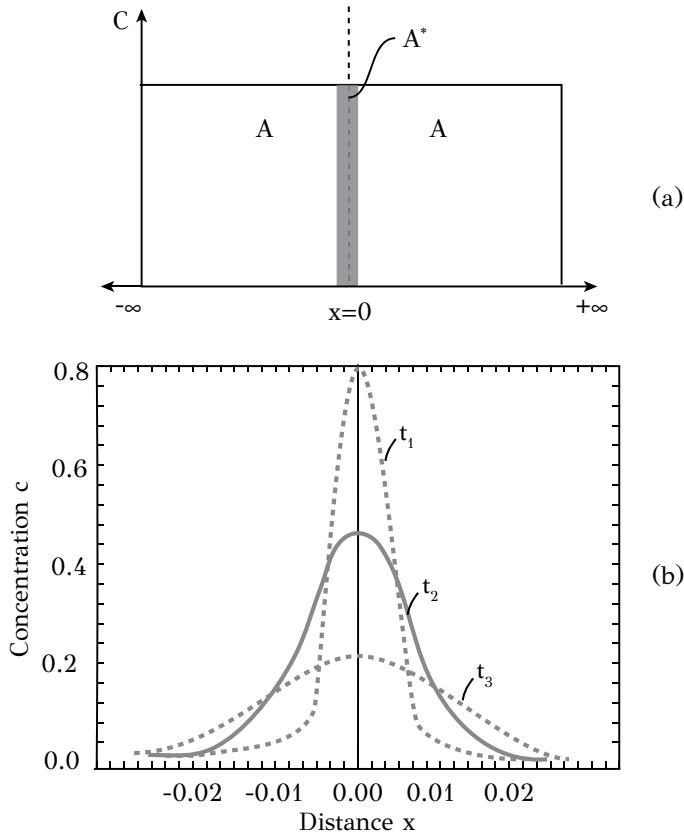


Figure 5.20: (a) Instantaneous source of radioactive tracer isotope A^* in an infinite medium, (b) resulting concentration profiles of A^* at several diffusion times.

Using the above conditions, the constant B is identified and the general solution becomes

$$c(x,t) = \frac{S}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) \quad (5.33)$$

For $x = 0$ the above relation gives

$$c(0,t) = \frac{S}{2\sqrt{\pi Dt}}$$

and (5.33) becomes

$$\frac{c(x,t)}{c(0,t)} = \exp\left(-\frac{x^2}{4Dt}\right) \quad (5.34)$$

Solution (5.34) is shown in Figure 5.20b. The above relation can be used to determine experimentally the diffusion coefficient D of a component from its concentration profile $c(x,t)$.

Constant surface concentration. One of the most common diffusion problems is to determine the concentration profile in a semi-infinite medium with a constant surface concentration. Consider for example the carburizing of steel specimen of length L , in Figure 5.21.

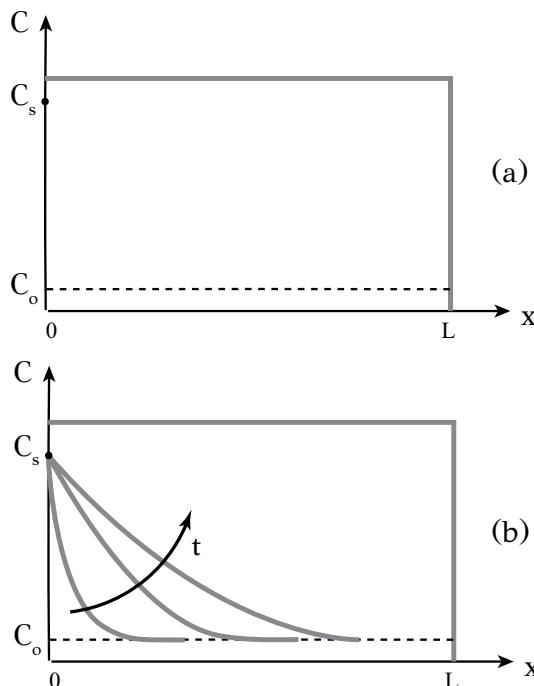


Figure 5.21: (a) Diffusion in a semi-infinite medium with a constant surface concentration, (b) resulting concentration profiles from the error function solution.

The specimen has an initial carbon content c_o and constant surface composition in the left end equal to c_s . Assuming that the diffusion coefficient is independent of composition, the diffusion equation is

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2}$$

The initial condition is

$$c(x, 0) = c_o$$

and the boundary conditions are

$$c(0, t) = c_s$$

$$c(\infty, t) = c_o$$

The general solution is

$$c(x, t) = A + B \operatorname{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (5.35)$$

where A and B are constants and $\operatorname{erf}(z)$ is the error function. Before finding the constants A and B , we should say a few words regarding the error function, which is defined as

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-\eta^2) d\eta$$

Its graph is shown in Figure 5.22 from which it is depicted that the error function is symmetric and asymptotic at -1 and 1 for $z = -\infty$ and $z = +\infty$ respectively.

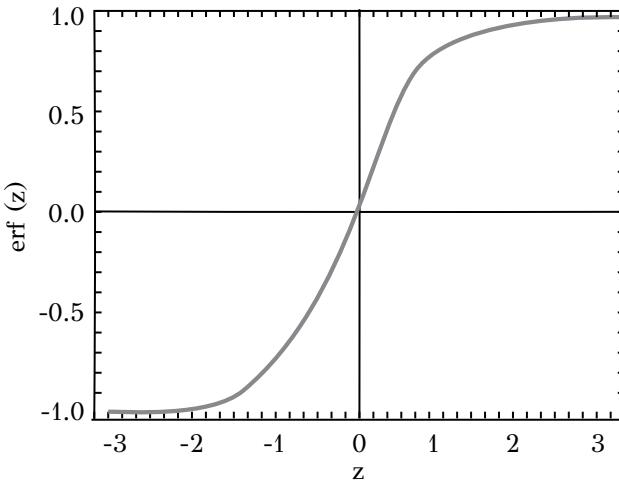


Figure 5.22: Graph of the error function $\operatorname{erf}(z)$.

Values of the error function are given in Table 5.2. It can be observed that up to $z = 0.6$, $\operatorname{erf}(z) \cong z$. It is also worth noting that $\operatorname{erf}(0) = 0$, $\operatorname{erf}(+\infty) = 1$ and $\operatorname{erf}(-\infty) = -1$.

Returning to the example of carburizing under constant surface composition, the constants A and B can be found from the boundary conditions

$$c(0, t) = c_s = A + B \operatorname{erf}(0) = A$$

$$c(\infty, t) = c_o = A + B \operatorname{erf}(\infty) = A + B$$

from which we get $A = c_s$ and $B = c_o - c_s$ and the solution (5.35) becomes

$$c(x, t) = c_s + (c_o - c_s) \operatorname{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (5.36)$$

Table 5.2
Values of the error function.

z	$\text{erf}(z)$	z	$\text{erf}(z)$
0	0	0.85	0.7707
0.025	0.0282	0.90	0.7970
0.05	0.0564	0.95	0.8209
0.10	0.1125	1.0	0.8427
0.15	0.1680	1.1	0.8802
0.20	0.2227	1.2	0.9103
0.25	0.2763	1.3	0.9340
0.30	0.3286	1.4	0.9523
0.35	0.3794	1.5	0.9661
0.40	0.4234	1.6	0.9763
0.45	0.4755	1.7	0.9838
0.50	0.5205	1.8	0.9891
0.55	0.5633	1.9	0.9928
0.60	0.6039	2.0	0.9953
0.65	0.6420	2.2	0.9981
0.70	0.6778	2.4	0.9993
0.75	0.7112	2.6	0.9998
0.80	0.7421	2.8	0.9999

In the special case of zero initial concentration of carbon, $c_o = 0$, the solution becomes

$$c(x, t) = c_s \left[1 - \text{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \right] \quad (5.37)$$

which can be written as

$$c(x, t) = c_s \text{erfc} \left(\frac{x}{2\sqrt{Dt}} \right)$$

where $\text{erfc}(z) = 1 - \text{erf}(z)$ is the complementary error function.

In order to obtain the above solutions, it was assumed that the specimen is semi-infinite. We should therefore check if the specimen length L is large enough for the general solution (5.35) to be valid. When the concentration profile reaches the right end of the specimen, there will be a leakage of carbon from the right end. We will require that this leakage is small enough, say 1/1000 of the total quantity of carbon in the specimen. This procedure is called “virtual leak test” and is expressed by

$$\frac{\int_L^\infty c(x, t) dx}{\int_0^\infty c(x, t) dx} = 10^{-3}$$

from which we get that $\text{erfc}(\ell/2\sqrt{Dt}) = 0.001$ and $L = 4.7\sqrt{Dt}$. This means that

the solution (5.36) can be used for specimen lengths longer than $4.7\sqrt{Dt}$. It is worth noting that the length L , required for the validity of (5.36), depends on time. Thus for a given length L , solution (5.36) can be used for short times.

5.11.3 LONG DIFFUSION TIMES OR FINITE MEDIA

For the treatment of diffusion problems concerning long diffusion times or finite media (t long, L short), the general solution (5.35) cannot be used. In this case the diffusion equation can be solved by the method of separation of variables. As an example we will study the diffusion of a component from a thin sheet with thickness L and initial composition c_o . The composition at the two sides of the sheet is zero. An application of such a case is the hydrogen outgassing during post weld heat treatment of high-strength steels. Hydrogen in these steels causes hydrogen embrittlement and, therefore, the steel is heated to remove hydrogen by diffusion from the alloy to the environment. The initial condition is

$$c(x, 0) = c_o$$

while the boundary conditions are

$$c(0, t) = 0$$

$$c(L, t) = 0$$

The solution can be written in the form

$$c(x, t) = X(x)T(t)$$

and the diffusion equation becomes

$$X(x)T'(t) = DX''(x)T(t)$$

Divide both sides by $X(x)T(t)$, then

$$\frac{T'}{DT} = \frac{X''}{X} \quad (5.38)$$

Since the above relation should be valid for all values of x and t , (5.38) should be equal to a constant. In order to get real solutions the constant is set to $-k^2$. Two differential equations are obtained from (5.38)

$$X'' + \kappa^2 X = 0 \quad (5.39)$$

$$T' + \kappa^2 DT = 0 \quad (5.40)$$

The solution of (5.39) is

$$X(x) = A \cos \kappa x + B \sin \kappa x$$

Setting $X(0) = X(L) = 0$ we get $A = 0$ and $B \sin \kappa L = 0$. For $\sin \kappa L = 0$ we get $\kappa = n\pi/L$ where $n = 1, 2, \dots$ then the above relation becomes

$$X(x) = B \sin \frac{n\pi x}{L} \quad (5.41)$$

The solution of (5.40) is

$$T(t) = C \exp(-\kappa^2 D t) = C \exp\left(-\frac{n^2 \pi^2 D t}{L^2}\right) \quad (5.42)$$

Combining the solutions (5.41) and (5.42) we get

$$c(x, t) = \sum_{n=1}^{\infty} B_n \exp\left(-\frac{n^2 \pi^2 D t}{L^2}\right) \sin \frac{n\pi x}{L} \quad (5.43)$$

In order to obtain B_n we use the initial condition

$$c(x, 0) = c_o = \sum_{n=1}^{\infty} B_n \sin \frac{n\pi x}{L}$$

from which we obtain

$$B_n = \frac{2c_o}{L} \int_0^L \sin \frac{n\pi s}{L} ds = -\frac{2c_o}{n\pi} \cos n\pi$$

The above relation gives $B_n = 0$ for even values of n . For odd values ($n = 2j + 1$, $j = 0, 1, 2, \dots$) the Relation (5.43) becomes

$$c(x, t) = \frac{4c_o}{\pi} \sum_{j=0}^{\infty} \frac{1}{2j+1} \sin \frac{(2j+1)\pi x}{L} \exp\left[-\left(\frac{(2j+1)\pi}{L}\right)^2 D t\right] \quad (5.44)$$

which converges for long times. The solution (5.44) can be used to find the average concentration $c(t)$ in the sheet, which is defined by

$$\bar{c}(t) = \frac{1}{L} \int_0^L c(x, t) dx$$

Integrating (5.44), the average concentration is

$$\bar{c}(t) = \frac{8c_o}{\pi^2} \sum_{j=0}^{\infty} \frac{1}{(2j+1)^2} \exp\left[-\left(\frac{(2j+1)\pi}{L}\right)^2 D t\right] \quad (5.45)$$

which is a function of time.

5.11.4 APPLICATION OF THE LAPLACE TRANSFORM IN DIFFUSION PROBLEMS

The Laplace transform is a mathematical tool for the solution of *PDEs*. With the Laplace transform we remove the time variable and the *PDE* is transformed to an *ODE*. We then solve the *ODE* and apply the reverse Laplace transform to obtain the solution of the *PDE*. If $F(t)$ is a function of t , for $t > 0$, the Laplace transform of $F(t)$ is

$$L\{F(t)\} = f(s) = \int_0^{\infty} e^{-st} F(t) dt$$

For example if $F(t) = 1$ then $L\{F(t)\} = \int_0^{\infty} e^{-st} dt = \frac{1}{s}$

Consider now a semi-infinite solid with a constant surface concentration c_s and zero initial concentration. The initial condition is

$$c(x, 0) = 0$$

and the boundary conditions are

$$c(0, t) = c_s$$

$$c(\infty, t) = 0$$

We set

$$c(x, t) \equiv U(x, t)$$

The Laplace transform of $U(x, t)$ is

$$L\{U(x, t)\} = u(x, s)$$

Fick's second law becomes

$$\frac{\partial U}{\partial t} = D \frac{\partial^2 U}{\partial x^2} \quad (5.46)$$

We have

$$L\left\{\frac{\partial U}{\partial t}\right\} = su - U(x, 0)$$

$$L\left\{D \frac{\partial^2 U}{\partial x^2}\right\} = \int_0^{\infty} e^{-st} D \frac{\partial^2 U(x, t)}{\partial x^2} dt = D \frac{\partial^2}{\partial x^2} \int_0^{\infty} e^{-st} U(x, t) dt = D \frac{d^2 u}{dx^2}$$

The Equation (5.46) becomes

$$\frac{d^2 u}{dx^2} - \frac{s}{D} u = 0$$

which is an *ODE* with solution

$$u(x, s) = A e^{x\sqrt{s/D}} + B e^{-x\sqrt{s/D}} \quad (5.47)$$

Using the second boundary condition, we get $A = 0$. The Laplace transform of the first boundary condition is $u(0, s) = \frac{c_s}{s} = \frac{U_s}{s} = B$ and (5.47) becomes

$$u(x, s) = \frac{U_s}{s} e^{-x\sqrt{s/D}}$$

We now apply the inverse Laplace transform

$$L^{-1} \left\{ \frac{U_s}{s} e^{-x\sqrt{s/D}} \right\} = U_s erfc(\frac{x}{2\sqrt{Dt}})$$

and the solution is

$$c(x, t) = c_s erfc(\frac{x}{2\sqrt{Dt}})$$

which is the same as (5.37).

5.11.5 MOVING BOUNDARY PROBLEMS

The problems discussed above were concerned with single-phase diffusion. However, in several cases, diffusion takes place in two-phase regions with the concurrent migration of the interface between the two phases. Examples are the carburizing of steel, the growth of a new phase and the coarsening or dissolution of precipitate phases. These problems are called moving-boundary problems. We will describe the analytical solution to such problems taking the carburizing of steel as an example. The carbon-rich layer, which forms at the surface, transforms to austenite during the process. The relevant section of the $Fe - C$ phase diagram is shown in Figure 5.23.

The initial carbon content of the steel is c_o , corresponding to point A on the diagram. Carburizing is performed at $800^\circ C$. The chemical potential of carbon is defined by the CO/CO_2 ratio in the carburizing gas through the reaction $2CO = C + CO_2$. The chemical potential of carbon fixes the surface concentration of carbon c_s , which corresponds to point B. With the diffusion of carbon in the steel, the composition moves from point A towards point B. In the two-phase region ($\alpha + \gamma$) the compositions at the interface are c_α^* and c_γ^* . Assuming local thermodynamic equilibrium at the interface, these values are taken directly from the phase diagram. If s is the displacement of the α/γ interface, we are interested in developing an expression for the position of the interface $s(t)$. The problem can be solved under the following assumptions:

1. The interface velocity is controlled by carbon diffusion
2. Local thermodynamic equilibrium at the interface so that

$$\begin{aligned} c &= c_\gamma^* \text{ for } x = s^- \\ c &= c_\alpha^* \text{ for } x = s^+ \end{aligned}$$

3. The diffusion takes place in a semi-infinite solid
4. The diffusion coefficients of carbon in ferrite and austenite, D_α and D_γ , are independent of carbon concentration

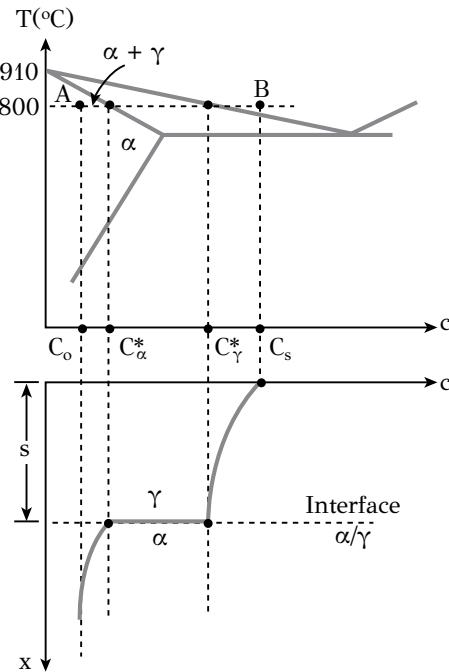


Figure 5.23: Carburizing of steel as a moving boundary diffusion problem: section of the $Fe - C$ phase diagram and associated carbon concentration profile in the two phases, ferrite and austenite (details in text).

5. The $\alpha \rightarrow \gamma$ transformation is not accompanied by a volume change.

Under the above conditions the diffusion equations are

$$\frac{\partial c_\alpha}{\partial t} = D_\alpha \frac{\partial^2 c_\alpha}{\partial x^2} \quad \text{for } x > s$$

$$\frac{\partial c_\gamma}{\partial t} = D_\gamma \frac{\partial^2 c_\gamma}{\partial x^2} \quad \text{for } 0 < x < s$$

The initial condition is

$$c(x, 0) = c_o \quad \text{for all } x > 0$$

The boundary conditions for the ferrite are

$$c = c_\alpha^* \quad \text{for } x = s^+ \tag{5.48}$$

$$c = c_o \quad \text{for } x = \infty \tag{5.49}$$

The boundary conditions for austenite are

$$c = c_\gamma^* \quad \text{for } x = s^- \tag{5.50}$$

$$c = c_s \quad \text{for } x = 0 \quad (5.51)$$

The mass balance at the α/γ interface gives

$$(J_\gamma - J_\alpha) dt = (c_\gamma^* - c_\alpha^*) ds$$

Using Fick's first law the above relation becomes

$$\left[-D_\gamma \left(\frac{\partial c_\gamma}{\partial t} \right)_{x=s} + D_\alpha \left(\frac{\partial c_\alpha}{\partial t} \right)_{x=s} \right] = (c_\gamma^* - c_\alpha^*) \frac{ds}{dt} \quad (5.52)$$

The carbon concentration profiles in ferrite and austenite are respectively

$$c_\alpha = A + Berfc \left(\frac{x}{2\sqrt{D_\alpha t}} \right)$$

$$c_\gamma = A' + B' erf \left(\frac{x}{2\sqrt{D_\gamma t}} \right)$$

From (5.49) we have $c_\alpha(\infty, t) = c_o = A$, then

$$c_\alpha(x, t) = c_o + Berfc \left(\frac{x}{2\sqrt{D_\alpha t}} \right)$$

Using (5.51) we get

$$c_\gamma(x, t) = c_s + B' erf \left(\frac{x}{2\sqrt{D_\gamma t}} \right)$$

At the interface $c_\gamma = c_\gamma^*$, then the argument of the error function above is constant for $x = s$. Setting $\frac{s}{2\sqrt{D_\gamma t}} = \beta$ we get

$$s = 2\beta \sqrt{D_\gamma t} \quad (5.53)$$

From (5.50) we have

$$c_\gamma(s, t) = c_\gamma^* = c_s + B' erf \left(\frac{s}{2\sqrt{D_\gamma t}} \right) \quad (5.54)$$

From (5.48) we have

$$c_\alpha(s, t) = c_\alpha^* = c_o + Berfc \left(\frac{s}{2\sqrt{D_\alpha t}} \right)$$

The above expression with the aid of (5.53) becomes

$$c_\alpha(s, t) = c_\alpha^* = c_o + Berfc \left[\frac{2\beta \sqrt{D_\gamma t}}{2\sqrt{D_\alpha t}} \right]$$

We set $\phi = D_\gamma/D_\alpha$ and the above relation becomes

$$c_\alpha(s, t) = c_o + Berfc(\beta\phi^{1/2}) = c_\alpha^* \quad (5.55)$$

Using (5.54) and the above relation, the expression (5.52) becomes

$$\frac{-D_\gamma B'}{2\sqrt{D_\gamma t}} \frac{2}{\sqrt{\pi}} \exp\left(-\frac{x^2}{4D_\gamma t}\right)_{x=s} - \frac{D_\alpha B}{2\sqrt{D_\alpha t}} \frac{2}{\sqrt{\pi}} \exp\left(-\frac{x^2}{4D_\alpha t}\right)_{x=s} = (c_\gamma^* - c_\alpha^*) \frac{ds}{dt}$$

Because $s = 2\beta\sqrt{D_\gamma t}$ we have $\frac{ds}{dt} = \beta\frac{\sqrt{D_\gamma}}{\sqrt{t}}$. And the above expression for $x = s$ becomes

$$c_\gamma^* - c_\alpha^* = \frac{-B'e^{-\beta^2}}{\sqrt{\pi}\beta} - \frac{Be^{-\beta^2\phi}}{\sqrt{\pi}\beta\sqrt{\phi}} \quad (5.56)$$

Using (5.54) and (5.55) we have

$$c_s + B' erf\beta = c_\gamma^*$$

$$c_o + Berfc\beta\phi^{1/2} = c_\alpha^*$$

We use the above relations to eliminate B and B' from (5.56), which becomes

$$c_\gamma^* - c_\alpha^* = \frac{c_s - c_\gamma^*}{\sqrt{\pi}\beta e^{\beta^2} erf\beta} - \frac{c_\alpha^* - c_o}{\sqrt{\pi}\beta\phi^{1/2} e^{\beta^2\phi} erfc(\beta\phi^{1/2})} \quad (5.57)$$

The Equation (5.57) can be solved for β . Then s can be determined from (5.53).

5.12 NUMERICAL METHODS – COMPUTATIONAL KINETICS

In the previous section, solutions to the diffusion equation, using analytical techniques were discussed. These solutions were obtained under the assumption of a diffusion coefficient independent from concentration. In most real problems and especially in problems concerning multicomponent diffusion, the diffusion coefficient depends on concentration. In this case the diffusion equation

$$\frac{\partial c_i}{\partial t} = \frac{\partial}{\partial x} \left(D_i \frac{\partial c_i}{\partial x} \right)$$

is a non-linear PDE and is solved with numerical methods, such as the finite differences and the finite element methods. The most important element in the solution of diffusion problems in multicomponent systems is the concentration dependence of the diffusion coefficient through the thermodynamic factor

$$D_i = D_i^* \left(1 + \frac{\partial \ln \gamma_i}{\partial \ln X_i} \right)$$

During the evolution of the concentration profile $c(x,t)$, the thermodynamic factor and consequently, the diffusion coefficient change. Thus a link with solution thermodynamics should be established for the calculation of the thermodynamic factor. The combination of numerical methods for the solution of the non-linear diffusion equation with the coupling to alloy thermodynamics is the core of the so-called computational kinetics approach. By employing this methodology, complex diffusion problems such as moving boundary problems and simulations of phase transformations in multicomponent systems can be treated. The interested reader is advised to consult the relevant literature at the end of this chapter. In addition, several simulations regarding diffusive phase transformations are treated in Chapter 11.

5.13 SYNOPSIS

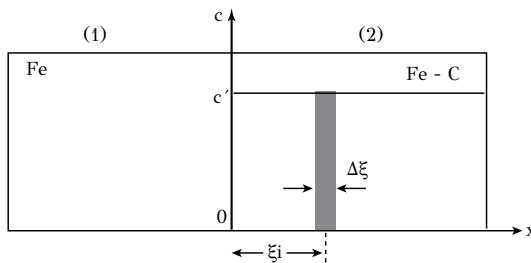
1. Diffusion is the transfer of mass through the material and constitutes one of the major mechanisms with which metallurgical processes in the solid state are accomplished.
2. Diffusion in substitutional solid solutions is accomplished with jumps of atoms in vacancies. Interstitial diffusion is accomplished with jumps of interstitial atoms in vacant interstitial sites.
3. In addition to lattice or bulk diffusion, diffusion takes place through high-diffusivity paths such as grain boundaries and dislocations.
4. According to Fick's first law, the diffusional flux of a component is proportional to its concentration gradient. The proportionality coefficient is called the diffusion coefficient and is a function of temperature and composition.
5. Fick's second law is a consequence of conservation of mass. It is also the diffusion equation, the solution of which provides the evolution of the concentration profile of a component during diffusion.
6. The driving force for diffusion is the gradient of the chemical potential of the component. The diffusive flux is then proportional to the chemical potential gradient. The proportionality coefficient is the atomic mobility, which is a function of temperature and composition.
7. The diffusion coefficient is related to mobility through the thermodynamic factor, which expresses the non-ideality of the solid solution. Therefore, the dependence of the diffusion coefficient on composition is caused by the non-ideality of the solid solution.
8. In substitutional solid solutions exhibiting concentration gradients, interdiffusion takes place. The diffusion of each component is described with the intrinsic diffusion coefficient, referred to the C-frame, which is the crystal lattice itself.

9. The inequality of the intrinsic diffusion coefficients of the components causes a displacement of the lattice according to the Kirkendall effect. In this case diffusion is described by the interdiffusion coefficient, referred to the V -frame, which is fixed relative to the laboratory.
10. Fick's first law is an empirical law compatible with the Onsager principles of irreversible thermodynamics.
11. The basic postulates of irreversible thermodynamics are: (a) the rate of entropy production can be expressed as a sum of terms, each of which is a product of a flux and a force (potential gradient), (b) the rate of entropy production is non-negative, (c) near equilibrium the flux is linearly proportional to its conjugate force.
12. Non-chemical forces may cause diffusion in metallic alloys. Such forces are related to differences in interface curvature and mechanical stresses.
13. Diffusion is the main mechanism of diffusive phase transformations, such as precipitation, coarsening or dissolution of phases.
14. Diffusion plays an important role in the plastic deformation of metals at high temperatures (creep). Several thermal and surface treatments are accomplished with the aid of diffusion.
15. Analytical solutions to the diffusion equation can be obtained under the assumption that the diffusion coefficient is independent of composition. In the opposite case, numerical methods are available in the context of computational kinetics of alloys.

5.14 REVIEW QUESTIONS

1. The diffusion coefficient of lithium in silicon is $10^{-5} \text{ cm}^2/\text{s}$ at 1100°C and $10^{-6} \text{ cm}^2/\text{s}$ at 692°C . Calculate the activation energy for diffusion ΔH^* and the pre-exponential term D_o .
2. Consider diffusion of aluminum in a silicon single crystal. At which temperature the diffusion coefficient is $D = 5 \times 10^{-11} \text{ cm}^2/\text{sec}$? ($D_o = 1.7 \text{ cm}^2/\text{sec}$ and $\Delta H^* = 315 \text{ KJ/mol}$).
3. Consider that a certain quantity S , in atoms/m^2 , of radioactive tracer A^* is placed between two rods of metal A . The system is annealed at a high temperature for some time t and the concentration profile of tracer A^* is determined. Given the concentration profile, determine the self-diffusion coefficient of A .
4. Consider the diffusion couple shown below. The left side (1) is pure iron, while the right side (2) is a steel with carbon content c' . The system is brought to a high temperature for a certain time t . Considering the solution regarding instantaneous sources (section 5.11.2) and using the superposition method, show that the carbon

concentration profile $c(x,t)$ in the couple is given by the error function. Hint: Consider a thin slice $\Delta\xi$ in side (2) as an instantaneous source of carbon.

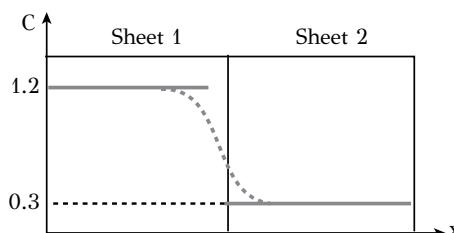


5. Doping of Silicon by Indium is required in a process so that the concentration of Indium at a depth $3\mu m$ from the surface is half of that at the surface. What is the required diffusion time? The diffusion coefficient of Indium in Silicon is $D = 8 \times 10^{-12} cm^2/sec$.
6. A steel specimen has been carburized at $930^\circ C$ for 10 hours. The carburization depth reached $400\mu m$. What is the required carburization time at the same temperature in order to double the carburization depth?
7. Consider a steel specimen with $2mm$ thickness, containing $0.005wt\%$ carbon. The specimen is subjected to carburization by placing it in a box containing graphite powder. The box is then heated to $900^\circ C$. Under these conditions the surface concentration of carbon becomes $0.8wt\%$ and remains constant throughout the carburization process.
 - a. Determine the carbon concentration profile $c(x)$ for diffusion times of 5, 50 and 500 s.
 - b. We aim at a carbon concentration of $0.4wt\%$ at $20\mu m$ depth from the surface. For which diffusion time, 5, 50 or 500s is this target achieved?
 - c. Assume that carburization is performed at a higher temperature, $1000^\circ C$. Compare the concentration profiles for $t = 5s$ between 900 and $1000^\circ C$.
 - d. Check if the requirement for $0.4wt\%$ carbon at $20\mu m$ depth is satisfied for diffusion time $t = 50s$ at $1000^\circ C$.
 - e. Calculate the diffusion time to achieve $0.4wt\%$ carbon at $20m$ depth at 900 and $1000^\circ C$.
 - f. Calculate the depth at which the carbon concentration reaches $0.5wt\%$ at $t = 500s$ at 900 and $1000^\circ C$. The diffusion coefficients of carbon in austenite are $D = 1.14 \times 10^{-12} m^2/sec$ at $900^\circ C$ and $D = 1.14 \times 10^{-11} m^2/sec$ at $1000^\circ C$. Assume semi-infinite medium.
8. Hydrogen in steel causes hydrogen embrittlement, especially in high-strength steels. For this reason, several engineering parts as well as welds are subjected to a bake out treatment to remove hydrogen. Consider a steel ball of a rolling bearing, with radius $0.5cm$. The ball is placed in a laboratory furnace, at $300^\circ C$, for bake out. The initial hydrogen concentration is $25ppmw$. Assume that heat transfer is so fast that it is not the rate limiting step of the bake out process.

- Estimate the hydrogen release rate (in gr/s) at the surface of the ball at time $t = 1s$.
- Estimate the time required to reduce the hydrogen concentration of the ball to $5ppmw$.

Given: Hydrogen diffusivity in iron $D = 10^{-4} cm^2/sec$. Density of iron $7.8 gr/cm^3$.

- Two large steel plates, with thickness $2cm$ and carbon content 1.2 and $0.3wt\%$ respectively, form the diffusion couple shown below. The couple is heated to $1000^\circ C$.



- At what time the concentration in the right side will reach 0.5% at a distance $0.15cm$ from the surface? Use the short time solution (error function).
- Above which time the error function solution is no longer valid. Use a "leak test" with a 1% margin.
- Derive an expression for the velocity of the plane with concentration $0.5wt\%$.
- Derive an expression for the velocity of the same plane, which is valid for long times.
- Make a plot of the position of the plane with 0.5% carbon as a function of time for $0 < x < 2cm$.

The diffusion coefficient of carbon is $D = 3 \times 10^{-7} cm^2/sec$.

- Consider diffusion in a semi-infinite medium with initial concentration c_o and a constant flux $J = A$ at the surface ($x = 0$). Calculate the concentration profile using the Laplace transform.
- Decarburization is an undesirable effect concerning the loss of carbon from engineering components made of carbon steel and subjected to hot working, such as rolling or forging. Decarburization also takes place during service at high temperatures, in the presence of oxygen, steam or CO_2 in the working environment. Decarburization leads to a reduction in hardness and degrades the wear and fatigue resistance of the steel. In a pearlite steel, decarburization causes the formation of a ferrite layer at the surface, which is actually the decarburized layer. Develop an expression of the decarburization depth vs time, under the following assumptions:
 - decarburization is controlled by the carbon diffusion in ferrite
 - the diffusion coefficient is independent of the carbon concentration
 - the concentration of carbon in pearlite is homogeneous and equal to c_o . The concentration c^* at the ferrite/pearlite interface is taken from the $Fe - C$ phase diagram assuming local thermodynamic equilibrium.

- d. the surface concentration c_s is established by the decarburization conditions and remains constant during the decarburization process.

Hint: Consider this a moving boundary diffusion problem (see section 5.11.5).

12. Consider a specimen of solid solution A – B exhibiting a concentration gradient of B. Assuming that the concentration profile of B, $c(x)$, has been determined experimentally after a diffusion time t, derive a method for the calculation of the interdiffusion coefficient as a function of concentration (The method is called Matano analysis).

BIBLIOGRAPHY–FURTHER READING

1. Shewmon, P. G. (1963). *Diffusion in Solids*, McGraw-Hill, New York.
2. Crank, J. (1963). *The Mathematics of Diffusion*, Oxford University Press, New York.
3. Carslaw, H. S., Jaeger, J. C. (1959). *Conduction of Heat in Solids*, Clarendon Press, Oxford.
4. Bocquet, J.L., Brebec, G., Limoge, V. (1983). *Diffusion in Metals and Alloys*, in *Physical Metallurgy*, R.W. Cahn, P. Haasen (eds), Elsevier Science Publishers.
5. Sinha, A.K. (2003). *Physical Metallurgy Handbook, Chapter 2: Diffusion in Metals and Alloys*, McGraw Hill, New York.
6. Porter, D.A., Easterling, K.E., Sherif, M.Y. (2009). *Phase Transformations in Metals and Alloys*, CRC Press.
7. Kirkaldy, J.S., Young, D.J. (1987). *Diffusion in the Condensed State*, Institute of Metals, London.
8. Poirier, D.R., Geiger, G.H. (1994). *Transport Phenomena in Materials Processing*, The Metallurgical Society (TMS), Warrendale, PA, USA.
9. Glicksman, M.E. (2000). *Diffusion in Solids*, John Wiley & Sons, New York.
10. Balluffi, R.W., Allen, S.M. & Carter, W.C. (2005). *Kinetics of Materials*, Wiley Inter-science.

6 Phase transformations

6.1 INTRODUCTION

Phase transformations play a key role in physical metallurgy, since most industrial alloys are subjected to some kind of heat treatment for the enhancement of their properties. Every alloy is made up of several phases, which are either solid solutions or intermetallic compounds. Heat treatment causes a change in the structure of the alloy through phase transformations. Let us consider an example: the precipitation hardening of an alloy (Figure 6.1).

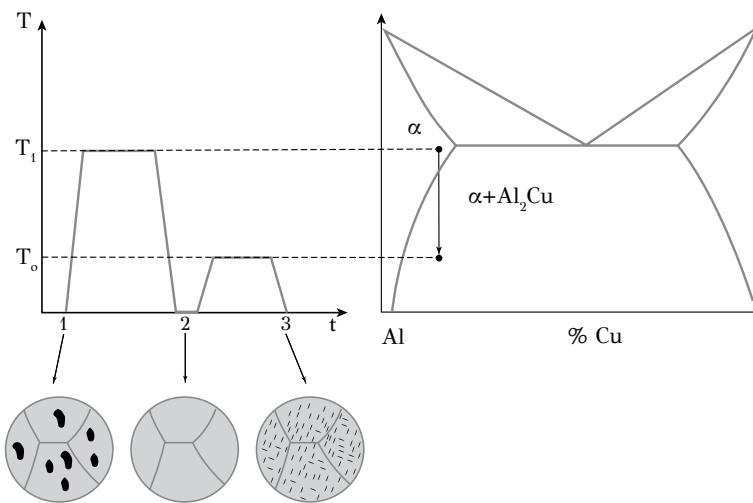


Figure 6.1: Dissolution and precipitation of Al_2Cu phase in $Al - Cu$ alloys: microstructure and relation with the phase diagram.

The alloy $Al - Cu$ has a starting microstructure composed of a coarse dispersion of particles of the intermetallic compound Al_2Cu in the solid solution of phase α . The alloy is heated to a temperature T_1 (just below solidus) and remains at this temperature for some time. During isothermal holding, the intermetallic phase Al_2Cu is dissolved and the Cu atoms are incorporated into the solid solution of phase α . The dissolution of the Al_2Cu phase is a phase transformation. Then the alloy is quenched (cooled rapidly) to room temperature. Due to the limited time during cooling, the Cu atoms do not have enough time to diffuse and reform the intermetallic phase Al_2Cu , as predicted by the phase diagram. Therefore, at room temperature, a supersaturated $Al - Cu$ solid solution is retained. The alloy is then heated again at an intermediate temperature T_0 and remains at that temperature for some time. During isothermal holding at T_0 , the intermetallic phase Al_2Cu forms as a fine particle dispersion in

the solid solution of the phase α . The formation of the intermetallic compound is a solid-state phase transformation called precipitation. Precipitation leads to hardening of the alloy, since the precipitates act as obstacles to dislocation glide. We see, therefore, that the microstructure and properties of alloys are shaped through phase transformations.

There are several types of phase transformations and a classification is shown in Figure 6.2. The majority of solid-state phase transformations are accomplished by thermally activated movement of atoms, i.e., by diffusion. There are transformations, however, where diffusion is not a prerequisite (e.g., the martensitic transformation). We can distinguish, therefore, phase transformations as *diffusive* and *non-diffusive*. Regarding the diffusive phase transformations, we take the example of the formation of a new phase β from phase α and we distinguish two classes: Nucleation and growth (*NGT*) and continuous transformations (*CT*).

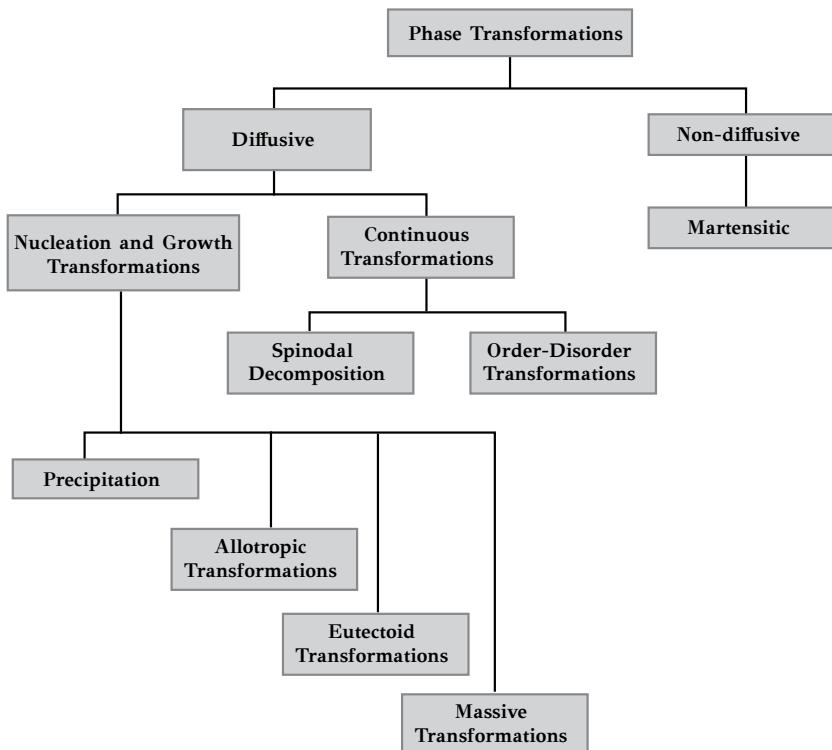


Figure 6.2: Classification of solid-state phase transformations.

Nucleation and Growth Transformations (NGT). These transformations begin with a composition fluctuation, which is highly localized and intense (Figure 6.3a). The system is metastable with respect to the formation of a small region with the composition of the new phase (nucleation), and the transformation proceeds with the

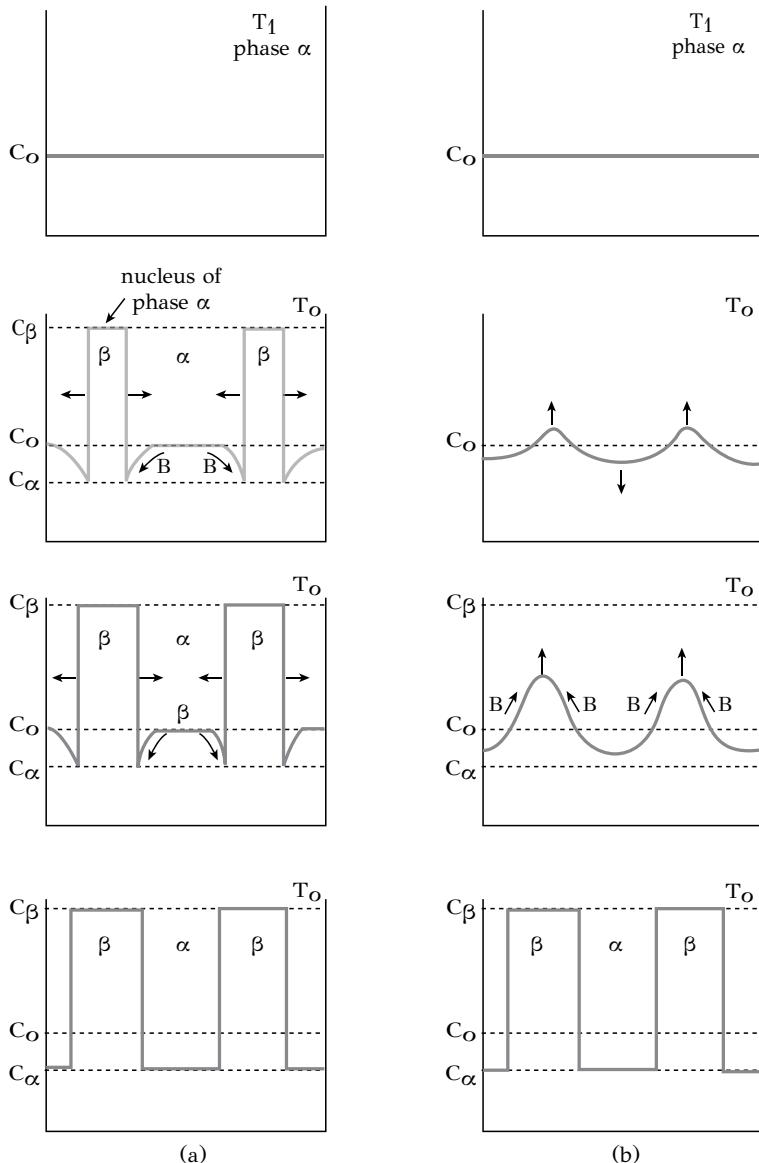


Figure 6.3: Evolution of composition fluctuations during nucleation and growth transformations (NGT) in (a) and during continuous transformations (CT) in (b).

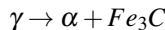
Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). *Phase Transformations in Metals and Alloys*, with permission from CRC press.

nucleation and growth of the new phase β . The evolution of the *NGT*, in terms of the composition profile of element B , is shown in Figure 6.3a for a binary alloy $A - B$. Initially at temperature T_1 , the element B possesses a uniform composition profile in phase α . The system is brought to temperature T_o , where nuclei of phase β form and the system consists of two phases $\alpha + \beta$. The phase β grows by diffusion of B from phase α to phase β . Growth of phase β seizes when its fraction reaches the equilibrium lever-rule value

$$f_{\beta}^{eq} = \frac{c_o - c_{\alpha}}{c_{\beta} - c_{\alpha}} \quad (6.1)$$

The following transformations can be considered as *NGT*:

- Precipitation of a new phase β from a metastable and supersaturated solid solution of phase α .
- Allotropic transformations in pure metals.
- Eutectoid transformations, where a metastable phase γ transforms to a more stable mixture of phases $\alpha + \beta$. A very well known example is the formation of pearlite in steels:



- Massive transformations, where the parent phase transforms to a new phase possessing the same composition but different crystal structure from the parent phase. These transformations are accomplished with the thermally activated movement of interphase boundaries, which involves atomic diffusion. The massive transformations should not be confused with martensitic transformations where the new phase forms by deformation of the parent phase.

Continuous Transformations (CT). These transformations start with a composition fluctuation, which is not localized (diffused) and has low intensity. The transformation proceeds with the amplification of these initial fluctuations. The evolution of the *CT*, in terms of the composition profile of element B , is shown in Figure 6.3b for a binary alloy $A - B$. Initially at temperature T_1 , the element B possesses a uniform composition profile in phase α . The system is brought to temperature T_o , where a low-intensity composition fluctuation of element B forms, the amplitude of the fluctuation being significantly lower than c_{β} , the composition of phase β . The fluctuation is continuously amplified until it reaches the value c_{β} . This amplification is accomplished by up-hill diffusion of B (where the diffusion coefficient D is negative) and seizes when the phase fraction of β reaches the equilibrium lever-rule value given by Equation (6.1).

6.2 NUCLEATION AND GROWTH TRANSFORMATIONS (NGT)

We will now study in more detail the *NGT*. We will see that the transformation starts with the formation of small regions of the new phase, which are called *nuclei*. The nuclei possess the composition of the new phase β . The process is termed *nucleation*.

The nuclei then increase in size through a process termed *growth*, which is controlled by atomic diffusion. Growth ceases when the phase fraction of β reaches the value given by the lever rule in the phase diagram.

6.3 NUCLEATION

6.3.1 THE CHEMICAL DRIVING FORCE FOR NUCLEATION

Consider the binary alloy $A - B$ with composition c_o . The respective phase diagram is depicted in Figure 6.4. At temperature T_1 the alloy exists as a single-phase solid

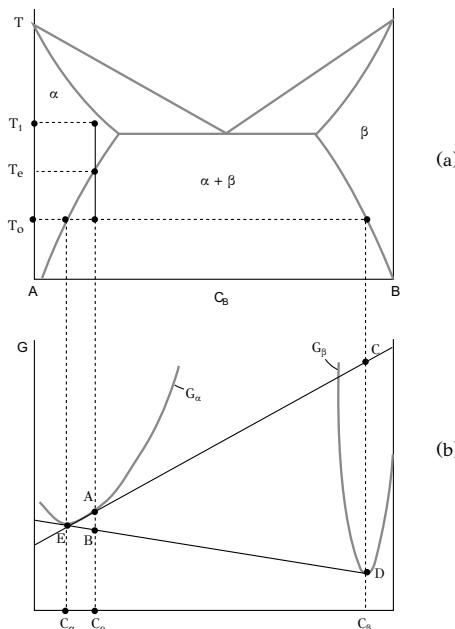


Figure 6.4: Binary phase diagram in (a) and corresponding free energy - composition diagram at temperature T_0 in (b). The total free energy change due to precipitation of β phase is indicated by AB , while the driving force for the nucleation of β phase is indicated by CD .

solution of phase α . Then the alloy is quenched to temperature T_0 where it remains so that the transformation $\alpha \rightarrow \alpha + \beta$ can take place isothermally. Initially at T_0 the system is composed of a supersaturated solid solution of phase α . Therefore the transformation involves the precipitation of phase β (rich in component B) from the supersaturated solid solution of phase α . The free energy curves of phases α and β at T_0 are shown in Figure 6.4b. At thermodynamic equilibrium the system consists of two phases with compositions c_α and c_β corresponding to the contact points E and D of the common tangent to the free energy curves. The total free energy change of the system due to the precipitation of phase β is given by $\Delta G_0 = AB$ in Figure 6.4b. Point

A corresponds to the free energy of the supersaturated solid solution of phase α with composition c_o , while point B corresponds to the two-phase mixture α and β with compositions c_α and c_β respectively. However it should be noted that ΔG_o is not the driving force for nucleation, but instead it is the driving force for the precipitation of phase β , i.e., for the transformation $\alpha \rightarrow \alpha + \beta$, which includes both nucleation and growth of β . In order to determine the *driving force for nucleation*, we will assume that the composition of phase α does not change considerably by the formation of the first nuclei of β with composition c_β . This allows us to draw the tangent to the free energy curve of phase α at point A and determine the intersection point C with the composition line at c_β . The segment $CD = \Delta G_n$ is the molar chemical driving force (in J/mol) for the nucleation of phase β with composition c_β . The chemical driving force per unit volume (in J/m^3) is obtained from ΔG_n through the relation

$$\Delta G_v = \frac{\Delta G_n}{V_m}$$

where V_m is the molar volume (in m^3/mol) of phase β . For dilute solid solutions it can be shown that

$$\Delta G_v = -\frac{RT}{V_m} \ln \frac{c_o}{c_\alpha}$$

The ratio c_o/c_α is the *supersaturation* ratio of phase α . From Figure 6.4a we see that for the alloy with composition c_o , the driving force for nucleation, being proportional to the natural logarithm of the supersaturation, increases with undercooling below the solvus temperature T_e .

6.3.2 HOMOGENEOUS NUCLEATION

We will consider first the case of *homogeneous nucleation*, i.e., the formation of nuclei of phase β within the crystal lattice of phase α , without the assistance of any structural imperfections, such as vacancies, dislocations or grain boundaries. The nucleation of phase β requires diffusion of B atoms in phase α for the formation of small regions with the composition c_β . The nucleation process is accompanied by a change in free energy of the system ΔG , which consists of the following terms:

- The generation of a volume V of phase β at temperature T_o results in a reduction of free energy equal to $V\Delta G_v$, where ΔG_v is the chemical driving force for nucleation per unit volume of the new phase.
- The nucleation of phase β results in the formation of new surface, i.e., the interface between phases α and β (interphase boundary). If this interface has a surface area A , then the increment of the free energy of the system is $\gamma_{\alpha\beta}A$, where $\gamma_{\alpha\beta}$ is the interfacial energy of the interphase boundary α/β .
- The nucleation of the new phase is accompanied by a deformation of the lattice of α and/or β . These strains are either coherency strains, arising from lattice misfit, or volumetric strains arising from differences in volume between the new phase β and the volume of phase α that was replaced by β .

The increment in the free energy of the system is then $V\Delta G_\varepsilon$, where ΔG_ε is the strain energy per unit volume.

Taking into consideration the contributions stated above, the total free energy change due to nucleation is

$$\Delta G = V\Delta G_v + \gamma_{\alpha\beta} A + V\Delta G_\varepsilon$$

If we consider that the particles of the new phase are spherical with a radius r the above relation becomes

$$\Delta G(r) = \frac{4}{3}\pi r^3 \Delta G_v + 4\pi r^2 \gamma_{\alpha\beta} + \frac{4}{3}\pi r^3 \Delta G_\varepsilon \quad (6.2)$$

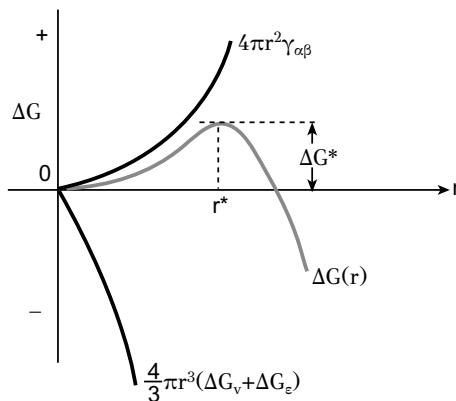


Figure 6.5: Variation of the free energy of the new phase as a function of particle radius during nucleation.

Equation (6.2) is shown in Figure 6.5. The growth of particles with radius smaller than r^* (called *embryos*) increases the free energy. Embryos, therefore, cannot grow and redissolve in phase α . However the growth of particles with radius larger than r^* lowers the free energy. These particles are called *nuclei*. The nuclei are thermodynamically stable because their growth lowers the free energy. The radius r^* is called *critical radius for nucleation*. As shown in Figure 6.5, nucleation requires overcoming of the energy barrier ΔG^* , called the *activation energy* for nucleation arising from the interfacial energy and the strain energy. For $r > r^*$ the reduction of free energy arising from the chemical driving force dominates the positive contributions from the interfacial energy and the strain energy (the first term in Equation (6.2) is negative while the other two are positive). The activation parameters r^* and ΔG^* are obtained by setting $\partial\Delta G/\partial r = 0$ in Equation (6.2) so that

$$r^* = -\frac{2\gamma_{\alpha\beta}}{\Delta G_v + \Delta G_\varepsilon} \quad (6.3)$$

$$\Delta G^* = \frac{16\pi\gamma_{\alpha\beta}^3}{3(\Delta G_v + \Delta G_e)^2} \quad (6.4)$$

We can see that the greater the driving force, the smaller the nucleation size of the new phase. In addition the energy barrier has a strong dependence on the interfacial energy (3^{rd} power), meaning that the interfacial structure (coherent, semicoherent or incoherent interface) plays an important role in the nucleation process, something that we will study in detail later in this chapter. The homogeneous nucleation rate I_{hom} (in number of nuclei per unit volume and time) can be calculated if we consider the density of critical nuclei N^* , which is

$$N^* = N_v \exp\left(-\frac{\Delta G^*}{kT}\right)$$

where N_v is the nucleation site density. The rate R of the diffusional transport of atoms to the critical nucleus (in atoms/sec) is

$$R = v \exp\left(-\frac{\Delta G_D^*}{kT}\right)$$

where v is the Debye frequency ($\approx 10^{13}/\text{sec}$) and ΔG_D^* the activation energy for diffusion. The homogeneous nucleation rate is $I_{\text{hom}} = NR$ and by combining the above relations we get

$$I_{\text{hom}} = v N_v \exp\left(-\frac{\Delta G^*}{kT}\right) \exp\left(-\frac{\Delta G_D^*}{kT}\right) \quad (6.5)$$

Setting typical values $\Delta G_D^* = 120\text{KJ/mol}$, $N_v = 6 \times 10^{29}/\text{m}^3$, $T = 1000\text{K}$ in Equation (6.5) we get

$$I_{\text{hom}} = (10^{36}\text{m}^3/\text{s}) \exp\left(-\frac{\Delta G^*}{kT}\right)$$

The observable nucleation rate is of the order of $10^6/\text{m}^3\text{sec}$, which corresponds to 1 nucleus per cm^3 and sec . In this case the activation energy is $\Delta G^* \approx 70kT$ which is very high and cannot be surmounted by the available thermal energy. This explains the fact that homogeneous nucleation is very rare. In most cases, nucleation proceeds heterogeneously, i.e., the formation of nuclei of the new phase is assisted by structural defects, such as vacancies, dislocations and grain boundaries. Part of the energy associated with these defects is used to reduce the activation energy for nucleation. The details of heterogeneous nucleation are discussed below.

6.3.3 HETEROGENEOUS NUCLEATION

In the previous section it was stated that nucleation proceeds, in most cases, heterogeneously. Consider the following examples. The first example is concerned with

the precipitation of Mo_2C carbides during tempering of a $Fe - 1C - 4Mo$ steel. A TEM picture of the carbide dispersion is shown in Figure 6.6a. This fine precipitation is due to the heterogeneous nucleation of the carbides on dislocations. An atom probe field ion microscope (*AP/FIM*) image of the carbides is also shown in Figure 6.6b. The second example is concerned with the precipitation of phase β during artificial aging of aluminum alloy 6061 ($Al - Mg - Si$). The phase β is the intermetallic compound Mg_2Si , which nucleates heterogeneously on dislocations. A dispersion of β rod-shaped particles is shown in Figure 6.7a. A higher magnification shows a Mg_2Si particle surrounded by dislocations, depicting the role of dislocations in the nucleation of this phase. Finally, the third example is concerned with heterogeneous nucleation at grain boundaries. Nucleation of ferrite at the prior austenite grain boundaries during cooling of austenite in a $Fe - 0.4C$ proeutectoid steel is depicted in Figure 6.8.

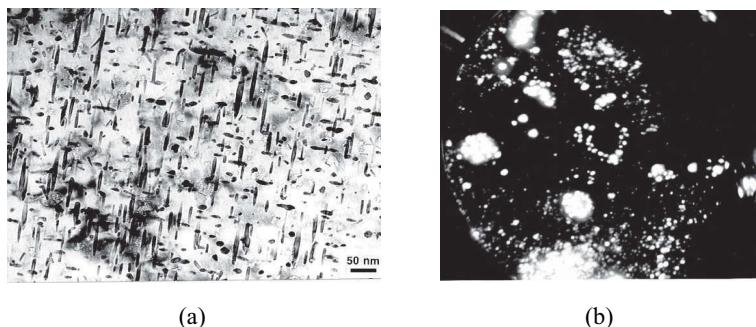


Figure 6.6: Heterogeneous nucleation: (a) precipitation of Mo_2C carbides on dislocations during tempering of a Mo -containing steel, (b) an *AP/FIM* image of the Mo_2C carbides.

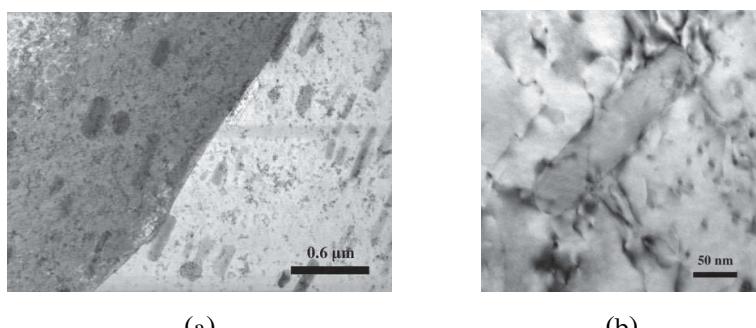


Figure 6.7: Heterogeneous nucleation: (a) a dispersion of $\beta - Mg_2Si$ particles in a 6061 alloy ($Al - Mg - Si$), (b) higher magnification image indicating nucleation of a $\beta - Mg_2Si$ particle on dislocations.

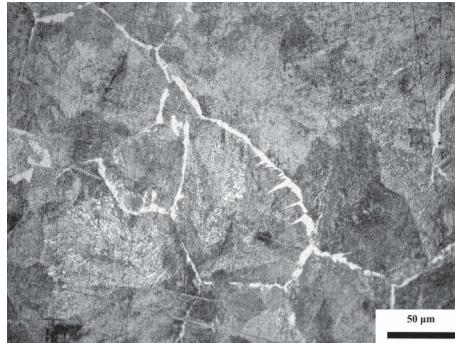


Figure 6.8: Heterogeneous nucleation of ferrite on prior austenite grain boundaries during cooling of austenite in a $Fe - 0.4C$ steel.

We will study in more detail the heterogeneous nucleation at grain boundaries. We make the assumption that the particles of the new phase β have incoherent interfaces with the matrix and, therefore, there are no coherency strains. In this case the strain energy can be neglected. The nucleus is then a lens-shaped particle with spheroidal surfaces in order to minimize the interfacial energy (Figure 6.9a).

In addition we consider that the interfacial energies $\gamma_{\alpha\alpha}$ and $\gamma_{\alpha\beta}$ are isotropic. The contact angle θ is given by the relation

$$\gamma_{\alpha\alpha} = 2\gamma_{\alpha\beta} \cos \theta$$

The free energy change ΔG_{het} for heterogeneous nucleation is

$$\Delta G_{het} = V\Delta G_v + A\gamma_{\alpha\beta} - A_{\alpha\alpha}\gamma_{\alpha\alpha} \quad (6.6)$$

where A is the surface area of the particle of phase β , $A_{\alpha\alpha}$ is the grain boundary area that was consumed for the formation of the particle, and V is the volume of the particle. Taking into account that $d = r \sin \theta$ and $c = r(1 - \cos \theta)$, the volume and the area of the particle are

$$V = \frac{2\pi r^3}{3} (2 - 3 \cos \theta + \cos^3 \theta)$$

$$A = 4\pi r^2(1 - \cos \theta)$$

Implementing the above relations, Equation (6.6) becomes

$$\Delta G_{het} = \left(\frac{2}{3}\pi r^3 \Delta G_v + 2\pi r^2 \gamma_{\alpha\beta} \right) (2 - 3 \cos \theta + \cos^3 \theta)$$

The quantity in the first parenthesis is $1/2$ the free energy change for homogeneous nucleation and therefore

$$\Delta G_{het} = \frac{1}{2} (2 - 3 \cos \theta + \cos^3 \theta) \Delta G_{hom}$$

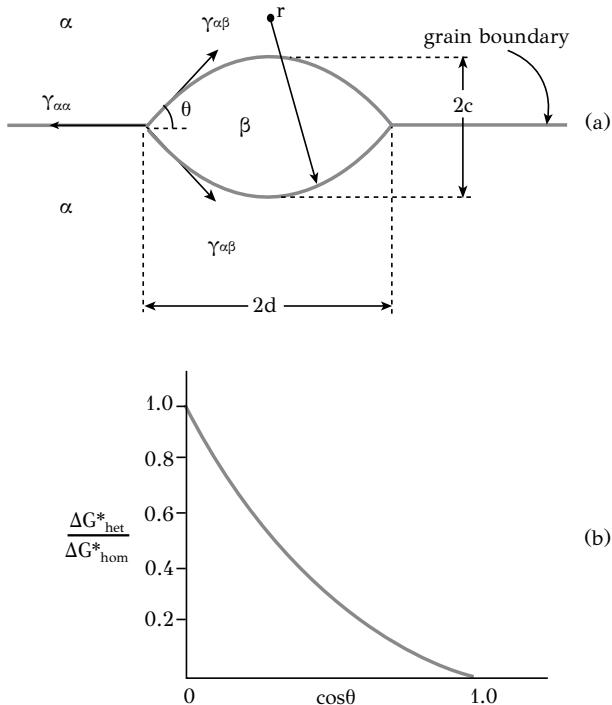


Figure 6.9: (a) Heterogeneous nucleation of β -phase on a grain boundary of the matrix α -phase. (b) Variation of the heterogeneous activation energy with the contact angle θ .

Setting $f(\theta) = \frac{1}{2}(2 - 3\cos\theta + \cos^3\theta)$ we get

$$\Delta G_{het} = f(\theta)\Delta G_{hom}$$

The activation parameters for heterogeneous nucleation are

$$r_{het}^* = r_{hom}^* = -\frac{2\gamma_{\alpha\beta}}{\Delta G_v}$$

$$\Delta G_{het}^* = f(\theta)\Delta G_{hom}^*$$

One can observe that the critical nucleation size r^* is the same for both homogeneous and heterogeneous nucleation and also is independent from the contact angle θ . However the potency of the grain boundaries to act as heterogeneous nucleation sites, i.e., the reduction of the energy barrier, depends on the contact angle through the factor $f(\theta)$. The variation of $f(\theta)$ with θ is depicted in Figure 6.9b. For $\theta = 90^\circ$ we get $\Delta G_{het}^* = \Delta G_{hom}^*$ while for $\theta = 0$ the energy barrier vanishes. Since the contact angle depends on the ratio $\gamma_{\alpha\alpha}/\gamma_{\alpha\beta}$, it can be shown that the activation energy

vanishes for values of that ratio above 2. The grain boundaries are very potent nucleation sites due to their high interfacial energy. Despite the fact that the activation energy for heterogeneous nucleation is lower than that for homogeneous nucleation, the available nucleation sites are considerably less than those available for homogeneous nucleation. This has an effect on the heterogeneous nucleation rate, which is

$$I_{het} = v N_v^{het} \exp\left(-\frac{\Delta G_D^*}{kT}\right) \exp\left(-\frac{\Delta G_{het}^*}{kT}\right)$$

Dividing by the homogeneous nucleation rate we get

$$\frac{I_{het}}{I_{hom}} = \frac{N_v^{het}}{N_v^{hom}} \exp\left(\frac{\Delta G_{hom}^* - \Delta G_{het}^*}{kT}\right)$$

For heterogeneous nucleation at grain boundaries we can consider that

$$N_v^{het} = N_v^{hom} \frac{s}{d}$$

where s is the thickness of the grain boundary and d the average grain diameter. Therefore we get

$$\frac{I_{het}}{I_{hom}} = \frac{s}{d} \exp\left(\frac{\Delta G_{hom}^* - \Delta G_{het}^*}{kT}\right)$$

and because

$$\ln \frac{I_{het}}{I_{hom}} = \ln \frac{s}{d} + \left(\frac{\Delta G_{hom}^* - \Delta G_{het}^*}{kT}\right)$$

it can be concluded that heterogeneous nucleation dominates over homogeneous nucleation when

$$\Delta G_{hom}^* - \Delta G_{het}^* > kT \ln \frac{d}{s} \quad (6.7)$$

For a given grain size of the matrix phase α (given d/s) the potency of the grain boundaries must be sufficiently large in order to satisfy Equation (6.7). In addition, it can be concluded that heterogeneous nucleation dominates when the matrix phase is grain refined. The smaller the grain size, the higher the nucleation rate.

6.3.4 NUCLEATION STRAIN ENERGY AND THE MORPHOLOGY OF THE NEW PHASE

As stated above, the strains accompanying the nucleation process contribute to the energy barrier for nucleation. There are two basic contributions to the strain energy ΔG_ε , which are caused by:

1. Coherency strains, due to the lattice misfit between the two phases α and β when these phases are separated by a coherent interface
2. Volumetric strains due to the difference in volume or elastic modulus between the two phases

The coherency strains can be reduced with the introduction of misfit dislocations in the boundary (see Chapter 3). In this case the boundary becomes a semicoherent interface. It possesses lower strain energy but higher interfacial energy than before. The coherency strains vanish only if the interface becomes incoherent. In this case the interfacial energy becomes large and the energy barrier for nucleation (proportional to γ^3) increases considerably. Due to these considerations, nucleation always begins with the formation of particles of phase β having coherent interfaces with phase α . The shape of the new phase depends on the magnitude of both the interfacial energy and the strain energy. If we consider that the particle is bounded by surfaces of area A_i and interfacial energy γ_i then the shape of the new phase is the one that satisfies the relation

$$\Delta G_\varepsilon + \sum A_i \gamma_i \rightarrow \min \quad (6.8)$$

i.e., the shape, which minimizes the sum of strain and interfacial energies. We will consider now the effect of particle shape on strain energy. First consider the case of fully coherent particles of phase β . In addition we make the assumption that the elastic constants of phases α and β are the same and that the β particles have the shape of an ellipsoid of revolution with axes R , R and y . In this case, when $y = R$ the particle is spherical, when $y/R \rightarrow 0$ the particle is disc-shaped and when $y/R \rightarrow \infty$ the particle is rod-shaped. For triaxial coherency strains, i.e., $\varepsilon_{11} = \varepsilon_{22} = \varepsilon_{33} = \frac{2}{3}\delta$, the strain energy is

$$\Delta G_\varepsilon = f_1(v)EV\varepsilon^2 = f_1(v)EV\delta^2$$

where $f_1(v)$ is a function of Poisson ratio, E is the modulus of elasticity and δ is the lattice misfit. If, as stated above, $E_\alpha = E_\beta$, then the strain energy is independent of the shape of the particle β . However, if $E_\beta > E_\alpha$ then the strain energy is minimized when the β particles are discs ($y/R \rightarrow 0$). In the case of shear coherency strains, i.e., $\varepsilon_{13} = \varepsilon_{31}$, the strain energy becomes

$$\Delta G_\varepsilon = f_2(v)EV\varepsilon_{13}^2 \frac{y}{R}$$

and, therefore, the strain energy is minimized for $y/R = 0$, i.e., when the β particles are discs. When the β particles are not coherent with the matrix phase α , there are no coherency strains. In this case the strain energy is entirely due to volumetric strains, which are caused by the volume difference between phases α and β , expressed by the parameter $\Delta = \Delta V/V$. In the case of ellipsoidal particle, defined by the relation $x^2/a^2 + y^2/a^2 + z^2/c^2 = 1$, the strain energy is given by Nabarro¹ as

$$\Delta G_\varepsilon = \frac{2}{3}\mu V\Delta^2 f(c/a)$$

where μ is the shear modulus and $f(c/a)$ a shape factor. The dependence of the shape factor on c/a is depicted in Figure 6.10. The shape factor $f(c/a)$ takes very low values for disc-shaped particles, is maximized for spherical particles and obtains intermediate values for needle-shaped particles.

¹F.R.N. Nabarro, Proc. Royal Soc. A175 (1940) 519.

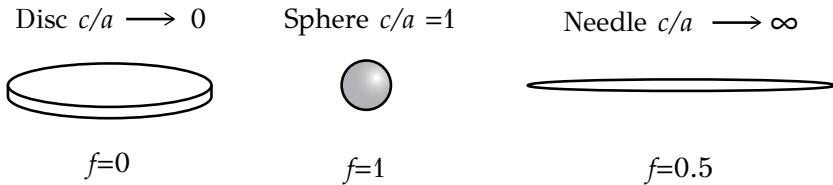


Figure 6.10: Variation of the strain energy (shape factor) with the shape of a particle.

According to the above considerations, the shape of the β particles should satisfy the relation 6.8. In the case of fully coherent particles, the shape depends on the lattice misfit δ . When δ is small, then the shape is defined by the minimization of the interfacial energy and the particles are in the form of spheres. When δ is large, then the shape is defined by the minimization of strain energy and the particles are discs. In the case of incoherent particles the shape depends on the volumetric strain Δ . When Δ is small, then the shape is defined by the minimization of the interfacial energy and the particles are spherical. When Δ is large then the shape is defined by the minimization of strain energy and the particles are in the form of discs.

We will now consider the coherency loss, which takes place in a particle dispersion of the new phase as the particles evolve with nucleation and growth. As discussed above, when the particles are fully coherent, they have low interfacial energy but high strain energy (due to coherency strains). On the other hand, semicoherent particles have a higher interfacial energy but lower strain energy. Consider a spherical β particle of radius r having a lattice misfit δ with phase α . The increase in the free energy of the crystal for these two cases is

$$\Delta G_{coh} = \frac{4}{3}\pi r^3 4\mu \delta^2 + 4\pi r^2 \gamma_{coh}$$

$$\Delta G_{sem} = 0 + 4\pi r^2 (\gamma_{coh} + \gamma_{sem})$$

where γ_{coh} and γ_{sem} are the interfacial energies for coherent and semicoherent interfaces (discussed in Chapter 3) and the strain energy per unit volume has been approximated with $4\mu \delta^2$. For a given lattice misfit δ the above relations are depicted in Figure 6.11. For low values of r the particles maintain coherency in order to minimize ΔG . However, when the particles grow in size, then they lose their coherency with the matrix in order to minimize ΔG and continue their growth as semicoherent particles. By equating the two relations above, the critical radius for coherency loss can be determined as

$$r_c = \frac{3\gamma_{sem}}{4\mu \delta^2}$$

Because γ_{sem} is proportional to the lattice misfit δ , we conclude that r_c is inversely proportional to δ . Therefore in the cases of large lattice misfits, the particles lose coherency at smaller sizes (at shorter times in the growth process) in order to reduce the free energy. For the above discussion we conclude that the morphology of a

particle dispersion, which evolves by *NGT*, depends on interfacial energy as well as strain energy. The shape of the particles and their coherency status with the matrix are defined by the minimization of the total free energy according to the relation 6.8 and affect the rate of the *NGT*.

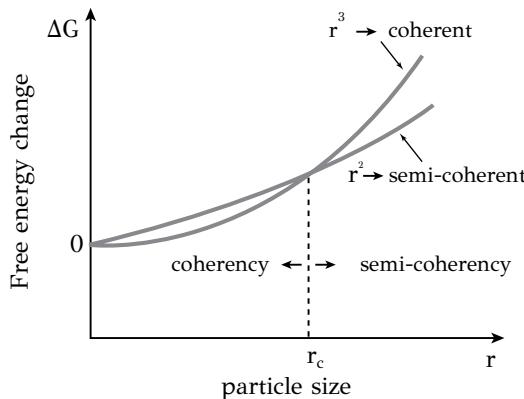
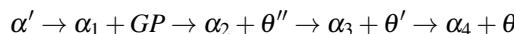


Figure 6.11: Variation of free energy with particle size for coherent and semicoherent particles.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). Phase Transformations in Metals and Alloys, with permission from CRC press.

A good example exhibiting the competition between interfacial energy and strain energy during *NGT* is the precipitation of the strengthening phase in heat-treatable $Al - Cu$ alloys. The process is called aging and was briefly discussed in the beginning of this chapter (Figure 6.1). The process consists of (a) heating in the single-phase field of phase α , termed dissolution, (b) rapid cooling to room temperature, where a supersaturated $Al - Cu$ solid solution (phase α') is obtained and (c) heating at an intermediate temperature (aging) where precipitation of the phase $\theta(Al_2Cu)$ takes place. During aging the equilibrium phase θ does not form directly but follows the formation of a series of transition phases. The precipitation sequence is



GP zones form first from the supersaturated solid solution α' . They are zones rich in copper and fully coherent with the matrix phase α_1 . They were discovered by Guinier and Preston in 1938. The reason the GP zones form instead of the equilibrium phase θ is (a) their low interfacial energy due to coherency and (b) their low strain energy since they form as discs, both these effects resulting in a low energy barrier for nucleation. So despite the higher driving force for the precipitation of θ , the lower activation energy of GP zones leads to their formation in the early stages of the *NGT*. The formation of transition phases θ'' and θ' follows the formation of GP zones. They too possess a lower activation energy for nucleation than the equilibrium phase θ as they have an intermediate crystal structure between the phases α

and θ , which results in a high level of coherency. From the two transition phases, θ'' is fully coherent while θ' is semicoherent. Therefore during the evolution of the *NGT* the coherency is reduced. The particles start as fully coherent (*GP* zones and θ''), become semicoherent (θ') and finally evolve as incoherent (θ). This coherency loss during precipitation is caused by the strain energy, which increases with the size of the particles as they grow. Therefore, the system prefers to loose coherency in order to minimize the total free energy. More details regarding the effect of the precipitation sequence on the strengthening of the alloy will be discussed in Chapter 8.

6.4 GROWTH

6.4.1 GROWTH PROCESSES

Growth is a process, which follows the nucleation of the new phase. Growth seizes when the new phase obtains its equilibrium fraction, defined by the lever rule in the phase diagram. During growth the free energy is dissipated until it is minimized at thermodynamic equilibrium. In order to study the growth process, consider the formation of a new phase β within the matrix phase α . The growth of β is accomplished with the migration of the α/β interface in the matrix phase, which involves transfer of atoms from phase α to phase β . As long as growth is concerned, the α/β interfaces are distinguished in glissile and non-glissile interfaces.

Growth processes involving the movement of glissile interfaces have the following characteristics: the interfaces consist of dislocation arrays and the movement of these interfaces is accomplished by dislocation glide. No thermal activation is required for interface movement (athermal migration). The new phase forms by a shear deformation of the matrix phase. No diffusion is required and, therefore, there is no change in composition. The atoms cross the interface with cooperated displacements. This type of growth is the characteristic of martensitic transformations and mechanical twinning.

Growth processes involving the movement of non-glissile interfaces have the following characteristic: the interface movement is accomplished by random atomic jumps from one phase to the other, i.e., diffusion. The growth process requires thermal activation.

We will consider two growth cases. The first case is growth with no change in composition (as in massive transformations). The diffusion is fast and the growth rate is controlled by how fast the atoms cross the interface. Growth takes place under interface control. The second case is growth with a change in composition. The new phase β forms with a different composition than phase α . In this case growth can take place under interface control (when diffusion is fast) or under diffusion control (when diffusion is slow). Growth can also take place under mixed control, when atoms diffuse and cross the interface with similar rates.

Consider now the conditions of interface control as well as diffusion control for two cases involving (a) allotropic transformation and (b) precipitation of a new phase. Figure 6.12 depicts the phase diagram for the allotropic transformation $\alpha \rightarrow \beta$ (Fig-

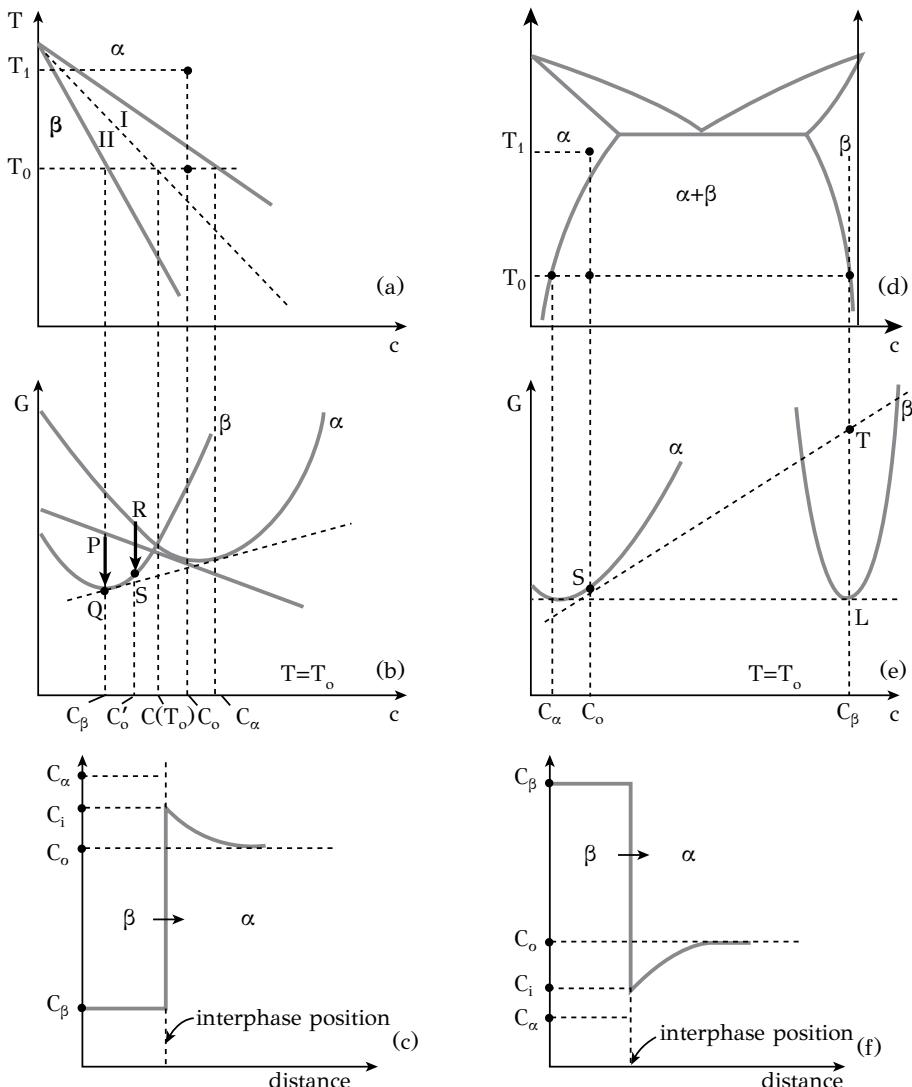


Figure 6.12: Growth during allotropic transformations: (a) phase diagram, (b) free energy curves, (c) concentration profile. Growth during precipitation: (d) phase diagram, (e) free energy curves, (f) concentration profiles.

ure 6.12a), the corresponding free energy-composition diagram for the two phases α and β (Figure 6.12b) as well as the composition profile of element B in the two phases (Figure 6.12c). The same figure depicts the phase diagram for the precipitation of phase β (Figure 6.12d), the corresponding free energy-composition diagram (Figure 6.12e) as well as the composition profile of element B (Figure 6.12f) during the growth of phase β . We start with the allotropic transformation $\alpha \rightarrow \beta$. In Figure 6.12a the alloy with composition c_o is cooled from temperature T_1 (phase α) to temperature T_o in the two-phase region $\alpha + \beta$. In Figure 6.12b the composition $c(T_o)$ is defined where $G_\alpha = G_\beta$ at T_o . If the alloy composition is lower than $c(T_o)$, e.g., c_o' , then it is possible for the transformation $\alpha \rightarrow \beta$ to take place with no change in composition. The driving force for this transformation is given by segment RS . In the phase diagram of Figure 6.12a, the region where it is possible to get such a transformation, upon cooling from T_1 is area II . In the actual case we have $c_o > c(T_o)$ and cooling is performed in area I of the phase diagram. In this case the transformation $\alpha \rightarrow \beta$ can proceed only by NGT and the nucleation driving force is given by segment PQ , where point P is on the tangent of the G_α curve at composition c_o . In Figure 6.12c the composition profile and the position of the interface α/β are depicted. When $c_i = c_o$ growth of phase β is interface controlled. Let c_i be the concentration of element B at the interface. When $c_i = c_a$ growth is diffusion controlled and when $c_o < c_i < c_\alpha$ the growth of phase β takes place under mixed control. In the case of diffusion-controlled growth, the compositions of α and β at the interface are c_α and c_β respectively, which correspond to the equilibrium compositions at T_o . Growth in this case proceeds under local equilibrium (LE). Consider now the case of precipitation of phase β from the supersaturated solid solution of phase α . In Figure 6.12d, the alloy with composition c_o is rapidly cooled from temperature T_1 (phase α) to temperature T_o where phase α is a supersaturated solid solution with composition c_o and a free energy given by point S in Figure 6.12e. At T_o precipitation of phase β will take place with a nucleation driving force given by TL . The composition profile of element B during growth is depicted in Figure 6.12f. When $c_i = c_o$ growth takes place under interface control, when $c_i = c_a$ growth is diffusion-controlled and when $c_\alpha < c_i < c_o$ growth proceeds under mixed control. It should be noted that in the case of allotropic transformation (Figure 6.12a-c) phase β is rich in component A and the interface moves with the diffusion of B from phase β to phase α . In the case of precipitation (Figure 6.12d-f) phase β is rich in component B and the growth of β proceeds with the diffusion of B from phase α to phase β .

6.4.2 INTERFACE-CONTROLLED GROWTH

Consider the case where the growing phase β has the same composition with phase α ; see for example (Figure 6.12b). The composition is in the area II of the phase diagram (Figure 6.12a). In this case since there is no change in composition, growth takes place under interface control. The growth rate depends on how fast the atoms cross the interface and move from one phase to the other. The growth process is thermally activated. The variation of free energy across the interface is depicted in Figure 6.13. The free energy change, due to the movement of one atom from phase

α to phase β , $\Delta G_{\beta\alpha}$, is the driving force for growth (segment RS in Figure 6.12b). $\Delta G_{\beta\alpha}^*$ is the activation energy for the atomic movement across the interface from phase α to phase β . We will consider the effect of the driving force on the growth rate. The flux of atoms from phase α to phase β is

$$J_{\alpha\beta} = n_\alpha v A \exp\left(-\frac{\Delta G_{\beta\alpha}^*}{kT}\right)$$

The flux of atoms from phase β to phase α is

$$J_{\beta\alpha} = n_\beta v A \exp\left(-\frac{\Delta G_{\beta\alpha} + \Delta G_{\beta\alpha}^*}{kT}\right)$$

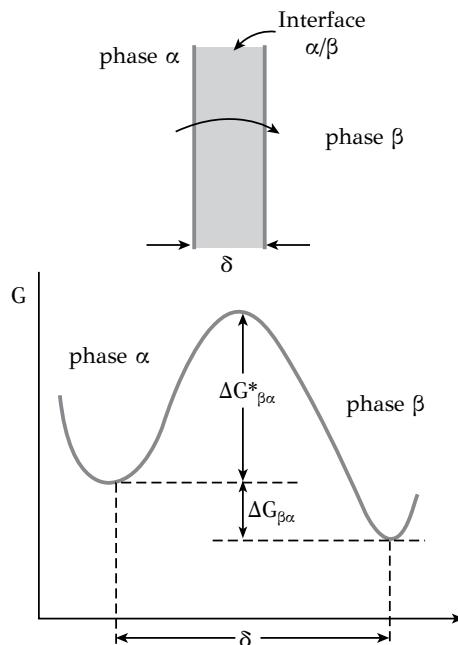


Figure 6.13: Variation of free energy across the interface between two phases for interface-controlled growth.

Where n_α, n_β the concentrations of phases α and β (in atoms/m^2), v is the Debye frequency ($\approx 10^{13} \text{ Hz}$) and A is the accommodation factor, which expresses the fraction of atoms departing from one phase and finding a position in the lattice of the other phase. For incoherent interfaces the accommodation factor is unity, while for semicoherent and coherent interfaces the accommodation factor is less than unity.

The net flux is $J = J_{\alpha\beta} - J_{\beta\alpha}$ and because $n_\alpha = n_\beta$ we get from the above relations

$$J = n_\alpha vA \exp\left(-\frac{\Delta G_{\beta\alpha}^*}{kT}\right) \left[1 - \exp\left(-\frac{\Delta G_{\beta\alpha}^*}{kT}\right)\right]$$

For low values of the driving force $\Delta G_{\beta\alpha}$ and using the approximate relation $e^x \approx 1 + x$ we get

$$J = n_\alpha vA \frac{\Delta G_{\beta\alpha}}{kT} \exp\left(-\frac{\Delta G_{\beta\alpha}^*}{kT}\right)$$

The growth rate u is proportional to the flux J according to the relation

$$u = J \frac{V_m}{N_A}$$

where V_m is the molar volume and N_A is the Avogadro number, then we get

$$u = \frac{n_\alpha v A V_m}{N_A k T} \Delta G_{\beta\alpha} \exp\left(-\frac{\Delta G_{\beta\alpha}^*}{kT}\right)$$

Setting

$$M_b = \frac{n_\alpha v A V_m}{N_A k T} \exp\left(-\frac{\Delta G_{\beta\alpha}^*}{kT}\right)$$

the growth rate becomes

$$u = M_b \Delta G_{\beta\alpha} \quad (6.9)$$

The quantity M_b is termed *interfacial mobility* of the interface α/β and should not be confused with the diffusional mobility M of Chapter 5. From Equation (6.9) it is concluded that the growth rate depends linearly on the driving force and is independent of time. The interfacial mobility M_b is exponentially dependent on temperature through an Arrhenius relation. In addition M_b depends on the accommodation factor and, therefore, on the interfacial structure (coherent, semicoherent or incoherent). If x is the position of interface α/β , the growth rate is

$$u = \frac{dx}{dt}$$

and from Equation (6.9) the position of the interface becomes

$$x = (M_b \Delta G_{\beta\alpha}) t \quad (6.10)$$

from which it is concluded that the size of the new phase is a linear function of time.

6.4.3 DIFFUSION-CONTROLLED GROWTH

Consider now the case where the particles of the growing phase β have different composition than phase α . In addition consider that the B atoms cross the interface α/β very fast and thus the growth rate depends on the diffusion rate of B towards the interface α/β . In this case growth takes place under diffusion control. We will study growth, taking place during the precipitation of the new phase β from the matrix phase α , as depicted in the phase diagram of Figure 6.12d. We make the assumption that the α/β interface is flat, incoherent and that local equilibrium (LE) conditions are established, so that the compositions of the two phases at the interface are given directly from the phase diagram. This condition is depicted in Figure 6.14a. Initially the composition of the alloy is c_o , but after cooling to T_o , the phase β nucleates with composition c_β . Consider that at time t the interface is at position x . We will calculate the interface velocity $u = dx/dt$ and the position of the interface $x(t)$ by considering the mass balance for component B . If it is supposed that the interface moves a distance dx in time dt then the quantity n_β of B that is incorporated in phase β per unit area (gr/cm^2) is

$$n_\beta = (c_\beta - c_\alpha)dx$$

while the corresponding quantity of B that is removed from phase α is

$$n_\alpha = D \frac{dc}{dx} dt$$

Equating the above quantities we get

$$(c_\beta - c_\alpha)dx = D \frac{dc}{dx} dt$$

so the interface velocity is

$$\frac{dx}{dt} = \frac{D}{c_\beta - c_\alpha} \frac{dc}{dx}$$

Assuming that the profile of B in phase α is linear and that L is the diffusion distance, the concentration gradient of B is

$$\frac{dc}{dx} = \frac{c_o - c_\alpha}{L}$$

and the interface velocity becomes

$$\frac{dx}{dt} = \frac{D}{c_\beta - c_\alpha} \frac{c_o - c_\alpha}{L}$$

The diffusion distance L can be determined by equating the hatched areas in Figure 6.14b

$$2x(c_\beta - c_o) = (c_o - c_\alpha)L$$

From the above two equations the velocity is

$$\frac{dx}{dt} = \frac{D}{c_\beta - c_\alpha} \frac{(c_o - c_\alpha)^2}{2x(c_\beta - c_o)}$$

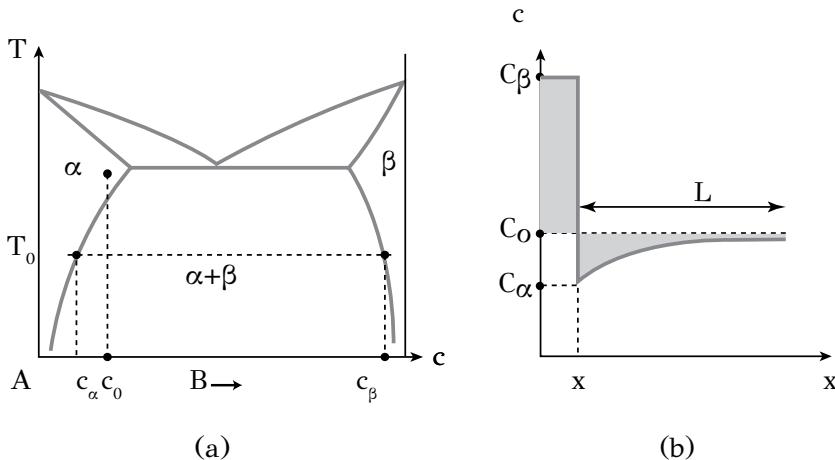


Figure 6.14: Diffusion-controlled growth: (a) binary phase diagram, (b) concentration profile for growth of β -phase under local equilibrium (LE) conditions.

Under the assumption

$$c_\beta - c_\alpha \simeq c_\beta - c_o$$

the above relation becomes

$$\frac{dx}{dt} = \frac{D}{2x} \left(\frac{c_o - c_\alpha}{c_\beta - c_\alpha} \right)^2$$

Taking the definition of the supersaturation Ω as

$$\Omega = \frac{c_o - c_\alpha}{c_\beta - c_\alpha}$$

and upon integrating the above relation, the position of the interface is

$$x = \Omega \sqrt{Dt} \quad (6.11)$$

and the velocity of the interface

$$u = \frac{\Omega}{2} \sqrt{\frac{D}{t}} \quad (6.12)$$

The relations 6.11 and 6.12 provide the position and the velocity of the interface α/β respectively. It can be seen from Equation (6.11) that growth depends on two terms. One is a thermodynamic term, the supersaturation Ω , which can be considered as the driving force for growth. The other term is a kinetic one, equal to \sqrt{Dt} , which is characteristic for diffusion control. From the relation 6.12 it is depicted that growth rate is reduced with time as the system approaches equilibrium. Summarizing the

above discussion, it is important to note that in growth under interface control, the size of the particle is a linear function of time and the growth rate is proportional to the driving force and interface mobility. In growth under diffusion control, the size of the particle is a parabolic function of time and the growth rate depends on the supersaturation and the diffusion coefficient. In both cases we have considered isothermal growth. Growth ceases when the fraction of phase β becomes equal to the equilibrium fraction predicted by the equilibrium phase diagram.

6.5 OVERALL RATE OF CONCURRENT NUCLEATION AND GROWTH

In most cases the formation of a new phase by *NGT* occurs with concurrent nucleation and growth. While particles of the new phase grow, new particles nucleate in the untransformed regions. In addition, growth can be hindered either due to a reduction in supersaturation or due to impingement of particles which nucleate close to each other. In the previous sections we have considered nucleation and growth independently. However when nucleation and growth operate concurrently, the overall transformation rate is affected by the interaction between these two processes. For the precipitation reaction $\alpha \rightarrow \alpha + \beta$ the overall transformation rate is expressed by the fraction f_β of the new phase as a function of time. The normalized fraction $f = f_\beta / f_\beta^{eq}$ will be used in the analysis, f_β^{eq} being the equilibrium fraction predicted by the phase diagram. Consider first the dependence of transformation rate on temperature. This relationship is shown schematically in Figure 6.15. When the precipi-

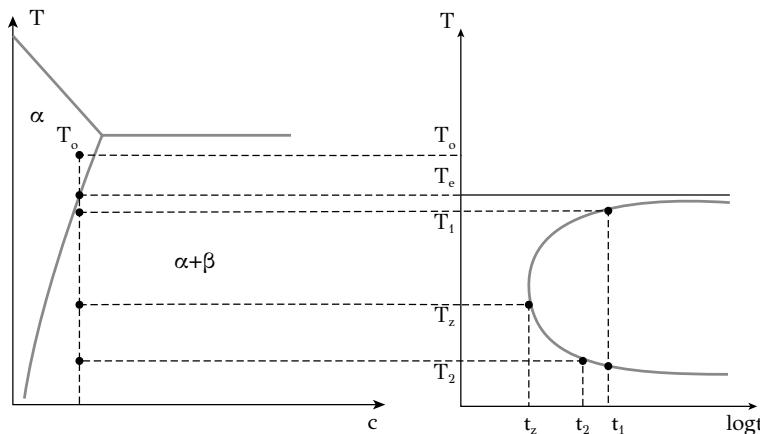


Figure 6.15: Temperature dependence of diffusional transformation rate: *C*—curve kinetics.

tation takes place at a relatively high temperature T_1 , under low supercooling below the solvus temperature T_e , diffusion is fast but the supersaturation, and therefore the driving force for nucleation and growth, is low. This leads to a low transformation rate (long transformation time t_1). When the precipitation takes place at low temper-

ature T_2 , then supersaturation is high but diffusion is slow. Again the transformation rate is low. The maximum transformation rate is accomplished at intermediate temperatures, e.g., T_z , where there is an optimum combination of supersaturation and diffusion. The curve depicting the overall transformation rate is C -shaped and therefore the kinetics of the transformation is usually termed C -curve kinetics. C -curves form the basis of the well-known isothermal transformation diagrams (TTT or IT), which depict the fraction f as a function of temperature and time. Such a diagram is depicted in Figure 6.16a with two C -curves, one corresponding to the start ($f = 0$) and the other to the end ($f = 1$) of the transformation. The fraction of transformation $f(t)$ is depicted in Figure 6.16b for two temperatures, T_1 where the transformation rate is low, and T_z where the transformation rate is maximized. The isothermal transformation diagrams find useful applications in the description of phase transformation in steels, and will be discussed in detail in another chapter of this book.

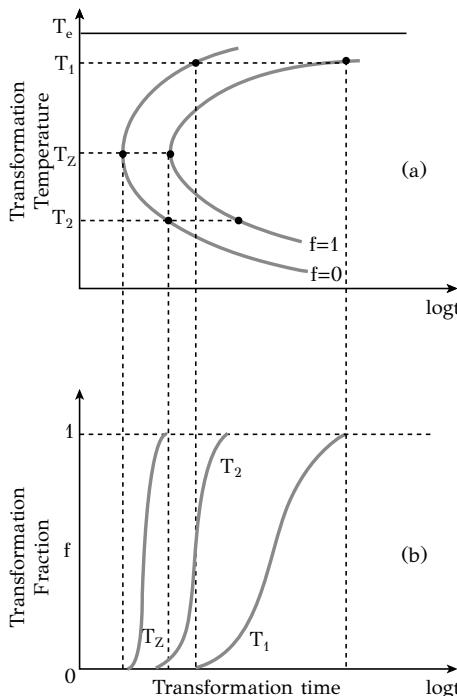


Figure 6.16: (a) Isothermal transformation diagram, (b) transformation curves for temperatures T_1 , T_2 and T_z .

Consider now the development of a model for the evolution of phase fraction of the new phase vs time $f(t)$ as in Figure 6.16b. The curve is sigmoidal in shape (Figure 6.17a) and spans three time periods: (a) induction, (b) growth and (c) impingement. The induction period corresponds to the time $t = \tau$ for nucleation of the

new phase. The impingement period corresponds to the retardation of growth due to the interaction of particles of phase β (Figure 6.17b). The evolution of phase fraction $f(t)$ depends on several factors including nucleation rate, distribution of nucleation sites and growth rate. Consider a volume V of material, which initially is occupied by phase α , $V = V_\alpha$. During transformation $V = V_\alpha + V_\beta$ and finally at the end of transformation $V = V_\beta$ so that the volume fraction $f = V_\beta/V$ of phase β varies from 0 to 1. Assume also a completely random distribution of nucleation sites. If I is the nucleation rate, then the number of particles forming in the period between t and $t + d\tau$ is $IV_\alpha d\tau$, where V_α is the volume of the untransformed phase α . At time t , a particle nucleated at the induction period τ has grown to a volume

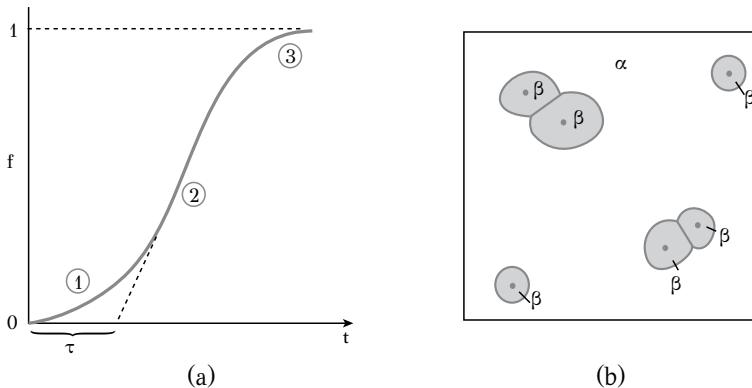


Figure 6.17: (a) Isothermal transformation curve depicting the periods of induction (1), growth (2) and impingement (3). (b) Graphical representation of impingement of β -phase particles.

$$v_\tau = \frac{4}{3}\pi[u(t-\tau)]^3 \quad (6.13)$$

where u is the growth rate. The assumption here is that growth is isotropic and linear, so the particle radius at time t can be expressed by the product $u(t - \tau)$. In the early stages of transformation $V_\beta \ll V_\alpha$ and impingement can be ignored, so the increase in volume of phase β due to particles formed in the period τ is

$$dV_\beta = v_\tau IV_\alpha d\tau$$

Setting $V_\alpha \approx V$, the volume of phase β is given by integration

$$V_\beta = \frac{4}{3}\pi V \int_0^t Iu^3(t-\tau)^3 d\tau$$

If now we assume that the nucleation rate is constant then

$$f = \frac{\pi}{3}Iu^3 t^4$$

The above relation describes the evolution of the volume fraction at the early stages of the transformation ($f \ll 1$). In the later stages of transformation, impingement of particles of phase β has to be taken into account. When two particles of phase β meet during growth, forming an interface β/β , their growth is retarded. The effect of impingement on $f(t)$ has been studied by Johnson and Mehl¹ and also by Avrami². Consider that in time $d\tau$, in addition to the $IV_\alpha d\tau$ nuclei formed in phase α , there is formation of $IV_\beta d\tau$ fictitious nuclei of phase β in the already transformed region, i.e., within phase β . These fictitious nuclei represent nuclei, which could form if no nucleation had already taken place in this region. The extended volume $V_{\beta e}$ of phase β can be defined by

$$dV_{\beta e} = v_\tau IV_\alpha d\tau + v_\tau IV_\beta d\tau$$

The extended volume $V_{\beta e}$ differs from the real volume V_β due to the fictitious nuclei formed in the transformed region of phase α . We also assumed in this case that growth takes place with no impingement, therefore the extended volume is greater than V_β . A relation is therefore needed between $V_{\beta e}$ and V_β . Consider an untransformed region with fraction $1 - f = 1 - V_\beta/V$ at time t . During the next period $d\tau$ the extended volume increases by $dV_{\beta e}$ while the real volume by dV_β . A fraction $1 - V_\beta/V$ of the extended volume corresponds to untransformed regions and therefore can contribute to the real volume dV_β so

$$dV_\beta = \left(1 - \frac{V_\beta}{V}\right) dV_{\beta e}$$

and by integration

$$V_{\beta e} = -V \ln \left(1 - \frac{V_\beta}{V}\right) = -V \ln (1 - f)$$

finally

$$-\ln (1 - f) = \frac{4}{3} \pi u^3 \int_0^t I(t - \tau)^3 d\tau$$

The above equation is known as the Johnson-Mehl equation, which can be integrated under certain assumptions regarding the nucleation rate. If the nucleation rate is constant then the above equation becomes

$$f = 1 - \exp \left(-\frac{\pi}{3} u^3 I t^4\right) \quad (6.14)$$

which is valid for all times. However in most cases the nucleation rate is not constant, especially in alloy systems where there is a consumption of nucleation sites, as in heterogeneous nucleation at grain boundaries. If N_v^α is the initial nucleation site

¹W. Johnson and R. Mehl, Trans. AIME 135 (1939) 416

²M. Avrami, J. Chem. Phys. 7 (1939) 1103

density (per unit volume of phase α) and N_v the nucleation site density remaining after time t , then in the time period dt there will be a consumption of $dN_v = -N_v v_1 dt$ nucleation sites. Here the rate v_1 denotes the rate at which a nucleation site is activated forming a nucleus of phase β . The number of nucleation sites at time t is therefore

$$N_v = N_v^0 \exp(-v_1 t)$$

and the nucleation rate is

$$I(t) = -\frac{dN_v}{dt} = N_v^0 v_1 \exp(-v_1 t)$$

Introducing the nucleation rate above into the Johnson-Mehl equation and integrating, the fraction becomes

$$f = 1 - \exp \left\{ \left(\frac{-8\pi N_v^0 u^3}{v_1^3} \right) \left[\exp(-v_1 t) - 1 + v_1 t - \frac{v_1^2 t^2}{2} + \frac{v_1^3 t^3}{6} \right] \right\} \quad (6.15)$$

Consider now two limiting cases concerning the term $v_1 t$. If $v_1 t$ is small then the consumption of nucleation sites is slow. Using the relation

$$e^{-v_1 t} \approx 1 - v_1 t + \frac{v_1^2 t^2}{2} - \frac{v_1^3 t^3}{6} + \frac{v_1^4 t^4}{24}$$

the above fraction becomes

$$f = 1 - \exp \left(\frac{-8\pi N_v^0 u^3}{v_1^3} \frac{v_1^4 t^4}{24} \right)$$

Setting

$$N_v^0 v_1 = I$$

the fraction becomes

$$f = 1 - \exp \left(-\frac{\pi}{3} I u^3 t^4 \right) \quad (6.16)$$

If the term $v_1 t$ is large then the nucleation rate drops very fast and the term

$$v_1^3 t^3 / 6$$

dominates in Equation (6.15) and the fraction becomes

$$f = 1 - \exp \left\{ -\frac{4}{3} \pi N_v^0 u^3 t^3 \right\} \quad (6.17)$$

Summarizing, Equation (6.14) describes $f(t)$ for constant nucleation rate, while Equations (6.16) and 6.17 are valid for slow and fast consumption of nucleation sites respectively. The above equations can be generalized as

$$f = 1 - \exp(-kt^n) \quad (6.18)$$

which is known as the Johnson-Mehl-Avrami-Kolmogorov (*JMAK*) equation for transformation kinetics. The parameter k depends on nucleation and growth rates and $3 \leq n \leq 4$. Equation (6.18) has a general validity and describes all stages of concurrent nucleation and growth transformations under linear growth conditions. It also describes the early stages of transformations under conditions of parabolic growth, where the particle radius is proportional to $t^{1/2}$. The exponent n for various cases of growth is given in Table 6.1.

Table 6.1
Values of the exponent in the *JMAK* kinetic equation

**Interface-controlled growth,
allotropic transformations, eutectoid transformations**

Constant nucleation rate	4
Nucleation rate increasing	> 4
Nucleation rate decreasing	3 – 4
Saturation of nucleation sites	3
Nucleation at grain edges	2
Nucleation at grain surfaces	1

Diffusion-controlled growth

Constant nucleation rate	2.5
Nucleation rate increasing	> 2.5
Nucleation rate decreasing	1.5 – 2.5
Saturation of nucleation sites	1.5
Nucleation of needles and plates	1
Nucleation on dislocations (early stages)	0.75

Reprinted from: J. W. Christian, The theory of transformations in metals and alloys, Pergamon Press, 1975, with permission from Elsevier.

6.6 COARSENING

As discussed in the previous section, the precipitation of a new phase β from a supersaturated solid solution of phase α takes place by nucleation and growth. As depicted in Figure 6.4 the alloy is cooled at T_o where phase α is a supersaturated solid solution. Precipitation of phase β takes place by nucleation and growth. When equilibrium has been reached, phase β has formed as a dispersion of particles within phase α . The phase fractions are given by the lever rule as $f_{\alpha}^{eq} = (c_{\beta} - c_o)/(c_{\beta} - c_{\alpha})$, $f_{\beta}^{eq} = (c_o - c_{\alpha})/(c_{\beta} - c_{\alpha})$ while the total free energy change is $\Delta G_o = AB$ (Figure 6.4). However in constructing the equilibrium phase diagram, the interfacial energy α/β is not taken into account. In reality, the microstructure, at the end of the growth process, remains metastable, since the system (matrix and particles) has not mini-

mized its total interfacial energy. A microstructure with a fine dispersion of particles (a large number of small particles) will therefore transform to a microstructure with a coarse dispersion of particles (a small number of large particles) in order to minimize the total interfacial area α/β . The transformation of a fine to a coarse dispersion under constant equilibrium volume fraction f_{β}^{eq} of phase β is termed coarsening. In precipitation-strengthened systems, such as high strength aluminum-copper alloys, coarsening has an undesirable effect on strength. During coarsening the interparticle spacing, as well as particle size increase. This reduces the ability of the particle dispersion to pin dislocations and the strength drops. In addition, the coarsening rate increases with temperature and this should be taken into account when designing alloys for high-temperature service. It is important, therefore, to study the thermodynamics and kinetics of coarsening and identify the key parameters affecting the coarsening rate. The driving force for coarsening is the reduction of interfacial energy, through the reduction of the total interfacial area α/β . Consider now the effect of curvature of the α/β interface on free energy (for a relevant discussion see also Chapter 5, diffusion under non-chemical forces). Assuming that phase β has spherical shape, Figure 6.18a depicts free energy - composition curves of phases α and β for particle radius $r = \infty$, $r = r_1$ and $r = r_2$, with $r_2 < r_1$. The $r = \infty$ curve is the one used for the construction of the equilibrium phase diagram with no curvature effects. As depicted in the figure, the free energy depends on the curvature of the α/β interface and $G^{\beta}(r_2) > G^{\beta}(r_1) > G^{\beta}(\infty)$ indicating that the free energy increases with a reduction of particle radius. This has been called the Gibbs-Thomson effect. A direct consequence is that the composition of phase α in equilibrium with phase β also depends on the particle radius. This is because the common tangent between the curve of phase α and the curves of $\beta(r_1)$ and $\beta(r_2)$ define a different contact point in the curve of phase α (Figure 6.18a). In this case, the composition of phase α in equilibrium with $\beta(r = \infty)$ is c_{∞} , while the composition of phase α in equilibrium with $\beta(r_1)$ and $\beta(r_2)$ is c_1 and c_2 respectively. The increase in free energy of phase β with a radius of curvature r is

$$\Delta G = \frac{2\gamma V_m}{r}$$

where γ are the interfacial energy of the α/β boundary and V_m the molar volume of phase β .

The composition c_r of phase α in equilibrium with phase $\beta(r)$ is

$$c_r = c_{\infty} \exp \left(\frac{2\gamma V_m}{kTr} \right)$$

where c_{∞} is the composition of phase α in equilibrium with phase $\beta(r = \infty)$. Due to the Gibbs-Thomson effect a concentration gradient develops between particles r_1 and r_2 (Figure 6.18b). Because $c_2 > c_1$ there is diffusion of component B from the small to the large particle through the matrix of phase α . In this way the small particles get even smaller while the large particles get larger and the mean radius r increases

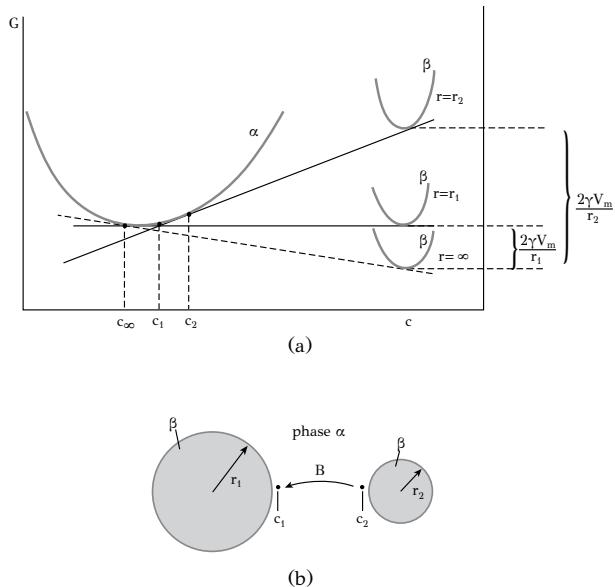


Figure 6.18: Particle coarsening: (a) free energy curves depicting equilibrium of α -phase with β -phase possessing different curvatures, (b) diffusion of alloying element B from the small to the large particles.

with time. The kinetics of coarsening has been studied by Greenwood¹ who gives the change of particle radius as

$$\frac{dr}{dt} = Dc_\infty \frac{2\gamma V_m}{kT} \frac{1}{r} \left(\frac{1}{\bar{r}} - \frac{1}{r} \right)$$

The above equation is shown in Figure 6.19 for two values of the mean radius \bar{r}_1 and \bar{r}_2 . Particle with radius $r < \bar{r}$ have $dr/dt < 0$, and those particles shrink. On the other hand, particles with $r > \bar{r}$ have $dr/dt > 0$ and these particles grow. Particles with $r = 2\bar{r}$ grow with the maximum rate. In addition, finer particle dispersions exhibit a higher coarsening rate, as for example the dispersion with mean radius \bar{r}_1 relative to the dispersion with mean radius \bar{r}_2 . The evolution of the mean radius of particle dispersions during coarsening has been studied by Lifshitz and Slyozov², as well as by Wagner³. Their theory, known as the LSW coarsening theory, describes the evolution of particle size distribution with time and provides the kinetics of coarsening as

$$\bar{r}^3 - \bar{r}_o^3 = \frac{8D\gamma V_m c_\infty}{9kT} t \quad (6.19)$$

¹G.W. Greenwood, Acta Metall. 4 (1956) 243

²I.M. Lifshitz and V.V. Slyozov, J. Phys. Chem. Solids 19 (1961) 35

³C. Wagner, Z. Electrochem. 65 (1961) 581

where \bar{r}_o is the mean curvature before coarsening. As depicted from the above expression, the kinetics of coarsening depends on the diffusion coefficient, the composition c_∞ of phase α and interfacial energy γ . The coarsening rate depends strongly on temperature, since, both D and c_∞ , increase exponentially with temperature. Equation (6.19) depicts that coarsening exhibits a $t^{1/3}$ kinetics compared with the $t^{1/2}$ kinetics exhibited by diffusional growth. An example of particle coarsening is shown in Figure 6.20a, which depicts the same dispersion of Mo_2C carbides of Figure 6.6a but at a higher temperature where coarsening is taking place. A later stage of coarsening is depicted in Figure 6.20b, where the initial fine dispersion has been transformed to a coarser dispersion with a severe loss of strength. Alloys for high-temperature service should exhibit resistance to coarsening. A good example is the Ni -base superalloys strengthened by a dispersion of Ni_3Al particles of phase γ' . This phase has *FCC* structure as the Ni matrix and, therefore, the interface boundaries between the precipitates and the matrix are coherent. This leads to a low interfacial energy γ and low coarsening rate even at high temperatures. A second example is the strengthening of Ni alloys by ThO_2 . These oxides are practically insoluble in Ni and the value c_∞ in Equation (6.19) is very low, indicating a correspondingly low coarsening rate. Therefore these alloys can maintain high strength even at high temperatures.

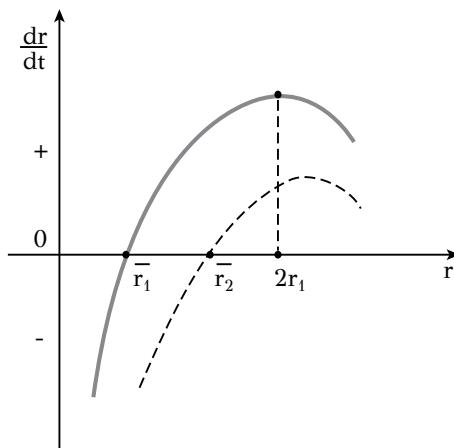


Figure 6.19: The coarsening rate for two values of the mean radius according to the Greenwood theory.

6.7 CONTINUOUS TRANSFORMATIONS

In this section we will discuss continuous transformations, which are accomplished with a gradual change in composition. Characteristic examples of this class of reactions are spinodal decomposition and order-disorder transformations. First we will discuss the distinction between sharp and diffuse interfaces between two phases.

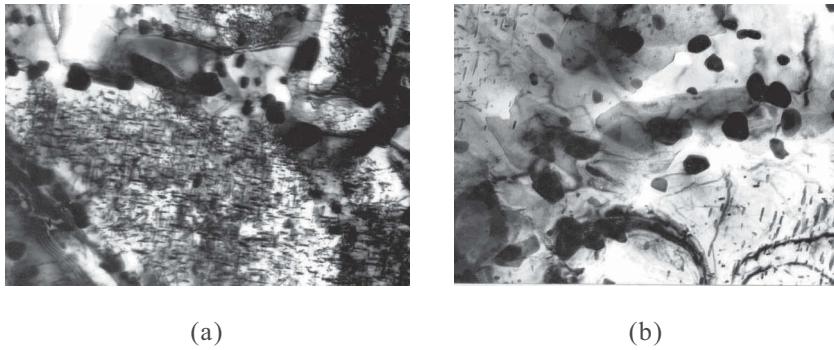


Figure 6.20: (a) Coarsening of a dispersion of Mo_2C carbides in a Mo -containing steel. (b) Advanced stage of coarsening where the fine dispersion in (a) has evolved to a much coarser dispersion.

6.7.1 SHARP AND DIFFUSE INTERFACES

Consider Figure 6.3, which depicts the decomposition of a homogeneous solid solution into two phases α and β with nucleation and growth (Figure 6.3a) and with continuous transformation (Figure 6.3b). A snapshot from the evolution of these transformations is shown in Figure 6.21. In nucleation and growth transformations, the phases α and β are separated by a sharp interface with an abrupt change in composition, while in continuous transformations the two phases are separated by a diffuse interface with a gradual change in composition. We can say that sharp interfaces possess zero thickness while diffuse interfaces possess a finite thickness. We have seen, that sharp interfaces are characterized by a specific interfacial energy γ , the value of which depends on the interfacial structure (coherent, semicoherent or incoherent). We have also seen that the interfacial energy has a direct influence on the energy barrier for nucleation (and hence on nucleation kinetics) and the morphology of the new phase. In diffuse interfaces, a corresponding “interfacial energy” must be defined in order to express the energetics of the interface. This “interfacial energy” is the result of compositional fluctuations with a short wave length, when the concentration profile $c(x)$ exhibits intense gradient and curvature (Figure 6.21b). In this way the region at position x_0 “feels” the variation of composition at its immediate environment (point c' instead of point c) due to the curvature of $c(x)$. As a result, the free energy is different from the one that the system could possess if the composition was uniform. The free energy of a solid solution, which exhibits compositional fluctuations in one direction, $c(x)$, can be calculated if we consider the Taylor expansion in two variables around $G(c_0)$, with c_0 being the average composition

$$G(\xi, \eta) = G(c_0) + \xi \frac{\partial G}{\partial \xi} + \eta \frac{\partial G}{\partial \eta} + \frac{1}{2} [\xi^2 \frac{\partial^2 G}{\partial \xi^2} + \eta^2 \frac{\partial^2 G}{\partial \eta^2} + 2\xi \eta \frac{\partial^2 G}{\partial \xi \partial \eta}] + \dots$$

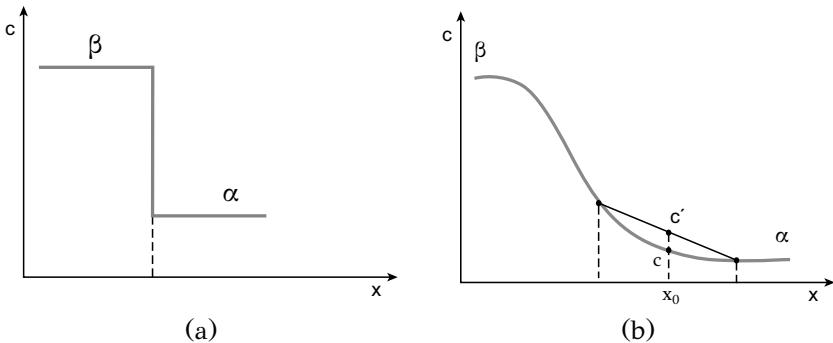


Figure 6.21: (a) Sharp interface between two phases in nucleation and growth transformations (NGT), (b) diffuse interface between phases α and β during continuous transformations (CT).

where the variables ξ and η are the first and second derivatives of $c(x)$ respectively

$$\xi = \frac{dc}{dx}$$

and

$$\eta = \frac{d^2c}{dx^2}$$

and the above equation becomes

$$G = G(c_0) + a_1 \frac{dc}{dx} + a_2 \frac{d^2c}{dx^2} + a_3 \left(\frac{dc}{dx} \right)^2$$

where

$$a_1 = \frac{\partial G}{\partial (dc/dx)}$$

$$a_2 = \frac{\partial G}{\partial (d^2c/dx^2)}$$

$$a_3 = \frac{1}{2} \frac{\partial G}{\partial (dc/dx)^2}$$

Due to the symmetry of cubic crystals, the coefficient a_1 should have a zero value since a change of sign of the x axis could change the value of G . The total free energy is obtained by integrating over the volume V

$$G = \int_V \left[G(c_0) + a_2 \frac{d^2c}{dx^2} + a_3 \left(\frac{dc}{dx} \right)^2 \right] dV$$

Integration of the second term above gives

$$\int_V a_2 \frac{d^2c}{dx^2} dV = a_2 \frac{dc}{dx} \Big|_V - \int_V \frac{da_2}{dc} \left(\frac{dc}{dx} \right)^2 dV$$

For the same reasons mentioned above, the first term in the *rhs* is zero and the free energy of the heterogeneous system becomes

$$G = \int_V \left[G(c_0) + K \left(\frac{dc}{dx} \right)^2 \right] dV \quad (6.20)$$

The term $K(\partial c / \partial x)^2$ is called *gradient energy* and expresses the additional free energy of the system arising from compositional fluctuations. In a sense, it represents the interfacial energy of the diffuse interface of Figure 6.21. We will use this term in the discussion of spinodal decomposition in the next section.

6.7.2 SPINODAL DECOMPOSITION

According to the discussion in section 6.1, a phase transformation can be initiated by a low-intensity composition fluctuation. The amplification of this fluctuation can lead to decomposition of a homogeneous solid solution in two phases (Figure 6.3b). We will discuss first the conditions under which such a transformation can take place. Consider the free energy curve of phase α (solid solution), with a positive enthalpy of mixing ΔH_m in Figure 6.22. At a high temperature T_o the solid solution is homogeneous with a negative free energy of mixing ΔG_m at all compositions, because of the entropic contribution $-T\Delta S_m$ (Figure 6.22a). The system is then brought to a lower temperature T_1 where the entropic contribution is smaller. Due to the positive enthalpy of mixing, the ΔG_m curve exhibits a characteristic arch shape (Figure 6.22b). In this case the solid solution is decomposed in two phases α_1 and α_2 while the phase diagram exhibits a miscibility gap. However the decomposition mechanism depends on the composition of phase α . Consider the case where the alloy composition is c_{o1} . A small composition fluctuation around c_{o1} increases the free energy and therefore, composition c_{o1} corresponds to an unstable state. Decomposition in this case requires an intense composition fluctuation, i.e., regions rich in B should form with composition c_{a2} , leading to a decrease in free energy. In other words the decomposition takes place by thermally-activated nucleation and growth (*NGT*). Now consider the case where the alloy composition of phase α is c_{o2} . Any small composition fluctuation around c_{o2} decreases the free energy and decomposition can proceed spontaneously. The fluctuations grow until the system decomposes in two phases α_1 and α_2 , with compositions $c_{\alpha1}$ and $c_{\alpha2}$ respectively. The two phases do not form by nucleation and growth but rather with a continuous change of composition (Figure 6.3b), which is the key feature of continuous transformations (*CT*). Composition c_{o1} , in Figure 6.22b, is in a region where the free energy curve of phase α is convex (positive curvature) while composition c_{o2} is in a region where the free energy curve is concave

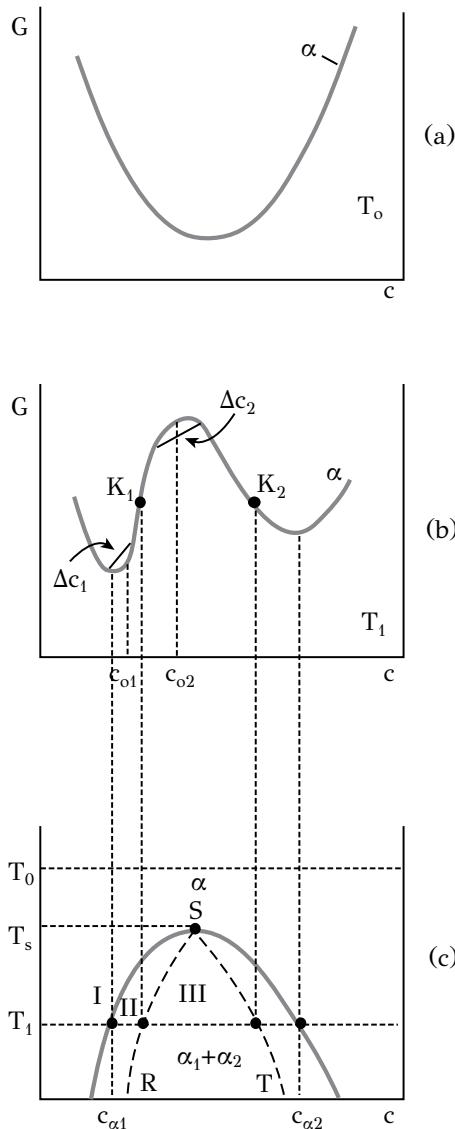


Figure 6.22: The thermodynamics of spinodal decomposition: (a) free energy at T_o . (b) Free energy curve at T_1 with K_1 and K_2 being the spinodal points on the G -curve. (c) Corresponding phase diagram with a miscibility gap. The RST curve (dashed line) is the chemical spinodal within the miscibility gap.

(negative curvature), between points K_1 and K_2 . These points are the spinodal points of the free energy curve where the second derivative of G with respect to composition is zero, i.e., $G'' = 0$. Between K_1 and K_2 , $G'' < 0$ while outside the region $K_1 - K_2$, $G'' > 0$. The spinodal points K_1 and K_2 are defined for the specific temperature T_1 . The locus of spinodal points with temperature defines the spinodal curve RST inside the miscibility gap of Figure 6.22c. The spinodal curve, defined by $G'' = 0$, is termed the chemical spinodal. In this case three regions are distinguished in the phase diagram. In region I the solid solution of phase α is homogeneous and thermodynamically stable. In region II, with $G'' > 0$, the solid solution is decomposed in two phases α_1 and α_2 by nucleation and growth. In region III, where $G'' < 0$, the decomposition proceeds with a continuous change in composition. This transformation is termed spinodal decomposition. Examples of alloy systems exhibiting spinodal decomposition are the binaries $Au - Ni$, $Bi - Zn$, $Al - Zn$ and the ternaries $Al - Ni - Co$ and $Cu - Ni - Fe$. The theory of spinodal decomposition has been developed by Cahn and Hilliard and is based on the solution of the diffusion equation. According to the discussion in Chapter 5, the diffusional flux is proportional to the chemical potential gradient

$$J = -M \frac{\partial \mu}{\partial x}$$

where M is the diffusional mobility. Because $\mu = \partial G / \partial c$, we get $\partial \mu / \partial x = (\partial \mu / \partial c)(\partial c / \partial x)$ and finally $\partial \mu / \partial x = G''(\partial c / \partial x)$. The above equation becomes

$$J = -MG'' \frac{\partial c}{\partial x}$$

Considering Fick's first law

$$J = -\tilde{D}(\partial c / \partial x)$$

the interdiffusion coefficient is

$$\tilde{D} = MG''$$

and because $M > 0$, the sign of \tilde{D} depends on the sign of G'' . Inside the spinodal (region III in Figure 6.22c) $G'' < 0$ and therefore, $\tilde{D} < 0$, meaning that spinodal decomposition proceeds with up-hill diffusion, against the concentration gradient. Component A diffuses from regions poor in A to regions rich in A and correspondingly for component B. The composition fluctuations of Figure 6.3b, during spinodal decomposition, can be determined from the solution of Fick's second law, which is written as

$$\frac{\partial c}{\partial t} = \frac{MG''}{N_V} \frac{\partial^2 c}{\partial x^2}$$

where N_V is the number of atoms per unit volume. The solution of the above equation takes the form

$$c - c_o = A(\beta, t) \cos(\beta x) \quad (6.21)$$

where $A(\beta, t)$ is the intensity of the fluctuation with exponential time dependence

$$A(\beta, t) = A(\beta, 0) \exp[R(\beta)t] \quad (6.22)$$

and $R(\beta)$ is the amplification factor

$$R(\beta) = -\frac{M}{N_V} G'' \beta^2 \quad (6.23)$$

β is the wavenumber, $\beta = 2\pi/\lambda$, where λ is the wavelength of the composition fluctuation. Inside the spinodal $R(\beta) > 0$ for all values of β , i.e., all fluctuations are amplified irrespectively of wavelength. However this is not the case since fluctuations with a short λ do not grow because they are associated with high gradient energy. Outside the spinodal, $R(\beta) < 0$, and all composition fluctuations decay. Therefore Equation (6.23) cannot predict the decomposition by nucleation and growth and a modification is needed. A first modification concerns the free energy change due to very short wavelength composition fluctuations, where the composition profile $c(x)$ exhibits intense gradient and curvature (Figure 6.23). It is the gradient energy term $K(\partial c / \partial x)^2$, discussed in the previous section, and corresponds to the “interfacial energy” of the diffuse interface between the two phases. The modified free energy is given by Equation (6.20) as

$$G = \int_V \left[G(c_0) + K \left(\frac{dc}{dx} \right)^2 \right] dV$$

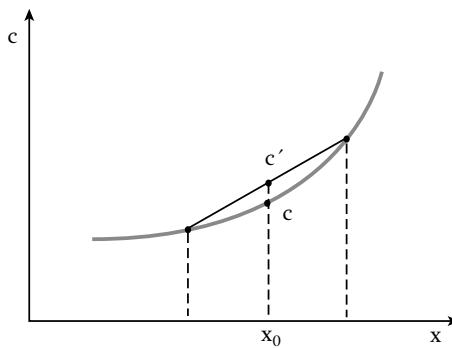


Figure 6.23: Compositional profile with intense curvature generating an “interfacial energy” of the diffuse interface corresponding to the gradient energy.

and the modified chemical potential

$$\mu = \frac{\partial G}{\partial c} - 2K \frac{\partial^2 c}{\partial x^2}$$

The above equation shows that when the composition profile is convex, the chemical potential is lower than it would be for a uniform composition profile. The diffusional flux becomes

$$J = -M \left[G'' \frac{\partial c}{\partial x} - 2K \frac{\partial^3 c}{\partial x^3} \right]$$

and Fick's second law becomes

$$\frac{\partial c}{\partial t} = \frac{M}{N_V} \left[G'' \frac{\partial^2 c}{\partial x^2} - 2K \frac{\partial^4 c}{\partial x^4} \right]$$

A second modification concerns the strain energy due to the maintenance of coherency between the two phases α_1 and α_2 in the early stages of spinodal decomposition. The coherency strain energy is $\eta Y(c - c_o)$ with $\eta = (1/\alpha)(d\alpha/dc)$ expressing the change of lattice parameter with composition and $Y = E/(1 - \nu)$. The diffusion equation becomes

$$\frac{\partial c}{\partial t} = \frac{M}{N_V} \left\{ \left[G'' + 2\eta^2 Y \right] \frac{\partial^2 c}{\partial x^2} - 2K \frac{\partial^4 c}{\partial x^4} \right\}$$

and using the interdiffusion coefficient

$$\frac{\partial c}{\partial t} = \tilde{D} \left\{ \left[1 + \frac{2\eta^2 Y}{G''} \right] \frac{\partial^2 c}{\partial x^2} - \frac{2K}{G''} \frac{\partial^4 c}{\partial x^4} \right\}$$

The solution of the above equation is given by Equations (6.21) and (6.22) but with an amplification factor

$$R(\beta) = -\frac{M}{N_V} [G'' + 2\eta^2 Y + 2K\beta^2] \beta^2 \quad (6.24)$$

or using the diffusion coefficient

$$R(\beta) = -\tilde{D} \left[1 + \frac{2\eta^2 Y}{G''} + \frac{2K}{G''} \beta^2 \right] \beta^2$$

For a given composition and temperature inside the spinodal, there is a critical wavenumber β_c (or a critical wavelength λ_c) giving $R(\beta_c) = 0$. Setting $R(\beta) = 0$ in the above equation the critical wavenumber is

$$\beta_c = \left(-\frac{G'' + 2\eta^2 Y}{2K} \right)^{1/2}$$

The graphical representation of $R(\beta)$ and $R(\lambda)$ are given in Figure 6.24. For $\lambda < \lambda_c$ we get $R(\lambda) < 0$ and the fluctuations decay. For $\lambda > \lambda_c$ we get $R(\lambda) > 0$ and the composition fluctuations are amplified. The fluctuation exhibiting the maximum amplification possesses a wavelength λ_{\max} . This wavelength dominates in the final microstructure consisting of a mixture of phases α_1 and α_2 (Figure 6.25). The

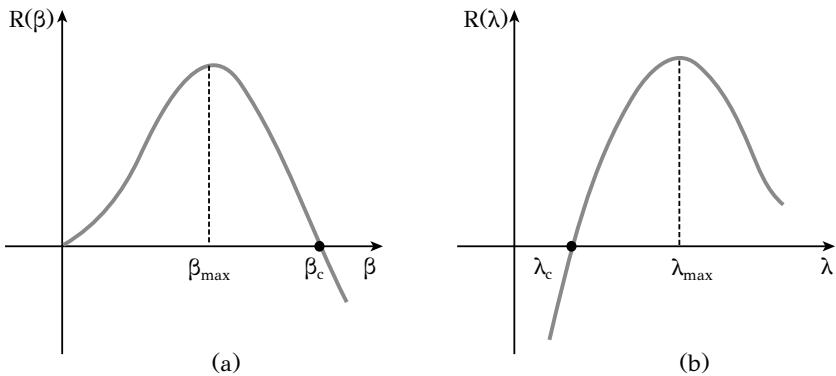


Figure 6.24: Variation of amplification factors of composition fluctuations with wavenumber and wavelength.

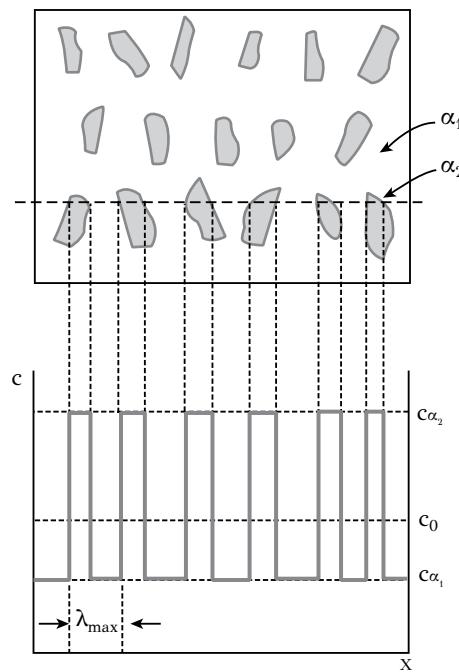


Figure 6.25: Schematic representation of microstructure after separation in two phases a_1 and a_2 with spinodal decomposition and corresponding composition profile across the dotted line.

wavelengths λ_c and λ_{\max} are related by $\lambda_{\max} = \lambda_c \sqrt{2}$. Typical values for λ_{\max} are 50 Å for $Al - Zn$ and 100 Å for $Al - Ag$.

Following the theory of spinodal decomposition presented above, consider the evolution of the microstructure. It is depicted from Equation (6.23) that any sinusoidal composition fluctuation will be amplified or decayed according to whether $R(\beta)$ is positive or negative. For compositions outside the spinodal $R(\beta)$ is negative for all values of β , composition fluctuations decay and the system is homogenized. For compositions inside the spinodal, $R(\beta)$ is positive for $\beta < \beta_c$ or $\lambda > \lambda_c$. The fluctuations with $R(\beta) > 0$ that decay, exhibit either high strain energy or high gradient energy. For this reason only the fluctuations with $\lambda > \lambda_c$ develop. Regarding strain energy, we should consider that in cubic crystals the factor Y is lower in directions [100] and [111], the so-called soft directions. It is in these directions that the fluctuations grow preferentially in order to minimize free energy. Another important issue regarding strain energy is its influence on thermodynamic equilibrium. Regular phase diagrams do not take into account coherency strains, which increase the free energy. It is then possible to define the elastic miscibility gap and, correspondingly, the coherent spinodal Figure 6.26. While the chemical spinodal is defined by the relation $G'' = 0$, the coherent spinodal is defined by the relation

$$G'' + 2\eta^2 Y = 0$$

The coherent spinodal lies within the arch of the chemical spinodal. As depicted from Equation (6.23), a prerequisite for spinodal decomposition is the term corresponding to the gradient energy $2K\beta^2$ be smaller than

$$|G'' + 2\eta^2 Y|$$

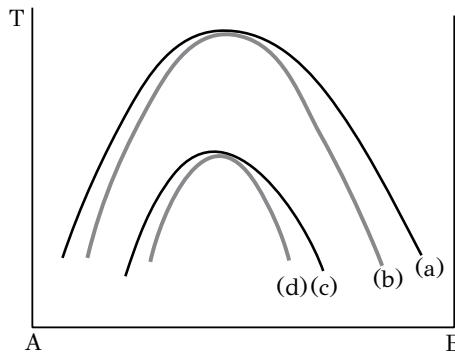


Figure 6.26: Schematic phase diagram with (a) miscibility gap, (b) chemical spinodal, (c) coherent miscibility gap and (d) coherent spinodal.

In this case the potential for spinodal decomposition is defined from the coherent spinodal, which is located at lower temperatures than the chemical spinodal. A good example is the system $Au - Ni$, which exhibits a wide miscibility gap (Figure 6.27).

In this system the term $2\eta^2Y$ is very large due to coherency strains, and the coherent spinodal is displaced 600°C lower than the chemical spinodal. Regarding the effect of temperature, it is depicted that at lower temperatures the decomposition takes place with shorter λ_{\max} and higher $R(\lambda)$ leading to a finer microstructure. In addition, because of the coherency between phases, the microstructure exhibits high resistance to coarsening. Spinodal decomposition has been observed in several systems including: Aluminum alloys: $Al - Cu$, $Al - Zn$, $Al - Li$, Ferrous alloys: $Fe - Mo$, $Fe - Al$, $Fe - Cr$, $Fe - Ni - Al$, $Fe - Ni - C$, $Fe - Al - Cr - Ni$, Copper alloys: $Cu - Ti$, $Cu - Zn$, $Cu - Co$, $Cu - Ni - Fe$, $Cu - Ni - Sn$, $Cu - Ni - Si$, $Cu - Ni - Cr$ and Nickel alloys: $Ni - Al$, $Ni - Cr$, $Ni - Mo$, $Ni - V$, $Ni - Si$.

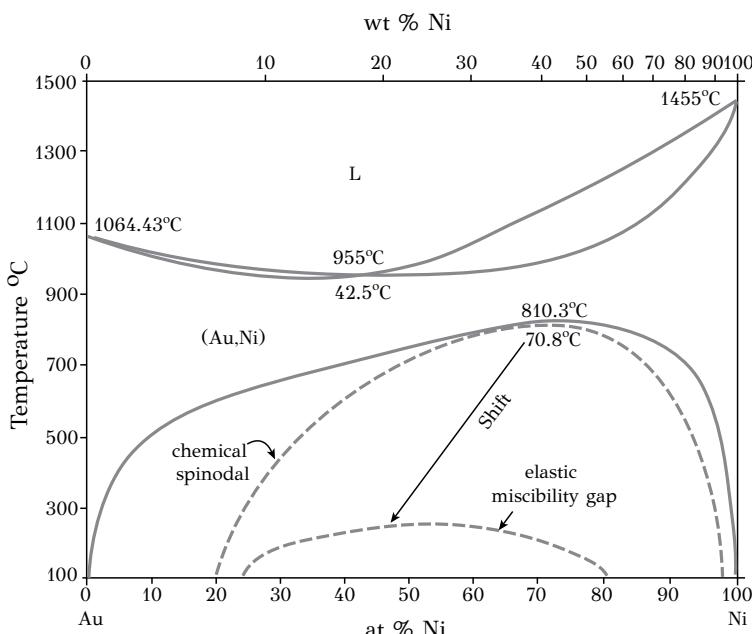


Figure 6.27: Binary phase diagram for the $Au - Ni$ system. The coherent spinodal is displaced by at least 600°C relative to the chemical spinodal.

An example of microstructure developed by spinodal decomposition is depicted in Figure 6.28. It concerns alloy $Cu - 27Ni - 9Fe$, which has been subjected to aging at 625°C for 1000 hours. The wavelength is 90nm. Several industrial alloys exhibit microstructures, developed by spinodal decomposition, with enhanced mechanical and physical properties. Examples include the $Al - Ni - Co$ magnetic alloys as well as $Cu - Ni - Fe$ and $Fe - Co - Cr$ alloys.

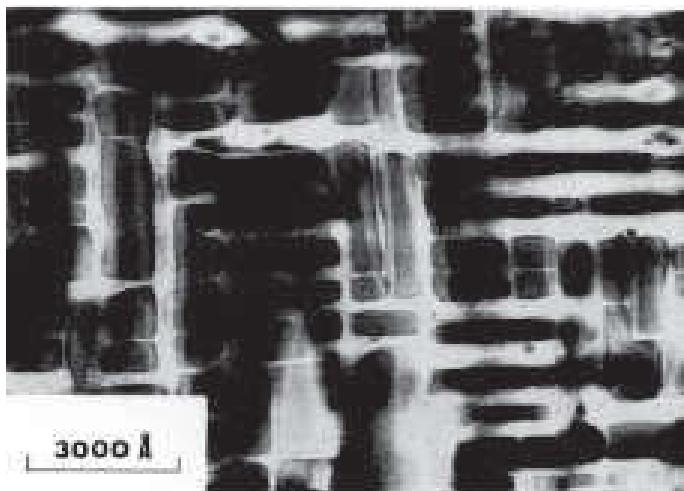


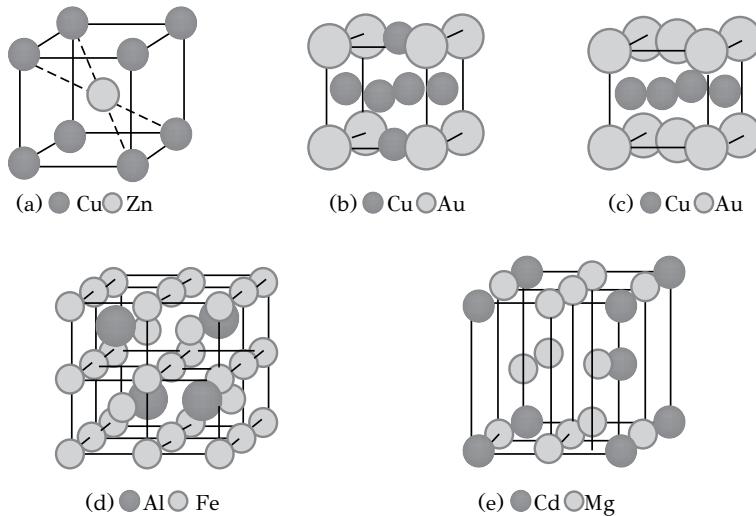
Figure 6.28: Spinodal microstructures in $Cu - Ni - Fe$ alloy with Ni -rich (black) and Cu -rich (white) regions. The wavelength is 90nm.

Reprinted from: R.J. Livuk and G. Thomas, Acta Metall. 19(1971)491, with permission from Elsevier.

6.7.3 ORDER-DISORDER TRANSFORMATIONS

In order-disorder transformations, an ordered phase is transformed to a disordered phase and vice versa. As discussed in Chapter 4, in solid solutions exhibiting a negative enthalpy of mixing ($\Delta H_m < 0$), $A - B$ bonding is preferred from $A - A$ and $B - B$ bonding. In this case there is a tendency to form ordered phases at low temperatures. Order is distinguished in short and long-range order. When the number of $A - B$ bonds is greater than $\frac{1}{2}ZNX_AX_B$ (see Chapter 4), the solid solution exhibits short-range order, *SRO*. On the other hand, in several solid solutions exhibiting simple proportionality ratios $x : y$ between A and B ($A : B = 1 : 1, 1 : 3, 3 : 1, \dots$) a long-range order, *LRO*, is observed. We will only consider phases with *LRO*. In such phases the atomic positions are not equivalent, as in disordered solutions, but refer to positions belonging exclusively to A or B atoms. In other words the A atoms occupy positions in the A sublattice while B atoms occupy positions in the B sublattice, these two sublattices forming a superlattice of the ordered phase A_xB_y . An example of a phase with *LRO* is phase $CuZn$ (Figure 6.29a) where the Cu atoms occupy the unit cell corners (A -positions) while the Zn atoms occupy the center of the unit cell (B -positions). Table 6.2 lists some of the most common ordered phases with the crystal structure of the corresponding disordered phase. The respective unit cells are depicted in Figure 6.29.

During order-disorder transformations, an ordered solid solution is transformed to a disordered one upon heating above a critical temperature T_c . Consider the phases $CuZn$ (β') and Cu_3Au at the respective phase diagrams $Cu - Zn$ (Figure 6.30) and $Cu - Au$ (Figure 6.31). For the $CuZn$ phase, the critical temperature is between 454

Figure 6.29: Ordered phases: (a) $CuZn$, (b) Cu_3Au , (c) $CuAu$, (d) Fe_3Al , (e) Mg_3Cd .

Reprinted from: R.E. Smallman, *Modern Physical Metallurgy*, Butterworths, London, 1970, with permission from Elsevier.

Table 6.2
The five most important groups of ordered phases

Str/richt notation	Crystal structure	Examples	Unit cell
$L2_0$	BCC (A2)	$CuZn, FeCo, FeAl, AlMg, NiAl$	Figure 6.29a
$L1_2$	FCC (A1)	$Cu_3Au, Au_3Cu, Ni_3Fe, Ni_3Mn, Ni_3Al, Pt_3Fe$	Figure 6.29b
$L1_0$	FCC (A1)	$CuAu, FePt, CoPt$	Figure 6.29c
$D0_3$	BCC (A2)	$Fe_3Al, Cu_3Al, Fe_3Si, Fe_3B$	Figure 6.29d
$D0_{19}$	HCP (A3)	$Mg_3Cd, Cd_3Mg, Ti_3Al, Ni_3Sn$	Figure 6.29e

Reprinted from: R.E. Smallman, *Modern Physical Metallurgy*, Butterworths, London, 1970, with permission from Elsevier.

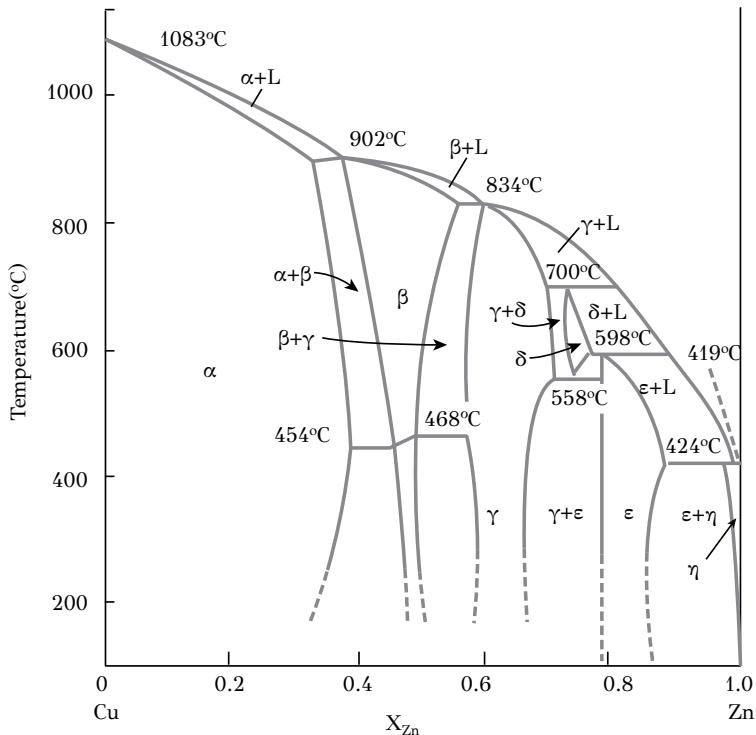


Figure 6.30: Phase diagram of the $Cu - Zn$ system. At $460^{\circ}C$ the disordered β phase (BCC) transforms to the ordered β' $CuZn$ phase.

and $468^{\circ}C$, while for the Cu_3Au phase the critical temperature is $390^{\circ}C$. The state of order is characterized by an order parameter L , which takes values between $L = 0$ for the disordered state and $L = 1$ for the ordered state. Consider the change of the order parameter with temperature for the two phases $CuZn$ and Cu_3Au . For $T = 0K$ the solid solutions in both cases are fully ordered ($L = 1$), corresponding to a minimum of internal energy and zero configurational entropy.

With increasing temperature, minimization of the free energy can be accomplished by an increase of the configurational entropy i.e., by the introduction of randomness and disorder in the system. In the case of the $CuZn$ phase, the Cu atoms diffuse and occupy positions in the Zn -sublattice and vice versa. As a result order is destroyed. The change of the order parameter for the two phases $CuZn$ and Cu_3Au is depicted in Figure 6.32. The order parameter L decreases and vanishes at the critical temperature T_c where the solid solution becomes disordered. The change of L for the case of $CuZn$ is continuous while for the case of Cu_3Au the order parameter changes abruptly at T_c . In the case of the $CuZn$ phase, the variation of enthalpy and internal energy is continuous and the order-disorder transformation $\beta' \rightarrow \beta$ is a 2^{nd} order transformation, since there is a discontinuity in the second derivative of the free

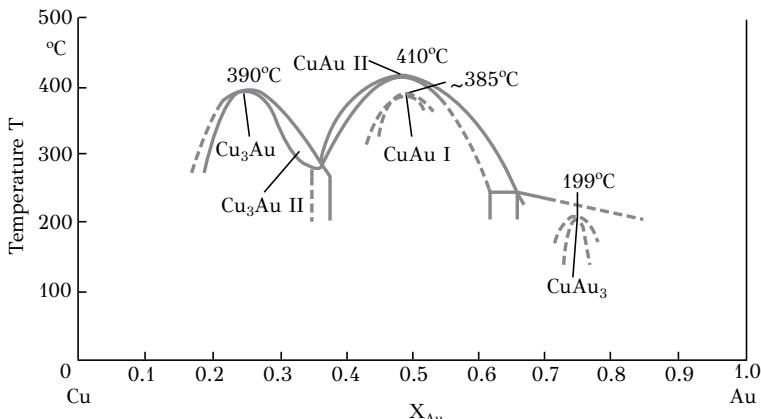


Figure 6.31: Phase diagram of the $Cu - Au$ system. At $390^{\circ}C$ the disordered $Cu - Au$ solid solution (FCC) transforms to the ordered Cu_3Au phase.

energy $\partial^2 G / \partial T^2$, expressing the specific heat. In the other case, the order-disorder transformation of Cu_3Au is a 1st order transformation, since the abrupt change of L at T_c is accompanied by an abrupt change in enthalpy and entropy. In this case the first derivative of the free energy, $\partial G / \partial T = -S$ exhibits a discontinuity at the critical temperature T_c .

The inverse transformation on cooling for the formation of an ordered solid solution from a disordered one, can take place in two ways. The first case concerns a continuous transformation during which the atoms rearrange locally, by diffusion, creating short range order, which then evolves to long-range order with $L = 1$. This transformation is equivalent with spinodal decomposition discussed in the previous section. The second case concerns the nucleation and growth of ordered domains, which exhibit full order ($L = 1$). The growth of ordered domains is depicted in Figure 6.33 for the case of Cu_3Au phase. Since nucleation is a random event, two ordered domains, nucleated independently, form a boundary when they meet during growth.

However the atomic arrangement at the boundary does not correspond to the order of phase Cu_3Au and the boundaries are termed antiphase boundaries, APB . These boundaries are depicted in Figure 6.34 for $AlFe$. The rate of the order disorder transformation depends on the alloy system. In the case of $\beta' \rightarrow \beta$ for phase $CuZn$ the transformation rate is large, because the transformation is continuous. On the other hand, the formation of the ordered Cu_3Au phase is slow because it requires nucleation and subsequent growth of ordered domains involving the sluggish migration of antiphase boundaries.

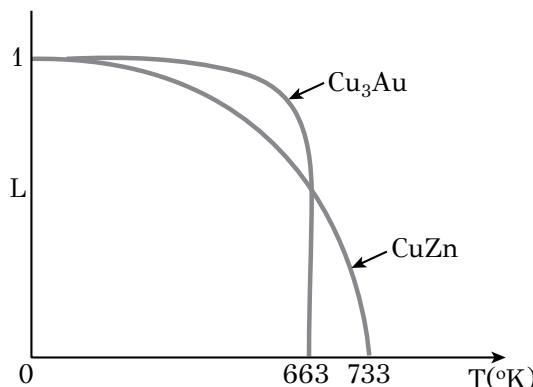


Figure 6.32: Variation of the order parameter L with temperature for the ordered $CuZn$ and Cu_3Au phases.

6.8 MARTENSITIC TRANSFORMATIONS

6.8.1 BASIC CHARACTERISTICS OF MARTENSITIC REACTIONS

Martensitic transformations (*MT*) are non-diffusive phase reactions with great technological importance. The strengthening of steel is based on these transformations and this explains the fact that *MT* is best known in steels, although they exist in other non-ferrous alloys. When steel is heated in the austenitic region and then cooled rapidly in order to impede diffusive phase transformations, then austenite (*FCC*) transforms to a new phase, martensite with a body-centered tetragonal (*BCT*) structure. Martensite in steels is very hard and forms as plates or laths in the austenite matrix. The rapid cooling process is termed the quenching of steel. The term martensite was adopted in honor of the German metallographer Adolf Martens. The terms *martensite* and *martensitic transformations* are also used in other non-ferrous systems as $Ti - Mo$, $Ti - Mn$, $Zr - Nb$, $Cu - Zn$, $Cu - Al$, $Cu - Sn$, $Au - Cd$, $Au - Mn$ and $Au - Cu$ alloys. In this section the basic characteristics of martensitic reactions, which are common for all alloy systems, will be discussed.

We will start from the observation that *MT* deforms the material. Actually, in several cases, *MT* is referred as an alternative deformation mechanism in addition to dislocation glide and mechanical twinning. The various components of deformation accompanying the transformation will be analyzed. One basic characteristic of *MT* is that it is a *diffusionless transformation*. Diffusion is not a prerequisite for the growth of martensitic crystals. On the contrary, martensitic growth takes place by the migration of glissile interfaces. These interfaces are composed of dislocation arrays, the glide of which causes the migration of the interface without the need of atomic diffusion. In a later section, the structure of the glissile interfaces will be presented in order to understand the non-diffusive character of interface migration. We will also discuss martensitic nucleation, which is heterogeneous, since lattice defects and dislocations play an important role in the nucleation process. We will finalize

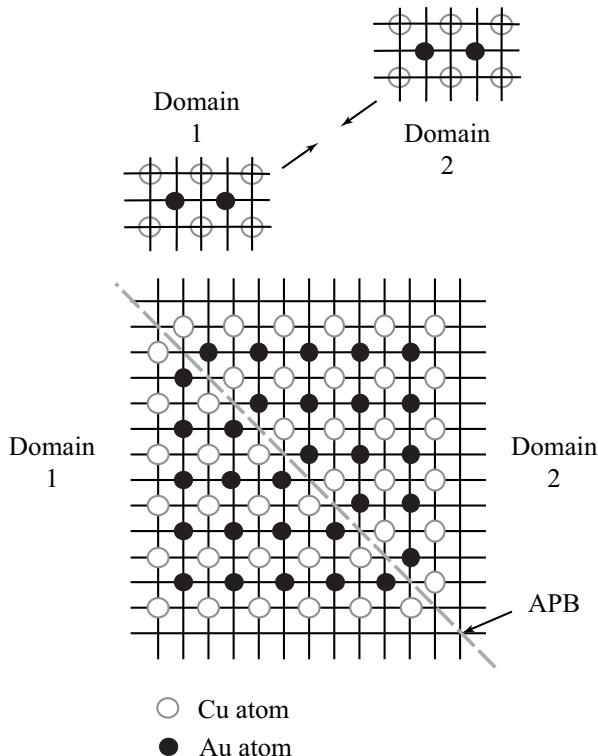


Figure 6.33: Growth of ordered domains (a) and formation of *APB* at the junction (b).

with the discussion of the mechanical effects of martensitic transformations, such as transformation plasticity, thermoelasticity and shape memory alloys. The basic characteristics of martensitic transformations are summarized below: The transformation is diffusionless and takes place at extremely high speeds approaching the speed of sound in the material. There is no change in composition and the austenitic and martensitic phases have the same composition at the end of the transformation. *MT* starts at a specific temperature on cooling, the M_s temperature. Martensite can be hard or soft, depending on the alloy system. In steels martensite is hard. *MT* is accompanied by large shear-like displacements. The martensitic crystals are shaped either as laths or lens-shaped plates possessing internal deformation in the form of dislocation glide or twinning.

From the characteristics stated above it is obvious that it is necessary to have a definition for martensitic transformations. In this book the following definition is adopted¹: “The martensitic transformation is a structural lattice transformation

¹M. Cohen, G.B. Olson and P.C. Clapp, Proc. ICOMAT-79, MIT, Cambridge, MA, 1979.

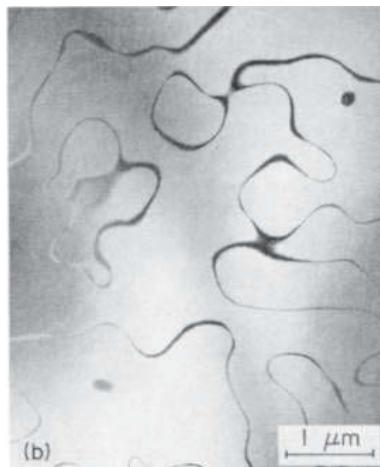


Figure 6.34: Antiphase boundaries in $Al - Fe$.

Reprinted from: S.M. Allen, J.W. Cahn, Acta Metall., 27(1979)1085, with permission from Elsevier.

with three basic characteristics: (a) the transformation is displacive (b) the transformation is diffusionless and (c) the transformation displays a large component of lattice-distortive shear such that the transformation kinetics and martensite morphology are dominated by strain energy". According to the above definition, in order for a transformation to be classified as martensitic, it should possess all three characteristics. The term *displacive* means that the transformation takes place by a cooperative movement of atoms so that there is a correspondence between initial and final lattice points. The change of crystal structure is accomplished by a homogeneous lattice deformation as we shall see below. The term *diffusionless* means that diffusion is not required for *MT* to take place. A consequence of the diffusionless character is that the martensitic phase inherits the composition, atomic order and lattice defects of the parent phase (austenite). Because there is no change in composition, the system can be regarded as a one-component system despite the fact that it may be a multicomponent system. The interface between the austenite and martensite crystals is a glissile interface, the migration of which is accomplished by the glide of the dislocation arrays composing the interface. In this case the dislocation glide is conservative, i.e., it takes place without the generation or annihilation of point defects. The shear-dominated strain energy has a major effect on transformation kinetics and morphology of the transformation product. The martensite crystals are lath or lens shaped in order to reduce the strain energy.

6.8.2 THE SHAPE DEFORMATION

In order to describe the shape deformation associated with martensitic reactions, some fundamental elements of the phenomenological crystallographic theory of

martensitic transformations by Bowles and Mackenzie will be presented.¹ The total shape deformation can be described by a 3×3 matrix $[F]$ termed the deformation gradient tensor, which in the case of homogeneous deformation defines the position vector $\{y\}$ of a material point after deformation as a function of the corresponding position vector $\{x\}$ before deformation

$$\{y\}_{3 \times 1} = [F]_{3 \times 3} \{x\}_{3 \times 1}, F_{ij} = \frac{\partial y_i}{\partial x_j} \quad (6.25)$$

where the numbers below the matrices indicate their dimensions. It is noted that during homogeneous deformation the plane surfaces remain plane after deformation. The shape deformation is revealed macroscopically by the surface tilts, which form on an originally plane and polished surface (Figure 6.35). The martensitic transformation is homogeneous and changes the shape and volume of the transformed region. The shape deformation can be revealed if we draw a straight line on a polished surface of the alloy, such as line XY in Figure 6.35. After transformation line

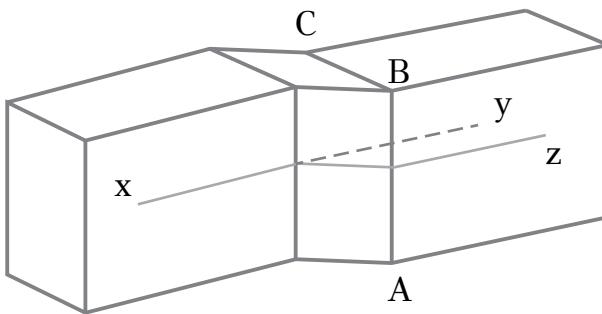


Figure 6.35: The shape deformation caused by the invariant plane strain (IPS). ABC is the habit plane.

XY is displaced to the new position XZ , however it remains continuous as it crosses the interface ABC between the martensite and parent (austenite) phases. Any deformation or rotation of plane ABC would require deformation of the parent phase in order to maintain coherency between the martensite crystal and the surrounding matrix. Therefore the plane ABC remains invariant during MT and is termed the *habit plane*. The habit plane is the interface between martensite and parent crystals. It follows that the shape deformation is an *invariant-plane strain, IPS*. We discuss next the displacements and deformations associated with MT . According to Continuum Mechanics, the deformation gradient $[F]$ in the martensite crystal can be written as²

¹J.S. Bowles and J.K. Mackenzie, Acta Metall. 2 (1954) 129.

²K. Bhattacharya, Microstructure of Martensite, Oxford University Press, 2003, p.28.

$$[F] = [I] + \zeta \{p\} [n]$$

where $[I]$ is the unit matrix, ζ is a scalar, $[n]$ is the row-vector corresponding to the unit normal \mathbf{n} to the habit plane and $\{p\}$ is the column-vector corresponding to the unit vector \mathbf{p} , which defines the direction of displacements of all material points during deformation. If we substitute the above expression for $[F]$ in 6.25 we get

$$\{y\} = \{x\} + \zeta \{p\} [n] \{x\}$$

The quantity $[n] \{x\}$ is a scalar equal to the distance d of point $\{x\}$ from the habit plane. Therefore the last equation can be written as

$$\{y\} = \{x\} + \zeta d \{p\}$$

From the above relation it is concluded that point $\{x\}$ is displaced in the direction of $\{p\}$ and the displacement has a magnitude ζd proportional to its distance from the habit plane. In Figure 6.36 the displacement of a point A to A' during MT is depicted. A new set of axes is then introduced, with origin at the habit plane Π , axis 3 is perpendicular to Π and axis 2 is parallel to the projection of \mathbf{p} in Π . Then $\{p\}$ and $[n]$ can be written as

$$\{p\} = \begin{Bmatrix} 0 \\ \cos \omega \\ \sin \omega \end{Bmatrix}$$

and

$$[n] = [0 \ 0 \ 1]$$

where ω is the angle of \mathbf{p} with the habit plane. With reference to the new set of axes $[F]$ takes the form

$$[F] = [I] + \zeta \{p\} [n] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \zeta \cos \omega \\ 0 & 0 & 1 + \zeta \sin \omega \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \gamma_0 \\ 0 & 0 & 1 + \varepsilon_0 \end{bmatrix} \quad (6.26)$$

where $\gamma_0 = \zeta \cos \omega$ and $\varepsilon_0 = \zeta \sin \omega$ are the shear and normal components of the deformation. Figure 6.37 depicts the deformation of a “unit square” in the plane defined by \mathbf{p} and \mathbf{n} during MT . Points C and D are displaced to C' and D' and the initial square is transformed to the parallelogram $ABC'D'$ with the superposition of two displacements: (a) one parallel to the habit plane causing a shear strain γ_0 and (b) one perpendicular to the habit plane causing a normal strain ε_0 . This superposition can be described analytically if we observe that

$$[F] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \gamma_0 \\ 0 & 0 & 1 + \varepsilon_0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 + \varepsilon_0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \gamma_0 \\ 0 & 0 & 1 \end{bmatrix} = [F_\varepsilon] [F_\gamma]$$

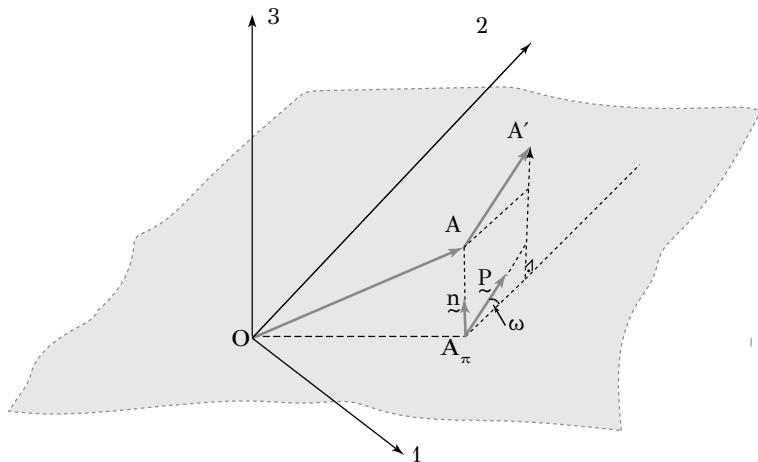


Figure 6.36: Displacements of a material point A associated with the martensitic transformation. Details in text. (Courtesy Prof. N. Aravas).

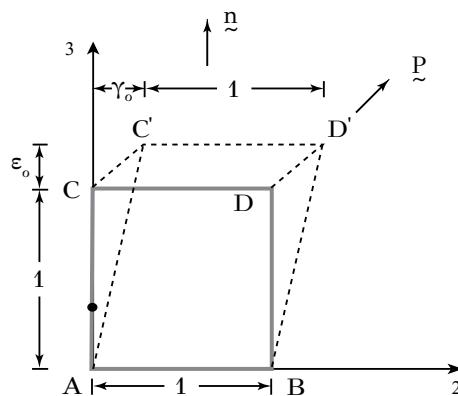


Figure 6.37: Deformation of a unit square during martensitic transformation. Superposition of shear and normal components of strain. (Courtesy Prof. N. Aravas).

and

$$\{y\} = [F] \{x\} = [F_\varepsilon] ([F_\gamma] \{x\})$$

According to the last equation the total deformation can be accomplished in two stages: (a) a shear strain, which brings the material points to an “intermediate position” $[F_\gamma] \{x\}$ and (b) a normal strain, which brings the material points to the final position $[F_\varepsilon] ([F_\gamma] \{x\})$. The volume change during *MT* is given by

$$\frac{V}{V_0} = \det[F]$$

where V_0 and V is the material volume before and after transformation and $\det[F]$ is the determinant of $[F]$. Using 6.26 it is concluded that

$$\frac{V}{V_0} = 1 + \varepsilon_0$$

meaning that the volume change is exclusively due to the normal strain ε_0 . We will discuss now the components of the shape deformation taking as an example the transformation $FCC \rightarrow BCC(BCT)$ in steels (the lattice ends up being *BCT* due to carbon trapped in interstitial positions). According to the phenomenological theory, the strain component, which causes the change in crystal structure from *FCC* to *BCC* is the Bain strain $[B]$. The Bain strain reveals the correspondence of lattice points between the parent crystal *FCC* and the martensitic crystal *BCC*. Figure 6.38 depicts the correspondence cell (bold lines) embedded in two unit cells of the austenitic crystal *FCC*. The correspondence cell will become the martensite *BCT* cell after transformation. The initial dimensions of this cell are $\frac{\alpha_0}{\sqrt{2}}$, $\frac{\alpha_0}{\sqrt{2}}$ and α_0 while the final dimensions are a , a and c . The Bain strain is, therefore, expressed by a matrix $[B]$

$$[B] = \begin{bmatrix} \eta_1 & & \\ & \eta_2 & \\ & & \eta_3 \end{bmatrix} = \begin{bmatrix} \frac{\alpha}{\alpha_0/\sqrt{2}} & & \\ & \frac{\alpha}{\alpha_0/\sqrt{2}} & \\ & & \frac{c}{\alpha_0} \end{bmatrix} = \begin{bmatrix} \frac{\alpha\sqrt{2}}{\alpha_0} & & \\ & \frac{\alpha\sqrt{2}}{\alpha_0} & \\ & & \frac{c}{\alpha_0} \end{bmatrix}$$

For the transformation $FCC \rightarrow BCT$ in steels the components of the Bain strain are $\eta_1 = \eta_2 = 1.136071$ and $\eta_3 = 0.803324$ corresponding to about 13 % tension along axes x' and y' and about 20 % compression along axis z . As discussed above, martensitic interfaces are glissile. This requires the presence of only one array of parallel dislocations, which are capable to glide freely without any intersections between them (Figure 6.39a). Dislocation intersections lead to jog formation that blocks their motion and impede the martensitic transformation. This important requirement means that all dislocations of the array should be parallel to the same direction on the interface plane. This direction should remain invariant (without deformation or rotation) during the transformation as the interface moves. This means that the deformation should possess an *invariant-line strain*, *ILS*, character. The Bain strain itself is not *ILS*. As depicted in Figure 6.40 we can consider a spherical austenite crystal being deformed to an ellipsoid by the Bain strain. The strain can leave

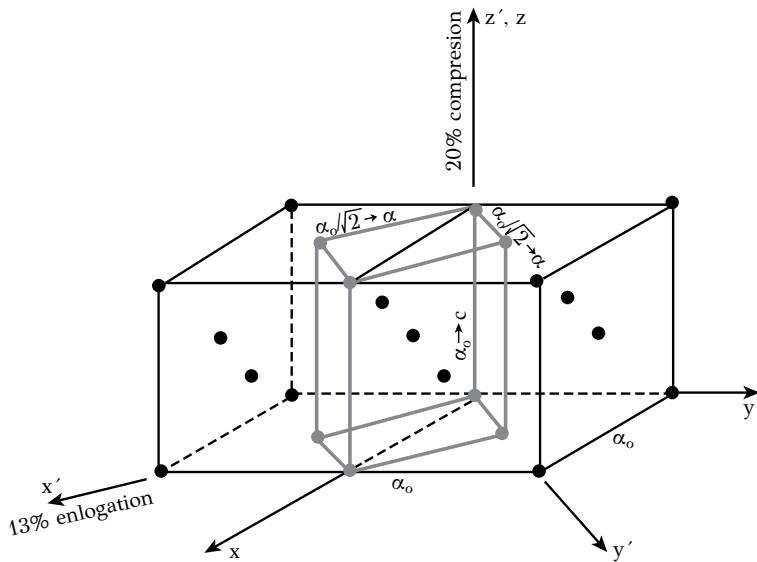


Figure 6.38: Lattice correspondence for the $FCC \rightarrow BCT$ martensitic transformation in carbon steels. Details in text.

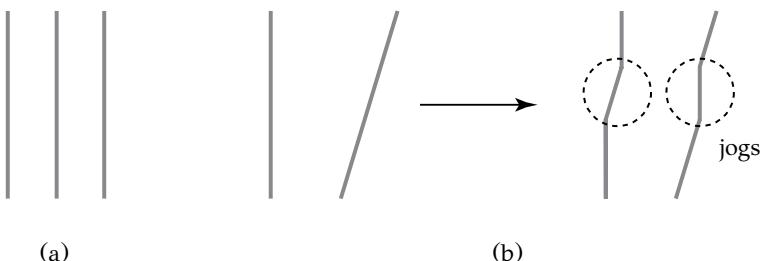


Figure 6.39: (a) Array of parallel dislocations, (b) Non-parallel dislocations, which form jogs upon intersection.

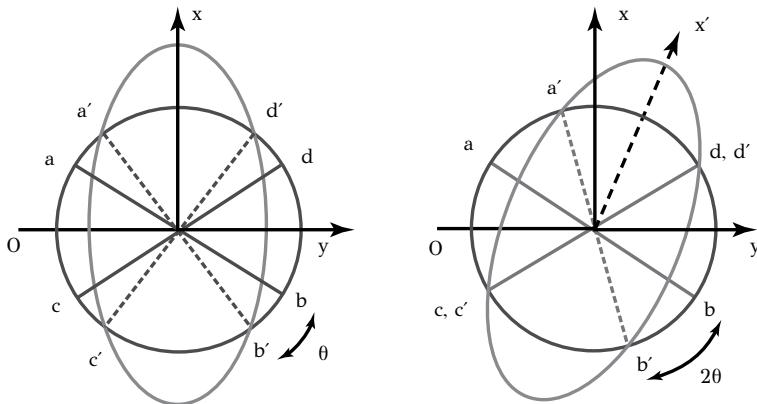


Figure 6.40: Deformation of a spherical austenite crystal during martensitic transformation: (a) application of the Bain strain, (b) application of the Bain strain plus a rigid body rotation results in ILS.

directions such as ab and cd invariant in magnitude, however they have rotated to the new positions $a'b'$ and $c'd'$. In order to get a real ILS, the martensite crystal should be rotated by an angle θ in order to have a coincidence of lines cd and $c'd'$ (Figure 6.40b). This simply means that it is the combination of the Bain strain $[B]$ with a rigid body rotation $[R]$ that can leave an invariant line. Therefore the deformation $[B][R]$ is ILS. However the deformation $[B][R]$ is much larger than the observable shape deformation $[F]$. For this reason an additional deformation takes place, in the opposite sense, in order to relieve the strain energy caused by $[B][R]$. This complementary shear $[P]$ does not change the crystal structure (the change has already taken place) and is termed *invariant-lattice deformation*. It takes place by dislocation glide or twinning of the martensitic crystal, as depicted in Figure 6.41. With the three components $[R]$, $[B]$ and $[P]$ the shape deformation can be written as

$$[F] = [R][B][P] \quad (6.27)$$

Concluding our discussion, the shape deformation $[F]$ has three components, the Bain strain $[B]$, which accomplishes the change in crystal structure, the rigid body rotation $[R]$, which is required for the ILS character and the complementary deformation $[P]$ of martensite by simple shear (slip or twinning). The volume change can be calculated as

$$\frac{V}{V_o} = \det[F] = \det[B]$$

where for the transformation $FCC \rightarrow BCT$ in steels is

$$\frac{V}{V_o} = \eta_1 \eta_2 \eta_3$$

which for the values stated above gives $V/V_o = 1.0368$ i.e., 3.68% dilatation.

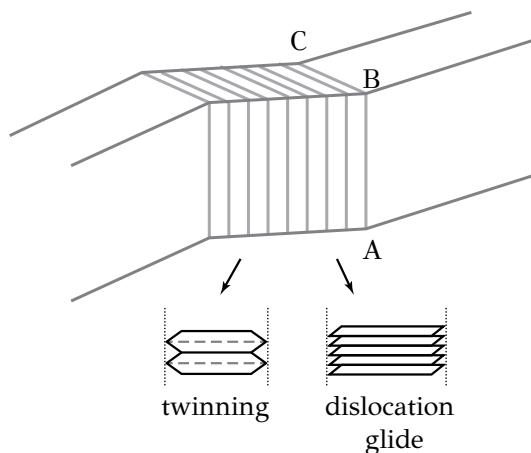


Figure 6.41: The two types of invariant-lattice deformation: dislocation glide and twinning.

6.8.3 HABIT PLANES AND ORIENTATION RELATIONSHIPS

Due to the rigid body rotation of the martensite crystal, discussed above, the habit plane possesses irrational Miller indices. At the same time close-packed planes and directions in the martensitic and parent lattices are only approximately parallel. In low carbon steels, the habit plane is tilted 5° from the close-packed planes $\{111\}_A$ of austenite and the orientation relationship between austenite and martensite is

$$(111)_A \approx // (011)_M$$

$$[\bar{1}11]_A \approx // [\bar{1}\bar{1}1]_M$$

In medium carbon steels the habit plane is close to $\{225\}_A$ and the orientation relationship is

$$(111)_A \approx 0.45^\circ \text{ from } (011)_M$$

$$[\bar{1}01]_A \approx 0.53^\circ \text{ from } [\bar{1}\bar{1}1]_M$$

In high carbon steels ($1.8\%C$) the habit plane is $\{259\}_A$ and the orientation relationship is

$$(111)_A \approx 0.86^\circ \text{ from } (011)_M$$

$$[\bar{1}01]_A \approx 4.42^\circ \text{ from } [\bar{1}\bar{1}1]_M$$

6.8.4 MORPHOLOGY OF MARTENSITE

In general, martensite acquires a lenticular shape, which minimizes the total strain energy. Two martensite morphologies are observed: (a) plate martensite and (b) lath

martensite depicted in Figure 6.42 and 6.43 respectively. The martensite crystals possess internal structure or substructure, which depends on the type of lattice-invariant deformation, twinning or slip. In case of twinning, the martensite crystal is internally twinned and in case of slip the martensite laths possess a high dislocation density (Figure 6.43). The morphology of martensite depends also on the alloy system and the chemical composition. For example steels with carbon content up to 0.4% exhibit lath martensite morphology while steels with higher carbon contents exhibit plate morphology.

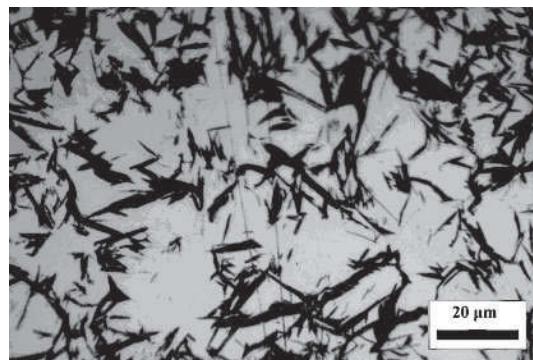


Figure 6.42: Plate martensite morphology.



Figure 6.43: Lath martensite morphology.

6.8.5 MARTENSITIC INTERFACES

As discussed in section 6.8.1 the interface between martensite and the parent phase should be glissile. Glissile interfaces migrate, during transformation, with the glide of interfacial dislocation arrays. Dislocation glide in this case is conservative, i.e., proceeds without the generation or annihilation of point defects. As discussed above, this requires that the dislocations of the array should be parallel. Figure 6.44 depicts two types of glissile interfaces (for a description of interfaces see also Chapter 3). The first is a coherent interface between two phases with lattice parameters α_α and

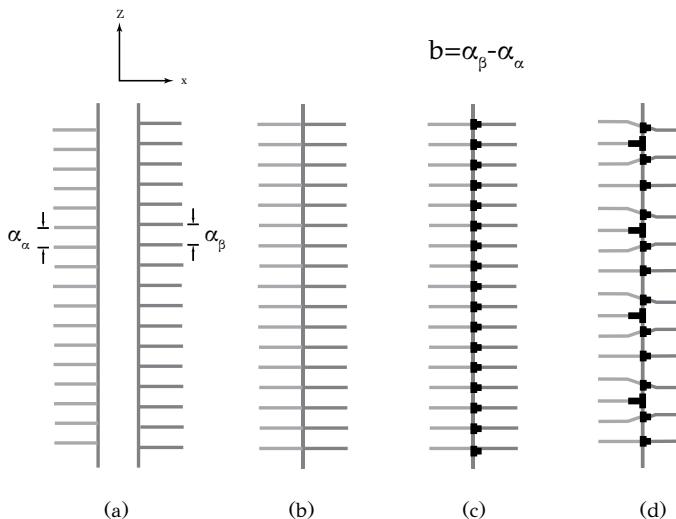


Figure 6.44: (a) Phases with different lattice parameters forming a coherent interface in (b). Introduction of coherency dislocations in (c) to account for the coherency strains. (d) Semicoherent interface with coherency and anticoherency dislocations.

Reprinted from: G.B. Olson, M. Cohen, Acta Materialia 27 (1979) 1907, with permission from Elsevier.

α_β . The continuity of crystal planes across the interface generates coherency strains. These strains can be described by an array of partial interfacial dislocations with Burger's vector $b = a_\beta - a_\alpha$. These are partial dislocations, since their Burger's vector is shorter than the lattice parameter, and are termed *coherency dislocations*. The glide of these dislocations causes a conservative interface migration. The second type of glissile interface is a semi-coherent interface, which also contains coherency dislocations. However in order to reduce the coherency strains, coherency across the interface is broken at certain intervals by the introduction of dislocations. These are perfect lattice dislocations and are termed *anticoherency dislocations*. The semicoherent interface is made up of two arrays of parallel dislocations (coherency and anticoherency dislocations), the glide of which causes the conservative migration of the glissile interface. It should be noted, that the glide of the coherency dislocations accomplishes the change in crystal structure while the glide of anticoherency dis-

locations accomplishes the lattice-invariant deformation. With the aid of the above description of interfacial structure we will discuss an equivalent to the phenomenological theory description of shape deformation of MT . It will be shown that

$$[F] = [E][P]^{-1}[P]$$

where $[E]$, $[P]^{-1}$ and $[P]$ are deformations caused by the different interfacial dislocation arrays (Figure 6.45).

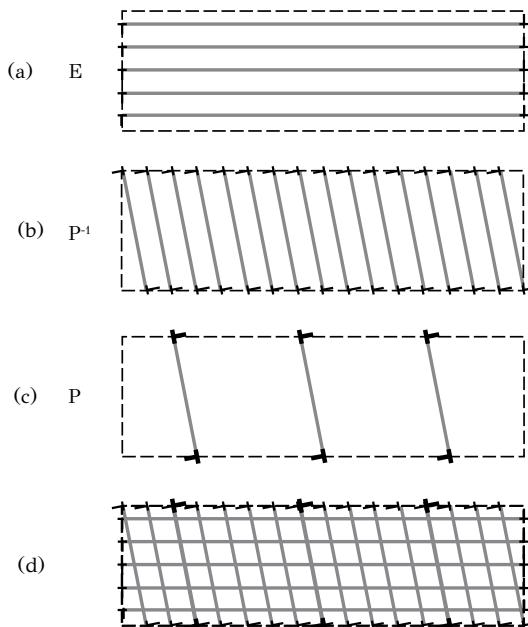


Figure 6.45: Interfacial structure of martensitic crystal: (a) side coherent interface with array of coherency dislocation loops achieving strain E , (b) top semicoherent interface with coherency dislocation loops achieving strain $[P]^{-1}$, and anticoherency dislocations achieving strain P in (c). Total interfacial structure in (d).

Reprinted from: G.B. Olson, M. Cohen, Dislocation of Martensitic Interfaces, in Solid-to Solid Phase Transformations, Aaronson, Laughlin, Sekerka, Wayman eds, AIME, 1982, with permission of The Minerals, Metals & Materials Society.

This figure shows that the two side interfaces of the martensite crystal are coherent interfaces and are composed of arrays of partial coherency dislocations, forming dislocation loops. The glide of these loops causes the growth of the martensitic crystal and the resulting strain $[E]$ has *IPS* character. The top and bottom interfaces are semicoherent interfaces and are composed of coherency and anticoherency dislocation arrays forming corresponding dislocation loops. The glide of the coherency dislocation loops causes growth of the martensitic crystal in the lateral direction and the resulting strain $[P]^{-1}$ has *IPS* character, while the glide of the anticoherency dislocation loops causes the lattice-invariant deformation $[P]$. It is important to note that

strain $[P]$ cancels out strain $[P]^{-1}$, relieving the strain energy. The shape deformation becomes

$$[F] = [E]$$

possessing *IPS* character. The strain $[P]^{-1}$ has been referred to as “invisible shear” since it is immediately canceled by $[P]$. It can be shown that

$$[E][P]^{-1} = [R][B]$$

indicating the correspondence between the phenomenological theory and the dislocation theory of MT. Summing up the above discussion, martensitic interfaces are made up of dislocation arrays, the glide of which accomplishes both (a) the growth of martensite crystal and (b) the deformation of martensite. The change in crystal structure (transformation) is caused by the glide of the coherency dislocations and the lattice-invariant deformation of martensite is caused by the glide of anticoherency dislocations.

6.8.6 NUCLEATION OF MARTENSITE

The martensitic transformation evolves by nucleation and growth. With the exception of thermoelastic transformation, to be discussed below, growth is extremely fast, approaching the speed of sound in the metal. The limiting step controlling transformation kinetics is, therefore, nucleation. We start our discussion by considering that nucleation is heterogeneous, meaning that it is assisted by structural defects in the austenitic lattice and we will show at the end of this section that homogeneous nucleation is energetically unfavorable. Following the Olson and Cohen¹ theory of martensitic nucleation, the nucleation process is related to the formation of loops of partial dislocations. As discussed in Chapter 3, partial dislocations are generated by the dissociation of perfect dislocations and in order for the dissociation to be energetically feasible two requirements should be met: (a) Frank’s rule should be obeyed, i.e., the total energy of the partials should be lower than the energy of the full dislocation and (b) the structural fault generated between the partials should decrease the total free energy. Here the fault possesses the structure of the martensitic crystal. The nucleation process is depicted in Figure 6.46.

Nucleation is heterogeneous and starts from a pre-existing defect in the austenite lattice, such as a dislocation pile-up in a grain boundary. The dislocations of the pile-up dissociate into partials generating a fault, which is the first martensitic embryo (Figure 6.46b,c). Whether or not the embryo will evolve into a martensitic nucleous and then into a growing martensitic crystal depends on the fault energy, designated as $\bar{\gamma}$. The formation of the fault will decrease the free energy if the new structure is more stable than the parent phase, i.e., if the fault energy is negative ($\bar{\gamma} < 0$). The fault energy can be expressed (per unit fault area) as

¹G.B. Olson and M. Cohen, Met. Trans. A, 7A (1976) 1897

$$\bar{\gamma} = n\rho_A(\Delta G_{ch} + E_{str}) + 2\gamma \quad (6.28)$$

where n is the fault thickness (in number of crystal planes), ρ_A the density of atoms in the fault plane (in mol/m^2), ΔG_{ch} the chemical driving force (in J/mol), E_{str} the strain energy (in J/mol) and γ the interfacial energy between the martensitic and austenitic phases.

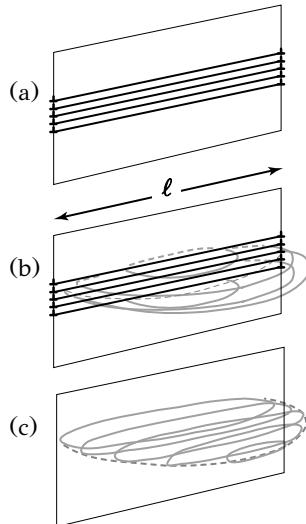


Figure 6.46: Heterogeneous martensitic nucleation: (a) pre-existing defect, a dislocation pile-up at a grain boundary, (b) dissociation into partial coherency dislocations generating the fault, (c) simultaneous generation of coherency and anticoherency dislocations at the top and bottom interfaces of the fault (see also Figure 6.45d).

Reprinted from: G.B. Olson, M. Cohen, Theory of Martensitic Nucleation, in Solid-to Solid Phase Transformations, Aaronson, Laughlin, Sekerka, Wayman eds, AIME, 1982, with permission of The Minerals, Metals & Materials Society.

The temperature dependence of the chemical driving force is depicted in Figure 6.47. At temperature T_0 the austenite and martensite have the same free energy, so $\Delta G_{ch} = 0$, and at $T < T_0$ the driving force is $\Delta G_{ch} < 0$ and becomes more negative with a decrease in temperature. ΔG_{ch} has a negative sign in Equation (6.28), while E_{str} and γ possess a positive sign. Therefore in order for $\bar{\gamma}$ to become negative, the driving force ΔG_{ch} should obtain a critical value. For this reason martensitic nucleation starts at a specific temperature on cooling, termed the M_s temperature, where the chemical driving force gets a critical value ΔG_{ch}^c and the fault energy becomes negative. The change of total free energy with the partial dislocation separation r is depicted in Figure 6.48. In the case of the formation of a stable fault (with $\bar{\gamma} > 0$) the

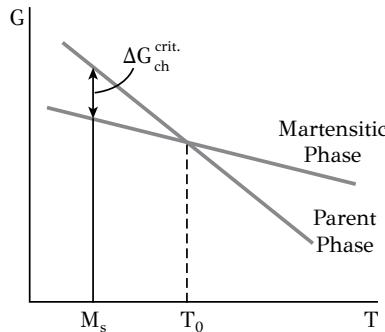


Figure 6.47: Temperature dependence of the free energies of the parent and martensite phases. The chemical driving force for martensitic transformation is also depicted.

partials repel each other and separate at an equilibrium distance r_{eq} where $\bar{\gamma}$ is minimized. With cooling, the chemical driving force increases (becomes more negative) and when $\bar{\gamma} = 0$ the fault becomes unstable. On further cooling $\bar{\gamma}$ becomes negative and the fault evolves as the martensitic nucleus. The critical condition is expressed by $\bar{\gamma} = 0$ at the M_s temperature.

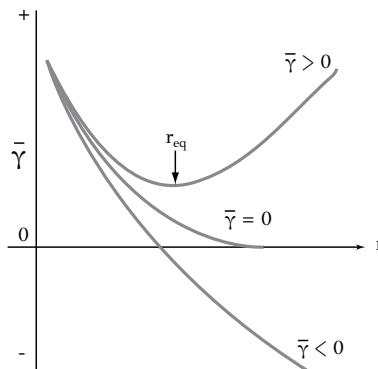


Figure 6.48: Variation of fault energy with the separation distance of partial dislocations. Details in text.

Reprinted from: G.B. Olson and M. Cohen, Metall. Trans. A, 7A(1976) 1897, with permission from Springer.

It was discussed above that possible heterogeneous nucleation sites are dislocation pile-ups at austenite grain boundaries. However the sites do not have the same potency to nucleate martensite. It follows from Equation (6.28) that the fault energy is a function of fault thickness n , i.e., the number of dislocations making up the fault. The nucleation site potency can then be expressed by n . Figure 6.49 shows the

change of fault energy $\bar{\gamma}$ as a function of temperature for three different values of n . Dislocation arrays with higher n dissociate and form martensite nuclei at higher temperatures. On the other hand, activation of sites with lower n requires a larger drop in temperature, in order to increase the driving force ΔG_{ch} . The macroscopic M_s temperature corresponds to a value of n for a site, which when activated produces a measurable quantity of martensitic crystal.

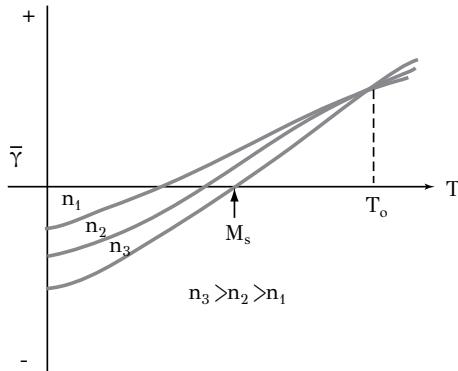


Figure 6.49: Variation of fault energy with temperature for three values of the fault thickness. At M_s the fault energy becomes negative.

Reprinted from: G.B. Olson and M. Cohen, Metall. Trans. A, 7A(1976) 1897, with permission from Springer.

We turn now our discussion in showing that homogeneous martensite nucleation is energetically unfavorable. The Gibbs free energy change due to the formation of a martensite particle in a perfect crystal is

$$\Delta G = \frac{4}{3}\pi r^2 c \Delta G_{ch} + \frac{4}{3}\pi r^2 c A \frac{c}{r} + 2\pi r^2 \gamma \quad (6.29)$$

In this case we have considered that the martensite particle is a thin ellipsoid with radius r and thickness $2c$, possessing volume $(4\pi/3)r^2c$ and surface area $2\pi r^2$. The term ΔG_{ch} is the chemical driving force as before, the quantity Ac/r is the elastic strain energy per unit volume and γ is the surface energy between martensite and austenite. If we further consider that the shape deformation has *IPS* character with a shear component γ_o and a normal component ϵ_o then

$$A = \frac{\pi(2-\nu)}{8(1-\nu)} \mu \gamma_o^2 + \frac{\pi}{4} \mu \epsilon_o^2 \quad (6.30)$$

where μ is the shear modulus and ν the Poisson ratio. μ and ν are considered the same for austenite and martensite. For martensitic transformations in steels A takes

a value of about $2.4 \times 10^3 \text{ MJ/m}^3$. The interface boundary is a semicoherent boundary with $\gamma = 100 - 200 \text{ mJ/m}^2$. For a typical steel composition, the chemical driving force at M_s is $\Delta G_{ch} = -170 \text{ MJ/m}^3$. Introducing these values in Equation (6.29) one obtains a relation for $\Delta G(r, c)$, from which we obtain the activation energy ΔG^* for homogeneous nucleation. ΔG^* has a value of about $8 \times 10^{-16} \text{ J}$ per nucleation event ($= 5 \times 10^3 \text{ eV}$). This value corresponds to $10^5 kT$ at temperatures where martensite nuclei form. It is obvious that the available thermal energy is not sufficient to overcome the energy barrier for homogeneous nucleation, and, therefore, martensitic nucleation is heterogeneous.

The heterogeneous nature of martensitic nucleation has been confirmed by the Cech and Turnbull¹ small particle experiments in Fe-30Ni. The fraction of particles transformed to martensite is depicted as a function of particle size and temperature in Figure 6.50.

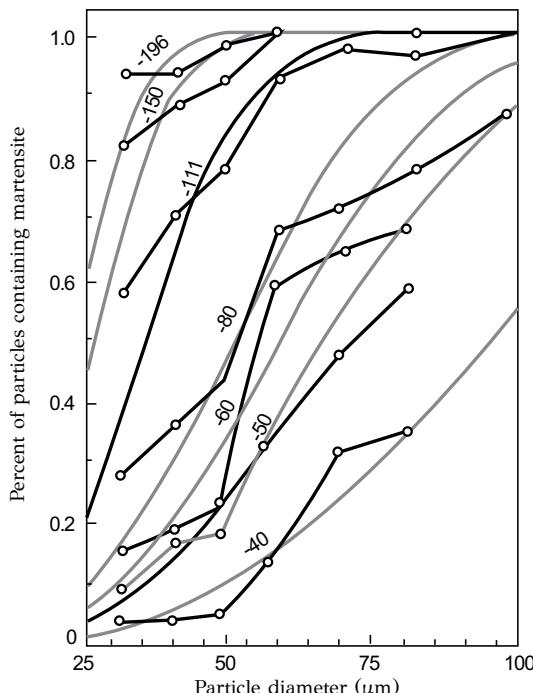


Figure 6.50: Martensitic transformation in particles of a $Fe - 30Ni$ alloy. The fraction of transformed particles as a function of particle diameter and quench temperature.

Reprinted from: R.E. Cech and D. Turnbull, Trans. AIME 206 (1956) 124, with permission of The Minerals, Metals & Materials Society.

At a given temperature the fraction of particles transformed to martensite in-

¹R.E. Cech and D. Turnbull, Trans. AIME, 206 (1956) 124

creases with particle size. This is attributed to the higher probability to find a nucleation site in a larger particle. In addition, for a given particle size, the fraction of particles transformed to martensite increases with a drop in temperature. This confirms the fact that the nucleation sites do not possess the same potency to cause martensitic nucleation. Sites with lower potency require higher driving force (lower temperature) to be activated.

6.8.7 MECHANICAL EFFECTS

Due to the shape deformation associated with *MT*, strong interactions are expected when external stresses are applied. Such effects include transformation plasticity, thermoelasticity, pseudoelasticity and shape memory.

Transformation plasticity. Mechanical effects in terms of applied stresses or strains in martensitic transformation can be discussed with the aid of Figure 6.51. Without any applied stress, martensitic transformation starts at pre-existing nucle-

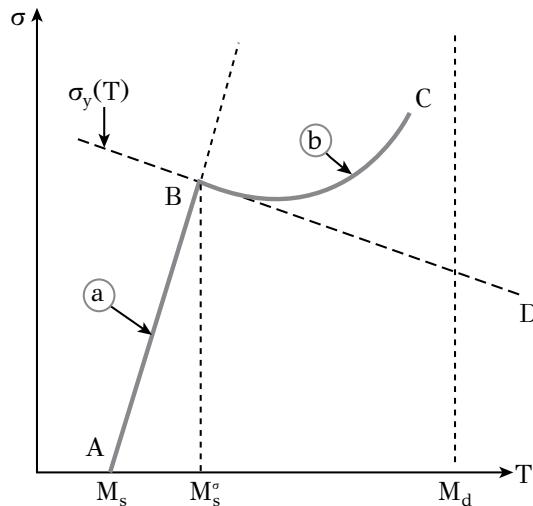


Figure 6.51: Stress effects on martensitic transformation and the regimes of stress-assisted (a) and strain-induced transformation (b).

Reprinted from: G.B. Olson and M. Cohen, J. Less Common Metals 28 (1972) 107, with permission from Elsevier.

ation sites on cooling to the M_s temperature. Under an applied tensile stress below the yield strength σ_y of the parent phase, martensitic nucleation commences at a temperature higher than M_s . In this case martensitic nucleation is termed as stress-assisted. The applied stress generates a mechanical driving force ΔG_σ , which adds to the chemical driving force ΔG_{ch} and contributes to the critical driving force for martensitic nucleation required for $T > M_s$. The higher the temperature above M_s , the higher the stress required for martensitic nucleation. The temperature dependence

of the required stress for stress-assisted nucleation is depicted by line *AB*. At the M_s^σ temperature, the stress reaches the yield strength of the parent phase (austenite), which exhibits a temperature dependence given by line *BD*. Due to plastic deformation, new and more potent nucleation sites are generated in the parent phase. In this case the stress required follows line *BC* and the nucleation is termed strain-induced. Above the M_d temperature the chemical driving force is so small that it is not possible to nucleate martensite by the application of stress or strain. The shape deformation of martensitic transformation contributes to the fundamental mechanism of plastic deformation, i.e., dislocation glide. This contribution is termed *transformation-induced plasticity (TRIP)*. In other words the transformation operates as a deformation mechanism as dislocation glide or mechanical twinning. The *TRIP* effect influences the constitutive behavior $\sigma - \varepsilon$ of the material, with significant consequences on the mechanical properties, the most important being the enhancement of ductility, fracture toughness and formability. The exploitation of transformation plasticity has led to the development of *TRIP* steels, either as fully austenitic *TRIP* steels or as low-alloy multiphase steels containing retained austenite. The first exhibit excellent fracture toughness due to the interaction of transformation plasticity with the stress field ahead of the crack tip. The second exhibit excellent formability due to the stabilization of plastic flow during cold forming.

Thermoelastic martensitic transformation. In most cases the shape deformation of the martensitic transformation is large compared to the strength of the parent phase and the accommodation of this strain takes place by plastic deformation. As a consequence the migration of the martensite/austenite interface is not reversible. In addition, a large part of the chemical driving force is dissipated during transformation and is not stored elastically in the system, making transformation reversal impossible. However there are cases where the accommodation of the shape deformation is elastic and the interface migration is reversible. During cooling of the material, the increasing chemical driving force is balanced by the strain energy, which is stored in the system in the form of elastic strain energy, as the interface migrates during growth of martensite. Due to this *elastic accommodation*, the interface remains glissile and maintains its interfacial structure. During heating, the stored elastic strain energy causes a reversal of the shape deformation with a reversed migration of the interface, maintaining a balance with the chemical driving force, which now decreases with the increase in temperature. The balance between chemical and mechanical forces during martensitic transformation is characterized as *thermoelasticity*¹. The thermoelastic behavior is depicted in Figure 6.52 during growth of a martensite plate. During cooling to the M_s temperature, a thin lenticular martensite plate forms (point a) and grows in the radial direction until it meets a barrier, such as a grain boundary (point b). Further cooling increases the volume of the plate to point c. Along line bc during cooling and line cbd during heating there is a continuous balance between the chemical driving force and the elastic strain energy expressed by the relation

¹G.V. Kurdjumov, J. Metals, 11 (1959) 449

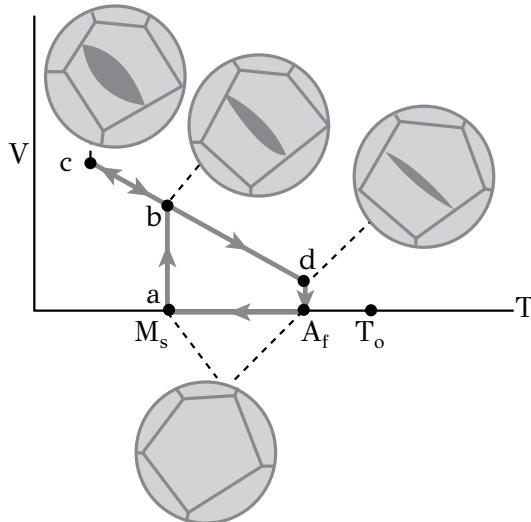


Figure 6.52: Thermoelastic behavior during growth of a martensitic plate. Details in text.

Reprinted from: G.B. Olson and M. Cohen, Scripta Metall. 11 (1977)345, with permission from Elsevier.

$$\Delta G_{ch} + 2A \frac{c}{r} = 0 \quad (6.31)$$

where A is given by Equation (6.30). For a given plate length r the semi thickness c increases during cooling (plate thickening), the inverse takes place during heating. The thermoelastic line cbd is extended to temperature T_o for $V = 0$. However the plate volume vanishes at a lower temperature than T_o , the A_f temperature, due to surface tension, which causes the plate to shrink as it detaches from the grain boundary. In order to reform the plate, the material should be cooled to M_s in order to nucleate a new martensite plate. The thermoelastic martensitic transformation in $Fe-28.1$ at% Pd is depicted in Figure 6.53 before and after the application of compressive stress. The thermoelastic behavior discussed above, generates two important effects: pseudoelasticity and shape memory.

Pseudoelasticity. In alloy systems exhibiting thermoelastic martensitic transformation, it is possible to activate the transformation above the M_s temperature by the application of mechanical stresses. This type of mechanically-induced transformation has been characterized as pseudoelasticity due to the large observed elastic strains (about 10%). Actually these strains are fully recoverable on unloading due to the operation of the reversed transformation. An example of pseudoelastic behavior is depicted in Figure 6.54 for the $Cu - 39.8Zn$ alloy at $-77^{\circ}C$. Although there are 24 available orientations of the habit plane, the applied stress during loading imposes

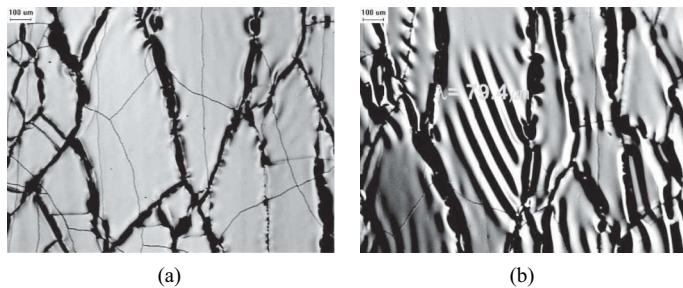


Figure 6.53: Thermoelastic martensitic transformation in $Fe - 28.1$ at% Pd . Martensite appears dark in the figure (a) before and (b) after loading.

Reprinted from: J.H. Han, J. Yoon, K.K. Jee, J.J. Vlassak, *Thin Solid Films* 546(2013)196, with permission from Elsevier.

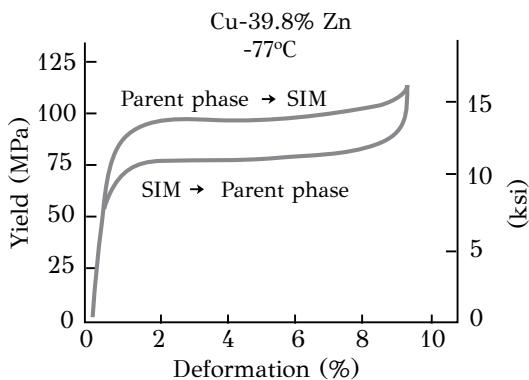


Figure 6.54: Stress-strain curves for pseudoelastic loading and unloading of a $Cu - 39.8\%$ Zn alloy at $-77^\circ C$.

Reprinted from: T.A. Schroeder and C.M. Wayman, *Acta Metall.* 27 (1979) 405, with permission from Elsevier.

a preferred habit orientation that maximizes the elongation in the direction of the tensile axis. Therefore, martensite has been fully formed and is fully oriented at the upper plateau of Figure 6.54. A lower plateau is also observed during unloading, where the transformation is reversed to form the parent phase. The stress difference between the two plateaus is attributed to the lattice resistance during interface movement. This difference can be expressed as $2\tau_f$, where τ_f is the friction stress. Taking into account the friction stress, the thermoelastic balance of Equation (6.31) becomes

$$(\Delta G_{ch} - \tau\gamma_o) + 2A\frac{c}{r} = \pm\tau_f\gamma_o$$

where γ_o is the shear component of shape deformation and τ the applied shear stress. The term $\tau\gamma_o$ is the mechanical work contribution to the chemical driving force. The term $\tau_f\gamma_o$ is negative when the martensite plate grows during loading and positive when the plate shrinks during unloading. An important feature depicted in Figure 6.54 is that relatively large reversible strains can be obtained under a practically constant load.

The pseudoelastic behavior $\tau - \gamma$ is depicted in Figure 6.55. At a given temperature $T > M_s$ when $\tau_f = 0$ (Figure 6.55a), the increase of stress τ causes the formation of a martensite plate at stress τ_s , accompanied by a shape deformation γ_s . The pseudoelastic equilibrium of forces is maintained along the inclined lines. Full reversal of the martensite plate back to the parent phase takes place during unloading at stress τ_r (corresponding to A_f in Figure 6.52). When $\tau_f > 0$, the hysteresis $2\tau_f$ results in an increase of τ_s during loading and a decrease of τ_r during unloading (Figure 6.55b). The stress τ_o is the applied stress that could nucleate martensite in the absence of the nucleation barrier and interfacial friction. In addition to the $Cu - Zn$ system, depicted in Figure 6.54, pseudoelastic behavior is also exhibited by the system $Ni - Ti$, and it is for this reason that these alloys have wide applications as orthodontic wires. The maintenance of a constant load for a large deformation range minimizes the doctor visits and associated “chair time” for the retightening of the wire. One of the main characteristics of thermoelastic martensite is the high mobility of the glissile martensitic interfaces. A martensitic specimen can be deformed below the M_f temperature with a low stress and the deformation is caused by the growth of favorably oriented habits (variants) in the expense of the less-favorably oriented ones. Under these conditions, the interfaces experience a restoring force and the specimen acquires the original shape during unloading. This phenomenon has been characterized as rubber-like behavior, to be distinguished from pseudoelastic behavior. The first is associated with exclusively internal movements of martensite interfaces while the second is associated with movements of the austenite/martensite interfaces during transformation.

Shape memory. As discussed above, the dominant deformation mechanism of thermoelastic martensite below the M_f temperature is the migration of glissile martensitic interfaces. Under the absence of rubber-like behavior, i.e., without any restoring forces, it is possible to cause permanent deformation of martensite with the application of relatively low stresses, without the activation of dislocation glide. The change in shape of the specimen is accomplished exclusively by the growth of favor-

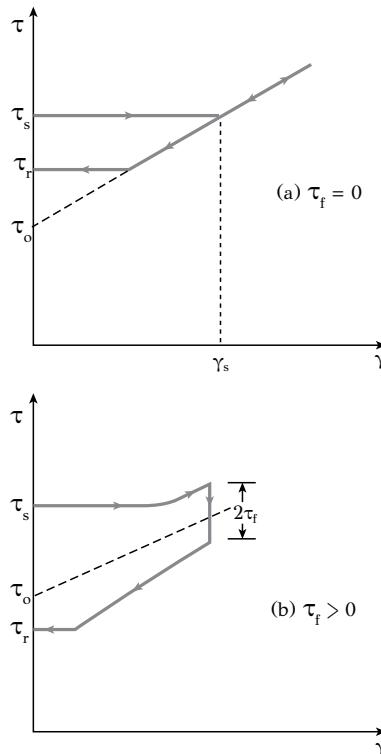


Figure 6.55: Pseudoelastic behavior of a martensite plate for two cases of zero and non-zero friction stresses. Details in text.

Reprinted from: M. L. Green, M. Cohen, G.B. Olson, Mat. Sci. Eng. 50 (1981) 109, with permission from Elsevier.

ably oriented martensite plates (habits) in the expense of the rest. Heating the specimen above the A_f temperature in order to form the parent phase, restores the original shape of the specimen (before deformation). The phenomenon has been termed shape memory effect. It is possible to achieve large permanent strains of the order of 6-8%, which can be fully restored by heating above A_f . Several alloys exhibit the shape memory effect, which is based on the reversible migration of martensitic interfaces. The shape memory effect is depicted schematically in Figure 6.56. When a single crystal of the parent phase undergoes thermoelastic martensitic transformation on cooling below the M_f temperature, it is possible to form martensite plates with 24 different orientations of the habit plane, which cancel out the individual shape deformation of each plate, thus leaving the shape of the specimen unchanged. The change of specimen shape is accomplished below M_f by deformation of martensite, which causes growth of favorably oriented variants. In most cases only one martensite plate variant dominates, resulting in single crystal martensite. Heating above the

A_f causes the formation of the parent phase, the deformation is restored and the specimen returns to its original shape. The most important shape memory alloy systems are $Cu - Zn - Al$, $Cu - Al - Ni$, $Ni - Ti$ and $Fe - Mn - Si$.

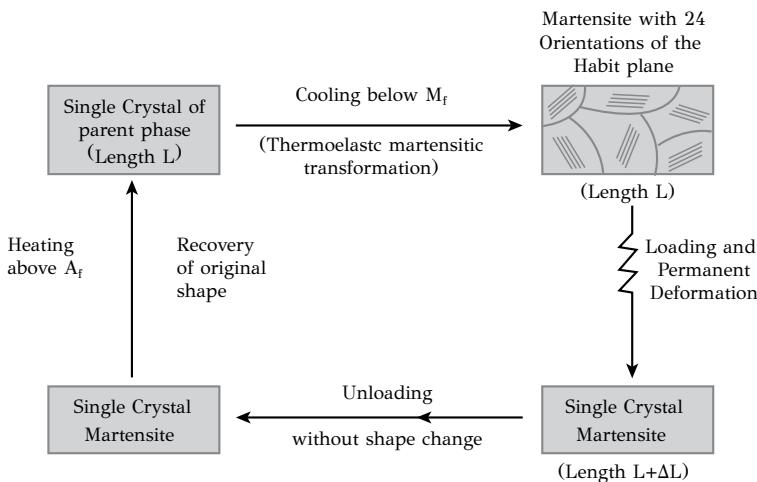


Figure 6.56: Schematic representation of the shape memory effect.

6.9 EFFECTS OF PHASE TRANSFORMATIONS

The most important effect of phase transformations is that they shape the microstructure and properties of metals and alloys. Several microstructural changes are caused by phase transformations: change of phase fractions and phase compositions, dissolution of some phases and precipitation of others, modification of the grain size, compositional segregation or homogenization, changes in dislocation density and other lattice defects, generation of new interfaces or modification of interfacial structure, generation of internal stresses and strains. All these transformations shape the microstructure, and, therefore, phase transformations are exploited in order to obtain desirable properties. Some examples follow: In steels, martensitic transformation is exploited for the development of high strength. As discussed above, the martensitic transformation is a diffusionless structural change, and takes place when the steel is quenched from the austenitic region. The cooling rate should be fast so that other competitive diffusive transformations, such as pearlite formation, are not activated. The martensite in steels is hard but brittle. Ductility is restored with a tempering treatment. Tempering causes aging of martensite with the precipitation of cementite (Fe_3C) through a precipitation sequence, as in the aging of heat treatable aluminum alloys. During tempering, transition phases such as ϵ -carbide form prior to the equilibrium cementite phase. Cementite forms with a nucleation and growth transformation, and the ϵ -carbides serve as a heterogeneous nucleation site for the precipitation of cementite. It is obvious that during the heat treatment of steel, several phase trans-

formations shape the final microstructure and properties of the material. Wrought aluminum alloys are distinguished in non-heat-treatable and heat-treatable alloys. The non-heat-treatable alloys are strengthened by strain hardening during forming, such as cold rolling. Heat treatable alloys, on the other hand, are strengthened by precipitation hardening with an aging heat treatment. In cast aluminum alloys, in addition to heat treatment for precipitation hardening, grain size refinement is applied for additional strengthening. A certain amount of inoculants is added to the liquid alloy, which during solidification serve as heterogeneous nucleation sites for the formation of the solid phase. In this case the nucleation rate is increased leading to grain refinement of the casting. It is important to note that phase transformations are activated also during service, in most cases with undesirable effects, such as softening or embrittlement. A good example is the sensitization of austenitic stainless steels during welding. The formation of *Cr* carbides at the grain boundaries causes *Cr* depletion and intergranular corrosion. One way dealing with sensitization is alloy additions with *Ti* or *Nb* for the precipitation of *TiC* or *NbC* carbides instead of *Cr* carbides. Another example is high-temperature service, where phenomena such as coarsening of precipitates and associated softening take place. For very high temperature applications, such as turbine blades, *Ni*-based alloys are used. Creep resistance in this case is a result of precipitation of Ni_3Al intermetallics. These intermetallics, possess *FCC* structure, and therefore, are coherent with the matrix *FCC* phase. The low interfacial energy of the coherent interface contributes to the high coarsening resistance of these precipitates and the maintenance of strength at high temperatures. The examples discussed above indicate that understanding of phase transformations, is the fundamental prerequisite towards (a) understanding material behavior and (b) development of alloy design methodologies for a knowledge-based development of new materials.

6.10 SYNOPSIS

1. Most metallic alloys undergo some phase transformation during processing, such as heat treatment and mechanical or thermomechanical treatment. These transformations shape the microstructure and properties of alloys.
2. Phase transformations are distinguished in diffusive and non-diffusive, depending on whether atomic diffusion is a prerequisite for the transformation to take place.
3. Nucleation, growth and coarsening are the three stages during the precipitation of a new phase from a supersaturated solid solution.
4. The nucleation of a new phase is distinguished in homogeneous and heterogeneous. Homogeneous nucleation takes place without the aid of lattice defects and the energy barrier for nucleation, consisting of the interfacial energy and strain energy, is large. Heterogeneous nucleation is assisted by lattice defects such as grain boundaries and dislocations, and the energy barrier is reduced.

5. The shape of the new phase depends on the balance between interfacial energy and strain energy. In the case of coherent precipitates, the shape depends on the lattice misfit. When the misfit is small, the shape is dominated by the minimization of surface energy and the particles adopt a spherical shape. When the misfit is large, the shape is dominated by the minimization of strain energy and the particles form as discs. In the case of incoherent precipitates the shape depends on the volumetric strain. If this is small, then the shape is spherical; if it is large, the particles adopt the shape of a disc.
6. Growth of a new phase can be either diffusion or interface controlled. In the first case the atoms cross the interface very fast and the limiting step is the diffusion of atoms towards the interface. In the second case, the diffusion is fast and the limiting step is crossing the interface.
7. The size of a phase, which grows under diffusion control, is a parabolic function of time. The growth rate is proportional to supersaturation and the diffusion coefficient. When the phase grows under interface control, the size is a linear function of time and the growth rate is proportional to the driving force and the interface mobility.
8. The kinetics of concurrent nucleation and growth transformations (*NGT*) is expressed by the time evolution of the phase fraction. It includes the induction, growth and impingement stages, and is described by the general Johnson-Mehl-Avrami-Kolmogorov equation (*JMAK* kinetics).
9. Temperature influences the kinetics of diffusive phase transformations through (a) the thermodynamic driving force and (b) the diffusion coefficient. Regarding precipitation reactions: at high temperatures, diffusion is fast but the precipitation driving force (supersaturation) is small resulting in a low transformation rate. At low temperatures, the driving force is large but diffusion is slow, resulting again in a low transformation rate. The maximum transformation rate is obtained at intermediate temperatures, resulting in *C*—curve kinetics, which is the base of *TTT* (*IT*) diagrams.
10. Coarsening of particle dispersions is a diffusive transformation caused by non-chemical forces. It is caused by capillarity or curvature effects (Gibbs-Thomson effect). During coarsening, the mean particle size increases with time under constant volume fraction.
11. Coarsening kinetics is described by the *LSW* theory, according which the particle size is a function of time. Coarsening leads to softening of precipitation-hardened alloys.
12. Spinodal decomposition is a continuous phase transformation, which takes place

in alloy systems exhibiting a miscibility gap and when the alloy composition is in the region where the second derivative of G with respect to composition is negative. The decomposition of a solid solution to two phases takes place by uphill diffusion, where the diffusion coefficient is negative.

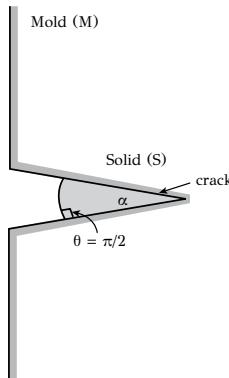
13. In continuous transformations, which evolve by the amplification of an initial compositional fluctuation, the interface between phases is not sharp but diffuse. In this case the interfacial energy is described by the gradient energy. The amplification or not of an initial compositional fluctuation during spinodal decomposition depends on the gradient energy and the coherency strains arising from the maintenance of coherency between the two phases.
14. Ordered phases are solid solutions possessing negative enthalpy of mixing, where positions of atoms A and B are not random as in disordered solutions. The atoms occupy specific positions at A and B sublattices, forming a superlattice, which is characterized by long range order (LRO).
15. During order-disorder transformations, an ordered solid solution with LRO transforms to a disordered one by heating above a critical temperature. In reverse transformation, order can evolve in two ways. The first is by a continuous transformation where local atomic rearrangements lead to short range order (SRO) and then to LRO . The second is by nucleation and growth of LRO regions, called ordered domains. These domains are separated by antiphase boundaries (APB).
16. The martensitic transformation is a displacive, diffusionless transformation. The strain energy arises from shear-like displacements and influences the transformation kinetics and the morphology of martensite.
17. The shape deformation associated with martensitic transformation is an invariant-plane-strain deformation (IPS). According to the phenomenological theory, the shape strain has three components: the Bain strain, the rigid body rotation and the complementary lattice-invariant deformation, by simple shear, of martensite. The invariant plane is the martensite habit plane.
18. Martensitic transformation is nucleation-controlled. The martensitic nucleation is heterogeneous and the nucleation sites are lattice defects in the parent phase, such as dislocation arrays at grain boundaries. Nucleation requires a critical chemical driving force, to overcome the energy barrier consisting of strain energy and interfacial energy contributions.
19. The growth of martensite is accomplished with the migration of the austenite/martensite interface. The interface is glissile and is composed of two arrays of parallel dislocations. The first array is composed of the partial coherency dislocations, the glide of which accomplishes the change of crystal structure (e.g.,

from *FCC* to *BCC*). The second array is composed of perfect anticoherency dislocations, the glide of which accomplishes the lattice-invariant shear.

20. The most important mechanical effects associated with martensitic transformation are transformation plasticity, thermoelasticity, pseudoelasticity and shape memory.
21. The characteristics of martensitic transformations are fundamental for several metallurgical technologies, such as ultra-high strength quenched and tempered steels, TRIP steels, maraging steels and shape memory alloys.

6.11 REVIEW QUESTIONS

1. Consider the nucleation of phase β at the grain boundaries of phase α . The interfacial energy is $\gamma_{\alpha\beta} = 600mJ/m^2$.
 - (a) If the grain boundary energy is $\gamma_{\alpha\alpha} = 800mJ/m^2$, calculate the reduction of the energy barrier with respect to homogeneous nucleation. (b) Calculate the value of $\gamma_{\alpha\alpha}$ for which the energy barrier vanishes.
2. Consider the solidification of an alloy in a mold at temperature T . The internal surface of the mold is not smooth and has small scratches or other geometrical discontinuities, which could act as heterogeneous nucleation sites for the solid phase. In the figure below, consider that a small crack-like discontinuity of unit length and angle α exists in the internal surface of the mold. The contact angle is θ .



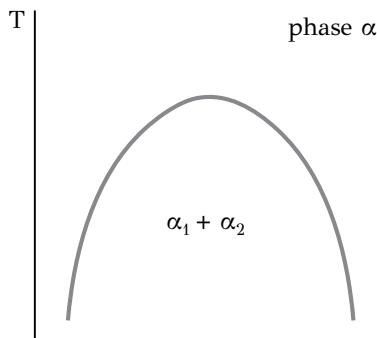
- (a) Show that the critical radius for nucleation r^* is independent of the contact angle.
- (b) Show that the activation energy for nucleation (energy barrier) varies linearly with the angle α of the discontinuity.
- (c) Compare the activation energies for the cases of a very acute angle of the discontinuity ($\alpha \ll \pi/2$) and a flat surface ($\alpha = \pi$).

3. Experimental studies on the nucleation of a new phase β indicate that when the undercooling $T = T_e - T_o$ in the $\alpha + \beta$ region (see Figure 6.4) is small, the β phase nucleates at the grain boundaries. On the other hand, if the undercooling is large, the β phase nucleates in the grain interior. Explain these observations based on your knowledge of diffusional phase transformations.
4. The precipitation sequence during aging of $Al - Ag$ alloys is



where $SSSS$ is the supersaturated solid solution of the α phase ($Al - Ag$).

- (a) Using thermodynamics arguments, explain why GP zones, which form at short aging times, exhibit a spherical shape.
- (b) Describe the experimental techniques that you would use to determine the size and volume fraction of GP zones after a certain aging time.
- (c) Explain, in quantitative terms, the way GP zones affect the yield strength of an $Al - Ag$ alloy, based on your knowledge of dislocation-particle interactions, discussed in Chapter 3.
5. Consider spinodal decomposition
 - (a) Draw and discuss the differences between the miscibility gap, the chemical spinodal and the coherent spinodal lines in the phase diagram shown below.
 - (b) Why is it necessary to modify Fick's second law in order to describe the kinetics of spinodal decomposition?
 - (c) Which factor is responsible for the growth of concentration fluctuations in preferred orientations during spinodal decomposition?
 - (d) Discuss which are the most important criteria in selecting an alloy system for strengthening through spinodal decomposition.



6. Discuss:
 - (a) the effect of plastic deformation on the kinetics of nucleation and growth phase transformations.
 - (b) if the above effects depend on the temperature where the deformation took place. Consider the cases of hot and cold working.
7. Consider the precipitation of β phase from a supersaturated solid solution of α phase.

- (a) Which factors influence the morphology of the β phase?
 - (b) Assuming that strain energy is negligible, what is the most possible shape of the β phase when (i) the interfacial energy $\gamma_{\alpha\beta}$ is isotropic and (ii) when $\gamma_{\alpha\beta}$ is highly anisotropic.
 - (c) What is the most possible shape, when the formation of the β phase is accompanied by a considerable volume change?
8. Differentiate between martensitic transformation and deformation twinning in metals. Discuss similarities and differences regarding crystallography, structure and kinetics.
9. Consider martensitic transformations.
- (a) Differentiate between invariant plane strain and invariant lattice deformation.
 - (b) Why the habit plane exhibits irrational Miller indices?
 - (c) Regarding martensitic interfaces, differentiate between coherency and anticoherency dislocations.
 - (d) Explain why martensitic nucleation is mostly heterogeneous.
 - (e) Is martensitic transformation, nucleation or growth controlled?
 - (f) What is the origin of stress effects, such as pseudoelasticity, in martensitic reactions?

BIBLIOGRAPHY—FURTHER READING

1. Christian, J.W. (1965). *The Theory of Transformations in Metals and Alloys*, Pergamon Press, New York.
2. Turnbull, D. (1956). in *Solid State Physics*, Vol. 3, Academic Press, New York.
3. Porter, D.A., Easterling, K.E., Sherif, M.Y. (2009). *Phase Transformations in Metals and Alloys*, CRC Press.
4. Doherty, R. D. (1983). *Diffusive Transformations in the Solid State*, in Physical Metallurgy, editors R. W. Cahn, P. Haasen, Elsevier Science Publishers.
5. Wayman, C.M. (1983). *Phase Transformations, NonDiffusive in Physical Metallurgy*, editors R. W. Cahn, P. Haasen, Elsevier Science Publishers.
6. Cohen, M., Wayman, C.M. (1981). *Fundamentals of Martensitic Reactions*, in Metallurgical Treatises, Eds J. Tien and J.F. Elliott, TMS-AIME, USA.
7. Reed-Hill, R.E. (1973). *Physical Metallurgy*, 2nd ed., Ch.13: *Nucleation and Growth Kinetics*, Brooks/Cole Engineering Division, USA.
8. Aaronson, H.I., Enomoto, M., Lee, J.K. (2010). *Mechanisms of Diffusional Phase Transformations on Metals and Alloys*, CRC Press.

7 Plastic deformation and annealing

7.1 INTRODUCTION

Metals are inherently ductile materials. They can withstand large plastic deformations before fracture, even at low temperatures. Of course there are exceptions to this rule. Structural steel, for example, exhibits a considerable drop in ductility with decreasing temperature. The phenomenon has been termed ductile-to-brittle transition and is of concern not only in steels but in other *BCC* metals as well. On the other hand, the *FCC* metals do not exhibit this temperature sensitivity. Those metals, such as copper, possess a greater ability for plastic deformation before fracture, compared to the *BCC* or even to the *HCP* metals. It is apparent that ductility is directly related to the ability of a metal to withstand large amounts of plastic deformation and this property is also related to crystal structure. Furthermore, other aspects of mechanical behavior besides fracture, such as fatigue and creep, are also affected by plastic deformation. Fatigue involves the accumulation of plastic strain during cyclic loading, which eventually leads to cracking and final fracture after a certain number of cycles. Creep involves plastic deformation at high temperatures, a process assisted by diffusion, leading to excessive plastic strains or even rapture. All the failure phenomena mentioned above share a common background: plastic deformation. However, in addition to the failure phenomena, plasticity plays an important role in mechanical processing and especially in the deformation processing of metals, such as rolling, extrusion, deep drawing and forging. In these cases the formability of the metal depends on its ability to withstand large amounts of plastic deformation. To be more exact, formability depends on the amount of uniform plastic deformation, which is related to the stability of plastic flow, i.e., the ability of the metal to delay plastic flow localization. Even in the recently developed joining process of friction-stir welding (*FSW*), plastic deformation plays a key role, since the formation of the weld zone is accomplished by severe plastic deformation of the two metals to be welded.

From the above discussion it is apparent that the study of plastic deformation is a prerequisite for the understanding of the mechanical behavior as well as the deformation processing of metals and alloys. The present chapter deals with the mechanisms of plastic deformation, focusing on the principal mechanism of dislocation glide. We will start with the fundamentals of single crystal plasticity and will generalize for polycrystalline aggregates. We will see that plastic deformation is accompanied by hardening, termed strain hardening or work hardening, which is one of the main strengthening mechanisms of metals. On the other hand ductility and formability decrease with the progress of plastic deformation, in other words the ability of the metal to deform decreases with strain hardening. We will also see that annealing can re-

move partially or totally strain hardening. Annealing leads to material softening and the metal can regain its ability for further plastic deformation. Annealing is a process involving several stages such as recovery, recrystallization and grain growth. Certain microstructural changes take place during these stages, involving reduction of dislocation density, rearrangement of dislocations as well as nucleation and growth of new grains. During both, plastic deformation and annealing, grains reorient themselves into specific orientations and the material develops deformation or annealing texture. The metal is not as isotropic as it was, and properties depend on the direction of measurement.

7.2 MECHANISMS OF PLASTIC DEFORMATION

The discussion will be limited to deformations at low temperatures ($T < 0.3T_m$) since creep deformation will be presented in Chapter 9. In principle three mechanisms cause plastic deformation in metals (Figure 7.1):

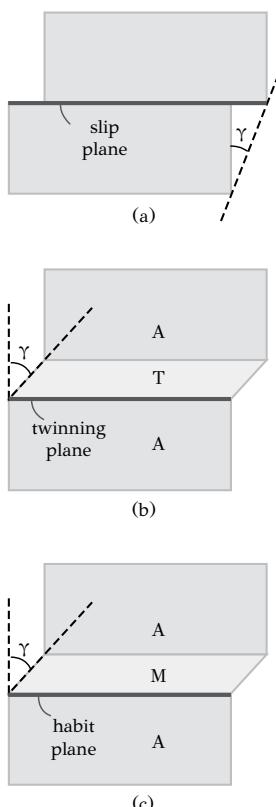


Figure 7.1: Mechanisms of plastic deformation in metals: (a) dislocation glide, (b) twinning, (c) martensitic transformation.

- slip
- mechanical twinning
- martensitic transformation

All the above mechanisms share a common characteristic. They involve plane shear deformation. In slip, shear is accomplished by dislocation glide in the slip plane. In mechanical twinning, shear is accomplished in the twinning plane by the formation of the twinned crystal. The crystal structure of the twinned crystal (T) is the same as the parent crystal (A). Finally martensitic transformation is characterized by a large shear strain in the habit plane. In this case, contrary to mechanical twinning, there is also a structural change taking place. The martensitic crystal (M) has a different crystal structure from the parent crystal (A). From the three mechanisms mentioned above, the most common is slip. Metals and alloys deform primarily with slip. The other mechanisms are activated, in most cases, when for some reason slip is impeded. At very low temperatures, for example, the dislocation mobility is reduced so that slip cannot take place. In this case the metal deforms by mechanical twinning. In austenitic stainless steels, at low temperatures, martensitic transformation takes place even at low stresses, well below the yield strength of the austenitic matrix. This stress-assisted martensitic transformation is accompanied by large shear deformations. Since we have already discussed martensitic transformations in Chapter 6, we will focus our discussion in this chapter on the mechanisms of slip and twinning.

7.3 DEFORMATION OF SINGLE CRYSTALS BY SLIP

We will discuss slip in a single grain in order to relate slip with crystal structure. We have seen that plastic deformation takes place by dislocation glide. It is important to note at this point that in order to understand the topics presented in this chapter, the reader should first be familiar with the geometrical characteristics of dislocations as well as the concepts of full, partial and mixed dislocations, stacking faults, dislocation loops, cross slip, dislocation climb, dislocation stress fields and dislocation energetics. These topics are treated in Chapter 3. The reader is advised to read this chapter first before proceeding with the study of plastic deformation. We have seen that, under the action of shear stresses, dislocations glide and deform the crystal. Dislocations are characterized by two vectors: the line vector \vec{s} , which indicates the direction and the sense of the dislocation line and the Burgers vector \vec{b} , which indicates the amount and direction of slip caused by the glide of a dislocation in the crystal. Dislocation glide takes place in the slip plane. If in Figure 7.2 the slip plane is characterized by the plane normal \vec{n} , then for edge dislocations:

$$\vec{b} \times \vec{s} = \vec{n}$$

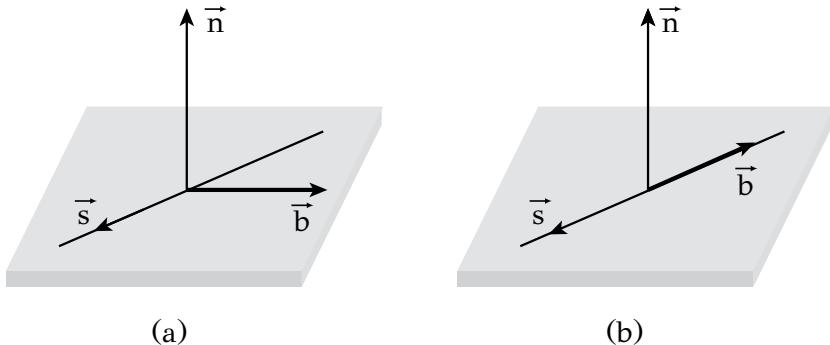


Figure 7.2: Slip plane of (a) edge and (b) screw dislocations.

while for screw dislocations:

$$\vec{b} \cdot \vec{s} = 1$$

We see that for edge dislocations the slip plane is fixed by \vec{b} and \vec{s} . On the contrary, screw dislocations can glide in multiple slip planes, since there are many planes containing \vec{b} and \vec{s} . This is very important in plastic deformation, since screw dislocations can change their slip system by cross slip, if glide in the current slip plane is impeded. It should be noted here that cross slip is easier to take place in *BCC* than in *FCC* or *HCP* crystals where full dislocations dissociate in partial dislocations, the dissociation being an obstacle to cross slip. We have also seen in Chapter 3 that slip takes place at specific slip systems, which include the slip plane and the slip direction. For example one such system is the $\{111\}\langle110\rangle$ in *FCC* referring to the close packed planes and close packed directions of the *FCC* crystal.

In order to activate slip in a crystal, the shear stress should reach a critical value. As depicted in Figure 7.3 the single crystal is subjected to a tensile force F , which acts at an angle ϕ with the slip plane and angle λ with the slip direction. The shear component of F in the slip plane is $F \cos \lambda$ while the slip plane has an area of $A / \cos \phi$. Therefore the resolved shear stress is

$$\tau = \frac{F}{A} \cos \phi \cos \lambda \quad (7.1)$$

Slip in the single crystal of Figure 7.3 is activated when the resolved shear stress reaches a critical value τ_c , the *critical resolved shear stress (CRSS)*. This is the so-called Schmid law, while the factor $\cos \phi \cos \lambda$ is termed the *Schmid factor* and expresses the orientation of the slip system relative to the applied stress. Because the tensile stress is $\sigma = F/A$, the yield strength in tension of the single crystal σ_o is related to the *CRSS* by:

$$\sigma_o = \frac{\tau_c}{\cos \phi \cos \lambda}$$

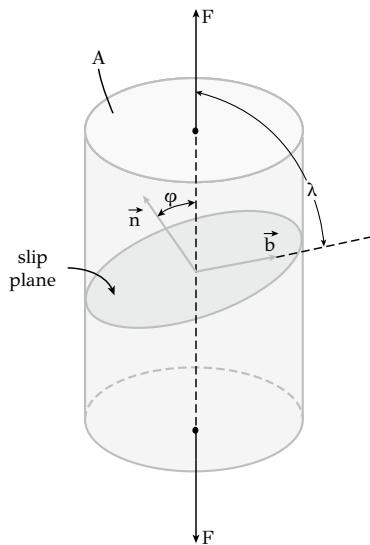


Figure 7.3: Single crystal in tension: the tensile axis forms an angle ϕ with the slip plane and an angle λ with the slip direction.

For a given value of τ_c the graph of the tensile yield strength is a hyperbola, depicted in Figure 7.4. The minimum value of σ_o appears when the Schmid factor takes the value 0.5, when $\phi = 45^\circ$. Therefore in planes oriented at 45° to the tensile axis the shear stress will reach the CRSS value and this is where slip will be activated first. If there is no plane with an exact 45° orientation, then slip will be activated in a plane with the closest orientation to 45° . These planes are said to be favorably oriented for slip. When in a single crystal, only one slip system is activated then the crystal is deformed by *simple slip*.

It has already been mentioned that a slip system includes the slip plane and the slip direction of the dislocations. The CRSS is lower in the close-packed planes and close-packed directions. In the FCC metals the close-packed directions are the $\langle 110 \rangle$ corresponding to the face diagonals, while the close-packed planes are the $\{111\}$ planes, as indicated in Figure 7.5a. Therefore the slip systems in FCC are $\{111\}\langle 110 \rangle$. Every close-packed plane of the $\{111\}$ family contains three close-packed directions, therefore in FCC there are $4 \times 3 = 12$ slip systems. In these systems the CRSS, which is required to activate slip, takes very low values as indicated in Table 7.1 for Cu, Ag, Au and Al.

In the BCC system (Figure 7.5b) the close-packed directions are the cube diagonals $\langle 111 \rangle$. However in the BCC system the closest-packed planes are the $\{110\}$ planes and, therefore, the principal slip systems in BCC are the $\{110\}\langle 111 \rangle$. However the $\{110\}$ planes in BCC do not possess the same close packed atomic arrangement with the $\{111\}$ planes of FCC. So the $\{110\}$ planes are not the exclusive slip planes in BCC. Other slip systems such as the $\{112\}\langle 111 \rangle$ and $\{123\}\langle 111 \rangle$ have

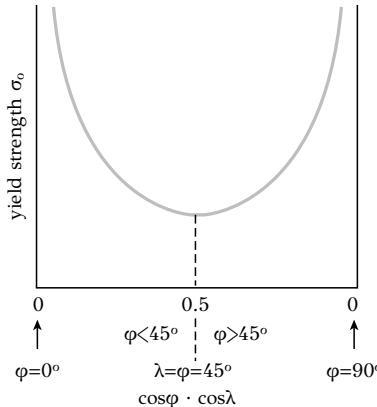


Figure 7.4: Variation of the tensile yield strength of a single crystal with the orientation of the slip plane (Schmid factor).

been observed and in general any plane containing the $\langle 111 \rangle$ directions could act as a potential slip plane in *BCC*. The *CRSS* in the *BCC* metals is also low, for example it is only 28 MPa for iron. In the metals of the *HCP* system (Figure 7.5c), the basal plane (0001) has the same close-packed arrangement with the $\{111\}$ planes of *FCC*. Therefore it would have been expected that slip in *HCP* should be as easy as in *FCC*. Actually, as indicated in Table 7.1, the *CRSS* takes very low values for the slip system $(0001)\langle 11\bar{2}0 \rangle$ for *Zn*, *Cd* and *Mg*. Slip in this system is termed *basal slip*. In two other important *HCP* metals, *Ti* and *Be*, the *CRSS* takes high values for basal slip. The system $(10\bar{1}0)\langle 11\bar{2}0 \rangle$ exhibits a much lower *CRSS* for *Ti*, while the same system is the principal system for *Zr*. The reason for this differentiation among the *HCP* metals is the deviation from the ideal $c/a = 1.63$ ratio. In *Mg*, which has a c/a ratio close to the ideal as well as in *Cd* and *Zn* with c/a ratio larger than the ideal value, the interplanar spacings of (0001) planes are larger than those of the metals *Ti*, *Be* and *Zr*, which have c/a ratios lower than the ideal value. As a consequence, *Ti*, *Be* and *Zr* have a high *CRSS* for basal slip and, therefore, these metals prefer to deform by *prismatic slip* using the system $(10\bar{1}0)\langle 11\bar{2}0 \rangle$, as indicated in Figure 7.5c.

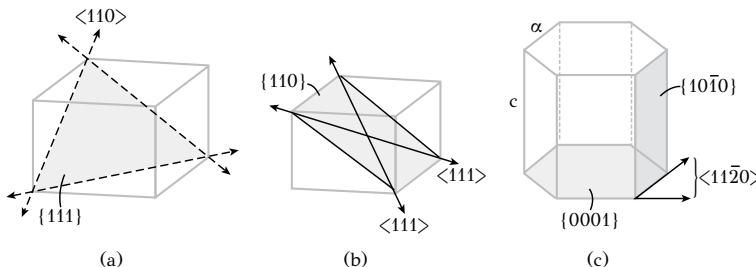


Figure 7.5: Principal slip systems in (a) *FCC*, (b) *BCC* and (c) *HCP* structures.

Table 7.1
Slip systems and CRSS of some high purity FCC, HCP and BCC metals at room temperature

Metal	Strain Field Type	Slip System	CRSS (MPa)
Cu	FCC	{111}⟨110⟩	0.63
Ag	FCC	{111}⟨110⟩	0.37
Au	FCC	{111}⟨110⟩	0.90
Al	FCC	{111}⟨110⟩	1.01
Zn	HCP $c/\alpha = 1.86$	(0001)⟨11̄20⟩ basal slip	0.18
Cd	HCP $c/\alpha = 1.88$	(0001)⟨11̄20⟩ basal slip	0.56
Mg	HCP $c/\alpha = 1.62$	(0001)⟨11̄20⟩ basal slip	0.43
Ti	HCP $c/\alpha = 1.59$	(0001)⟨11̄20⟩ basal slip	110.2
Ti	HCP $c/\alpha = 1.59$	(10̄10)⟨11̄20⟩ prismatic slip	48.9
Be	HCP $c/\alpha = 1.56$	(10̄10)⟨11̄20⟩ prismatic slip	39.2
Zr	HCP $c/\alpha = 1.59$	(10̄10)⟨11̄20⟩ prismatic slip	6.2
α -Fe, W, Mo	BCC	{110}⟨111⟩ the most important system of BBC	28
α -Fe, W, Mo	BCC	{211}⟨111⟩	
α -Fe	BCC	{321}⟨111⟩	

Now consider the tensile deformation of a single crystal. It will be shown that slip is accompanied by a reorientation of the crystal, which is caused by the rotation of the crystal relative to the tensile axis. It will be shown that the length and the direction of a random line in the crystal, changes during slip. This change depends on the magnitude of shear and the original orientation of the line relative to the slip system. In Figure 7.6a a single crystal is subjected to simple shear in a slip system defined by \vec{b} and \vec{n} . Consider now a line AB defined by the vector \vec{l} . After a certain amount of slip the crystal takes the form of Figure 7.6b. The shear strain is given by

$$\gamma = \frac{|\vec{u}|}{\vec{l} \cdot \vec{n}}$$

where \vec{u} is the displacement of point B in the slip direction, shown in Figure 7.6c,

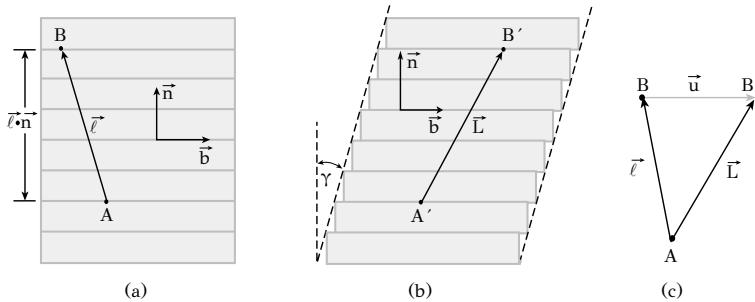


Figure 7.6: Length and orientation change of a linear segment AB in a single crystal due to slip: (a) before slip, (b) after slip, (c) displacement in the slip direction.

and $\vec{l} \cdot \vec{n}$ is the vertical distance between points A and B, which includes all the crystal planes which slipped and caused the change of AB to $A'B'$.

The displacement \vec{u} is then, for small values of γ

$$\vec{u} = \gamma(\vec{l} \cdot \vec{n})\vec{b}$$

After deformation the line $A'B'$ is defined by the vector \vec{L} , where

$$\vec{L} = \vec{l} + \vec{u}$$

which becomes

$$\vec{L} = \vec{l} + \gamma(\vec{l} \cdot \vec{n})\vec{b} \quad (7.2)$$

It is apparent that the line AB has changed in length and direction (from \vec{l} to \vec{L}) and that this change depends on shear strain and the initial orientation (\vec{l}). The only case where the line AB remains invariant during deformation is when the second term on the rhs of (7.2) becomes zero, i.e.,

$$\vec{l} \cdot \vec{n} = 0$$

that is when line AB is on the slip plane. Therefore we can say that all lines on the slip plane remain invariant during simple slip, while all other lines reorient themselves relative to the slip system.

Now consider the tensile deformation of a single crystal along line $AB = \vec{l}$, as depicted in Figure 7.7. Assume that the crystal deforms by simple slip in a slip system defined by \vec{b} and \vec{n} . The tensile axis is at an angle ϕ with \vec{n} (with the slip plane) and at an angle λ with the slip direction, as we have seen in Figure 7.3. According to the above discussion, the tensile axis changes orientation relative to AB and becomes $A'B' = \vec{L}$, forming angles ϕ' and λ' with the slip plane and the slip direction respectively.

As a consequence the tensile axis rotates relative to the crystallographic axes. Of course, during tensile testing in a servohydraulic machine, the loading direction

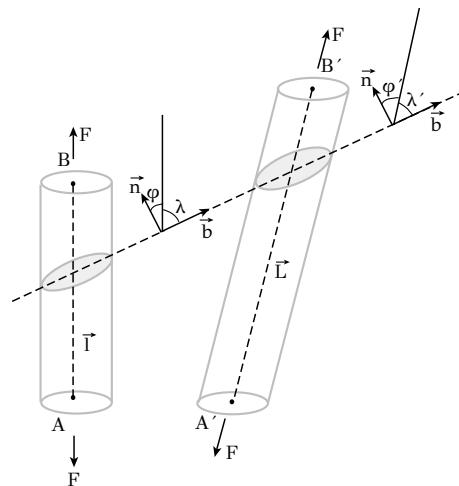


Figure 7.7: Change of the orientation of the tensile axis relative to the slip plane during the deformation of a single crystal.

does not change relative to the laboratory. This simply means that it is the crystal that rotates relative to the tensile axis. This is shown in greater clarity in Figure 7.8. Before deformation, the tensile axis forms angles ϕ and λ with the slip plane and the slip direction respectively (Figure 7.8a).

After deformation (Figure 7.8b) the slip of the crystal planes causes the rotation of the tensile axis to its new orientation. However since the tensile axis remains invariant relative to the laboratory, this corresponds to a rotation of the crystal relative to the tensile axis. Now, after deformation (Figure 7.8c) the tensile axis forms angles ϕ' and λ' with the slip plane and the slip direction respectively. If the tensile axis before deformation is defined by \vec{l} and after deformation by \vec{L} then the rotation is expressed by a vector \vec{r} which is the cross product of \vec{l} and \vec{L}

$$\vec{r} = \vec{l} \times \vec{L}$$

which due to Equation (7.2) becomes

$$\vec{r} = [\vec{l} + \gamma(\vec{l} \cdot \vec{n})\vec{b}] \times \vec{L}$$

and after some algebra becomes

$$\vec{r} = \gamma(\vec{l} \cdot \vec{n})(\vec{b} \times \vec{l})$$

It is apparent that the rotation axis is perpendicular to both the tensile axis \vec{l} and the slip direction \vec{b} . The sense of rotation is such that the angle between \vec{b} and \vec{l} , i.e., the angle λ is always reduced. In other words, during simple slip, the crystal rotates and the slip direction approaches the tensile axis. The opposite rotation takes

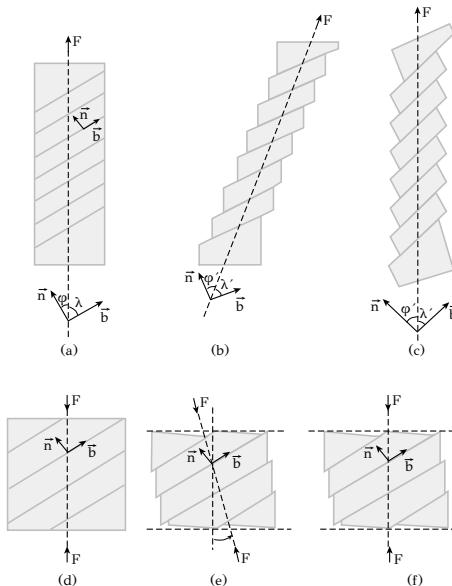


Figure 7.8: Crystal rotation during slip: (a) before uniaxial tension, (b) rotation of tensile axis, (c) corresponding crystal rotation, (d,e,f) corresponding rotation during uniaxial compression.

place during compression. As depicted in Figure 7.8d,e,f, during deformation the slip direction moves away from the compression axis. Due to crystal rotation during simple slip, there is a variation of the Schmid factor ($\cos \phi \cos \lambda$) in the rest of the slip systems. In another system the CRSS is reached and slip is activated also in that system. Simple slip turns into *multiple slip*, meaning that slip is activated in several slip systems which have reached the required CRSS. An important consequence of multiple slip is the intersection of dislocations which glide in non-parallel slip planes. These intersections impede dislocation motion and cause strain hardening, as will be discussed below.

Another consequence of multiple slip is that it affects the deformation of polycrystals. It will be shown that the deformation of a grain cannot be performed independently but it has to adapt to the deformation of neighboring grains.

7.4 DEFORMATION IN POLYCRYSTALS

The plastic deformation of polycrystalline aggregates is more complicated than the deformation of single crystals. The basic differences are the following:

- a) Single crystals exhibit elastic and plastic anisotropy. On the other hand, polycrystals can be considered as isotropic materials.

- b)** Single crystals can be deformed significantly with the activation of only one slip system (simple slip). As the grains in a polycrystal have different orientations they deform directly with multiple slip. Polycrystals, therefore, cannot be deformed by simple slip.
- c)** The grains in a polycrystal cannot deform independently like a single crystal but, instead, since they must remain in contact during deformation they should adapt to the deformation of the neighboring grains. In other words, the deformation of one grain must be compatible with the deformation of the neighboring grains.
- d)** The presence of grain boundaries impedes dislocation glide and the propagation of slip from one grain to the next. This leads to the dependence of the flow stress from the grain size, a matter to be discussed in the next chapter.

The *compatibility criterion* (point c above) was studied originally by Von Mises¹, who showed that the homogeneous plastic deformation of polycrystals requires the activation of at least five (5) independent slip systems. From the nine components of the deformation tensor, three are fixed by the relation $\varepsilon_{ij} = \varepsilon_{ji}$ and a fourth by the relation $\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33} = 0$, since the volume is constant during plastic deformation. This leaves five components of the deformation tensor to be defined independently. An independent slip system is one that produces a shape change, which cannot be accomplished by any other combination of the remaining slip systems. The *FCC* metals have twelve $\{111\}\langle110\rangle$ slip systems. These systems correspond to $12!/(5!7!) = 798$ combinations of five systems. From these combinations, 384 are independent. In *BCC* metals there are also 384 combinations of five systems of the type $\{110\}\langle111\rangle$. In *HCP* metals there are three systems of which only two are independent. This is the reason for the limited ductility of *HCP* metals, especially at low temperatures, and for the activation of alternative deformation mechanisms, such as mechanical twinning.

Now the question is: which group of five slip systems will be activated and cause a shape change under constant volume? Taylor² studied the 384 combinations of slip systems in *FCC* and *BCC* metals and defined a criterion, known as the *Taylor criterion*, stating that the operational combinations of slip systems are those, which minimize the strain energy. If the stress σ_{ij} causes an increment of plastic strain $\delta\varepsilon_{ij}$ then the strain energy is

$$\delta W = \sigma_{ij} \delta \varepsilon_{ij}$$

However δW is equal to the work that is performed in the five slip systems, which are activated by the same shear stress τ_c

$$\delta W = \tau_c \sum_k \delta \gamma_k$$

¹R. Von Mises, Z. Angew. Math. Mech. 8 (1928) 161.

²G.I. Taylor, Plastic strain in metals, J. Institute of Metals 62 (1938) 307.

where $\delta\gamma_k$ is the shear strain in the slip plane k of the group ($k = 1 \dots 5$). The above relations lead to

$$\sigma_{ij}\delta\varepsilon_{ij} = \tau_c \sum_k \delta\gamma_k$$

and after rearranging terms

$$\frac{\sigma_{ij}}{\tau_c} = \frac{\sum_k \delta\gamma_k}{\delta\varepsilon_{ij}} = M_T$$

The term M_T is termed the *Taylor factor* and plays the same role in multiple slip of polycrystals as the Schmid factor in single crystals. It relates the applied tensile stress with the shear stress in the active slip systems. For cubic crystals, Taylor found that $M_T = 3.06$. If the tensile yield strength of a polycrystal is σ_y then the above relation becomes

$$\sigma_y = M_T \tau_c \quad (7.3)$$

which expresses the relation between the tensile yield strength of a polycrystal to the CRSS of a single crystal. The corresponding relation for strains is

$$\varepsilon = \frac{\gamma}{M_T} \quad (7.4)$$

The relations (7.3) and (7.4) relate the tensile stress-strain curve of a polycrystal with the shear stress-strain curve of a single crystal of the same metal.

7.5 STRAIN HARDENING

When a metal is plastically deformed, its resistance to further plastic deformation increases. This phenomenon is termed *strain hardening* (or work hardening). During plastic deformation the density and the distribution of dislocations change. The dislocation density increases, rapidly, due to the activation of dislocation sources, which produce new dislocations. On the other hand, dislocation interactions, caused by multiple slip impede plastic deformation. Dislocations are immobilized and the mobile dislocation density, i.e., the density of those dislocations, which could contribute to plastic deformation, decreases. Strain hardening is expressed by the increase of the applied stress required to continue plastic deformation. Strain hardening is a complex phenomenon since the density and the distribution of dislocations depend on several factors such as crystal structure, temperature and deformation rate. In the next sections, the basic characteristics of strain hardening, including dislocation multiplication and dislocation interactions will be presented. Strain hardening is one of the most important strengthening mechanisms, considering that several metals have limited other options for developing high strength, for example with precipitation hardening. These metals and alloys develop their strength during cold working, such as cold rolling, stretching or deep drawing. In most cases strain hardening is accompanied by a reduction in ductility. It is for this reason that cold working is followed

by annealing, for the reduction of hardness and increase of ductility. In fact, strain hardening is removed or canceled by annealing. The mechanisms of annealing will be discussed later in the present chapter.

7.5.1 DISLOCATION MULTIPLICATION

Plastic deformation raises the dislocation density. While in an annealed metal the dislocation density is of the order of $\rho = 10^4 \text{ mm}^{-2}$, it increases to $\rho = 10^9 \text{ mm}^{-2}$ with plastic deformation. This simply means that certain mechanisms of dislocation multiplication, i.e., dislocation sources, are activated in the course of plastic deformation. The most common dislocation source is the *Frank-Reed source*, shown in Figure 7.9. A dislocation line is pinned in two obstacles *A* and *B* on the slip plane. These obstacles could be second-phase particles or other dislocations crossing the slip plane. The distance between obstacles is *L*. Under the action of the applied shear stress (Figure 7.9b) the dislocation bows between the two obstacles. As the radius of curvature *R* decreases, the stress required for further bowing increases. The maximum stress is obtained when the dislocation assumes a semicircular shape with $R = L/2$. It has been shown, in Chapter 3, that the maximum stress is $\tau = \mu b/L$. With the continuous application of the shear stress the dislocation bows further with a larger *R* (Figure 7.9c) and forms a loop since the segments *x* and *y* are annihilated when they meet (Figure 7.9d). This is because they move to opposite directions, having the same Burgers vector *b* and opposite line vector *s*. The result is the formation of a dislocation loop, which extends with the sustained application of stress. At the same time a new dislocation segment *AB* repeats the process for the formation of a new loop (Figure 7.9e). In effect, the Frank-Reed source emits dislocation loops when activated with the stress $\tau = \mu b/L$.

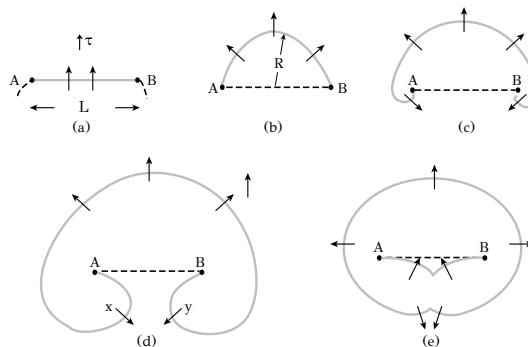


Figure 7.9: Consecutive positions of the dislocation segment *AB* and loop formation during the activation of a Frank-Reed source.

7.5.2 MECHANISMS OF STRAIN HARDENING

Typical stress-strain curves for single crystals of *FCC*, *BCC* and *HCP* metals are depicted in Figure 7.10. Regarding the *FCC* crystals, three distinct deformation stages can be observed. In stage *I*, only one of the twelve equivalent slip systems is activated, this being the one, which is most favorably oriented to the applied stress.

Slip in this system produces the observed plastic strain in stage *I*. Dislocation sources are also activated, producing dislocations, which also contribute to the deformation in stage *I*. In general, the deformation during stage *I* is not impeded, as expressed by the low strain hardening rate (slope λ_I). For this reason, stage *I* has been termed *easy glide*. The extent of easy glide depends on the orientation of the crystal in relation to the applied stress and on crystal structure. In *FCC* crystals the deformation in stage *I* is limited to few percent (up to 10%), because slip is activated in other equivalent systems causing multiple slip, indicating the start of stage *II*. On the other hand, in *HCP* crystals and especially in *Zn*, *Cd* and *Mg* the extent of stage *I* can be very large due to the activation of only one slip system. In *BCC* metals, stage *I* is absent since there are several slip systems with a common $\langle 111 \rangle$ direction, causing multiple slip. The termination of stage *I* and beginning of stage *II* coincides with the activation of other equivalent or secondary slip systems, in which the shear stress reaches the *CRSS*. An example of slip system activation sequence is shown in Figure 7.11. A Nickel single crystal is loaded in tension along the $[1\bar{3}5]$ axis. The first system, where stress reaches the *CRSS* is $(\bar{1}\bar{1}1)[101]$, depicted by (a) on the stress-strain curve. Multiple slip is denoted by the activation of other systems denoted by b, c, d and e. In *FCC* crystals, multiple slip in equivalent and intersecting slip planes causes dislocation interactions, which lead to the formation of obstacles to plastic deformation. These obstacles will be discussed below. The strain hardening rate λ_{II} remains constant during deformation and stage *II* has been termed as the stage of *linear strain hardening*. The dislocation density increases rapidly in stage *II* approaching values of the order of 10^9 mm^{-2} . Furthermore, the dislocations form groups of entangled dislocations, incapable for further glide. The only way for plastic deformation to proceed is the change of slip plane. This coincides with the beginning of stage *III*, termed *dynamic recovery*. Dislocations can switch slip planes by two mechanisms, discussed in Chapter 3, climb and cross-slip. Dislocation climb requires diffusion of point defects and is activated at high temperatures. On the other hand cross-slip does not require thermal activation and, therefore, the deformation in stage *III* takes place with the cross-slip of screw dislocations. The strain hardening rate λ_{III} in stage *III* is reduced relative to stage *II*. As shown in Figure 7.10, *BCC* metals exhibit mainly stage *III* deformation, while in *HCP* metals stage *III* is totally absent.

The basic mechanisms, which impede dislocation glide and cause strain hardening are the following:

Dislocation intersections. As discussed in Chapter 3, when dislocations intersect they form kinks and jogs. Both possess the same Burgers vector with the dislocation to which they belong. Kinks share the same slip plane and, therefore, they do not impede the glide of dislocations. On the other hand, jogs impede dislocation glide. The jogs formed at a screw dislocation, in particular, have an edge character, making

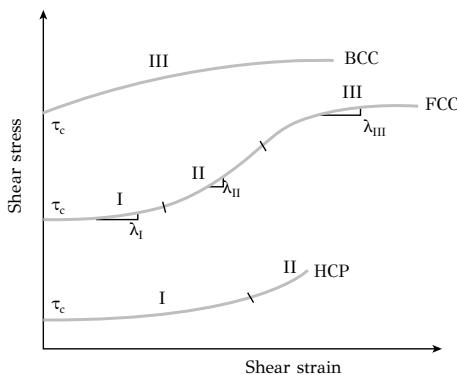


Figure 7.10: Stress-strain curves in shear for *BCC*, *FCC* and *HCP* single crystals.

Table 7.2
Stacking fault energies of various metals

Metal	Stacking Fault Energy (mJ/m^2)
Austenitic Stainless Steel (<i>Fe-Cr-Ni</i>)	10
Brass (<i>Cu-Zn</i>)	10
Silver (<i>Ag</i>)	25
Gold (<i>Au</i>)	75
Copper (<i>Cu</i>)	90
Nickel (<i>Ni</i>)	200
Aluminum (<i>Al</i>)	250

the glide of such a screw dislocation very difficult, since its glide requires the climb of the jog. In effect, the screw dislocation is immobilized. Multiple slip, accompanied by dislocation intersections leads to the formation of dislocation tangles, which contain dislocations with limited mobility and, thus, contribute to strain hardening.

Dislocation dissociation. In several cases the dissociation of a dislocation into partial dislocations with the simultaneous formation of a stacking fault is energetically favored (Chapter 3). The dissociation process is more common in metals with low stacking fault energies (*SFE*). Values of the *SFE* appear in Table 7.2.

FCC metals possess lower *SFE* than *BCC* metals and the dissociation process is easier. However when dislocations dissociate they cannot cross-slip, since the dislocations acquire a specific slip plane, which is defined by the stacking fault. Therefore dislocation dissociation hinders the cross-slip of screw dislocations and the dissociation into partials contributes to strain hardening.

Lomer-Cottrell locks. During multiple slip dislocations glide in intersecting slip planes and interact, forming immobile dislocations called *sessile dislocations*. One of

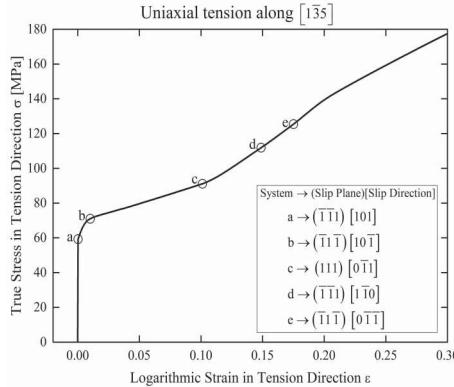


Figure 7.11: Sequence of activation of slip systems in a nickel single crystal loaded in tension.

Courtesy of N. Aravas, N. Vasiou and I. Bellas, University of Thessaly.

the most common mechanisms for the formation of sessile dislocations is the Lomer-Cottrell lock, indicated in Figure 7.12 for the *FCC* system. The dislocation $a/2[101]$ glides on the $(11\bar{1})$ plane while the dislocation $a/2[\bar{1}\bar{1}0]$ glides on the $(1\bar{1}\bar{1})$ plane. When the two dislocations intersect, a new dislocation forms by the reaction

$$\frac{a}{2}[101] + \frac{a}{2}[\bar{1}\bar{1}0] = \frac{a}{2}[0\bar{1}1]$$

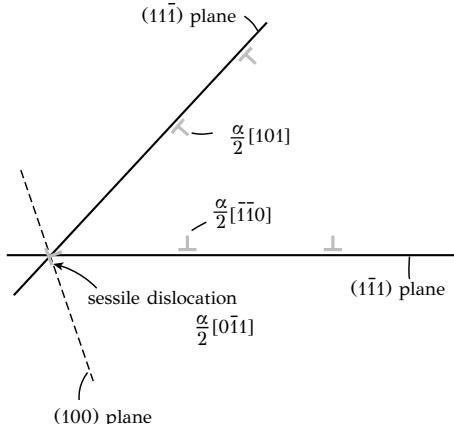


Figure 7.12: Formation of sessile dislocation at a Lomer-Cottrell lock.

The reaction is energetically favored since Frank's rule is obeyed. However the new dislocation $a/2[0\bar{1}1]$ can only glide on the (100) plane, which is not a close-packed plane, and therefore it is not a slip plane in *FCC*. The dislocation $a/2[0\bar{1}1]$ is,

therefore, a sessile dislocation. The dislocations arriving at the lock, cannot continue their glide and are immobilized, forming dislocation pile-ups. In this way, Lomer-Cottrell locks contribute to strain hardening.

Forest hardening. An additional mechanism, which contributes to strain hardening, is the interaction of dislocations gliding on the slip plane with dislocations intersecting the slip plane. These dislocations form a forest and are called forest dislocations (see Figure 7.13). A dislocation gliding on the slip plane should cross the forest of immobile dislocations. The forest dislocations have a mean separation distance L and their density is $\rho_f = 1/L^2$. Intersection points act like a Frank-Reed source and the stress required to cross the intersection points is, therefore, $\tau_f = \mu b/L$, and

$$\tau_f = \mu b \sqrt{\rho_f}$$

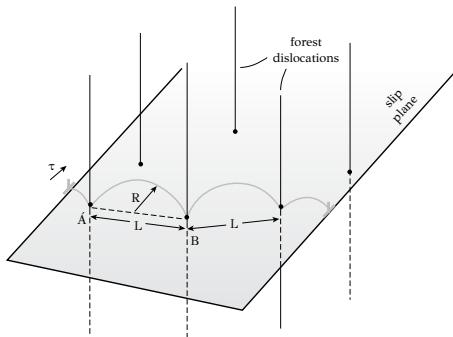


Figure 7.13: Dislocation glide on a slip plane in the presence of forest dislocations.

The density of the forest dislocations ρ_f is related to the total dislocation density ρ via a factor α , then the stress required to cross the forest is

$$\tau_f = \alpha \mu b \sqrt{\rho} \quad (7.5)$$

where α takes the value of about 0.5. As indicated by the Relation (7.5) the increase of dislocation density due to dislocation multiplication leads to an increase of the flow stress, termed *forest hardening*. Strain hardening in stage II is directly related to forest hardening. The increase in dislocation density with deformation is shown in Figure 7.14. With increasing plastic deformation, the dislocations form cell structures where the dislocations arrange in cell walls, while the cell interior is essentially dislocation free.

7.5.3 STRAIN HARDENING IN POLYCRYSTALS

In polycrystalline metals the resistance to further plastic deformation is caused by the interactions between dislocations and, therefore, the same mechanisms of strain hardening, discussed above for single crystals, are activated. However due to the requirement for compatibility of the deformation between neighboring grains, multiple

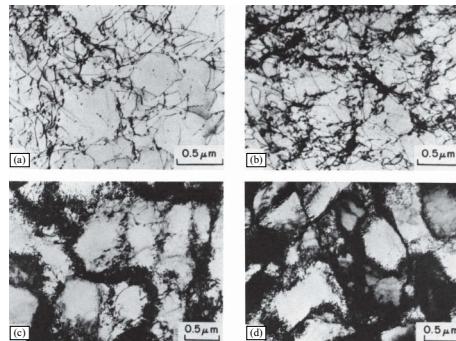


Figure 7.14: Dislocation arrangements in the ferrite phase of a Dual Phase steel during plastic deformation: (a) 1 %, (b) 2%, (c) 7%, (d) 14% strain.

Reprinted from: D.A. Korzekwa, D. K. Matlock, and G. Krauss, Metall. Trans. A, Vol. 15A (1984) 1221-1228, with permission from Springer.

slip starts immediately at the beginning of deformation and, therefore, strain hardening develops early in the deformation process. In general, polycrystalline metals do not exhibit stage I and stage II hardening but deform with stage III hardening from the beginning of deformation. A typical stress-strain curve for a polycrystalline metal is shown in Figure 7.15. It has a parabolic shape and can be expressed by

$$\sigma = K \varepsilon^n \quad (7.6)$$

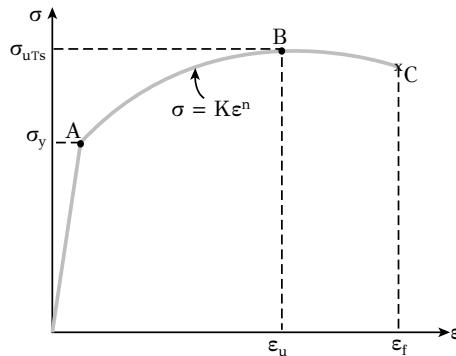


Figure 7.15: Typical stress-strain curve of a polycrystalline metal.

where K is a constant and n is the strain-hardening exponent. K varies between 10^{-3} and $10^{-2} \mu$ while n varies between 0.1 and 0.5 for most metals. It should be pointed out that the Relation (7.6) is an empirical relation and does not describe the physical mechanisms of strain hardening. This is because: (a) in different parts of the curve there are different strain hardening mechanisms operating and (b) the plastic strain

increment depends on the deformation history, a fact, which does not allow us to define a unique relationship between the plastic strain and the state of dislocations in a polycrystal. The stress-strain curve of Figure 7.15 contains some additional interesting features. The *yield strength* σ_y is the stress required to activate generalized plastic deformation in the polycrystal. In this sense the yield strength (point A) separates the elastic from the plastic regime in the stress-strain curve. The stress σ_{UTS} (Point B) is the *ultimate tensile strength*. It is the highest stress the material can withstand before failure. Between σ_y and σ_{UTS} the material strain hardens and deforms uniformly. This means that a cylindrical specimen elongates with a simultaneous reduction in cross sectional area, which is uniform across the specimen length. At point B the specimen forms a neck and the deformation is localized at the neck. The cross section of the neck is continuously decreased up to fracture (point C). At point B plastic flow is destabilized. While between points A and B the flow is stable and uniform, at point B the flow is destabilized and is localized in the neck region of the specimen. It is important to understand the *localization* and *destabilization* of plastic flow. During deformation from point A to point B the cross section of the specimen is decreased uniformly. This causes a corresponding increase in stress. The material responds to the increased stress exhibiting a higher resistance, which stems from strain hardening. However the strain-hardening rate decreases with deformation and the strain hardening capacity is exhausted. At point B there is no more increase in strength to offset for the increased stress. The plastic flow is destabilized and localizes at a certain geometrical discontinuity at the specimen surface, forming the neck. At this point we should differentiate between *uniform elongation* ϵ_u and *total elongation* ϵ_f . In deformation processing, the deformation limits are defined by the appearance of surface defects, which are caused by the localization of plastic flow. Therefore, the uniform elongation characterizes the *formability* of the material. The total elongation expresses the total plastic deformation before fracture and, therefore, characterizes the *ductility* of the material.

7.6 MECHANICAL TWINNING

As discussed in section 7.2, in addition to inhomogeneous deformation by slip, metal crystals can be deformed by homogeneous shear, forming twins. Mechanical twinning does not change the crystal structure (as martensitic transformation) but it changes the orientation of the twinned crystal. Twinning is not the primary deformation mechanism in metals exhibiting several slip systems, like the *FCC* metals. It is, however, an important mechanism for metals with a limited number of slip systems, like the *HCP* metals, in which slip is normally limited to basal slip. It has also been discussed that in polycrystalline metals, for compatibility reasons, the simultaneous activation of several slip systems is required. Again *HCP* metals do not have the required number of slip systems and twinning constitutes an important alternative deformation mechanism. The metals of the *FCC* and *BCC* systems deform by twinning when, for some reason, slip is impeded. This is usually the case at low temperatures, where the dislocation mobility is very low. Mechanical twins are shown in Figure 7.16 for an *Mg-Al-Zn* alloy. The twins have formed after mild plastic deformation.

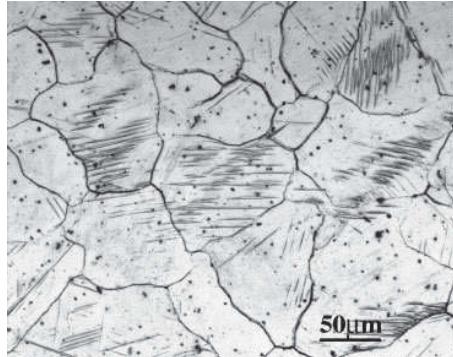


Figure 7.16: Mechanical twins in a Mg-3Al-1Zn alloy after mild plastic deformation.

Twining takes place by a cooperative movement of atoms, where each atom is displaced by a fraction of the interatomic spacing. In this way, twinning results in homogeneous shear of the lattice. The atomic displacements during twinning are shown in Figure 7.17. The twinning shear takes place on the twinning plane, K_1 , in the twinning direction, η_1 .

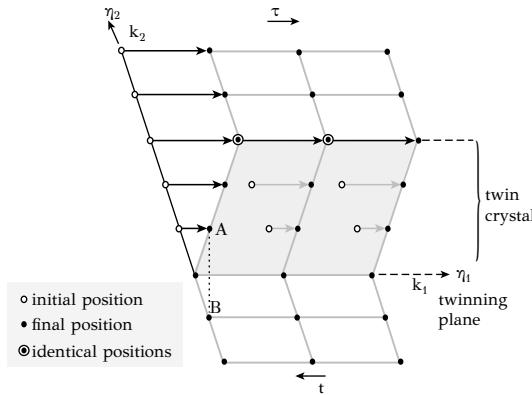


Figure 7.17: Deformation twinning: cooperative atomic displacements for the formation of the twin crystal. K_1 is the twinning plane and η_1 is the twinning direction.

The twinned crystal has a different orientation than the matrix (parent crystal). It is actually the mirror image of the parent crystal. Point A of the twinned crystal, for example, is the mirror position of point B of the parent crystal relative to the twinning plane. Contrary to slip, which involves atomic displacements of one or more interatomic spacings, twinning is the result of shorter displacements, which take place at successive planes of the lattice, as seen in Figure 7.17. The geometric characteristics of twinning are shown in Figure 7.18. Consider a spherical single crystal with the twinning plane defined by the equator plane K_1 and twinning direction η_1 . The

upper hemisphere is deformed by twinning, and turns into an ellipsoid. There are only two planes, which remain invariant during twinning. The twinning plane K_1 and plane K_2 . The K_2 plane forms an angle 2ϕ with the twinning plane before and after twinning. If it is assumed that the sphere has a unit radius, then the shear strain due to twinning is $\gamma = AA'$ and then

$$\gamma = 2 \cot 2\phi$$

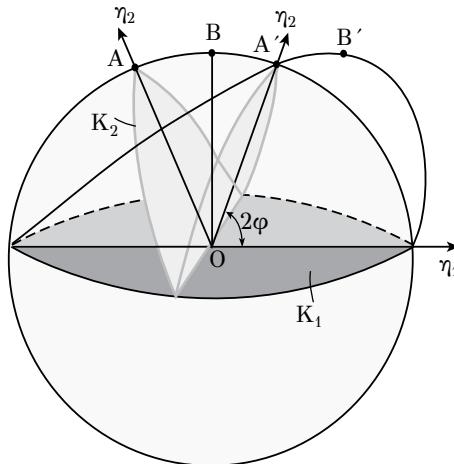


Figure 7.18: Deformation twinning of a spherical single crystal depicting the magnitude of shear. Planes K_1 and K_2 remain invariant during twinning.

If the invariant planes K_1 and K_2 are defined, then the twinning shear can be calculated from the angle between those planes. The crystallographic characteristics and the twinning shear are given for various metals in Table 7.3. Although it appears that in the *FCC* and *BCC* metals the twinning shear is large, the contribution of twinning to the total deformation is rather low. The reason is that the volume fraction of twins is low. There is, however, a larger contribution of twinning in the *HCP* metals.

In this case the mode of deformation, i.e., either tension or compression, influences twinning. Furthermore, this behavior depends on the c/a ratio. As depicted in Figure 7.19a for the case of *Zn*, with $c/a = 1.86$, the twinning plane $(10\bar{1}2)$ forms an angle $\theta = 47^\circ$ with the basal plane (0001) . The twinning direction is $[10\bar{1}1]$.

Twinning results in extension along the basal plane and contraction along the *c*-axis. In other words, twinning in *Zn* is activated either by tension (*T*) parallel to the basal plane or by compression (*C*) along the *c*-axis of the crystal. The case of *Mg* is depicted in Figure 7.19b. *Mg* has $c/a = 1.62$ and the twinning plane $(10\bar{1}2)$ forms an angle $\theta = 43^\circ$ with the basal plane. Twinning results in contraction along the basal plane and in extension along the *c*-axis. Therefore, twining in *Mg* is activated

Table 7.3
Characteristic twinning parameters for various metals

Metal	Str/re	Ratio c/a	K_1	η_1	K_2	η_2	Shear γ
<i>Cu, Ag, Ni</i>	<i>FCC</i>		(111)	[112̄]	(11̄1)	[112]	0.707 ^a
α -Fe	BCC		(112)	[11̄1]	(112̄)	[111]	0.707
Cd	HCP	1.886	(10̄12)	[10̄1̄1]	(10̄11)	[10̄1̄1]	0.171 ^b
Zn	HCP	1.860	(10̄12)	[10̄1̄1]	(10̄11)	[10̄1̄1]	0.139
Mg	HCP	1.624	(10̄12)	[10̄1̄1]	(10̄11)	[10̄1̄1]	0.129
Ti	HCP	1.587	(1012)	[10̄1̄1]	(1̄012)	[10̄1̄1]	0.167

Reprinted from: Reed-Hill, R.E., Deformation Twinning, Gordon and Breach, 1964, New York, with permission from Taylor & Francis.

$$^a 1/\sqrt{2} = 0.707$$

$$^b [(c/a)^2 - 3]/(c/a)\sqrt{3} = 0.171$$

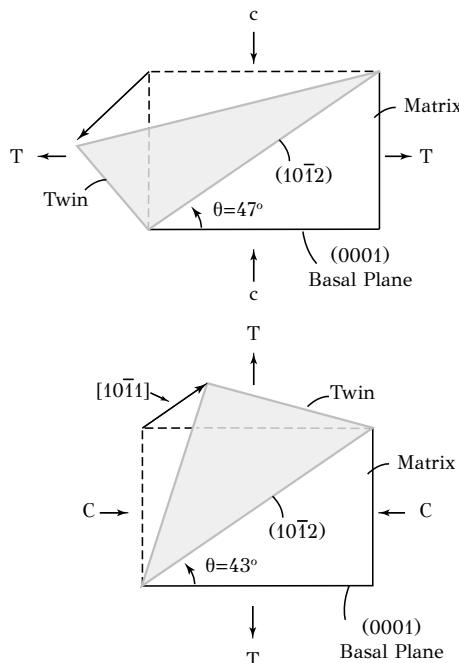


Figure 7.19: Twinning in HCP metals: (a) twinning of Zn with $c/a = 1.86$, (b) twinning in Mg with $c/a = 1.62$. T: tension, C: compression. Details in text.

either by compression (C) parallel to the basal plane or by tension (T) along the c-axis. In generalizing, twinning in *HCP* metals with $c/a > \sqrt{3}$, such as *Zn* and *Cd*, is activated by compression along the c-axis, while in *HCP* metals, with $c/a < \sqrt{3}$, such as *Mg* and *Be*, twinning is activated by tension along the c-axis. An additional role of twinning in *HCP* metals is the reorientation of the crystal so that they can be deformed by basal slip.

7.7 ANNEALING

7.7.1 ANNEALING PROCESSES

As discussed in the previous sections, strain hardening which takes place during plastic deformation is associated with the increase of dislocation density as well as the interaction between dislocations. Now consider a metal undergoing plastic deformation during cold working, as in cold rolling or deep drawing. A significant part of the work performed by the external loads is dissipated in the form of heat. The rest remains in the material as *stored energy*. This energy consists mainly of the elastic strain energy of the dislocations and point defects generated during plastic deformation. At low temperatures ($T < 0.3T_m$) this energy remains stored in the material and is termed *cold work*. The cold-worked metal is thermodynamically unstable with respect to a metal, which has not been deformed plastically. There is, therefore, a driving force to reduce the free energy. The process by which the metal reduces the free energy of cold work is called *annealing*. Annealing removes the strain hardening effects. This means that during annealing, important changes take place, which eventually lead to softening, such as reduction of dislocation density as well as rearrangement of dislocations into lower energy configurations. Annealing is a thermally activated process. Three annealing stages are observed with the increase in temperature: (a) recovery, (b) recrystallization and (c) grain growth.

During *recovery*, the dislocations rearrange into lower energy configurations. In this course, the dislocations form low-angle grain boundaries, which have a significantly lower energy than the dislocation tangles formed during plastic deformation. Another mechanism, which is activated during recovery is dislocation climb, by which dislocations are released from various obstacles and become mobile.

In heavily cold-worked metals, recovery is followed by *recrystallization*, which leads to a significant reduction of dislocation density and hardness. Recrystallization involves the formation of new dislocation-free grains and is usually activated above a certain critical temperature called the *recrystallization temperature*.

At even higher temperatures, recrystallization is followed by *grain growth*, which takes place by grain boundary migration. The driving force for grain growth is the reduction of the interface energy of the high-angle grain boundaries.

The recrystallization temperature, T_{RX} , sets an important operational limit in the deformation processing of metals. When the material is processed below the recrystallization temperature, the process is termed *cold working*, like, for example, cold rolling. When processed above T_{RX} , the process is termed *hot working*. Hot rolling of metals is a good example.

The annealing stages discussed above can be either static or dynamic. For example, during hot working, it is possible for recovery or even recrystallization to take place simultaneously with plastic deformation. In this case we talk about *dynamic recovery* and *dynamic recrystallization*.

7.7.2 STORED ENERGY

New crystal defects, mostly vacancies and dislocations, are generated during plastic deformation. The energy of these defects constitutes the 80 – 90% of the total stored energy of cold work. The magnitude of stored energy depends on the following factors:

- a) *Metal cleanliness and alloying.* Alloying elements and impurities impede dislocation motion and activate mechanisms of dislocation multiplication. Therefore, the addition of solute atoms increases the stored energy.
- b) *Stress-state of deformation.* In deformation processing involving complex stress states, such as in deep drawing, more slip systems are activated than in processing under uniaxial stress states, such as wire drawing. The activation of additional slip systems leads to higher dislocation densities.
- c) *Grain size.* For the same macroscopic deformation, a fine-grained metal stores more energy than a coarse-grained one. This is due to the interaction of dislocations with grain boundaries, leading to increased dislocation density.
- d) *Deformation temperature.* The stored energy increases when the deformation takes place at low temperatures. In this case the dislocations are not aided by thermal activation to overcome obstacles. As a result there are more dislocation interactions leading to an increase of dislocation density.

7.7.3 PROPERTY CHANGES DURING ANNEALING

The main change during annealing is softening, which is attributed to the reduction of dislocation density. However other physical properties change as well, such as the density and the electrical resistivity of the metal. These property changes are shown as a function of annealing temperature and are correlated to the reduction of stored energy in Figure 7.20. The hardness drops drastically in the recrystallization stage, where there is a large decrease in dislocation density, coming from the formation of new grains. The drop in electrical resistivity is due to the reduction of the density of point defects, vacancies and interstitials, which normally impede electron migration. During plastic deformation, the density of a metal decreases due to the generation of new vacancies and dislocations. During annealing, the population of these defects decreases and the density of the metal increases. As shown in Figure 7.20, the largest part of the stored energy is released during the recrystallization stage.

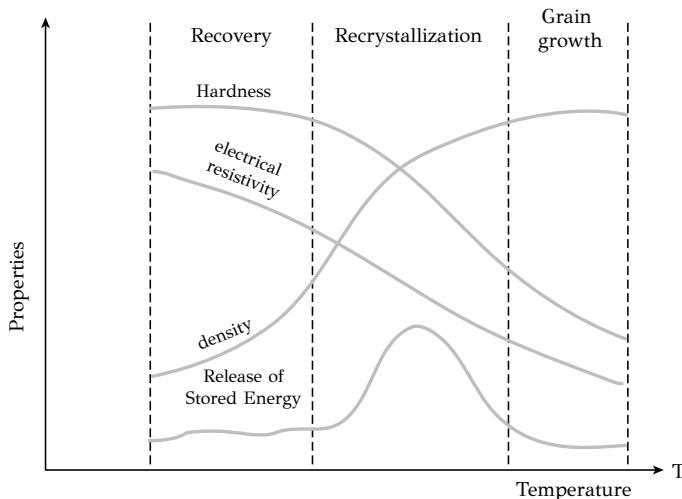


Figure 7.20: Property changes in a metal with annealing temperature.

7.7.4 RECOVERY

Recovery is the first stage of annealing. The mechanisms activated during recovery depend on annealing temperature since each mechanism exhibits a different activation energy. After cold working and due to the dislocation interactions, the microstructure is characterized by a large density of sessile dislocations in the grain interiors (Figure 7.21a). At relatively low annealing temperatures, the main mechanism activated is vacancy diffusion. The fraction of vacancies is reduced since they are absorbed at grain boundaries or at dislocation sites. At intermediate annealing temperatures the dislocations become mobile. They rearrange and form *cells*. The interior of the cells is relatively free of dislocations, which now concentrate at the cell walls (Figure 7.21b). This is accompanied by a slight decrease of dislocation density, which is attributed to a mutual annihilation of dislocations of opposite sign. At high annealing temperatures dislocation climb is activated, aided by diffusion. The dislocations are released from obstacles and with combined glide and climb transform the cell walls into low-angle grain boundaries with lower energy (Figure 7.21c and Figure 7.22). The mechanism has been termed *polygonization* and is observed in single crystals, which have been deformed by bending with the activation of a single slip system (Figure 7.23a). During annealing the dislocations rearrange into low-angle grain boundaries, separating the crystal into polygon segments (Figure 7.23b).

7.7.5 RECRYSTALLIZATION

In a heavily cold worked metal, recovery releases only a small fraction of stored energy associated with the rearrangement of dislocations into lower energy configurations. The most drastic release of stored energy is accomplished with recrystalliza-

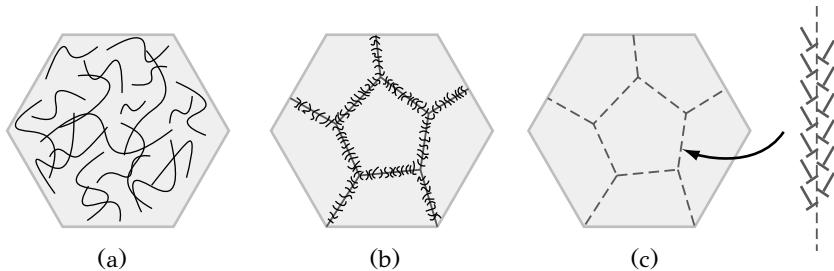


Figure 7.21: Schematic drawing showing the change in the dislocation structure during recovery: (a) high sessile dislocation density after cold working, (b) formation of cells at low annealing temperatures, (c) formation of low-angle grain boundaries at higher annealing temperatures.

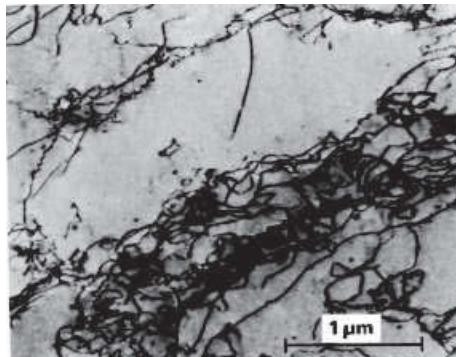


Figure 7.22: Rearrangement of dislocations and cell formation during recovery of a titanium alloy.

Reprinted from: C. T. Young, T. J. Headly and J.L. Lytton, Mat. Sci. Eng. 81, (1986), p.391, with permission from Elsevier.



Figure 7.23: Mechanism of polygonization: (a) dislocations in a crystal after plastic bending, (b) formation of low-angle tilt boundaries separating the crystal into undeformed segments.

tion, i.e., with the formation of new undeformed grains. The hardness of the material is reduced significantly and most of the strain-hardening effects are removed. Recrystallization is a structural change, which is accomplished by nucleation and growth, just like the phase transformations discussed in Chapter 6. New grains are nucleated, which grow at the expense of the old dislocation-containing grains. The process is depicted in the TEM micrographs of Figure 7.24. The microstructure immediately after recovery is shown in Figure 7.24a. The dislocations have rearranged in cell walls, however the total dislocation density is still high. The microstructure after recrystallization is shown in Figure 7.24b. New dislocation-free grains have formed. If annealing is continued, these grains grow in size (Figure 7.24c). Grain growth will be further discussed below. The driving force for recrystallization is the reduction of stored energy. Recrystallization kinetics follow the JMAK isothermal transformation kinetics, presented in Chapter 6. The recrystallization fraction is given by the Johnson-Mehl equation

$$f = 1 - \exp\left(-\frac{\pi}{3}u^3It^4\right) \quad (7.7)$$

which is valid for a constant nucleation rate of new grains. In the above equation, I is the nucleation rate and u is the growth rate of the new grains. In the general case, when the nucleation rate is not constant, recrystallization kinetics is described by the Avrami equation

$$f = 1 - \exp(-\kappa t^n) \quad (7.8)$$

The curves $f(t)$ described by Equations (7.7) and (7.8) are sigmoidal in shape and consist of three periods: (a) induction, (b) growth and (c) impingement, just like the isothermal *NGT* phase transformations discussed in Chapter 6. The recrystallization of a cold-worked metal starts at a characteristic temperature, called the *recrystallization temperature*, T_{RX} . It is defined as the temperature at which recrystallization takes a certain time (e.g., 1 hour) to be completed. Setting $f = 0.99$, the recrystallization time t_{RX} can be calculated from (7.7) as

$$t_{RX} = \left[\frac{c_1}{Iu^3} \right]^{1/4} \quad (7.9)$$

where c_1 is a constant ($c_1 = 4.4$). The recrystallization grain size d_{RX} can be estimated by setting $d_{RX} = ut_{RX}$ in (7.9)

$$d_{RX} = c_2 \left(\frac{u}{I} \right)^{1/4} \quad (7.10)$$

where c_2 is another constant ($c_2 = 1.44$). From the above equations we see that the recrystallization time (or recrystallization temperature) and the recrystallization grain size depend on nucleation and growth rate. The factors affecting I and u are the following:

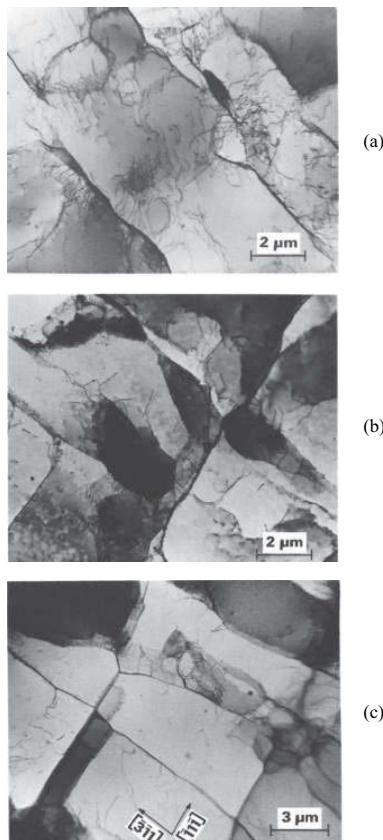


Figure 7.24: Microstructure of aluminum during annealing: (a) after recovery, (b) after recrystallization, (c) after grain growth.

Reprinted from: C. T. Young, T. J. Headly and J.L. Lytton, Mat. Sci. Eng. 81, (1986), p.391, with permission from Elsevier.

- a) *Plastic deformation.* Increase of prior plastic deformation or cold work raises I and u and lead to a decrease of the recrystallization temperature. Since I increases faster than u , more cold work leads to a finer grain size after recrystallization.
- b) *Initial grain size.* A fine-grained material stores more energy during cold working. I and u increase, leading to a reduction of the recrystallization temperature and finer grain size after recrystallization.
- c) *Metal cleanliness.* Impurities retard grain boundary movement and lead to an increase of the recrystallization temperature. For example, the recrystallization temperature of high purity copper is 120°C, while that of commercial quality copper exceeds 250°C. In addition, more energy is stored during cold working. The ratio I/u increases leading to a finer grain size d_{RX} after recrystallization.

7.7.6 GRAIN GROWTH

After the completion of recrystallization, most of the stored energy has been released. However the free energy of the metal can be further dissipated by the reduction of the grain boundary area. This is accomplished by grain growth. Grain growth is subject to two limitations regarding: (a) topology and (b) surface tension. Topological requirements are related to the shape of grains and their potential to fill space. In two dimensions, for example, circular grains do not fill space, while square or hexagonal grains do. In three dimensions there are only four solid shapes, which have the appropriate symmetry for space filling: cube, hexagonal prism, rhombic dodecahedron and tetrakaidecahedron. Surface tension requirements are related to the equilibrium between surface tensions at the vertices. Consider three faces meeting at point A in Figure 7.25. If the surface energies of the three faces are equal then the angles θ_1 , θ_2 and θ_3 are equal to 120° .

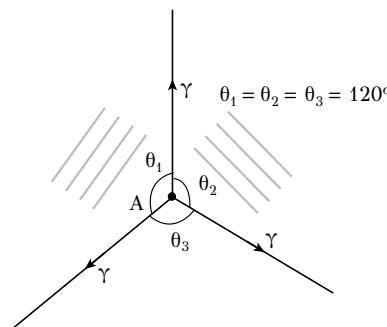


Figure 7.25: Local equilibrium of surface tensions.

However the requirement that the faces (grain boundaries) conform to both the topological and the surface tension limitation is satisfied only if the faces are curved. This is shown schematically in Figure 7.26 where in the first case (straight sides) only the topological requirement is satisfied, while in the second case (curved sides) the surface tension requirement ($\vartheta = 120^\circ$) is satisfied as well. From the solid shapes mentioned above, the one satisfying the topological requirement for space filling and at the same time exhibits the lowest surface area per unit volume is Kelvins tetrakaidecahedron (Figure 7.27). It has 14 faces (6 quadrilateral and 8 hexagonal) and 24 vertices. In order to satisfy the surface tension requirement, the four hexagonal and the two quadrilateral faces meeting at point A are curved so that the angle becomes $\vartheta = 109.5^\circ$. Metallic polycrystals are therefore composed of tetrakaidecahedral grains with curved faces.

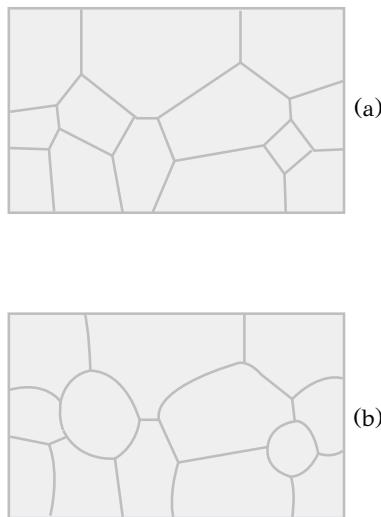


Figure 7.26: (a) topological requirement met by polygons with straight sides $\theta \neq 120^\circ$, (b) surface tension requirement met by polygons with curved sides ($\theta = 120^\circ$).

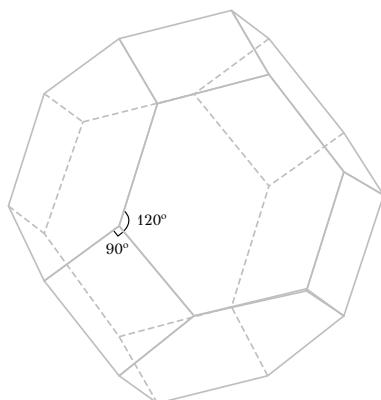


Figure 7.27: Kelvin's tetrakaidecahedron.

Curved grain boundaries experience a force acting towards the center of curvature. In a polycrystal where the grain boundaries exhibit different curvatures, the forces have different magnitudes and are not in equilibrium. The structure is, therefore, unstable. At high temperature, during annealing, these forces will cause the grain boundaries to migrate towards the center of curvature. This is shown schematically for a two-dimensional grain structure in Figure 7.28. In a grain surrounded by $n = 6$ boundaries, the surface tension requirement ($\theta = 120^\circ$) is associated with straight edges and zero force (Figure 7.28a). When the grain is surrounded by $n < 6$ boundaries (Figure 7.28b) the boundaries migrate towards the center of the grain and the grain shrinks. In the opposite case, when $n > 6$, the grain grows. The net result of grain boundary migration is the reduction in the number of grains and an increase of the mean grain size. This is associated with a reduction in interfacial energy. This phenomenon has been termed grain growth and is considered as the third stage of annealing.

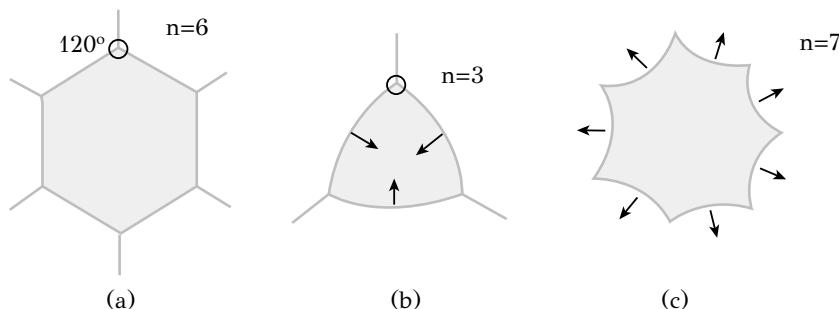


Figure 7.28: Grain boundary migration during grain growth: (a) $n = 6$ no forces acting on the boundaries, (b) $n < 6$ the grain shrinks, (c) $n > 6$ the grain grows.

Reprinted from: D.A. Porter, K.E. Easterling, M.Y. Sherif (2009). Phase Transformations in Metals and Alloys, with permission from CRC press.

The pressure difference between two grains separated by a boundary with curvature r is

$$\Delta p = 2\gamma/r$$

This pressure difference Δp is associated with a difference in free energy ΔG

$$\Delta G = 2\gamma V_m/r$$

where V_m is the atomic volume. It could be considered that the pressure difference Δp pushes the boundary towards the grain with a larger free energy. As indicated in Figure 7.29, the migration of the grain boundary is the result of atomic jumps from grain 2 to grain 1. The growth of grain 1 takes place under interface control, and can be described by the theory of interface-controlled growth with no change in composition, discussed in Chapter 6.

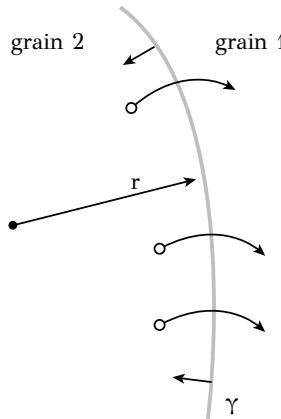


Figure 7.29: Atomic jumps from grain 2 to grain 1 causes grain boundary migration and grain growth.

The velocity u of grain boundary migration is

$$u = M\Delta G$$

where M is the interfacial mobility of grain boundaries, which is temperature-dependent through the relation

$$M = \frac{v\alpha A}{kT} \exp\left(-\frac{\Delta G^*}{kT}\right)$$

where v is the vibration frequency of the atoms (Debye frequency), A is the accommodation factor, which expresses the fraction of atoms which find a position in the new grain, α is the lattice parameter and ΔG^* is the activation energy. During grain growth the mean grain diameter \bar{D} increases. If we set $u = d\bar{D}/dt$, the above equations give

$$u = \frac{d\bar{D}}{dt} = \frac{4M\gamma V_m}{\bar{D}}$$

As indicated from the above relation, the velocity of grain boundary migration is inversely proportional to \bar{D} . Integrating the above relation, we get

$$\bar{D}^2 - \bar{D}_0^2 = kt \quad (7.11)$$

where $k = 4M\gamma V_m$. The mean grain size is a parabolic function of time.

7.7.7 GRAIN BOUNDARY PINNING AND GRAIN REFINEMENT

Grain boundary mobility is reduced by the drag exerted by solute atoms and precipitates. Regarding the effect of solute atoms, microsegregation results in the increase

of the local composition c_{gb} at the boundary region relative to the grain interior c_o , described by the relation

$$c_{gb} = c_o \exp\left(-\frac{Q_{gb}}{RT}\right)$$

where Q_{gb} is an interaction energy between the solute atoms and the boundary. Grain boundary migration is then possible only by the diffusion of the segregated atoms. In this way the segregated atoms exert a drag, termed *solute drag*, on the boundary reducing the grain-boundary mobility and grain growth rate. More important than the solute drag, discussed above, is the drag exerted by second-phase particles and intermetallic compounds, such as the AlN precipitates in steels deoxidized with Al , or the carbides and carbonitrides in *HSLA* steels. This has been termed *Zener pinning*. In this case a moving grain boundary is pinned at these particles, which exert a resisting force at the boundary and impede its motion (Figure 7.30). On contact with the particle the boundary is curved in order to remain perpendicular to the particle surface. The contact length is $2\pi r_p \cos\theta$, where r_p is the particle radius, while the component of the surface tension is $\gamma \sin\theta$. So the pinning force exerted by the particle at the boundary is

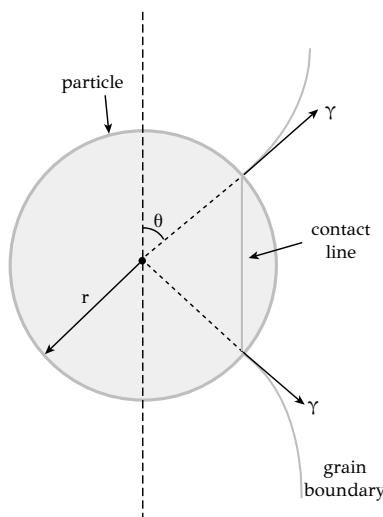


Figure 7.30: Grain boundary pinning by particles.

$$F = (2\pi r \cos \theta)(\gamma \sin \theta) = \pi r \gamma \sin 2\theta$$

The force is maximized at $\theta = 45^\circ$ so

$$F_{\max} = \pi r_p \gamma$$

If the volume fraction of particles is f_p then the number of particles per boundary area N_s is

$$N_s = \frac{3f_p}{2\pi r_p^2}$$

The resistive force per boundary area (stress) is

$$\sigma = N_s F_{max}$$

which becomes

$$\sigma = \frac{3f_p \gamma}{2r_p} \quad (7.12)$$

With the help of Equation (7.12) several issues are highlighted:

- (a) the pinning increases with the surface energy of the boundary,
- (b) the pinning increases with the volume fraction of the particles,
- (c) the pinning decreases with the particle size.

The pinning expressed by Equation (7.12) acts opposite to the driving force $2\gamma/\bar{D}$ for grain boundary migration, where \bar{D} is the mean grain diameter. At equilibrium of forces we get

$$\sigma = \frac{3f_p \gamma}{2r_p} = \frac{2\gamma}{\bar{D}}$$

and grain growth stops. The maximum grain size \bar{D}_{max} can then be obtained by the above equation as

$$\bar{D}_{max} = \frac{4r_p}{3f_p} \quad (7.13)$$

which holds for grain growth in the presence of a particle dispersion with volume fraction f_p and mean particle size r_p . Grain refinement is a significant strengthening mechanism of metallic alloys. In certain cases, alloying and heat treatment aim towards the formation of a grain refining dispersion. This dispersion forms either by precipitation or by mechanical alloying. The effectiveness of the dispersion depends on the f_p/r_p ratio as indicated in Equation (7.13). A high volume fraction of very small particles is required to maintain a fine grain size. At high temperatures the effectiveness of the grain refining dispersion is reduced due to the coarsening or even dissolution of the particles. So in the example discussed above, the AlN particles dissolve above $1000^\circ C$ and grain boundary pinning is not possible above this temperature. As a consequence rapid grain growth takes place.

7.8 TEXTURE IN POLYCRYSTALLINE METALS

7.8.1 DEVELOPMENT OF ANISOTROPY

Single crystals are inherently anisotropic. Physical and mechanical properties depend on crystallographic orientation. It would have been expected that due to the random orientation of grains, polycrystals are isotropic. However deformation processing causes strong anisotropy in polycrystalline metals due to the development of preferred orientations. In these cases we say that metals develop *deformation texture*. During annealing, recrystallization takes place. The new grains exhibit anisotropy as well, called *recrystallization texture*, which could be stronger than the deformation texture. In both cases the material becomes anisotropic and the mechanical and physical properties, such as strength, magnetic and thermal properties depend on the direction of measurement. On the other side, texture development has found some significant technological applications, for example in magnetic alloys and in texture strengthening.

7.8.2 DEFORMATION TEXTURE

As discussed in Section 7.3, the tensile deformation of a single crystal causes rotation of the crystal so that the slip direction approaches the tensile axis. Similar rotations take place in every grain of a polycrystalline metal. These rotations lead to the development of preferred orientations, i.e., texture. The type of texture depends on crystal structure and the stress field developed during deformation processing. Two types of texture will be discussed: (a) fibrous texture, which develops in uniaxial processing, such as wire drawing or extrusion and (b) rolling texture, which develops in plane deformation processing, such as cold rolling. The most important deformation textures are depicted in Table 7.4.

In fibrous texture the grains elongate in the direction of deformation and a low index crystallographic direction, e.g., $\langle 110 \rangle$ in *BCC*, becomes parallel to the deformation axis. In *FCC* metals, texture development is related to the formation of stacking faults. In metals with a high stacking fault energy (*SFE*), such as aluminum, the $\langle 111 \rangle$ texture is observed, while in metals with a low *SFE* the $\langle 100 \rangle$ texture dominates. In *HCP* metals the basal plane rotates as to become parallel to the deformation axis, which coincides with the $\langle 10\bar{1}0 \rangle$ direction. Rolling texture involves both a crystallographic direction and a crystallographic plane being parallel to the rolling plane. This plane is characterized by small Miller indices. In *FCC* metals rolling texture is characterized by the $\{110\}$ planes being parallel to the rolling plane and the $\langle 112 \rangle$ directions being parallel to the rolling direction. In *BCC* metals the cube faces $\{001\}$ are oriented parallel to the rolling plane. In *HCP* metals the basal plane $\{0001\}$ is oriented parallel to the rolling plane and the close packed directions $\langle 11\bar{2}0 \rangle$ become parallel to the rolling direction.

Table 7.4
Examples of deformation textures

Structure	Processing	Texture
FCC	drawing/extrusion	$\langle 100 \rangle$ parallel to the wire axis $\langle 100 \rangle$
BCC	drawing/extrusion	$\langle 110 \rangle$ parallel to the wire axis
HCP	drawing/extrusion	$\langle 10\bar{1}0 \rangle$ parallel to the wire axis
FCC	rolling	$\{110\}$ parallel to the rolling plane $\langle 112 \rangle$ parallel to the rolling direction
BCC	rolling	$\{001\}$ parallel to the rolling plane $\langle 110 \rangle$ parallel to the rolling direction
HCP	rolling	$\{0001\}$ parallel to the rolling plane $\langle 11\bar{2}0 \rangle$ parallel to the rolling direction

7.8.3 RECRYSTALLIZATION TEXTURE

As discussed above, the new grains formed during recrystallization do not grow with the same rate since the grain boundaries do not exhibit the same mobility. This leads to the unilateral growth of certain orientations and the material develops recrystallization texture. This type of texture is influenced by the deformation texture existing in the material as a result of deformation processing. In several *FCC* metals, like copper or aluminum, rolling and annealing leads to the development of the so-called cube texture. This is a $\{100\}\langle 001 \rangle$ texture, which is quite strong, causing the material to behave as a single crystal. Grain boundary pinning has been discussed above as a method to develop fine grain size by limiting the grain boundary mobility. Alloys with grain refining dispersions do not develop recrystallization textures but maintain the deformation texture developed during cold deformation processing.

7.8.4 TEXTURE EFFECTS

The most important effect of texture is the anisotropy of the material in terms of physical or mechanical properties. This can have both advantages and disadvantages. A basic disadvantage is the degradation of formability. A good example of this is the earing developed at the periphery of a cylinder during deep drawing (Figure 7.31). Earing is caused by the anisotropy of mechanical properties. Due to anisotropy metal flow depends on direction and the cup height is high in a direction that the sheet exhibits higher drawability. On the other side a metal can be made stronger in one useful direction. An example is the *HCP* metals, like titanium and magnesium with a $\langle 0001 \rangle\langle 11\bar{2}0 \rangle$ rolling texture. These metals have the basal plane and the close-packed directions on the rolling plane. As a consequence all strains during tensile deforma-

tion will take place on the rolling plane and the material will resist thinning. This enhanced resistance, in the thickness direction, provides enhanced strength during biaxial tensile loading and, therefore, *HCP* metals with a strong texture can find applications in pressure vessels.

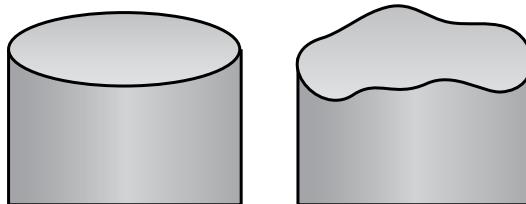


Figure 7.31: Earing during deep drawing caused by the texture of the material.

7.9 SYNOPSIS

1. The basic mechanisms of plastic deformation in metals are slip, twinning and martensitic transformation. However the principal mechanism is slip, which proceeds with dislocation glide.
2. Slip takes place in specific slip systems characterized by the slip plane and the slip direction.
3. Slip in a specific slip system is activated when the shear stress reaches a critical value, the critical resolved shear stress (*CRSS*). The orientation of a slip system relative to the applied tensile stress is expressed by the Schmid factor.
4. In *FCC* metals slip takes place in close-packed planes and close-packed directions, the major slip system being $\{111\} < 110 >$.
5. In *BCC* metals, where there is no close-packed plane, the most common slip system is $\{110\} < 111 >$, however any slip system containing the close-packed direction $< 111 >$ is a potential slip system.
6. In *HCP* metals the slip system depends on the c/a ratio. For *Mg*, *Cd* and *Zn* with c/a ratios higher than the ideal value (1.63), basal slip dominates and the slip system is $\{0001\} < 11\bar{2}0 >$. For *Ti*, *Be* and *Zr* with c/a ratios lower than the ideal value, prismatic slip is activated and the slip system is $\{10\bar{1}0\} < 11\bar{2}0 >$.
7. During simple slip the crystal rotates relative to the applied tensile or compressive stress. In tension, the slip direction moves towards the tensile axis. The opposite takes place in compression.

8. Crystal rotation causes activation of other slip systems and leads eventually to multiple slip. During multiple slip, dislocations in intersecting slip planes interact. These interactions impede dislocation glide and cause strain hardening.
9. The plastic deformation of polycrystalline aggregates is more complicated than the deformation of single crystals. The compatibility criterion states that the grains cannot deform independently but the deformation of one grain should be compatible with the deformation of neighboring grains.
10. According to Von Mises, homogeneous plastic deformation of polycrystals requires the activation of at least five (5) independent slip systems, in order for the compatibility criterion to be satisfied. *FCC* and *BCC* metals possess more than five independent slip systems. However *HCP* metals possess only two independent slip systems and other alternative mechanisms of plastic deformation are activated, such as twinning.
11. As a metal is plastically deformed, its resistance to further plastic deformation increases. This is termed strain hardening.
12. Plastic deformation raises the dislocation density. Dislocation multiplication is accomplished by the activation of dislocation sources. One of the principal sources is the Frank-Reed source.
13. *FCC* single crystals exhibit a three-stage deformation behavior. Stage I-easy glide, associated with the activation of a single slip system, stage II linear hardening, associated with multiple slip and stage III dynamic recovery, associated with cross slip. *BCC* crystals exhibit mainly stage III, while *HCP* crystals exhibit stages I and II.
14. Among the basic mechanisms of strain hardening, which impede dislocation glide are: (a) dislocation intersections forming dislocation tangles with limited dislocation mobility, (b) dissociation of dislocations, which hinders cross slip of screw dislocations, (c) Lomer-Cottrell locks, which form sessile dislocations and (d) forest hardening, which is the interaction of dislocations gliding on the slip plane with dislocations intersecting the slip plane.
15. Twinning in metals takes place by a cooperative movement of atoms resulting in homogeneous shear of the lattice. The twinning shear takes place on the twinning plane in the twinning direction. The twinned crystal is the mirror image of the parent crystal.
16. In *FCC* and *BCC* metals the twinning shear is large, however the contribution of twinning to the total deformation is rather low. The reason is that the volume fraction of twins is low. There is a larger contribution of twinning in the *HCP* metals

- including the reorientation of the crystal so that they can be deformed by basal slip.
17. The stored energy of cold work consists mainly of the elastic strain energy of the dislocations and point defects generated during plastic deformation. The magnitude of stored energy depends on (a) metal cleanliness and alloying, (b) the stress-state of deformation, (c) grain size and (d) deformation temperature.
 18. Annealing removes strain hardening and releases the stored energy of cold work. Physical and mechanical properties change during annealing, the most important are softening, decrease of electrical resistivity and increase of density.
 19. Annealing comprises of three stages: recovery, recrystallization and grain growth.
 20. During recovery, the dislocations rearrange into lower energy configurations. In this course, the dislocations form low-angle grain boundaries, which have a significantly lower energy than the dislocation tangles formed during plastic deformation. Another mechanism, which is activated during recovery, is dislocation climb, by which dislocations are released from various obstacles and become mobile.
 21. In heavily cold-worked metals, recovery is followed by recrystallization, which leads to a significant reduction of dislocation density and hardness. Recrystallization involves the formation of new dislocation-free grains and is usually activated above a certain critical temperature called the recrystallization temperature.
 22. At even higher temperatures, recrystallization is followed by grain growth, which takes place by grain boundary migration. The driving force for grain growth is the reduction of the interface energy of the high-angle grain boundaries.
 23. Grain boundary mobility is reduced by the drag exerted by solute atoms and precipitates leading to grain refinement.
 24. Deformation processing causes strong anisotropy in polycrystalline metals due to the development of preferred orientations. In these cases metals develop deformation texture. During annealing, recrystallization takes place. The new grains exhibit anisotropy as well, called recrystallization texture. In both cases the material becomes anisotropic and the mechanical and physical properties, such as strength, magnetic and thermal properties depend on the direction of measurement.

7.10 REVIEW QUESTIONS

1. The *CRSS* of a Titanium single crystal is about 110 MPa for basal slip while it is only 49 MPa for prismatic slip. Explain this difference based on crystallographic arguments.
2. Discuss the differences between single crystal and polycrystal behavior regarding plastic deformation by dislocation glide.
3. Two sheets of a $\text{Cu} - 10\text{Zn}$ (wt%) alloy of identical thickness and grain size are subjected to cold rolling. The first sheet undergoes a 15% reduction while the second sheet undergoes a 10% reduction. The two sheets are then annealed. Compare the recrystallization temperatures and the recrystallized grain size of the two sheets.
4. Two sheets (*A*) and (*B*) of a $\text{Al} - 3\text{Mn}$ (wt%) alloy have grain sizes $15\mu\text{m}$ (*A*) and $30\mu\text{m}$ (*B*). The sheets are subjected to 20% reduction in a cold rolling operation and then are annealed.
 - a. Which sheet will recrystallize faster?
 - b. Which sheet will exhibit a larger recrystallized grain size?
5. The control of recrystallization is important in the thermomechanical controlled processing (*TMCP*) of *HSLA* steels. A significant step in the process chain involves rolling just below the T_{NRX} temperature, which is the non-recrystallization temperature. Discuss the phenomena that could take place during rolling above and below T_{NRX} . Include dislocation activity in your discussion.
6. During easy glide of *FCC* crystals, only one slip system is operating and there are no dislocation interactions. As a result the strain hardening rate is low. Explain what causes this strain hardening.
7. Assuming that during plastic deformation the dislocation density increases with a rate $\dot{\rho} = d\rho/dt$ due to the operation of Frank-Read sources, derive an expression for $\dot{\rho}$ as a function of applied shear stress and deformation rate $\dot{\gamma}$.

BIBLIOGRAPHY–FURTHER READING

1. Meyers, M. A., & Chawla, K. K. (1984). *Mechanical Metallurgy. Principles and Applications*, Prentice Hall, Englewood Cliffs, New Jersey.
2. Hayden, H. W., Moffatt, W. G., & Wulff, J. (1965). *The Structure and Properties of Materials, Vol. 3: Mechanical Behavior*, John Wiley & Sons, USA.
3. Dieter, G.E. (1986) *Mechanical Metallurgy*, McGraw-Hill, 3 rd edition.
4. Cottrell, A. H. (1963). *Dislocations and Plastic Flow in Crystals*, Oxford University Press.
5. Honeycombe, R.W.K. (1984). *The Plastic Deformation of Metals*, Edward Arnold, 2 nd edition.
6. Cahn, R. W. (1983). *Recovery and Recrystallization, in Physical Metallurgy*, R. W. Cahn, P. Haasen editors, Elsevier Science Publishers.
7. Hertzberg, R.W. (1989). *Deformation and Fracture Mechanics of Engineering Materials*, John Wiley & Sons, USA.

8 Strengthening mechanisms

8.1 INTRODUCTION

Pure metals are soft materials. For most engineering applications they cannot provide the required strength. This is the reason alloys were invented. The strength of metals can be increased by alloying and by certain thermal, mechanical or thermomechanical processing. In the previous chapter we have seen that the plastic deformation of metals proceeds by the glide of dislocations, a process we called slip. Strengthening can be achieved by creating obstacles to dislocation glide. It is these obstacles that are created with the alloying and suitable processing discussed above. The strengthening mechanisms, which is the subject of the present chapter, are simply mechanisms of interaction between dislocations and various obstacles. The major mechanisms are: lattice resistance, strain hardening, solid solution strengthening, grain boundary strengthening and precipitation strengthening. In lattice resistance, the main obstacle to dislocation glide is the crystal lattice, the atomic bonds in particular. In strain hardening the main obstacle impeding dislocation glide is other dislocations. The interaction between dislocations forms sessile dislocations, which cannot contribute to plastic deformation. Strain hardening has been discussed in detail in the previous chapter and will not be considered further. In solid solution strengthening the obstacles are solute atoms, either substitutional or interstitial, the strain field of which interacts with dislocations. In grain boundary strengthening, the obstacles are high-angle grain boundaries, but also sub-boundaries and interfaces as the interfaces between ferrite and cementite in pearlite or the interfaces between martensite laths. Finally, in precipitation strengthening, the obstacles are precipitates, second-phase particles or intermetallic compounds formed during thermal processing. In most alloys the yield strength is the result of a superposition of strengthening mechanisms. The effectiveness of each mechanism is characterized by the specific obstacle strength, which is related to the stress required to overcome the obstacles at $T = 0\text{K}$. At higher temperatures, dislocations are assisted by thermal activation and they can overcome obstacles at lower stresses. This explains the temperature and strain rate dependence of the flow stress. We will begin our discussion with the issue of thermal activation of dislocation glide.

8.2 SLIP AS A THERMALLY ACTIVATED PROCESS

As discussed in Chapter 3, plastic deformation is produced by the glide of dislocations and the plastic strain rate $\dot{\gamma}$ is related to the mean dislocation velocity \bar{v} on the slip plane and the mobile dislocation density ρ_m by the relation

$$\dot{\gamma} = \rho_m b \bar{v} \quad (8.1)$$

The strength of metals can be increased by the introduction of barriers to dislocation glide. The strengthening mechanisms, to be discussed in this chapter, are related to the mechanisms by which various obstacles impede dislocation glide. Such obstacles are the crystal lattice, other dislocations, solute atoms, precipitates and grain boundaries. Each obstacle displays a specific resistance to dislocation glide, termed the *specific obstacle strength*. Either a higher stress or thermal activation are required for dislocations to overcome the various obstacles they encounter during glide. Thus the strain rate becomes stress and temperature dependent, as expressed by the following relation

$$\dot{\gamma} = \dot{\gamma}(\tau, T) \quad (8.2)$$

The specific obstacle strength influences the form of Equation (8.2), which will be derived by considering slip as a thermally-activated process. Since the process takes place on the slip plane, the yield strength in shear τ will be used. This stress is related to the tensile yield strength through the Taylor factor. The yield strength expresses the resistance of the material to plastic deformation and consists of two contributions, the thermal component τ^* and the athermal component τ_μ .

$$\tau = \tau^* + \tau_\mu$$

The athermal component is caused by long-range obstacles, such as the stress fields of other dislocations, which operate at large distances around the dislocation core. These obstacles cannot be surmounted by thermal activation, making the athermal component τ_μ nearly independent of temperature apart from a small variation of the shear modulus μ . On the other hand, the thermal component τ^* is caused by short-range obstacles, such as solute atoms and precipitates. As these obstacles have a local effect on dislocations, they can be surmounted by thermal activation. Thus the temperature dependence of the flow stress is attributed to the thermal component τ^* . At $T = 0K$, where there is no thermal activation, the thermal component takes its maximum value τ_o . At higher temperatures, dislocation glide can proceed at lower stress since thermal activation assists the dislocations to overcome obstacles, as indicated in Figure 8.1.

The higher the temperature the lower the thermal component of the flow stress. At a critical temperature T_c the thermal component vanishes and the obstacle does not provide any mechanical resistance to dislocation glide.

The specific obstacle strength defines the dependence of the flow stress on temperature and strain rate. If for example an obstacle exhibits a low specific strength, relative to the available thermal energy kT , then the vibrations of the crystal lattice assist the dislocations to overcome the obstacle at lower applied stresses in relation to those required at $T = 0K$. This means that increasing temperature or decreasing strain rate leads to a decrease of the flow stress. The interaction of a dislocation with

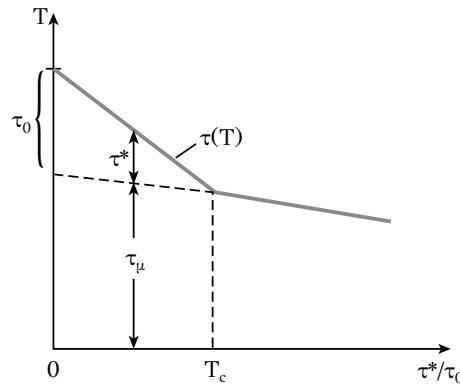


Figure 8.1: The temperature dependence of the flow stress in a metal.

obstacles can be studied by considering a dislocation gliding on the slip plane (Figure 8.2) under an applied shear stress τ^* . Two short-range obstacles are on the slip plane with specific strength $K(x)$. The maximum specific obstacle strength is K_{max} . If the obstacles are at a distance l apart, then the applied forward force on the dislocation line is τ^*bl . If $\tau^*bl > K_{max}$ then the dislocation overcomes the obstacles and continues its glide on the slip plane. If $\tau^*bl < K_{max}$ the dislocation stops at position x_1 (Figure 8.3a). In this case thermal activation is required in order for the dislocation to overcome the obstacles and move to position x_2 . The total obstacle energy is

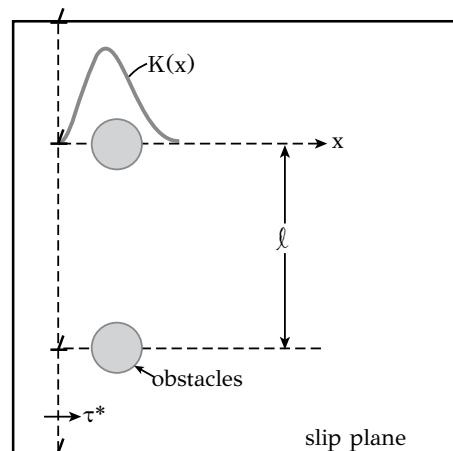


Figure 8.2: Obstacles with a specific resistance $K(x)$ impeding dislocation glide on the slip plane.

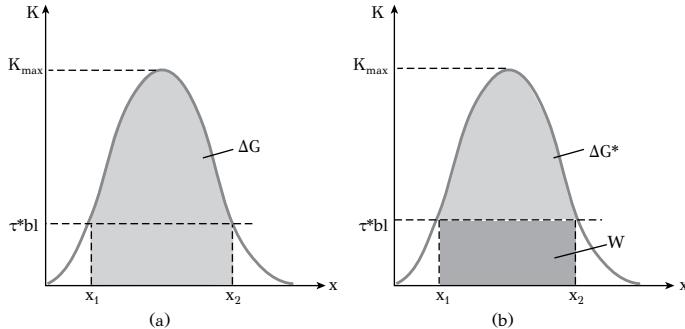


Figure 8.3: Slip as a thermally-activated process: (a) the total obstacle energy ΔG and (b) the mechanical component W and thermal component ΔG^* .

$$\Delta G = \int_{x_1}^{x_2} K(x) dx$$

As indicated in Figure 8.3b this energy consists of a mechanical contribution W related to the applied stress and a thermal contribution ΔG^* due to thermal activation. The mechanical contribution is equal to the work W of the applied stress produced by the movement of the dislocation from position x_1 to position x_2

$$W = \tau^{*bl} (x_2 - x_1) = \tau^{*} V^{*}$$

where

$$V^{*} = bl (x_2 - x_1)$$

is the *activation volume* of the process. The activation energy is then

$$\Delta G^* = \Delta G - W = \Delta G - \tau^{*} V^{*}$$

The activation energy is a function of stress, $\Delta G^* = \Delta G^*(\tau)$. The probability for a dislocation to overcome the energy barrier ΔG^* at temperature T is given by the Boltzmann factor $\exp(-\Delta G^*/kT)$. If the dislocation vibrates with a frequency v , then it overcomes $v \exp(-\Delta G^*/kT)$ obstacles per unit time and its velocity is then

$$\bar{v} = d v \exp(-\Delta G^*/kT)$$

where d is the displacement of the dislocation for each obstacle it overcomes. Using the Relation (8.1) the strain rate becomes

$$\dot{\gamma} = \rho_m b d v \exp\left[-\frac{\Delta G^*(\tau)}{kT}\right]$$

Setting $\dot{\gamma}_o = \rho_m b d v$, the above equation becomes

$$\dot{\gamma} = \dot{\gamma}_0 \exp \left[-\frac{\Delta G^*(\tau)}{kT} \right] \quad (8.3)$$

The above equation is in the form of Equation (8.2) $\dot{\gamma} = \dot{\gamma}(\tau, T)$, relating the strain rate with the applied stress and temperature. The influence of stress on strain rate comes from the stress dependence of the activation energy $\Delta G^*(\tau)$. This is shown in Figure 8.4. The higher the applied stress τ^* , the lower the thermal activation required to overcome the obstacles. At $\tau^* = \tau_0$ the required activation energy vanishes.

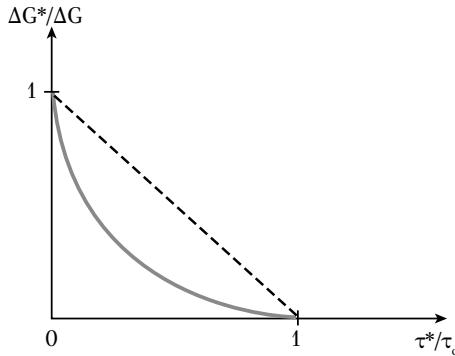


Figure 8.4: The dependence of activation energy ΔG^* on applied stress τ^* .

The temperature dependence of the flow stress can be derived if an expression for $\Delta G^*(\tau)$ is given. Under the assumption of a linear dependence (dotted line in Figure 8.4)

$$\frac{\Delta G^*}{\Delta G} = 1 - \frac{\tau^*}{\tau_0}$$

Equation (8.3) becomes

$$\frac{\tau^*}{\tau_0} = \frac{kT}{\Delta G} \ln \left(\frac{\dot{\gamma}}{\dot{\gamma}_0} \right) + 1 \quad (8.4)$$

which is actually the relation $\tau = \tau(\dot{\gamma}, T)$. Considering the temperature T_c as the temperature where $\tau^* = 0$, i.e., where the thermal component vanishes, we get

$$T_c = - \frac{\Delta G}{k \ln(\dot{\gamma}/\dot{\gamma}_0)}$$

and the Relation (8.4) becomes

$$\frac{\tau^*}{\tau_0} = 1 - \frac{T}{T_c} \quad (8.5)$$

providing the temperature dependence of the thermal component of the flow stress. The total flow stress is the sum of the thermal and the athermal components, as

discussed above. The temperature dependence has already been discussed in Figure 8.1. The temperature dependence below T_c depends on τ , which in turn depends on K_{max} , the specific obstacle strength. In the general case the relation $\Delta G^*(\tau)$ is not linear but has the form

$$\frac{\Delta G^*}{\Delta G} = \left[1 - \left(\frac{\tau^*}{\tau_0} \right)^p \right]^q \quad (8.6)$$

with $0 < p < 1$ and $1 < q < 2$ (full line in Figure 8.4). The obstacle energy ΔG ranges between $0.05\mu b^3$ for weak obstacles, such as substitutional solute atoms, and $2\mu b^3$ for strong obstacles, such as fine precipitates. The specific obstacles related to strengthening mechanisms will be discussed in the next section. The temperature dependence of the flow stress, such as Equation (8.5) is related to the thermal activation of dislocation glide and is valid for temperatures up to about $T = 0.3T_M$, where T_M is the melting point. At higher temperatures, other deformation mechanisms are activated with the contribution of diffusion and the Relation (8.5) is not valid. In this case atomic diffusion contributes to plastic deformation with creep mechanisms either by assisting dislocations to overcome obstacles by climb (dislocation creep) or by direct diffusional flow. Creep mechanisms will be discussed in Chapter 9 of this book.

8.3 OVERVIEW OF STRENGTHENING MECHANISMS

Strengthening in metals can be achieved by impeding the dislocation glide. The strengthening mechanisms are, therefore, related to the interaction between the dislocations and various obstacles present in the microstructure of the material. Among the obstacles to dislocation glide are the atomic bonds of the crystal lattice, other dislocations, interstitial or substitutional solute atoms, grain boundaries and precipitates or second-phase particles. The basic strengthening mechanisms are (see also Figure 8.5):

- a)** *Lattice resistance*, caused by the resistance exhibited by the lattice to dislocation glide. This mechanism is always operational and independent of the operation of any other strengthening mechanisms. In single crystals, the lattice resistance leads to the dependence of flow stress on the atomic density of the slip plane and slip direction.
- b)** *Strain hardening*, caused by the increase of dislocation density and the interaction between dislocations during plastic deformation. Strain hardening has been discussed in Chapter 7 and will not be considered further.
- c)** *Solid solution strengthening*, caused by the interaction between dislocations and solute atoms. Only the alloying elements in solid solution contribute to this type of strengthening, not the alloying elements present in the precipitates. The level of strengthening depends on the composition of the solid solution.

- d) *Grain boundary strengthening*, caused by the resistance of the grain boundaries to dislocation glide. Leads to the dependence of the flow stress on the grain size.
- e) *Precipitation strengthening*, caused by the resistance of precipitates or dispersed phases to dislocation glide. Leads to the dependence of flow stress on volume fraction and mean size of precipitates.

It is very common in alloys that more than one strengthening mechanism operate simultaneously. Actually it is a major aim in modern alloy design that alloying and processing should lead to a yield strength made up with contributions from more than one strengthening mechanism. An example is a $Fe - 4Mo - 0.2C$ steel. The heat treatment consists of (a) austenitization, during which all Mo and C are dissolved in austenite, (b) quenching to form martensite and (c) tempering of martensite, during which precipitation of Mo_2C takes place (secondary hardening). The strength of the alloy after the quench consists of three contributions: (i) the lattice resistance of the BCC lattice, (ii) solid solution strengthening from the $Fe - Mo - C$ BCC supersaturated solid solution, which has a strong contribution from the carbon atoms in interstitial sites and a weak contribution from the Mo atoms in substitutional sites, (iii) strain hardening due to the increased density of dislocations formed during the martensitic transformation. During tempering a recovery of the dislocation structure takes place and the strain hardening contribution decreases. In addition Mo and C go out of solution to form Mo_2C . This leads to a decrease of solid solution strengthening. However a new mechanism is activated, precipitation strengthening, associated with the Mo_2C carbides. It is clear that the strengthening mechanisms at play change with heat treatment. It should also be noted that subsequent processing, e.g., welding, could deactivate certain strengthening mechanisms. A good example are the heat treatable $Al - Cu$ alloys, which are strengthened by a fine dispersion of Al_2Cu particles formed during the aging treatment. When these alloys are welded by fusion welding, the heat affected zone adjacent to the weld metal becomes very soft. This is because during the weld thermal cycle, the Al_2Cu precipitates dissolve and the precipitation strengthening mechanism is deactivated.

8.4 LATTICE RESISTANCE

The term lattice resistance refers to the resistance provided by the crystal lattice to dislocation glide. Lattice resistance is related to atomic bonding and the periodicity of the lattice. Dislocation glide is not a continuous process but rather takes place in a step-wise fashion on the slip plane, since atomic positions are discrete. As shown in Figure 8.6a the dislocation is at the equilibrium position A , which is associated with an energy minimum. The dislocation can move to the next equilibrium position B , provided it breaks the bond between atoms 2 and 3.

This process is thermally activated. The activation energy and lattice resistance are, therefore, periodic functions of position on the slip plane. Based on lattice dis-

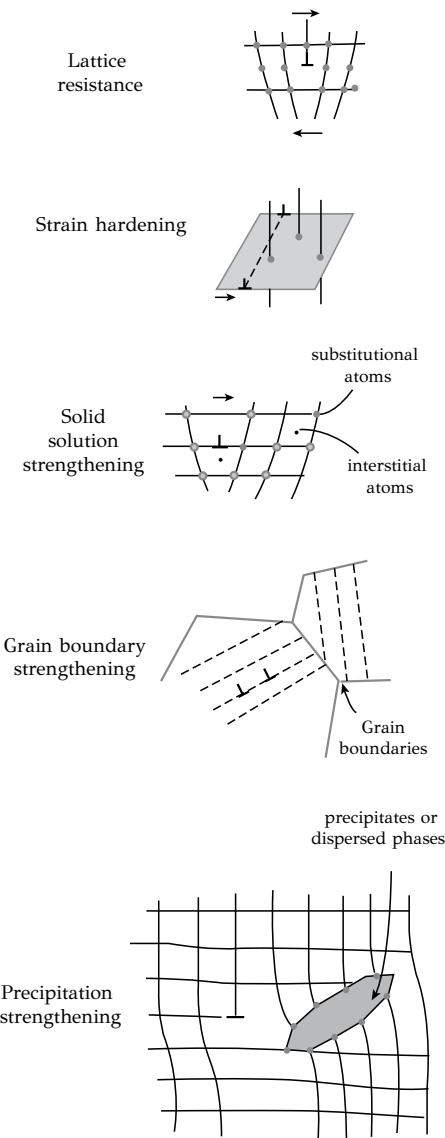


Figure 8.5: A schematic showing the most important strengthening mechanisms in metals.

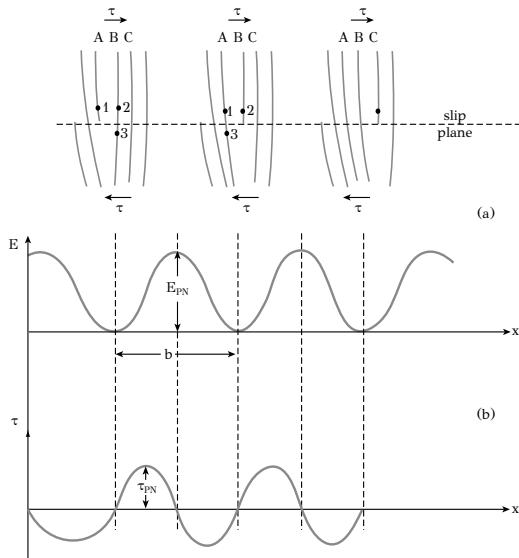


Figure 8.6: (a) The movement of an edge dislocation in three consecutive equilibrium positions on the slip plane, (b) the variation of energy and stress during dislocation glide.

creteness, Peierls¹ and Nabarro² considered the energy (potential) as a sinusoidal function of position (Figure 8.6b) and calculated the maximum energy and the maximum stress needed to overcome the lattice resistance. The Peierls-Nabarro energy is

$$E_{PN} = \frac{\mu b^2}{\pi(1-\nu)} \exp\left(\frac{-2\pi w}{b}\right)$$

and the Peierls-Nabarro stress ($P - N$ stress) is

$$\tau_{PN} = \frac{2\mu}{1-\nu} \exp\left(\frac{-2\pi w}{b}\right) \quad (8.7)$$

where μ is the shear modulus, ν is Poisson's ratio, b is the Burgers vector and w is the dislocation width, which expresses the distance around the dislocation, where atomic arrangement differs from that in a perfect crystal. Peierls related the dislocation width with the interplanar spacing d so that $w = d$ for screw dislocations and $w = d/(1-\nu)$ for edge dislocations. Edge dislocations, having a larger width, glide easier than screws since a lower $P - N$ stress is required. In addition, glide is easier in close-packed planes, possessing larger interplanar spacing, as the $\{111\}$ planes in FCC.

¹R. Peierls, The size of a dislocation, Proc. Phys. Soc., 52 (1940) 52.

²F.R.N. Nabarro, Dislocations in a simple cubic lattice, Proc. Phys. Soc. (1947) 59.

Table 8.1
Values of the Peierls-Nabarro stress for various crystals

Crystal	Structure	Slip system	τ_{PN} (Mpa)	μ (GPa)	$\tau_{PN}/\mu(10^3)$
Cu	FCC	$\langle 110 \rangle \{ 111 \}$	< 0.28	40.8	< 0.007
Ag	FCC	$\langle 110 \rangle \{ 111 \}$	< 9	25.4	< 0.035
Au	FCC	$\langle 110 \rangle \{ 111 \}$	< 0.9	23.8	< 0.038
Al	FCC	$\langle 110 \rangle \{ 111 \}$	< 1.4	24.8	< 0.056
Mg	HCP	$\langle 1210 \rangle \{ 0001 \}$	< 0.9	16.4	< 0.055
Cd	HCP	$\langle 1210 \rangle \{ 0001 \}$	< 0.8	19.6	< 0.041
Zn	HCP	$\langle 1210 \rangle \{ 0001 \}$	< 1.0	39.6	< 0.027
Fe	BCC	$\langle 111 \rangle \{ 011 \}$	390	71.0	5.2
Nb	BCC	$\langle 111 \rangle \{ 011 \}$	415	47.6	8.2
Mo	BCC	$\langle 111 \rangle \{ 011 \}$	730	134.0	5.4
Ta	BCC	$\langle 111 \rangle \{ 011 \}$	340	62.8	5.2
W	BCC	$\langle 111 \rangle \{ 011 \}$	960	159.8	5.9
K	BCC	$\langle 111 \rangle \{ 011 \}$	2.325	0.90	1.85

Reprinted from: Y. Kamimura, K. Edagawa, S. Takeuchi, *Acta Materialia* 61 (2013) 294-309, with permission from Elsevier.

The $P - N$ stress τ_{PN} is several orders of magnitude smaller than the theoretical strength of crystals, $\mu/2\pi$, and depends on bond strength. *FCC* and *HCP* metals have low values of the $P - N$ stress, between 10^{-6} and $10^{-4}\mu$. Higher values are observed in *BCC* metals, of the order of $10^{-3}\mu$. Finally in crystals with covalent bonding, such as in *Si* and diamond, the $P - N$ stress takes high values, of the order of $10^{-2}\mu$. Values of the $P - N$ stress are assembled in Table 8.1.

Regarding the effect of temperature, it should be noted that the $P - N$ stress given by Equation (8.7) is the stress required to overcome the lattice resistance at $T = 0K$. At higher temperatures, the movement of the dislocation from the equilibrium position *A* to position *B* (Figure 8.6a) does not take place along the whole dislocation length simultaneously but rather a short dislocation segment *PQ* (Figure 8.7) moves to the new equilibrium position *B* by forming two kinks *x* and *y* on the same slip plane. Kink formation requires thermal activation leading to the temperature dependence of lattice resistance. Strain rate depends on stress and temperature through Equation (8.3). The activation energy is given by Equation (8.6) with $p = 3/4$ and $q = 4/3$. The specific obstacle strength (energy) ΔG is $0.1\mu b^3$ for *Si*, *Fe* and *BCC* metals, while it is only $0.05\mu b^3$ for *FCC* metals.

8.5 SOLID SOLUTION STRENGTHENING

Solid solutions are formed by the introduction of alloying elements either at substitutional lattice sites (e.g., *Zn* atoms in the *Cu* lattice) or interstitial sites (*C* atoms

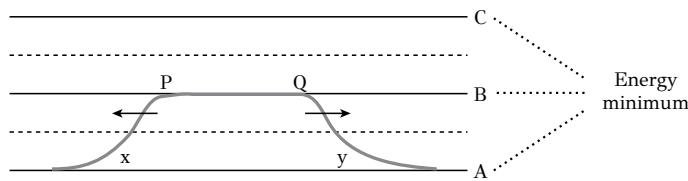


Figure 8.7: The formation of kinks at points x and y causes the displacement of dislocation segment PQ from position A to position B .

in the Fe lattice). In both cases strain fields develop around solute atoms, which interact with the stress fields of the dislocations and impede dislocation motion. The dislocation stress fields have been discussed in Chapter 3. A pure shear stress field develops around screw dislocations whereas a combination of shear and hydrostatic stresses develops around edge dislocations. Only on the slip plane of an edge dislocation there is a pure shear stress state. Above the slip plane there is a combination of shear and hydrostatic compression while below the slip plane there is a combination of shear and hydrostatic tension. Solute atoms distort the lattice and develop their own strain fields. In general there are two types of strain fields, symmetric and non-symmetric. Symmetric fields develop around substitutional solutes and are caused by the difference in size between the solute atom and the host atom, which they replace in the lattice site. The symmetrical volumetric strain field of a substitutional solute atom interacts with the hydrostatic stress field of an edge dislocation while it does not interact at all with the pure shear stress field of a screw dislocation. The substitutional solutes provide, therefore, a rather weak strengthening. On the other hand, non-symmetric strain fields develop around interstitial solutes. These fields interact with both, the stress fields of screw and edge dislocations. The best example is C in $\alpha - Fe$ (ferrite). Carbon occupies the octahedral interstitial sites in $\langle 100 \rangle$ directions (cube edges) causing the development of a non-symmetrical strain field. Thus the interaction with the stress field of screw and edge dislocations causes significant hardening in steels. For example, the most important contribution to the strength of martensite is attributed to this interaction. It should be noted that in $\gamma - Fe$ (austenite), the occupation of octahedral interstitial sites by carbon, causes the development of a symmetric strain field and, as a result, strengthening in austenite is weaker. It is apparent from the above discussion that solid solution strengthening depends on the type of the strain field, symmetric or non-symmetric, which develops around the solute atom. The details of the interactions of these strain fields with dislocations are discussed in the next section.

8.5.1 SUBSTITUTIONAL SOLUTES – SYMMETRIC FIELDS

The symmetric fields of solute atoms are caused by the difference in size between the substitutional solute atom and the host atom, which is replaced in the lattice site. If α_l and α_o are the radii of the solute and host atoms respectively, then the size

parameter δ is defined as:

$$\delta = \frac{\alpha_l - \alpha_o}{\alpha_o}$$

while the volume change associated with one substitutional atom is

$$\Delta V = \frac{4}{3}\pi(\alpha_l^3 - \alpha_o^3) = 3\Omega\delta$$

where Ω is the atomic volume. The hydrostatic pressure around an edge dislocation is

$$p = -\frac{1}{3}(\sigma_{rr} + \sigma_{\theta\theta} + \sigma_{zz})$$

The stresses are (Chapter 3)

$$\sigma_{rr} = \sigma_{\theta\theta} = -\frac{\mu b}{2\pi(1-\nu)} \frac{\sin\theta}{r}$$

$$\sigma_{zz} = \nu(\sigma_{rr} + \sigma_{\theta\theta})$$

The interaction energy between the edge dislocation and the solute atom is

$$E_i = p\Delta V$$

which becomes

$$E_i = -\frac{4}{3}\left(\frac{1+\nu}{1-\nu}\right)\mu b\delta\alpha_o^3 \frac{\sin\theta}{r}$$

indicating that the interaction energy depends on the size difference δ , while it decreases with the distance r of the substitutional solute atom from the dislocation. The solute atom exerts a resistance to dislocation glide given by

$$F_r = -\frac{\partial E_i}{\partial r}$$

The maximum resistance is then

$$F_{\max} \approx \mu b^2 \delta \quad (8.8)$$

Taking into account the Hume-Rothery rules, discussed in Chapter 2, the maximum value for δ is 0.15. This means that F_{\max} has a rather small magnitude and the resulting strengthening is weak. In order to derive an expression of solid solution strengthening, the maximum resistance of Equation (8.8) should be translated into a shear stress acting on the slip plane. This is accomplished by taking into account the distribution of the solute atoms on the slip plane as well as their positions relative to the dislocation line. This is a complicated statistical problem, which will not be considered here. From such an analysis, the CRSS corresponding to substitutional solid solution strengthening is

$$\tau = k_1 \mu b \delta^{3/2} X_S^{1/2} \quad (8.9)$$

where k_1 is a constant and X_s is the atomic fraction of substitutional solute atoms in the solid solution. The resulting strengthening is proportional to $X_s^{1/2}$. This relation is in general agreement with experimental results. There are, however, other treatments indicating that strengthening is proportional to $X_s^{2/3}$.

8.5.2 INTERSTITIAL SOLUTES – NON-SYMMETRIC FIELDS

An interstitial solute atom distorts the lattice and generates a non-symmetric strain field. The interaction of this field with dislocations is stronger than in the case of substitutional solutes. Consider the case of carbon in $BCC\alpha - Fe$. Carbon occupies the octahedral interstitial sites causing an extension in $\langle 100 \rangle$ directions and contraction in the transverse directions (see Chapter 2). The corresponding strains in the unit cell are $\varepsilon_{11} = 0.43$ and $\varepsilon_{22} = \varepsilon_{33} = 0.048$ generating an extremely non-symmetric strain field around the carbon atom. The interaction of this strain field with the stress field of a screw dislocation is characterized by an interaction energy E_i

$$E_i = \frac{A \cos \theta}{r}$$

where

$$A = \frac{\sqrt{2}\mu b\Omega}{3\pi} (\varepsilon_{11} - \varepsilon_{22})$$

and θ, r define the position of the carbon atom relative to the dislocation. As depicted in Figure 8.8 the interaction energy E_i is minimized in three positions, per 120° , around the dislocation, corresponding to carbon atoms in the directions $[100]$, $[010]$ and $[001]$. The energy minima express the strong binding of the screw dislocation by the non-symmetric strain field of the carbon atoms.

For the case of edge dislocations the respective interaction energy is

$$E_i = \frac{A \sin \theta}{r}$$

Here the interaction energy E_i is minimized below the extra half plane of the edge dislocation, as depicted in Figure 8.9, where there is more “free space” due to hydrostatic tension. The carbon atom exerts a resistance to dislocation glide, the maximum value of which is

$$F_{\max} = \mu b \zeta$$

where $\zeta = \varepsilon_{11} - \varepsilon_{22}$ is the tetragonality factor, which expresses the non-symmetric strain field generated by the carbon atom in an interstitial site. The CRSS corresponding to interstitial solid solution strengthening is then

$$\tau = k_2 \zeta^{3/2} X_c^{1/2} \quad (8.10)$$

where X_c is the atomic fraction of carbon atoms in $\alpha - Fe$. Strengthening depends on $X_c^{1/2}$, like in Equation (8.9). The difference is that strengthening is more intense due

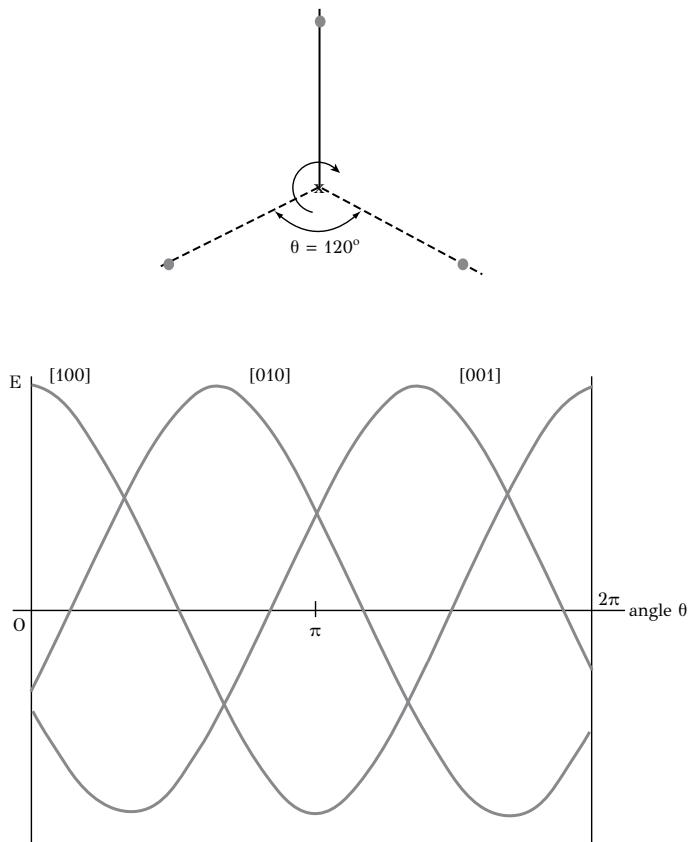


Figure 8.8: Positions of minimum interaction energy and variation of interaction energy of a carbon atom with a screw dislocation in iron.

to the tetragonality factor ζ . Interstitial solid solution strengthening is very important in steel technology. The major contribution to the strength of martensite comes from this mechanism. Martensite is a supersaturated interstitial solid solution with a *BCT* structure. The tetragonality of the crystal structure is caused by the occupation of octahedral interstitial sites by carbon atoms, generating a non-symmetric strain field, which interacts with both edge and screw dislocations. On the contrary, the occupation of the octahedral interstitial sites by carbon in austenite (*FCC* structure) generates a symmetric strain field, which interacts only with edge dislocations. As a result the strengthening of austenite by carbon is much less than in martensite. From the discussion above it is evident that solid solution strengthening depends on the type of strain field (symmetric or non-symmetric) developed around the solute atoms. Table 8.2 summarizes the potential of various metals for solid solution strengthening, expressed as $d\tau/dX$, the increase in strength with the addition of the alloying element. The case of *Al* is interesting in Table 8.2. There is a strong interaction between dis-

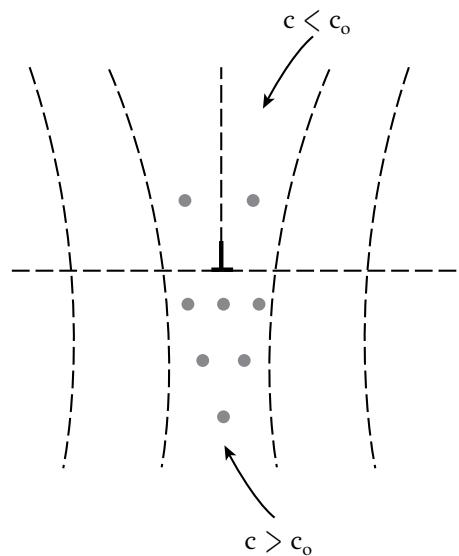


Figure 8.9: Segregation of carbon atoms below the extra half plane of an edge dislocation for the formation of a Cottrell atmosphere.

locations and vacancy discs. When Al is heated just below solidus (around $550^{\circ}C$) and quenched, a large population of vacancies is preserved (quenched-in vacancies). These vacancies arrange themselves in vacancy discs, which distort the aluminum lattice and generate a non-symmetric strain field. The interaction of this field with dislocations causes strengthening immediately after quenching.

Table 8.2

The potential of various metals for solid solution strengthening

Metal	Strain Field Type	Strengthening $d\tau/dX$
Cu	Symmetric fields-Substitutional solutes	$\mu/20$
Fe	Symmetric fields-Substitutional solutes	$\mu/16$
Ni	Symmetric fields-Substitutional solutes	$\mu/10$
Cu	Non-Symmetric field-Interstitial solutes (<i>Cu</i>)	9μ
Fe	Non-Symmetric field-Interstitial solutes (<i>Carbon</i>)	5μ
Al	Non-Symmetric field-Vacancy discs	2μ

8.5.3 COTTRELL ATMOSPHERES, YIELD POINT, AND STRAIN AGING

The interaction between dislocations and interstitial solute atoms generates some important phenomena, such as the formation of Cottrell atmospheres, the appearance of yield points in steels and strain aging.

Cottrell atmospheres. As discussed above the interaction energy between edge dislocations and interstitial solutes (such as C in $\alpha - Fe$) is minimized below the extra half plane of the edge dislocations. This leads to an increase of the concentration of carbon atoms in this area. If the average concentration in the solid solution is c_o then the concentration c at a point where the interaction energy is E_i is

$$c = c_o \exp\left(-\frac{E_i}{kT}\right) \quad (8.11)$$

Above the slip plane $c < c_o$ while below the slip plane $c > c_o$ as depicted in Figure 8.9. The carbon clusters below the extra half plane are called *Cottrell atmospheres*. The atmospheres impede dislocation glide, since the dislocation is attached to the atmosphere via the interaction energy. In steels, Cottrell atmospheres cause the appearance of yield points and strain aging.

Yield point and strain aging. In several metals, but specifically in low-carbon steels, the transition from elastic to plastic deformation is not smooth and continuous but exhibits a yield point, which is associated with a load drop. As depicted in Figure 8.10 the load drops from point A to point B and then follows a fluctuation around a constant stress value from B to C. The stress then rises again from C to D due to strain hardening. Point A is termed the *upper yield point* while point B is the *lower yield point*. The difference AB is the *yield drop*. The elongation ϵ_L from B to C is the *yield point elongation*. Plastic deformation in the area BC is localized in a zone of the material, termed the *Lueders band*. The Lueders band forms at point A and then propagates to the rest of the specimen as the specimen is elongated from B to C. The stress rises again from C to D after the Lueders band has propagated in the whole specimen. The formation of yield points was first noted in steels with low concentrations in carbon or nitrogen and was attributed to the pinning of dislocations by Cottrell atmospheres formed from carbon or nitrogen atoms. A higher applied stress is required to detach dislocations from the Cottrell atmospheres and this stress corresponds to the upper yield point. Once the dislocations are detached from the Cottrell atmospheres, glide can continue at a lower stress, corresponding to the lower yield point, from B to C. If now the specimen, after loading up to point D, is unloaded and then reloaded without delay, then the yield point will not reappear upon yielding at point E, since the dislocations are free from any Cottrell atmospheres. In this case yielding is smooth as depicted in Figure 8.10b. However if the specimen is unloaded at point Z, heated to $120^\circ C$ for 1 hour and then reloaded (Figure 8.10c), the yield point will reappear (point H). This is attributed to the re-establishment of the Cottrell atmospheres caused by the thermally-activated diffusion of carbon or nitrogen atoms towards the dislocations. The increase of yield strength due to a mild heat treatment after cold working (deformation EZ) has been termed *strain aging* and is usually accompanied by a reduction in ductility. *Bake hardening* in steels is

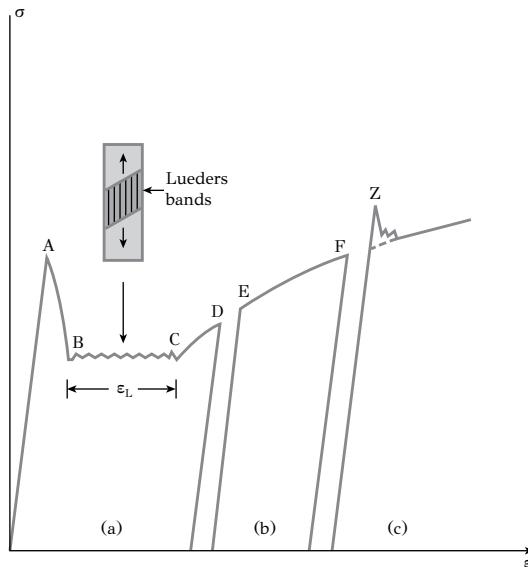


Figure 8.10: (a) Upper and lower yield point and yield point elongation, (b) immediate specimen reloading without the appearance of a yield point, (c) reappearance of a yield point after heating at 120°C for 1 hour (strain aging).

based on strain aging. It corresponds to the increase of strength through a combination of work hardening during part forming and strain aging during a subsequent thermal cycle such as a paint-baking operation. Bake-hardenable steels are ideal for applications where a high dent resistance is required.

In certain cases strain aging takes place during plastic deformation at relatively high temperatures. This phenomenon has been termed *dynamic strain aging* or *Portevin Le Chatelier effect (PLC)* and is associated with discontinuous plastic flow (Figure 8.11). The appearance of yield points is not limited to steels or other alloys containing interstitial solutes forming Cottrell atmospheres but it appears to other materials as well, such as *Ge* and *LiF*.

The yield drop can be analyzed further if we correlate the strain rate with the dislocation velocity via the expression (8.1)

$$\dot{\gamma} = \rho_m b \bar{v}$$

The mean dislocation velocity is related to the applied stress with the expression

$$\bar{v} = A \tau^m \quad (8.12)$$

During tensile testing the strain rate remains constant while the dislocation density increases leading to a reduction of mean dislocation velocity. According to Equation (8.12) this leads to a corresponding reduction of stress, which appears as a yield drop.

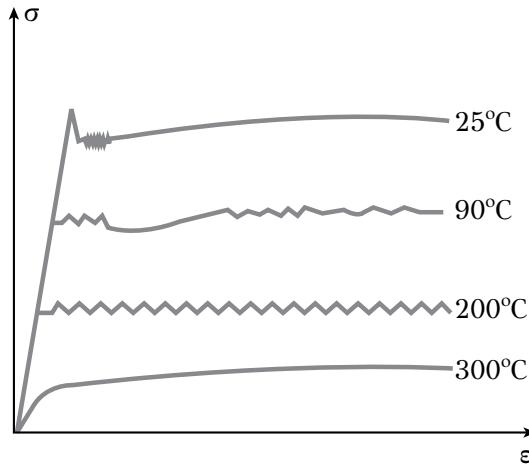


Figure 8.11: Dynamic strain aging in a low carbon steel.

The magnitude of the yield drop depends on the initial dislocation density and the exponent m in Equation (8.12). When the initial dislocation density (before deformation) is low, all dislocations are attached to Cottrell atmospheres and a relatively high stress, corresponding to the upper yield point, is required to release the dislocations. The density of dislocations rises after they are released from the Cottrell atmospheres. This leads to a decrease in their velocity and eventually to a load drop. When the initial dislocation density is high, the load drop is not so intense (Figure 8.12a). Based on these arguments, a low exponent m in Equation (8.12) could lead

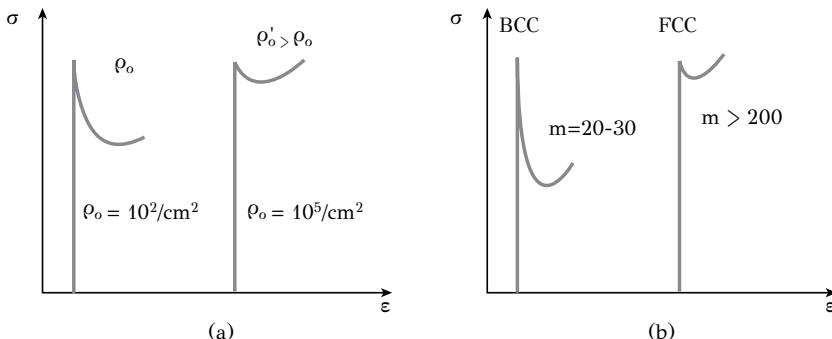


Figure 8.12: The effect of initial dislocation density ρ_0 and the exponent m on the magnitude of the yield drop in α -Fe.

to the appearance of an intense yield drop (Figure 8.12b). In *BCC* metals (such as steels) the exponent m takes values in the range $20 - 30$, leading to intense yield drops. In *FCC* metals the exponent m takes high values ($m > 200$) and as a result

these metals do not exhibit a yield drop.

8.6 GRAIN BOUNDARY STRENGTHENING

The orientation of slip planes changes from one grain to another at a grain boundary. This impedes dislocation glide in the material and leads to the grain size dependence of yield strength. Thus a fine-grained alloy, containing more grain boundaries, will exhibit a higher yield strength than a coarse-grained alloy. The relation between grain size and yield strength has been originally proposed by Hall¹ and Petch² as

$$\sigma_y = \sigma_i + k_y d^{-1/2} \quad (8.13)$$

where d is the mean grain diameter, k_y the Hall-Petch coefficient and σ_i the lattice resistance. The Relation (8.12) is known as the Hall-Petch relation and is followed by many metals and alloys as depicted in Figure 8.13. Values of σ_i and k_y are given

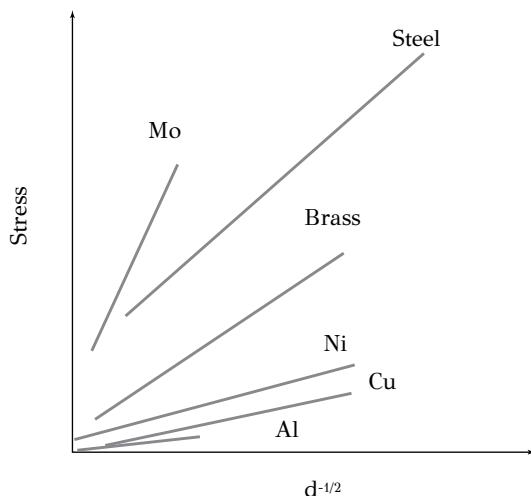


Figure 8.13: Hall-Petch diagram for several metals. (schematic)

for various metals in Table 8.3.

The Hall-Petch coefficient expresses the potential for grain boundary strengthening in a certain metal. According to the values in Table 8.3, grain size refinement leads to higher strengthening in steel than in aluminum.

Several theories have been proposed to rationalize Hall-Petch strengthening. Consider that a Frank-Reed source in grain 1 emits dislocations, which then form a pile-up at the grain boundary (Figure 8.14a). The first dislocation at the tip of the pile-up

¹E.O. Hall, Proc. Roy. Soc. (London), B64, p.474, 1951

²N.J. Petch, J. Iron Steel Inst., 174, p.25, 1953

Table 8.3**Values of σ_i and k_y for various metals**

Metal	$\sigma_i(MPa)$	$k_y(MN/m^{3/2})$
Mild steel	70	0.74
Al	16	0.07
Cu	25	0.11
Ti	80	0.40

is immobilized at the grain boundary. The interaction of the stress fields of the dislocations of the pile-up causes a stress concentration τ_s at the tip of the pile-up as depicted in Figure 8.14b. This stress is

$$\tau_s = n\tau_\alpha$$

where n is the number of dislocations forming the pile-up and τ_α the applied stress, which causes dislocation glide in grain 1. The number of dislocations in the pile-up is proportional to the applied stress and the grain size d

$$n = \frac{d\tau_\alpha}{\mu b} \pi (1 - \nu)$$

Combining the above relations we get

$$\tau_\alpha = \left(\frac{\mu b \tau_s}{\pi (1 - \nu)} \right)^{1/2} d^{-1/2}$$

Consider now that plastic deformation is transmitted from grain 1 to grain 2 with the activation of a Frank-Reed source in grain 2 (FR_2 in Figure 8.14c) by τ_s . The yield strength is $\sigma_y = \tau_\alpha M_T$, where M_T is the Taylor factor. If we include the lattice resistance the yield strength becomes

$$\sigma_y = \sigma_i + M_T \left(\frac{\mu b \tau_s}{\pi (1 - \nu)} \right)^{1/2} d^{-1/2}$$

We set

$$k_y = M_T \left(\frac{\mu b \tau_s}{\pi (1 - \nu)} \right)^{1/2} \quad (8.14)$$

and the above relation becomes

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

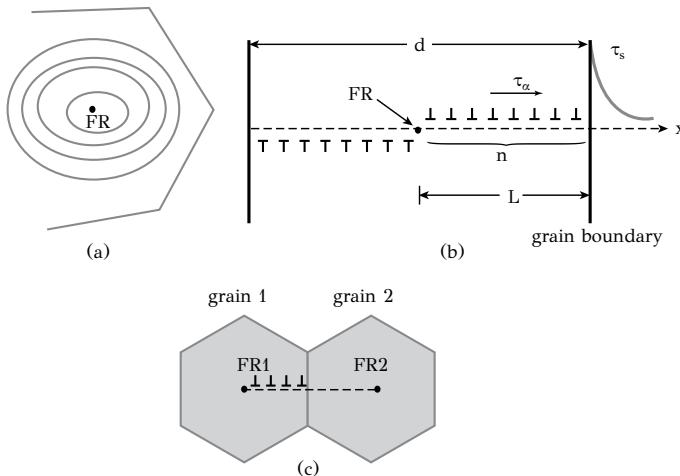


Figure 8.14: (a) Dislocation emission from a Frank-Reed source (*FR*), (b) formation of a dislocation pile-up and associated stress concentration, (c) activation of a Frank-Reed source (*FR2*) in grain 2 from a dislocation pile-up in grain 1.

which is the Hall-Petch Relation (8.12). From the Relation (8.13) the coefficient k_y can be written as

$$k_y = k_o \tau_s^{1/2}$$

The coefficient k_y is temperature dependent through k_o and τ_s . The temperature dependence of k_o arises from the temperature dependence of the shear modulus $\mu(T)$. Regarding the temperature dependence of τ_s we distinguish two cases. If τ_s represents the stress to activate the Frank-Reed source then it is independent of temperature since the activation of a Frank-Reed source does not require thermal activation. If τ_s represents the required stress for dislocations to be released from Cottrell atmospheres in grain 2, then the temperature dependence of k_y depends on whether the attachment of dislocations is strong or weak. If the dislocations are strongly attached to Cottrell atmospheres, e.g., after slow cooling, then τ_s does not depend on temperature. If on the other hand the attachment is weak (e.g., after quenching) then thermal activation can assist the dislocations to be released from the Cottrell atmospheres and τ_s is temperature dependent, as depicted in Figure 8.15.

Grain boundary strengthening or Hall-Petch strengthening has driven alloy development in steels and non-ferrous alloys. Grain refinement has been one of the most effective ways to achieve high strengths in metallic alloys. HSLA steels is a good example, where minor additions of carbonitride forming elements (*Ti*, *Nb*, *V*) cause the precipitation of fine dispersions of carbonitrides during thermomechanical forming, impeding grain growth. The Hall-Petch effect has driven the research and development of nanocrystalline metals as well. However when the size scale changes from micro to nano, a significant change in the Hall-Petch mechanism takes place. As the grain size gets very small, below say a threshold value of about 10nm, it is not pos-

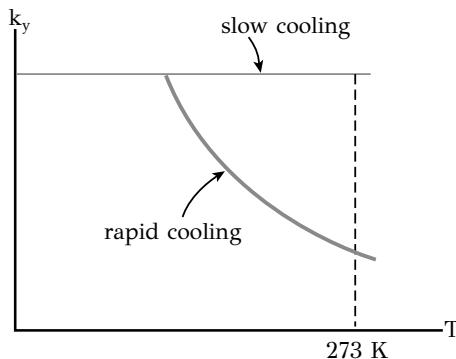


Figure 8.15: The temperature dependence of the coefficient k_y in the Hall-Petch relation.

sible for the dislocations to form pile-ups. The Hall-Petch coefficient k_y is reduced or becomes negative. This is known as the *inverse Hall-Petch effect (IHPE)* and has been associated to the fact that at very small grain sizes, in the nanocrystal range, the dislocations instead of forming pile-ups are absorbed by the grain boundaries¹. This causes the yield strength to deviate from the Hall-Petch expression under a certain grain size in nanocrystalline metals.

8.7 PRECIPITATION STRENGTHENING

Several alloy systems, among them steels, aluminum and titanium alloys, obtain their strength through the mechanism of precipitation strengthening. Dislocation glide is impeded by a fine dispersion of precipitates of a second phase or intermetallic compound, which forms by a precipitation reaction in the material. The resulting hardening can be significant and in several alloys precipitation strengthening dominates any additional strengthening mechanisms that might operate in the material. In our discussion we will use the $Al - Cu$ alloy as an example of precipitation strengthening. It forms the basis of the aeronautical alloys, such as alloy 2024, which obtains its strength through the precipitation of the intermetallic phase Al_2Cu through a process called aging. Therefore, aging will be discussed first in this chapter.

8.7.1 AGING

Precipitation of a second phase or an intermetallic compound from a supersaturated solid solution is one of the most important methods for alloy strengthening. The precipitate, in the form of fine particle dispersion, impedes dislocation glide. Precipitation is a diffusive phase transformation, which takes place by nucleation and

¹C.E. Carlton and P.J. Ferreira, Acta Mater. 55(2007) 3749

growth, processes that were discussed in detail in Chapter 6. We will discuss the basic characteristics of precipitation using the alloy $Al - 4Cu$ (wt %) as an example. The relevant section of the $Al - Cu$ phase diagram is shown in Figure 8.16. The heat

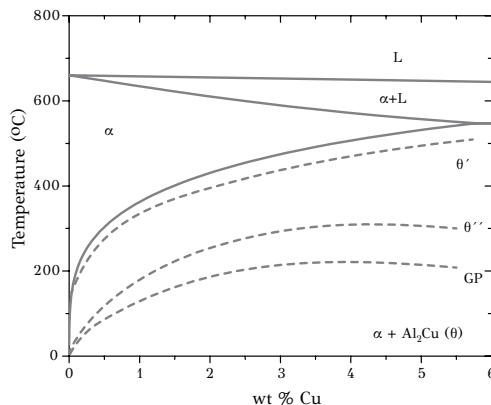


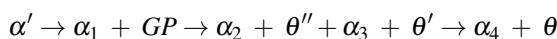
Figure 8.16: Section of the $Al - Cu$ phase diagram with the solvus lines of GP zones and phases θ'' , θ' and θ . (Calculated from Thermo-Calc)

treatment applied for the precipitation of the Al_2Cu (θ) phase consists of three steps:

1. Solution treatment, which involves heating above the solvus line ($540^\circ C$) for the dissolution of phase θ and the formation of a homogeneous $Al - Cu$ solid solution (phase α).
2. Quenching to room temperature, to avoid diffusive precipitation and obtain a supersaturated solid solution. The alloy develops a high driving force for the precipitation of the θ phase.
3. Heating at an intermediate temperature ($100 - 200^\circ C$). The activation of diffusion in combination with the high driving force results in the precipitation of phase θ . The process is called *aging*.

Aging is distinguished in artificial and natural aging depending on whether aging is carried out at a high temperature or at room temperature. Due to the aging process, the mechanism of precipitation hardening has also been termed *age hardening*.

As discussed in Chapter 6, the formation of the equilibrium phase θ during aging is preceded by the formation of transition phases following a precipitation sequence:



The supersaturated solid solution α' , which forms after quenching from the solution temperature, decomposes during aging. GP zones form first in the precipitation

sequence. The name *GP* comes from Guinier and Preston who identified them in 1938. They are *Cu*-rich zones, coherent with the matrix. There is a reason why *GP* zones form instead of the equilibrium phase θ . Due to coherency with the matrix they possess a very low interfacial energy, which results in a respectively low activation energy for nucleation. In addition their disk-like shape minimizes the coherency strains. So despite the fact that the θ phase has a higher driving force for precipitation, the *GP* zones form instead due to their lower activation energy for nucleation. The formation of *GP* zones is followed by the precipitation of the transition phases θ'' and θ' , which also possess a lower activation energy for nucleation than the equilibrium θ phase. Phase θ'' is coherent while phase θ' is semi-coherent with the matrix. Therefore during aging, as the precipitate particles grow, the state of coherency is reduced from fully coherent (*GP* zones and θ'') to semicoherent (θ') and to finally incoherent (θ) precipitates. As discussed in Chapter 6, the coherency loss is attributed to the strain energy associated with coherency and which increases as the coherent particles grow. The system, therefore, at some point during the aging process prefers to reduce the state of coherency in order to minimize the total free energy. The complete precipitation sequence is observed only if the aging treatment is performed below the *GP* solvus, as depicted in Figure 8.16. If aging is performed above the *GP* solvus but below the θ'' solvus then precipitation will start with the formation of phase θ'' . The nucleation of the *GP* zones as well as the transition phases (θ'' and θ') is heterogeneous. Typical nucleation sites for the *GP* zones are vacancy clusters, which form from the quenched-in vacancies formed at the solution temperature. *GP* zones then act as nucleation sites for the θ'' phase and so on. The precipitation sequences in other alloy systems are depicted in Table 8.4. The dispersion of the needle-shaped phase S' (Al_2CuMg) in a $Al - Cu - Mg$ alloy is shown in Figure 8.17, while the dispersion of phase β (Mg_2Si) in a $Al - Mg - Si$ alloy is depicted in Figure 8.18.

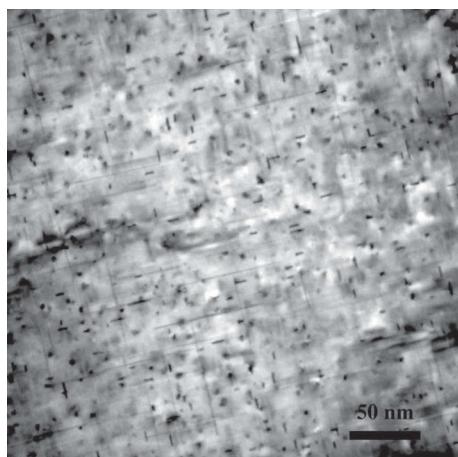


Figure 8.17: Dispersion of S' phase in an $Al - Cu - Mg$ alloy.

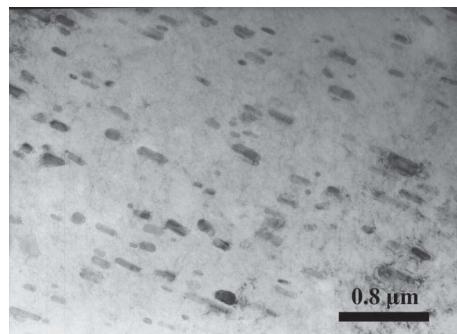


Figure 8.18: Dispersion of β phase in an $Al - Mg - Si$ alloy.

Table 8.4
The precipitation sequences in several alloy systems

Alloy	Precipitation sequence
Al-Cu	GP (discs) $\rightarrow \theta''$ (discs) $\rightarrow \theta'$ $\rightarrow \theta$ (Al_2Cu)
Al-Mg-Si	GP (rods) $\rightarrow \beta'$ (rods) $\rightarrow \beta$ (Mg_2Si) (plates)
Al-Mg-Si-Cu	clusters $\rightarrow GP \rightarrow \beta'' + \alpha' \rightarrow \alpha + \beta$ (Mg_2Si)
Al-Cu-Mg	GP (rods) $\rightarrow S'$ (needles) $\rightarrow S$ ($CuMgAl_2$) (needles)
Al-Zn-Mg	GP (spheres) $\rightarrow \eta'$ (plates) $\rightarrow \eta$ ($MgZn_2$) (plates/rods)
Cu-Be	GP (discs) $\rightarrow \gamma' \rightarrow \gamma$ ($CuBe$)
Fe-C	ε -carbide (discs) $\rightarrow (Fe_3C)$ (plates)
Ni-Cr-Ti-Al	γ' Ni_3Ti or Ni_3Al (cubes or spheres)

The transition phases, which form during the aging process, exhibit individual specific resistance to dislocation glide. Thus the strength of the material depends on aging time and temperature. The change in hardness of a $Al - 4Cu$ alloy during aging is depicted in Figure 8.19 for two aging temperatures 120 and $180^{\circ}C$. During aging at $120^{\circ}C$, which is lower than the GP zone solvsus, GP zones form. They are fully coherent with the matrix and develop coherency strains. It is the interaction of the dislocation stress fields with those coherency strains that leads to hardening. With aging time the θ'' phase forms, again fully coherent with the matrix and as it grows so do the coherency strains. The hardness, therefore, increases with aging time. At a later stage, as discussed above, coherency is reduced with the formation of the semi-coherent phase θ' . Peak hardness is obtained just before coherency loss. The formation of θ is associated with the coarsening of the dispersion. The mean size of θ' , as well as the interparticle spacing increase, making it easier for dislocations to bypass the obstacles. As a result, the hardness decreases with aging time, a phenomenon termed *overaging*. During aging at $180^{\circ}C$, GP zones do not form, since the temperature is above the GP zone solvsus. The θ'' phase forms instead and is followed by the formation of phase θ' . However the supersaturation and the associated precipitation driving force are lower at 180 compared to $120^{\circ}C$. As a result, a lower volume fraction of θ' forms, leading to lower peak hardness. In addition, due to the accelerated diffusion at $180^{\circ}C$, the peak hardness is obtained at shorter aging times. It is apparent that in order to obtain the peak hardness in an alloy, an optimum combination of aging temperature and aging time should be applied. In several cases a two-stage aging is performed. During the low-temperature step GP zones form, while during the high-temperature step the phase θ' forms heterogeneously, using the GP zones as nucleation sites. In this way a finer particle dispersion is formed, leading to enhanced hardening at shorter aging times.

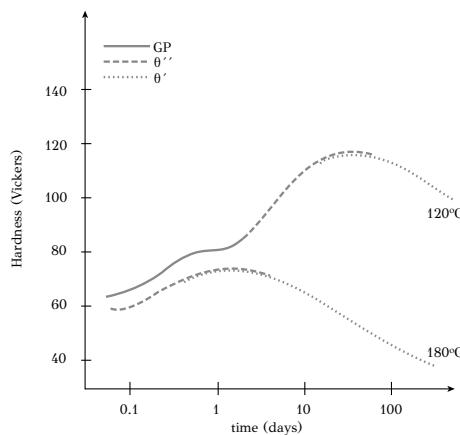


Figure 8.19: The change of hardness of alloy $Al - 4Cu$ with aging time for two aging temperatures 120 and $180^{\circ}C$.

8.7.2 PRECIPITATION STRENGTHENING MECHANISMS

The aging curve of Figure 8.19 can be interpreted by considering the interaction between dislocations and the various transition phases, which form during aging. *GP* zones and the θ'' phase are fully coherent with the matrix. As a result a dislocation can penetrate and glide through these particles, which are, therefore, called *penetrable particles*. On the contrary, the particles of phases θ' and θ are not coherent with the matrix and the dislocations can not glide through them. They are called *impenetrable particles*. As expected, the strengthening mechanism depends on the type of particles. The penetrable particles impede dislocation glide due to the interaction of the coherency strain fields with the dislocation stress fields. The resulting hardening has been termed *coherency hardening* and is given by

$$\Delta\tau = k_1 f^{1/2} r^{1/2}$$

where k_1 is a constant, f is the volume fraction and r the mean size of the coherent precipitates. Coherency hardening is a long-range interaction and the temperature dependence of the flow stress is rather weak, arising mainly from the lattice resistance. In some special cases the coherent precipitates exhibit an ordered structure. In this case a short-range interaction with dislocation operates. As the dislocation penetrates the ordered particle it creates an anti-phase boundary at the slip plane (see Chapter 6 for a discussion on anti-phase boundaries). The interfacial energy of the anti-phase boundary γ_{APB} is relatively large, leading to significant hardening given by

$$\Delta\tau = k_2 \gamma_{APB}^{3/2} f^{1/3} r^{1/2}$$

where f is the volume fraction and r the mean size of the precipitates. As stated above, this is a short-range interaction and the temperature dependence is stronger. Now consider impenetrable particles. Dislocations cannot glide through them. Instead they bow between the particles, as depicted in Figure 8.20 with particles of θ' phase in a $Al - Cu$ alloy. A pair of particles operates like a Frank-Reed dislocation source. The maximum stress required for the dislocations to pass between the particles by bowing is

$$\tau_{\max} = \frac{\mu b}{L}$$

where L is the mean interparticle distance, as depicted in Figure 8.21. As the dislocation bows between the particles and expands, segments *A* and *B* annihilate each other since they have opposite sign, forming dislocation loops around the particles. The particle by-passing has been termed *Orowan by-passing* and the resulting hardening, *Orowan hardening*. The hardening is given by

$$\Delta\tau = k_3 f^{1/2} r^{-1}$$

where k_3 is a constant, f is the volume fraction and r the mean size of the precipitates. The loop formation decreases the effective interparticle spacing.

This increases the stress required for the dislocations to by-pass the particles and leads to strain hardening. It is worth noting that in the case of penetrable particles,

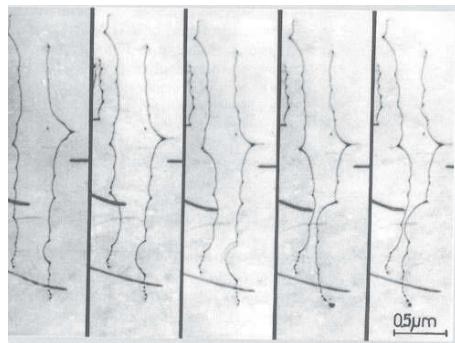


Figure 8.20: Interaction of dislocation with particles of θ phase in an $Al - Cu$ alloy.

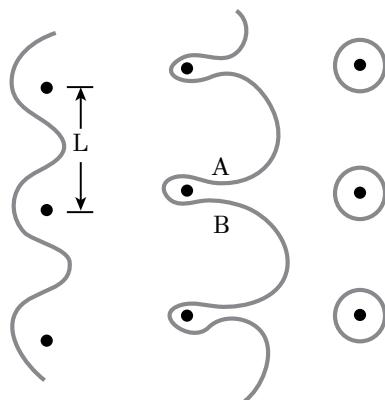


Figure 8.21: Orowan by-passing of impenetrable particles by dislocations.

strengthening is proportional to $r^{1/2}$, while in the case of impenetrable particles, strengthening is proportional to r^{-1} . These relations are depicted in Figure 8.22. The mean particle size increases during aging due to growth and coarsening processes. The strengthening arising from the penetrable particles of GP -zones and θ'' as well as the impenetrable particles of θ' and θ is depicted in Figure 8.22. The total hardening curve is similar to the aging curve of Figure 8.19, while the peak appears at the change of strengthening mechanism.

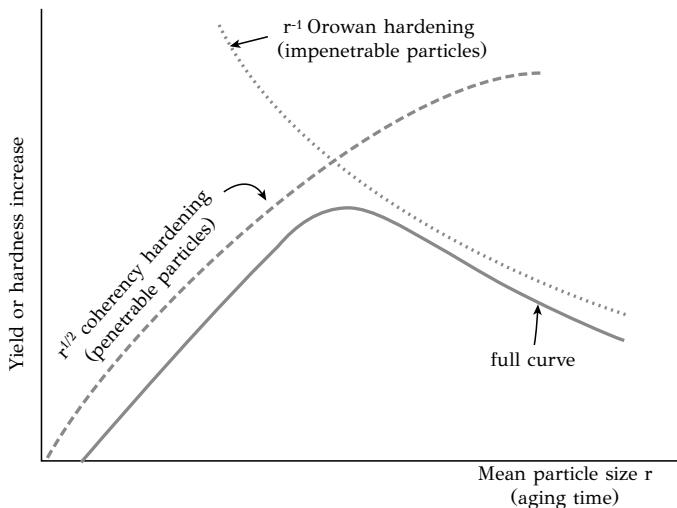


Figure 8.22: Strengthening by penetrable and impenetrable particles with particle size (aging time).

The strain hardening associated with precipitation strengthening depends on the specific operative mechanisms. Typical stress-strain curves are depicted in Figure 8.23 for aging at 180°C. Cases (a) and (b) correspond to coherency hardening. Strain hardening is mild since no loops are forming around the particles. On the other side, cases (c) and (d), dislocation loop formation leads to considerable strain hardening.

8.8 IMPLICATIONS OF STRENGTHENING MECHANISMS

The discussion on strengthening mechanisms will be concluded by referring to some implications and examples. Plastic deformation of metals takes place by one of the following mechanisms:

- dislocation glide
- deformation twinning
- martensitic transformation
- diffusional flow

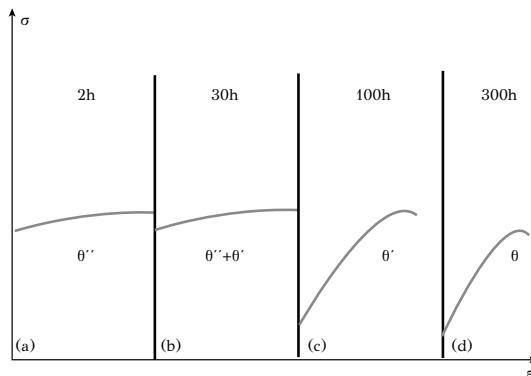


Figure 8.23: Stress-strain curves for alloy $Al - 4Cu$ for various aging times at $180^{\circ}C$.

From the above list, the dominant mechanism is dislocation glide. Twinning is activated in metals, which do not exhibit multiple slip systems or if for some reason dislocation glide is hindered (e.g., at low temperatures). Martensitic transformation takes place in specific alloy systems and is accompanied by significant shear-like deformations. Diffusional flow is the deformation caused by the diffusional transfer of mass and contributes to the creep deformation of a metal, when loaded at high temperatures. As discussed above, dislocation glide is the dominant deformation mechanism in metals. Accordingly, impeding dislocation glide, constitutes the basis for the strengthening of metals and alloys. The most important obstacles to dislocation glide and the respective strengthening mechanisms are depicted in Table 8.5. The information in Table 8.5 is important for two reasons:

- It helps to understand how a metal develops its mechanical strength, in other words it reveals the various strengthening mechanisms of metals and alloys
- It helps to design new alloys with enhanced strength. The word “design” encompasses both, chemical composition and processing, which is applied for the development of the desirable microstructure.

Table 8.5 contains several “elements of microstructure” of an alloy. The chemical composition influences directly the solid solution strengthening as well as precipitation strengthening, through the formation of precipitates. The volume fraction and the mean size of these precipitates affect precipitation strengthening. Grain size also affects strength through the mechanism of grain boundary strengthening (Hall-Petch strengthening). The dislocation density affects strain hardening. The mathematical relations of the third column of Table 8.5 are simple examples of the relation between structure and properties in metals.

A good example to illustrate the operative strengthening mechanisms are aluminum alloys. They are distinguished in heat-treatable and non heat-treatable alloys. The alloys of the 3xxx ($Al-Mn$) and 5xxx ($Al-Mg$) series are non heat-treatable. They

Table 8.5
Overview of the strengthening mechanisms in metals

Mechanism	Obstacle	Strengthening	Temperature dependence
Lattice resistance	Crystal lattice	$\frac{2\mu}{1-\nu} \exp\left(-\frac{2\pi w}{b}\right)$	large
Workhardening	Other dislocations	$\alpha\mu b\sqrt{\rho}$	small
Solid solution strengthening	Alloying elements	Substitutional solutes $\mu b\delta^{3/2} X_S^{1/2}$ Interstitial solutes $\zeta^{3/2} X_c^{1/2}$	moderate moderate
Grain boundary strengthening	Grain boundaries	$k_y d^{-1/2}$	moderate
Precipitation strengthening	Precipitates, second phase particles and intermetallic compounds	penetrable particles $f^{1/2} r^{1/2}$ penetrable ordered particles $f^{1/3} r^{1/2}$ impenetrable particles $f^{1/2} r^{-1}$	large large large

acquire their strength through strain hardening, after cold working operations. The alloys of the 2xxx (*Al-Cu*), 6xxx (*Al-Mg-Si*) and 7xxx (*Al-Zn-Mg*) are heat-treatable alloys and acquire their strength through precipitation hardening. An aging heat treatment is applied, after solution treatment and quenching, in order to form dispersions of fine precipitates.

In most alloys more than one strengthening mechanism is activated, i.e., several types of obstacles to dislocation glide contribute to strengthening. Martensite in steels is a good example. Intense solid solution strengthening caused by carbon in interstitial sites. Strain hardening due to the high dislocation density. Boundary strengthening involving not only grain boundaries but lath boundaries as well. Finally, with suitable alloying (e.g., *Cr*, *Mo*, *W*) to form carbides during tempering, martensite is additionally strengthened by precipitation hardening. The role of alloying elements becomes complicated if studied in the context of strengthening mechanisms. A good example is a ultrahigh-strength steel with composition *Fe-0.25C-14Co-10Ni-2Cr-1Mo*. Carbon causes strengthening of martensite (solid solution strengthening) while it forms carbides with *Cr* and *Mo* during tempering (precipitation strengthening). *Co* remains in solution since it is not a carbide former, while its contribution to solid solution strengthening is weak. Then why is cobalt added to this steel? Martensite possesses a high dislocation density after the quench from the austenitizing temperature. These dislocations contribute to strain hardening

but also operate as heterogeneous nucleation sites for the carbide precipitation during tempering. However the dislocation density quickly diminishes during tempering due to recovery processes. Cobalt restricts the recovery processes, maintaining a high dislocation density for heterogeneous nucleation of carbides. The resulting carbide dispersion is extremely fine contributing to significant precipitation strengthening. Several strengthening mechanisms associated with alloying elements contribute to the development of mechanical strength in this steel. In most cases the development of high strength is accompanied by a respective reduction in ductility. It is for this reason that *Ni* is added to the steel, to increase ductility and fracture toughness, especially at low temperatures.

8.9 SYNOPSIS

1. The strengthening mechanisms are related to the mechanisms by which various obstacles impede dislocation glide. Such obstacles are the crystal lattice, other dislocations, solute atoms, precipitates and grain boundaries. Each obstacle displays a specific resistance to dislocation glide, termed the specific obstacle strength.
2. Either a higher stress or thermal activation are required for dislocations to overcome the various obstacles they encounter during glide. Thus the plastic strain rate becomes stress and temperature dependent.
3. The yield strength, which expresses the resistance of the material to plastic deformation consists of two components, the thermal and the athermal component. The athermal component is caused by long-range obstacles, such as the stress fields of other dislocations, which operate at large distances around the dislocation core. These obstacles cannot be surmounted by thermal activation, making the athermal component nearly independent of temperature. On the other hand, the thermal component is caused by short-range obstacles, such as solute atoms and precipitates. As these obstacles have a local effect on dislocations, they can be surmounted by thermal activation. Thus the temperature dependence of the flow stress is attributed to the thermal component.
4. The major strengthening mechanisms are: lattice resistance, strain hardening, solid solution hardening, grain boundary strengthening and precipitation strengthening.
5. Lattice resistance refers to the resistance provided by the crystal lattice to dislocation glide. Lattice resistance is related to atomic bonding and the periodicity of the lattice. It is characterized by the Peierls-Nabarro stress.
6. Solid solution strengthening arises from the interaction of the stress fields of dislocations with the strain fields of solute atoms. These strain fields are distinguished in symmetric (caused by substitutional solutes) and non-symmetric (caused by interstitial solutes).

7. Cottrell atmospheres are concentrations of carbon atoms at dislocations. They generate several phenomena such as the appearance of yield points in steels and strain aging.
8. High-angle grain boundaries impede dislocation glide in the material and lead to the grain size dependence of the yield strength. Thus a fine-grained alloy, containing more grain boundaries, will exhibit a higher yield strength than a coarse-grained alloy. The effect of grain size is described with the Hall-Petch relation.
9. In precipitation strengthening dislocation glide is impeded by a fine dispersion of precipitates of a second phase or intermetallic compound, which forms by a precipitation reaction in the material. In most cases precipitates form with a heat treatment called aging. Precipitate particles are distinguished in two types: penetrable particles, exhibiting coherency hardening and impenetrable particles exhibiting Orowan hardening. In both cases the resulting hardening depends on the volume fraction and the mean size of the precipitates.
10. The mechanical strength of most alloy systems is developed through the operation of several strengthening mechanisms, which operate simultaneously. Alloying elements and material processing play a key role in the formation of the various obstacles, which are responsible for the operation of these strengthening mechanisms.

8.10 REVIEW QUESTIONS

1. Consider a line of carbon atoms attached below the half plane of an edge dislocation. The interaction energy is

$$E_i = A \frac{\sin \theta}{r}$$

Under the assumption that elasticity theory is valid close to the dislocation core, show that the required maximum shear stress to break away the dislocation from the atmosphere is

$$\tau = \frac{3\sqrt{3}A}{8b^4}$$

2. Estimate the maximum stress in the above question if

$$A = \frac{\sqrt{2}\mu b\Omega}{3\pi} (\varepsilon_{11} - \varepsilon_{22})$$

where Ω is the atomic volume of α -iron (BCC). Comment on your answer.

3. Consider the appearance of yield points in the stress-strain curve. Using the equations (8.1) and (8.12) show that sharp yield points are associated either with a low initial dislocation density ρ or with a small exponent m .

4. In a demanding application, it is required to develop a precipitation-hardened alloy with shear strength of $\mu/50$. Estimate the necessary mean spacing between precipitates. Discuss possible differences between coherent and incoherent precipitates.
5. HSLA steels with low carbon content (< 0.08) are subjected to controlled rolling and then, depending on cooling rate, develop either a ferritic or bainitic (acicular ferrite) structure. The first develops a yield point phenomenon while the second does not. Discuss the possible explanations for this behavior.
6. The phase diagram of the binary $Al - Mg$ alloys exhibits a sloping solvus. Discuss why these alloys do not exhibit a respective precipitation hardening, as do $Al - Cu$ alloys.
7. Consider the series of wrought aluminum alloys: 1xxx (unalloyed Al), 2xxx($Al - Cu$), 3xxx($Al - Mn$), 5xxx($Al - Mg$), 6xxx($Al - Mg - Si$) and 7xxx($Al - Zn - Mg$).
 - a. Which of the above series are heat treatable alloys and which non-heat treatable alloys?
 - b. Discuss at least one industrial application for each series.
 - c. The fuselage and wings of airplanes are manufactured from thin sheets of $Al - Cu$ alloys of the 2xxx series. Discuss the reasons that riveting is used instead of welding in the construction.
 - d. The rivets are also made from a 2xxx Al -alloy. Discuss why the rivets are refrigerated immediately after the solution heat treatment and remain there until the riveting process.
8. Identify the major strengthening mechanisms in the following alloys:
 - a. Quenched and tempered tool steel containing $0.39C - 1.0Si - 0.4Mn - 5.2Cr - 1.4Mo - 0.9V(wt\%)$
 - b. Wrought $Cu - 30Zn(wt\%)$ brass
 - c. 6061-T6 Al -alloy ($1Mg - 0.6Si - 0.5Fe - 0.2Cu, wt\%$)
 - d. 5083 Al -alloy ($4.5Mg - 0.4Si - 0.4Fe - 0.1Cu, wt\%$)

BIBLIOGRAPHY—FURTHER READING

1. Meyers, M. A., & Chawla, K. K. (1984). *Mechanical Metallurgy. Principles and Applications*, Prentice Hall, Englewood Cliffs, New Jersey.
2. Hayden, H. W., Moffatt, W. G., & Wulff, J. (1965). *The Structure and Properties of Materials, Vol. 3: Mechanical Behavior*, John Wiley & Sons, USA.
3. Cottrell, A. H. (1963). *Dislocations and Plastic Flow in Crystals*, Oxford University Press.
4. Kelly, A., & Nicholson, R. B. (1971). *Strengthening Methods in Crystals*, Halsted Press Division, John Wiley & Sons, New York.
5. Cahn, R. W. (1983). *Recovery and Recrystallization, in Physical Metallurgy*, R. W. Cahn, P. Haasen editors, Elsevier Science Publishers.
6. Hertzberg, R.W. (1989). *Deformation and Fracture Mechanics of Engineering Materials*, John Wiley & Sons, USA.

9 Fracture, fatigue, and creep of metals

9.1 INTRODUCTION – MECHANICAL BEHAVIOR OF METALS

The term *mechanical behavior* refers to the response of a metallic alloy to the application of mechanical loads. The material may be subjected to static or dynamic loading, at low or high temperatures, while at the same time it may be exposed to a corrosive environment. The response of a metallic alloy to these conditions is, in general, characterized by either deformation or fracture depending on the type and magnitude of applied loads on one hand and on its microstructure, which includes chemical composition, distribution and morphology of phases, grain size, dislocation density, interfaces, inclusions, etc., on the other. Each of the above microstructural elements affects, by specific individual mechanisms, the mechanical behavior of the metal.

The contribution of microstructure to mechanical behavior is expressed by the *mechanical properties* of the alloy. If these properties are known then the response of an alloy to mechanical loading can be determined. Take for example the yield strength. If a metal is loaded below its yield strength, the resulting deformation will be elastic. The magnitude of the elastic deformation depends on another property, the modulus of elasticity. However, if the metal is loaded above its yield strength, then the deformation will be plastic. The yield strength, therefore, characterizes quantitatively the resistance of the metal to plastic deformation. The yield strength is made up of several contributions arising from the crystal lattice, dislocations, alloying elements, precipitates, grain boundaries and other microstructural elements through the strengthening mechanisms, discussed in Chapter 8.

A central point in the study of the mechanical behavior of metals is the measurement of mechanical properties. This is accomplished by *mechanical testing*. The yield strength of a metal, for example, is measured by tensile testing.

The correlation between microstructure and properties is integrated with the effect of processing. Materials processing, including casting, deformation processing, heat treatment or surface treatment, affects the structure and consequently the mechanical properties and mechanical behavior of the metal.

Fracture, fatigue and creep are central issues regarding the mechanical behavior and will be discussed in the present chapter. Together with elastic and plastic deformation, they constitute the basic elements of mechanical behavior of metals. In summary, the basic characteristics of mechanical behavior are the following:

Elastic deformation. The metal is loaded below its yield strength and the deformation is reversible. The magnitude of elastic deformation depends on the elastic modulus. The major microstructural element influencing the modulus of elasticity is the strength of the metallic bond.

Plastic deformation. The material is loaded above its yield strength where generalized dislocation glide takes place. Loading results in permanent deformation. Plastic deformation is accompanied by strain hardening. Relevant mechanical testing: tensile and compression testing, hardness testing.

Fracture. There are two major modes of fracture. Ductile fracture, which takes place after considerable plastic deformation and brittle fracture, which takes place without considerable prior plastic deformation of the material. Besides mode of loading, fracture is influenced by pre-existing crack-like flaws and microstructure. Resistance to fracture is characterized by fracture toughness. Relevant mechanical properties: fracture toughness testing, impact testing.

Fatigue. The metal is subjected to cyclic loading. Fatigue is the progressive local plastic deformation and accumulation of damage, which includes the formation and propagation of cracks up to the final fracture. The resistance of the material to fatigue is characterized by the fatigue endurance limit and the crack propagation rate. Relevant mechanical properties: fatigue testing ($S - N$ curve), crack propagation testing (da/dN).

Creep. Loading, usually under constant stress, at a high temperature (above 40% of the melting point), causes permanent deformation, which increases with time. This time-dependent deformation is called creep. The resistance of the material to creep is characterized by the creep rate and the time to fracture. Relevant mechanical properties: creep testing, creep rupture testing.

The above constitute the core of the mechanical behavior of metals. The understanding of the mechanical behavior is a prerequisite for materials selection in engineering applications and for the design of new metallic alloys.

The discussion in the present chapter will be focused on fracture, fatigue and creep of metals. Due to space limitations, issues related to environmentally-assisted fracture (stress corrosion cracking, corrosion fatigue, high-temperature oxidation) will not be treated in the current version of this book.

9.2 FRACTURE

9.2.1 INTRODUCTION – THE PROBLEM OF BRITTLE FRACTURE

Metals play a significant role in construction due to the excellent strength/toughness combination. We use to say that metals are *ductile* relative to other materials since fracture takes place after significant plastic deformation. This type of fracture is called *ductile fracture*. A metallic structural component should fail by plastic deformation, when the applied stress reaches the yield strength of the material. However several metallic engineering structures, such as ships, pressure vessels and other mechanical components have failed from *brittle fracture*, which takes place without prior plastic deformation, at applied stresses well below the yield strength of the material. Brittle fracture is not so common as other types of failures of metallic structural components, such as fatigue or wear. However when brittle fracture takes place, usually the consequences are larger. The first reported incident of brittle fracture was the burst of a large molasses tank in Boston in 1919, killing 21 and injuring

150 people. From the investigation report at that time it is clear that the phenomenon of brittle fracture was not anticipated. Failures from brittle fracture increased during the 2nd world war, with the fracture of hundreds of ships (Liberty ships), several of which broke in two. After the war, the first regulations concerning the quality of steel, used in metallic constructions, were published and the scientific field of *Fracture Mechanics* started to evolve. Fracture mechanics provides the necessary tools to design metallic structures against the risk of brittle fracture. The basic elements of this field will be discussed in this chapter. However it is quite useful to see the common characteristics of some of the most important structural disasters attributed to brittle fracture. The failures, discussed below, took place between 1960 and 1980:

- Collapse of Kings Bridge in Melbourne Australia, July 1962. Cracks initiated from the heat-affected zone (*HAZ*) of fillet welds. The material of construction was a high strength steel according to an old 1941 specification, which did not require any fracture toughness testing.
- Collapse of a pressure vessel at the Chinon nuclear power station, France, 1962. Cracks initiated at the *HAZ* of a weld and propagated to the base metal, the ductility of which was deteriorated by a prolonged post weld heat treatment.
- Collapse of a large ammonia pressure vessel at Immingham, UK, December 1965, during a hydrotest. Fracture initiated from three cracks in the *HAZ* of a weld. As in the previous case, the vessel was not subjected to the proper post weld heat treatment.
- Collapse of the Sea Gem oil rig in North Sea, December 1965, with 19 victims. The brittle fracture initiated from a crack at a repair weld, while the steel used exhibited low toughness.
- Collapse of a NASA engine casing during hydrotest, 1965. Fracture was caused by a pre-existing crack at the *HAZ* of a weld, which was undetected, since no non-destructive inspection was performed between the heat treatment and the hydrotest. The length of the pre-existing crack was 35mm.
- Collapse of the Point Pleasant bridge at West Virginia, USA in December 1967, killing 45. The fracture initiated at a pre-existing crack at the *HAZ* of a weld. The crack propagated due to stress corrosion. The steel used for construction exhibited a low toughness at the temperature of the day of the incident. The crack causing the damage had a length of only 0.3mm.
- Collapse of the Alexander Kieland oil rig in Norway, March 1980, killing 123. The failure was attributed to a pre-existing crack in a weld of a hydophone. The crack propagated with fatigue. Several other cracks were detected, one of which had a length of 70mm. The steel used for construction exhibited low toughness.
- Collapse of a large crude oil storage tank in Saskatchewan, Canada, in December 1980. The fracture initiated from a crack at a secondary man-hole weld. The yield strength and the ultimate tensile strength of the steel were within specifications. However the impact test results indicated that the Ductile-to-Brittle transition temperature of the steel was too high

(10 – 15°C) while the temperature at the day of the incident was –27°C.

The failures described above share two common characteristics. The first is that all failures initiated from pre-existing cracks and the second is that the steel used for construction exhibited low fracture toughness.

Pre-existing cracks are the result of the construction process, especially welding procedures. It is known that welding is frequently accompanied by the formation of defects in the form of porosity, inclusions or even cracking in the weld metal and the HAZ. These defects and especially the crack-like defects play an important role in the fracture behavior of welded structures. It is therefore necessary to detect such flaws before the structure is put to service. For this purpose, several non-destructive testing (*NDT*) techniques have been developed, such as radiographic or ultrasonic testing. Another additional point of interest, indicated by the failures discussed above, is that cracks can grow by fatigue or stress corrosion before final fracture takes place. The second common characteristic of the failures discussed above is the low fracture toughness of the material of construction. In most cases (Kings Bridge and Point Pleasant Bridge, Alexander Kieland and Sea Gem oil rigs) the steel had inadequate fracture toughness. In other cases (Chinon and Immingham pressure vessels) the fracture toughness had been degraded by the application of improper post weld heat treatment. As will be discussed below the fracture toughness is a materials property and as such it is influenced by microstructure and processing. At the same time, especially in structural steels, the fracture toughness is reduced at low temperatures. It is not strange that most of the failures, discussed above, took place in wintertime. In general, the metals of the *BCC* system, including structural steels, exhibit a ductile-to-brittle transition with a drop in temperature, an issue that will be discussed in the following sections. From the presentation of several cases of failure above, it is apparent that the most important parameters influencing brittle fracture are: (a) loading, (b) pre-existing cracks and (c) fracture toughness. The correlation between these parameters is the subject of Fracture Mechanics, elements of which will also be discussed in the following section.

9.2.2 DUCTILE AND BRITTLE FRACTURE

Mechanisms of ductile and brittle fracture. There is a considerable difference between the two types of fracture. Brittle fracture takes place by *cleavage* along specific crystallographic planes, by breaking atomic bonds, e.g., the {100} planes in iron. Cleavage results in a flat fracture in the scale of a grain. In a polycrystalline material, however, the neighboring grains have different orientations and the advancing crack changes direction from grain to grain as it propagates along the cleavage planes. This results in *transgranular fracture*, which corresponds to a multifaceted fracture surface (Figure 9.1). A fresh fracture surface is usually shiny due to the reflection of light from the cleavage planes. The crack can also propagate along the grain boundaries of the metal resulting in *intergranular fracture*, as shown in Figure 9.2. Intergranular fracture is usually caused by the weakening of grain boundaries due to the segregation of certain elements, such as phosphorous and sulfur in steels.

Carbide precipitation at the grain boundaries, as in the sensitization of austenitic stainless steels, can also cause brittle intergranular fracture. Finally creep fracture at high temperatures can also be intergranular.

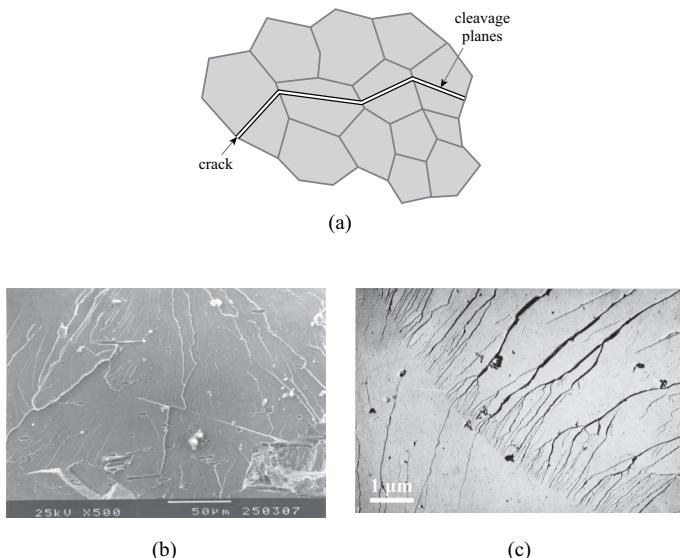


Figure 9.1: Transgranular cleavage fracture: (a) crack path, (b) and (c) fracture surfaces in the scanning electron microscope.

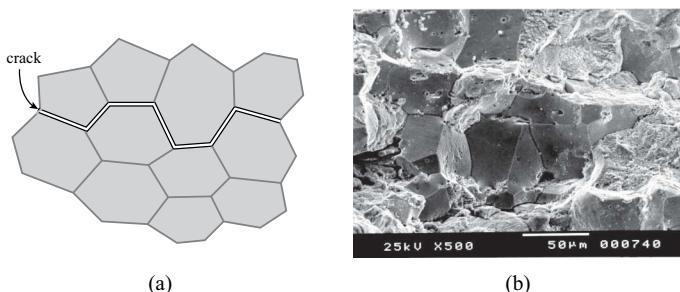


Figure 9.2: Intergranular fracture: (a) crack path, (b) fracture surface in the scanning electron microscope.

The mechanism of ductile fracture is quite different, since it involves considerable plastic deformation. All metallic alloys contain particle dispersions, either precipitates in the form of carbides, nitrides or other intermetallic compounds or inclusions in the form of oxides and sulfides. During plastic deformation either the fracture or decohesion of these particles from the surrounding matrix takes place. In both cases plastic deformation causes nucleation of microvoids, which with further plastic de-

formation grow and coalesce, forming a crack. Thus the mechanism of ductile fracture is the *nucleation, growth and coalescence of microvoids* formed around particles and inclusions during plastic deformation. The mechanism is depicted schematically in Figure 9.3. A typical ductile fracture surface is shown in Figure 9.4. The fracture surface is characterized by dimples, usually containing the particles responsible for the microvoid formation. The fracture surface is not shiny due to the scatter of light on the microvoids.

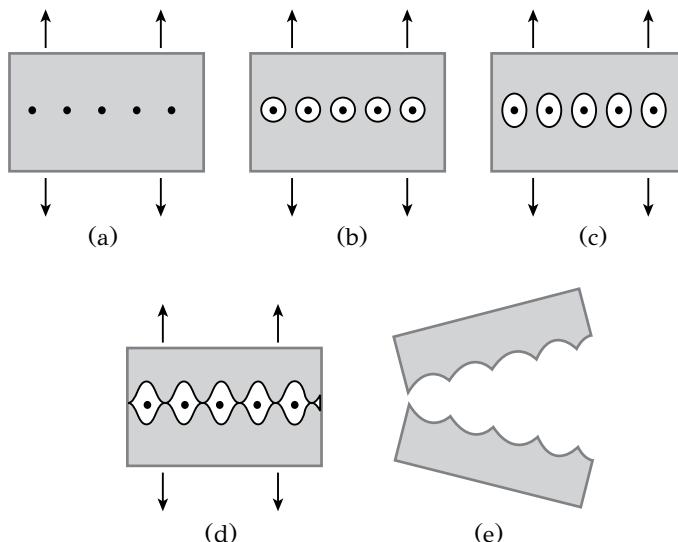


Figure 9.3: Mechanism of ductile fracture: (a) particle dispersion, (b) microvoid nucleation, (c) microvoid growth, (d) microvoid coalescence and (e) fracture.

The examination of the fracture surface either with the naked eye, a magnifying glass or the electron microscope can provide valuable information on the type of fracture. The study of fracture surfaces is called *fractography* and is one of the main procedures of a *failure analysis* investigation of structural components.

Ductile-to-Brittle transition. Metallic structures usually contain manufacturing defects such as notches, sharp corners, scratches and fillets causing stress concentration. These stress raisers can, under certain circumstances cause brittle fracture. The effort to correlate material behavior, either brittle or ductile, with the presence of notches started after the massive failures of the Liberty ships, discussed in the previous section. Several tests were developed for this purpose, the most important one is the Charpy-V notch impact test (*CVN* test). The test apparatus consists of a pendulum of known mass and length, which is dropped from a given height to impact a notched specimen of the material. The test and the notched specimen are standardized (ASTM E23). The energy absorbed for fracture is measured as a function of test temperature. Details of the test are shown in Figure 9.5, while typical results of the *CVN* test are depicted in Figure 9.6. High values of impact energy correspond

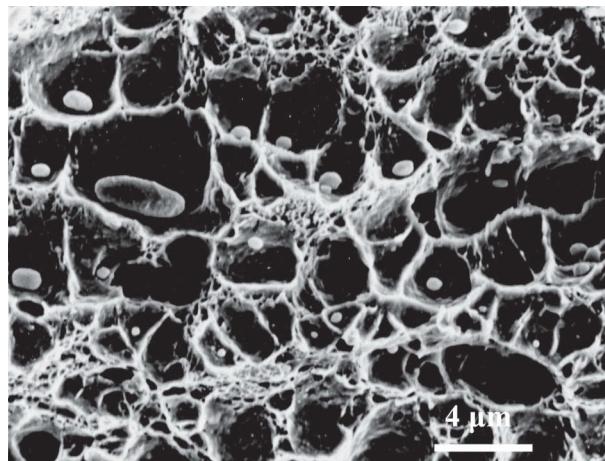


Figure 9.4: Ductile fracture surface in the scanning electron microscope with dimples and particles responsible for microvoid nucleation.

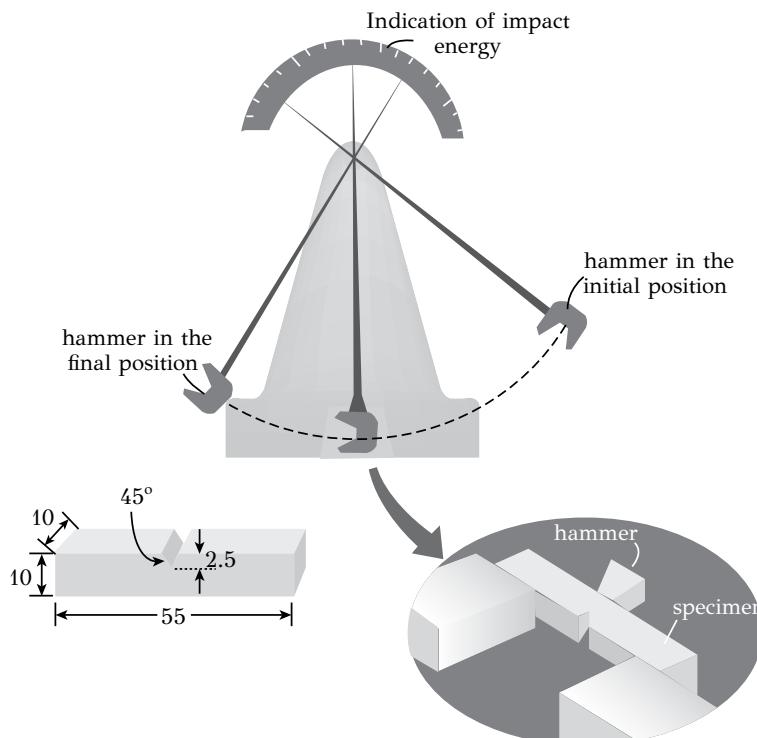


Figure 9.5: A schematic of the Charpy impact test with specimen details.

to ductile fracture while low values correspond to brittle fracture. Medium and low-strength *FCC* metals as well as most *HCP* metals exhibit high values of impact energy (curve 1). High-strength alloys, such as ultrahigh-strength steels and high-strength aluminum and titanium alloys, show low values of impact energy (curve 2). These materials fracture in a brittle manner, at low stresses in the elastic range, in the presence of notches. Structural steels as well as many *BCC* metals undergo a ductile-to-brittle transition with a drop in temperature (curve 3). The *ductile-to-brittle transition temperature (DBTT)* is used as a measure of the notch toughness or impact strength of the material. The lower the *DBTT* the higher the notch toughness of the material. There are several ways to define the *DBTT*. It can be defined as the temperature corresponding to a given value of impact energy (e.g., 20 J) or the temperature at which the fracture surface appears to be 50% ductile and 50% brittle. In order to avoid brittle fracture, several construction codes require that the material exhibits an impact energy above a certain value or the material should be used only above the *DBTT*.

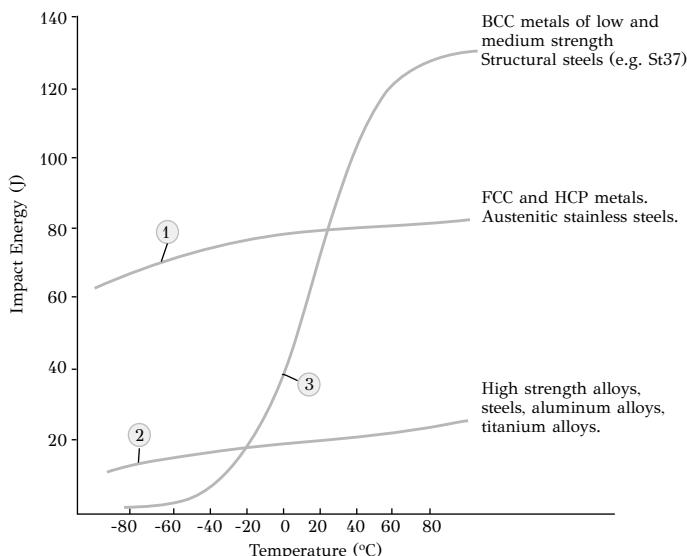


Figure 9.6: Typical curves of impact energy vs. temperature for several metallic alloys.

Mechanical and metallurgical factors influence *DBTT* and brittle fracture. The most important mechanical factor is the presence of notches. Notches act as stress concentration points and at the same time generate a triaxial stress field. Stress triaxiality imposes a *plastic constraint*, since the stress required to cause plastic deformation in a material with a notch is higher than in a material without a notch. Thus, with an increase of stress, the material may fail by brittle fracture before it is plastically deformed. Plastic constraint is more intense in plates with high thickness since the stress state is plane strain. This might explain the reason why most brittle fractures

were observed in heavy sections. Stress concentration causes corresponding concentration of plastic strain with associated local strain hardening of the material, leading to microcrack formation and brittle fracture. Finally, equally important is the strain rate. An increase in strain rate makes the material more prone to brittle fracture due to the associated increase of flow stress.

The most important metallurgical factors are chemical composition, grain size, the distribution and morphology of second phases and inclusions, all of which are developed by the mechanical and thermal treatments of the alloy. The discussion will be focused to structural steels, which exhibit a *DBTT* (curve 3 in Figure 9.6). Carbon raises and manganese lowers the *DBTT*. This is the reason that in most specifications the $Mn : C$ ratio is above 3. Phosphorous and sulfur also increase *DBTT* and the presence of these elements should be limited. The addition of manganese for desulfurization causes the formation of MnS inclusions, thus the $Mn : S$ ratio should be high. The addition of nickel lowers *DBTT*. Cryogenic steels used at extremely low temperatures contain up to 9% nickel. Oxygen and the associated deoxidation method influence *DBTT*. Rimmed steels exhibit high *DBTT*, even above room temperature, due to the significant amount of oxides they contain. Semi-killed steels, deoxidized by silicon, exhibit lower *DBTT* than rimmed steels. Finally, fully-killed steels, deoxidized with both silicon and aluminum, exhibit *DBTT* below $-50^\circ C$. Grain size has also an important effect. Grain size refinement has a beneficial effect in lowering *DBTT*. This has led to the development of high-strength low-alloy (*HSLA*) steels, which contain small quantities of carbide and carbonitride-forming elements such as *Ti* and *Nb*. These carbonitrides precipitate during the thermomechanical treatment of the steel and cause grain boundary pinning, thus retarding recrystallization and grain growth. The result is an excellent combination of strength and toughness.

Case study: How did the notch toughness of steel contribute to the Titanic disaster?

RMS Titanic sank in the North Atlantic Ocean on 15 April 1912 after colliding with an iceberg during her maiden voyage from Southampton to New York City. In 1985, 73 years later, Robert Ballard from the Woods Hole Oceanographic Institute, discovered the ship. This gave the opportunity to investigate the steel used for the construction of the Titanic. The results of this investigation, by Felkins, Leighly and Jankovic¹, indicated that the material exhibited similar yield strength, ultimate tensile strength and tensile ductility as A36 steel, which is used in today's modern shipbuilding. However the Titanic steel was inferior to notch toughness and exhibited a considerably higher *DBTT* than A36 steel, $32^\circ C$ for the Titanic steel compared with $-27^\circ C$ for A36 steel. This difference in *DBTT* was caused by the lower $Mn : S$ ratio in the Titanic steel, which was only 6.8 compared with 14.9 for A36 steel. The reduced notch toughness caused extended cracking of the shell, after the collision, leading to flooding of several compartments of the ship.

¹K. Felkins, H.P. Leighly, A. Jankovic, J. Metals 50 (1998) 12.

9.2.3 ELEMENTS OF FRACTURE MECHANICS

The analysis of brittle fractures, presented in section 9.2.1, indicated that the most important factors contributing to brittle fracture are the applied stresses, pre-existing cracks and the fracture toughness of the material. Fracture mechanics provides a systematic methodology with which these three factors can be correlated in engineering design. Thus today's engineer has the necessary tools to define allowable stress levels or required non-destructive testing for the detection of crack-like flaws for the design of safe metallic structures. Fracture mechanics formulates the condition for propagation of an existing crack and is based on the analysis of the stress state ahead of the crack tip (Figure 9.7). The fundamental concept is that the stress field can be

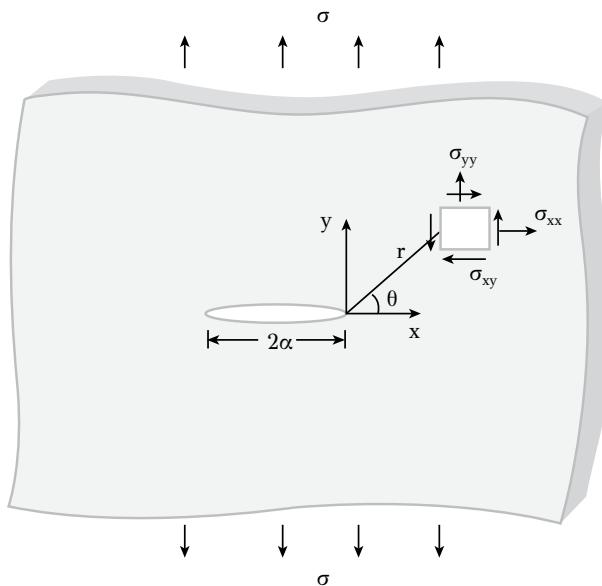


Figure 9.7: The components of the stress field ahead of a crack tip in a plate.

described by a single parameter K_I , the *stress intensity factor*, which depends on the applied stresses and crack length. For the through-thickness crack in an infinite plate of Figure 9.7, the stress intensity factor is given by

$$K_I = \sigma \sqrt{\pi \alpha} \quad (9.1)$$

The stress intensity factor has values $MPa\sqrt{m}$. The suffix I in K_I refers to the mode – I loading or opening mode. Mode – II and mode – III correspond to in-plane and out-of-plane shear respectively and are shown in Figure 9.8. Since mode – I is the most common, the discussion will be limited to the opening mode of fracture.

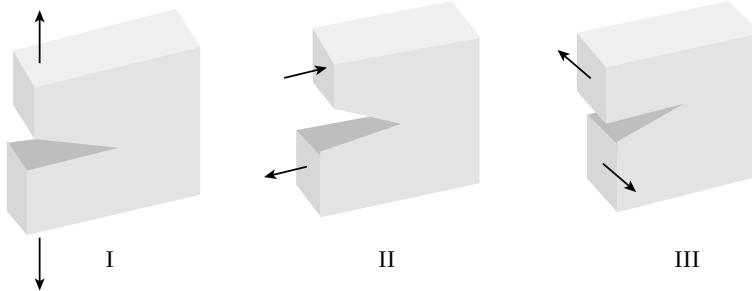


Figure 9.8: The three modes of crack loading: (I) tensile or opening mode, (II) in-plane shear and (III) out-of-plane shear.

The stresses at the tip of the crack of Figure 9.7 are given by

$$\begin{aligned}\sigma_{xx} &= \frac{K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left[1 - \sin \frac{\theta}{2} \sin \frac{3\theta}{2} \right] \\ \sigma_{yy} &= \frac{K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left[1 + \sin \frac{\theta}{2} \sin \frac{3\theta}{2} \right] \\ \sigma_{xy} &= \frac{K_I}{\sqrt{2\pi r}} \sin \frac{\theta}{2} \cos \frac{\theta}{2} \cos \frac{3\theta}{2}\end{aligned}\quad (9.2)$$

In the equations above, the *rhs* corresponds to the first term of an infinite series, which describes the stress field close to the crack tip, at low values of r . The stress $\sigma_{zz} = 0$ for plane stress while $\sigma_{zz} = v(\sigma_{xx} + \sigma_{yy})$ for plane strain. Equations (9.2) indicate that the stress field is characterized by the stress intensity factor, which acts as the driving force for crack advance. In fact the stress intensity factor plays a similar role to the stress in classical engineering design, where the failure criterion for plastic deformation in uniaxial tension (yield criterion) is simply

$$\sigma = \sigma_y \quad (9.3)$$

The respective fracture criterion can then be written as

$$K_I = \sigma \sqrt{\pi \alpha} = K_{IC} \quad (9.4)$$

according to which a pre-existing crack of length α will propagate and cause fracture when the stress intensity factor obtains the critical value K_{IC} . The K_{IC} operates in a similar fashion as the yield strength σ_y operates in the yield criterion of Equation (9.3). It is a material property and is called the fracture toughness. It has units of $MPa\sqrt{m}$ and is determined by a certain standardized test (ASTM E399). The importance of the fracture criterion is expressed by the plot of Equation (9.4) in Figure 9.9. Curve 1 corresponds to Equation (9.4) for a given fracture toughness K_{IC} . Point A corresponds to loading with a stress σ_A in a plate containing a through-thickness crack of length $2\alpha_0$. In this case

$$K_I^A < K_{IC} \quad (9.5)$$

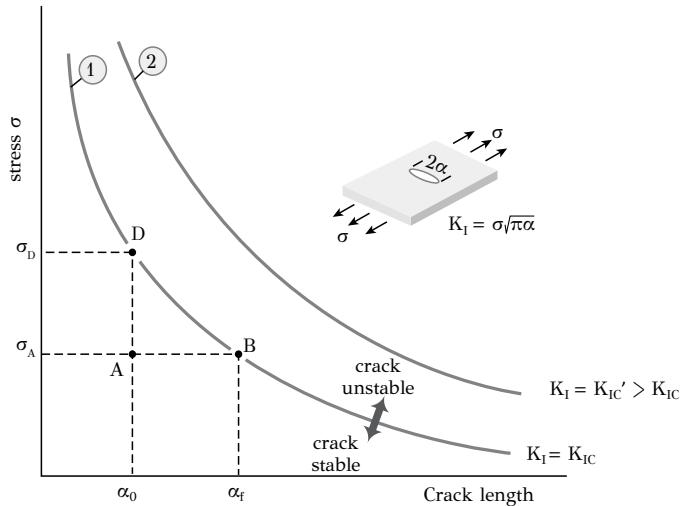


Figure 9.9: Plot of stress vs. crack length, indicating the fracture criterion.

and the crack is stable since the fracture criterion is not satisfied. Consider now that loading is increased to σ_D , then

$$K_I^D = K_{IC} \quad (9.6)$$

and the crack will advance since the fracture criterion is satisfied. In this case fracture could have been avoided if a material with a higher fracture toughness K'_{IC} was used instead, corresponding to curve 2 in Figure 9.9. Another scenario is the progressive increase of crack length from α_0 (point A) to α_f (point B), which can take place by either fatigue or stress corrosion cracking. During fatigue the material is subjected to cyclic loading and the crack advances by $d\alpha$ in every cycle. During stress corrosion cracking, the material is loaded in a corrosive environment and the crack advances with time. In both cases, fracture will take place when the crack reaches the critical size α_f . In these cases the use of a material with higher fracture toughness K'_{IC} would delay fracture, providing a life extension. Another way to state this is to say that a material of higher toughness exhibits a higher tolerance for larger cracks. This is the concept on which the *damage tolerance* design is based. Engineering design, based on damage tolerance, is applied in several structures, such as pressure vessels and airplanes. They are designed to tolerate the presence of cracks, the growth of which is monitored by non-destructive techniques.

The general form of Equation (9.1), which defines the stress intensity factor is

$$K_I = Q\sigma\sqrt{\pi\alpha} \quad (9.7)$$

where Q is a geometrical factor, which depends on geometry and crack loading conditions. For example $Q = 1.12$ for loading of an edge crack. Details for stress intensity equations are provided in the books at the end of this chapter (especially the book by Tada, Paris and Irwin).

Equations (9.2) are based on the assumption that the material behavior is elastic. This means that the validity of the stress intensity factor K_I for the description of the stress field ahead of a crack as well as the validity of the fracture criterion are limited by the plastic deformation, which takes place at the crack tip. The stress distribution, according to (9.2), is shown in Figure 9.10. When the stress exceeds the

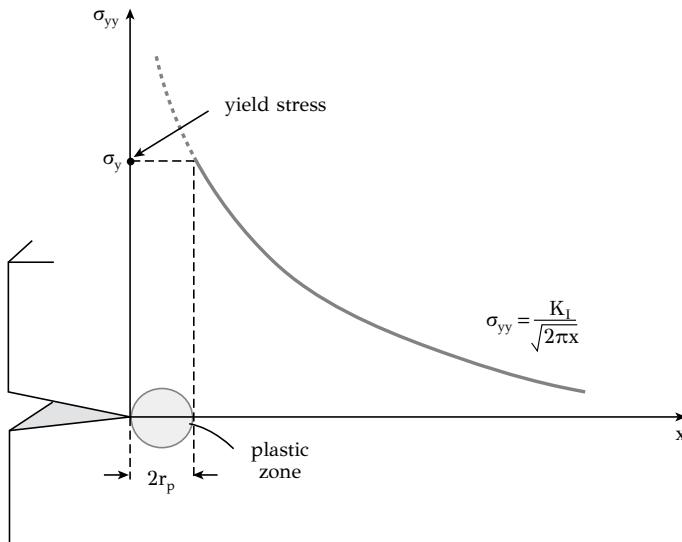


Figure 9.10: The plastic zone ahead of the crack tip.

yield strength, a plastic zone is formed ahead of the crack tip. The size of the plastic zone r_p is given by

$$r_p = \frac{1}{2\pi} \left(\frac{K_I}{\sigma_y} \right)^2 \quad (9.8)$$

The stress intensity factor can be used to describe the stress field ahead of the crack with an acceptable accuracy only outside the plastic zone. Setting $K_I = K_{IC}$ in (9.8) the plastic zone size becomes

$$r_p = \frac{1}{2\pi} \left(\frac{K_{IC}}{\sigma_y} \right)^2 \quad (9.9)$$

The above relation, which holds for plane stress conditions, indicates that the plastic zone size increases when the fracture toughness increases and the yield strength decreases. In soft and ductile materials (small σ_y and large K_{IC}) the plastic zone becomes very large, limiting the validity of K_I . In other words, the validity of fracture mechanics is greater in brittle than in ductile materials. In order for K_I and the fracture criterion to be valid, plasticity should be limited. The plastic zone size should

be considerably smaller than the specimen and crack dimensions (Figure 9.11), i.e.,

$$\begin{aligned} r_p &\leq \frac{1}{15}(\alpha, W - \alpha) \\ r_p &\leq \frac{1}{15}B \end{aligned} \quad (9.10)$$

The above relations correspond to the conditions of *small scale yielding*. An ad-

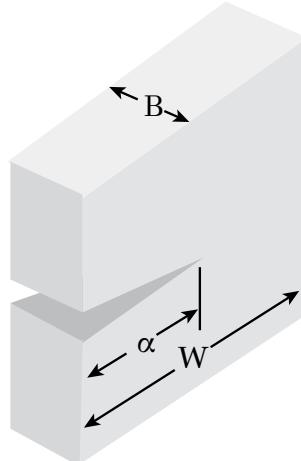


Figure 9.11: Characteristic dimensions of a specimen containing a crack.

ditional requirement is that the stress state should correspond to plane strain. This means that the crack tip will be situated in a triaxial stress field where shear stresses are significantly lower than normal stresses. The plane strain condition can be satisfied by increasing the specimen thickness. In the relevant standard for fracture toughness testing, the specimen thickness B should satisfy the condition

$$B \geq 2.5 \left(\frac{K_{IC}}{\sigma_y} \right)^2 \quad (9.11)$$

in accordance with the second Relation (9.10). It is apparent that the critical stress intensity factor K_C is a function of specimen thickness, as depicted in Figure 9.12. K_C decreases with the increase of B approaching a value K_{IC} , which is independent of thickness, corresponding to plane strain conditions. The limiting K_{IC} value is the *fracture toughness* of the material. The increase of thickness induces a corresponding change in the orientation of the crack plane, as indicated in Figure 9.12. When the thickness is small, plane stress conditions cause the crack to be inclined about 45° to the loading axis. At larger thickness the fracture surface exhibits a central flat area with shear lips at the edges. When the thickness is large enough for plane strain conditions to be established, no shear lips form and the fracture surface is macroscopically flat.

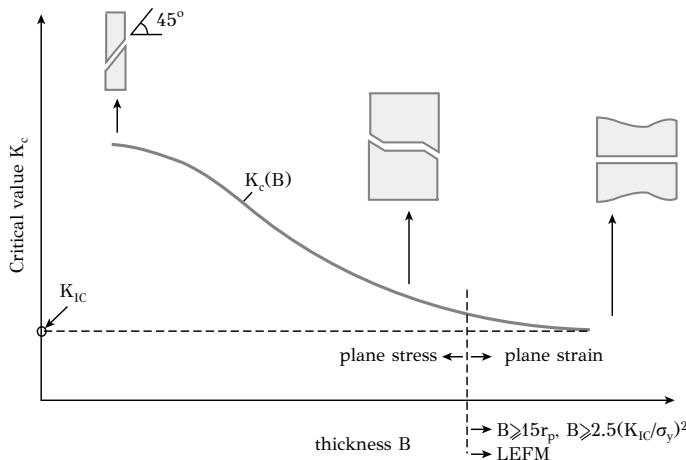


Figure 9.12: The effect of thickness on the critical stress intensity factor.

The methodology described above, together with the associated limitations, constitute the theory of Linear Elastic Fracture Mechanics (*LEFM*). The general concept of *LEFM* is that when the plastic zone is small, relative to the specimen and crack dimensions and when the specimen thickness is large for plane strain conditions to be established, then K_I can describe the stress field ahead of the crack tip. In addition the fracture criterion is valid and can be used for the design of structures against brittle fracture. In the case of a very soft or very ductile material however, it is possible that the requirements of the linear theory are not satisfied. In this case the theory of Elastic-Plastic Fracture Mechanics (*EPFM*) has been developed, with tools such as the J -integral and crack-tip opening displacement (*CTOD*) used to characterize fracture toughness. The description of *EPFM* is beyond the scope of this book. The interested reader is advised to see the reading list at the end of this chapter.

9.3 FATIGUE

9.3.1 INTRODUCTION

A mechanical component or structural element subjected to static or monotonic loading is safe against brittle fracture once the load is under a certain critical value, which depends, amongst others, on the magnitude of the load and the fracture toughness of the alloy. However when the same component is subjected to cyclic loading, it would fail at loads, which are much lower than those required for fracture under monotonic loading. In this case, failure is attributed to *metal fatigue*. There are two major attributes in fatigue. The first corresponds to the type of loading, which is cyclic. The material is subjected to an alternating stress between a minimum and a maximum value. The second corresponds to the damage, accumulated in the material, during fatigue. Damage is a general term that may be associated to plastic deformation, nu-

cleation of microcracks as well as the formation and propagation of a major crack. Fatigue causes local and progressive accumulation of damage in the material, which is characterized by the increase of crack length. The final fast fracture takes place when the fracture criterion is satisfied, i.e., when the crack length reaches the critical size. While in monotonic loading the fracture criterion is satisfied with the increase of stress, in fatigue, the fracture criterion is satisfied with the increase of crack length. While, in monotonic loading, fracture is instantaneous, in fatigue, the process takes time, for the crack to grow and reach the critical size for fracture. Fatigue damage is then characterized by the *fatigue life* N_f , measured in number of loading cycles.

Two periods can be distinguished in fatigue life. The first period corresponds to the number of cycles N_i , which are required for the formation or initiation of a crack. The second period corresponds to the number of cycles N_p required for the propagation of the crack to the critical size. The fatigue life is then

$$N_f = N_i + N_p \quad (9.12)$$

In general, there are two methodologies for the analysis and design of structural components against fatigue failure. In large welded structures, the assumption is made that the structure contains crack-like defects with size at least equal to the detection limit of the non-destructive tests (*NDT*), employed during construction. In this case $N_i=0$ and the fatigue life is equal to the number of cycles required for the propagation of the pre-existing crack to critical size. These structures are then designed according to the damage tolerance framework. The second methodology is concerned with mechanical components and machine elements, such as shafts, gears or bearings, where it is accepted that they do not contain any pre-existing crack-like defects. In these cases N_i can be a large fraction of N_f . The value of N_i increases with a decrease of applied cyclic loads. In addition N_i is very sensitive to the state or quality of the surface. While in smooth specimens, N_i can take up to 80% of N_f , in specimens containing notches or other stress raisers and corroded areas, N_i can be only 10% of N_f .

Several types of fatigue are encountered in engineering. *Mechanical fatigue* is associated with a material subjected to a cyclic stress or strain. When at the same time the material is exposed at high temperatures, then the behavior is characterized as *creep-fatigue*. When the material is subjected, concurrently, to both cyclic loading and cyclic heating or cooling, the behavior is characterized as *thermo-mechanical fatigue*. When cyclic loading takes place in a corrosive environment, then it is characterized as *corrosion fatigue*. Cyclic loading with simultaneous sliding or rolling between two metallic surfaces results in *sliding contact fatigue* or *rolling contact fatigue* respectively. Finally cyclic loading with simultaneous relative oscillatory sliding friction between two metallic surfaces causes *fretting fatigue*. The present chapter deals primarily with mechanical fatigue. For the other types of fatigue the interested reader is advised to consult the suggested reading list at the end of this chapter.

9.3.2 CYCLIC BEHAVIOR

Consider the cyclic loading depicted in Figure 9.13. The material is loaded under strain control between ε_{max} and ε_{min} with amplitude ε_α (Figure 9.13a). The mate-

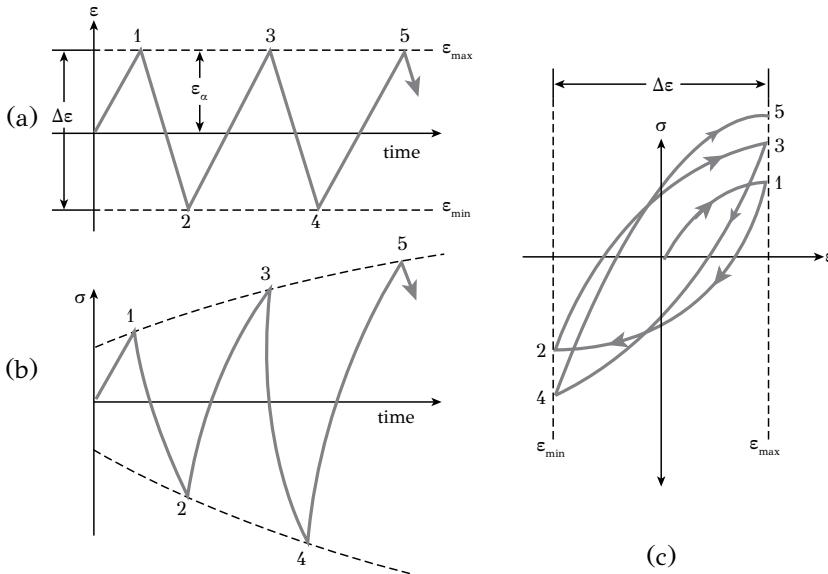


Figure 9.13: Cyclic hardening: (a) cyclic loading with strain range $\Delta\varepsilon$, (b) stress response, (c) hysteresis loops for cyclic hardening.

rial response is depicted in Figure 9.13b with the corresponding fluctuation of stress. The values of stresses and strains are combined in a single plot in Figure 9.13c, depicting the cyclic $\sigma - \varepsilon$ behavior. The stress required to maintain the given strain amplitude increases in every load reversal and the material exhibits *cyclic hardening*. This behavior is typical of metals in the annealed condition. The behavior of a material exhibiting *cyclic softening* is depicted in Figure 9.14. The stress required to maintain the given strain amplitude decreases at every load reversal. The behavior is typical of metals in the cold worked condition. Both cyclic hardening and cyclic softening are transient behaviors. After a certain number of load reversals, corresponding to 20–40% of fatigue life, the stress amplitude is stabilized, as shown in Figure 9.15. The corresponding hysteresis loop is also stabilized. The stable hysteresis loop is shown in Figure 9.16 for a strain range $\Delta\varepsilon = 2\varepsilon_\alpha$. The respective stress range is $\Delta\sigma = 2\sigma_\alpha$. The *cyclic stress-strain curve* can be constructed by connecting the tips of a family of stabilized hysteresis loops, which are obtained from different strain amplitudes. The procedure is shown in Figure 9.17 for strain amplitudes ε_1 , ε_2 and ε_3 . The cyclic stress-strain curve is compared to the respective monotonic stress-strain behavior in Figure 9.18. Cyclic softening exists when the cyclic curve is below the monotonic curve, as in Figure 9.18a. In this case, using monotonic proper-

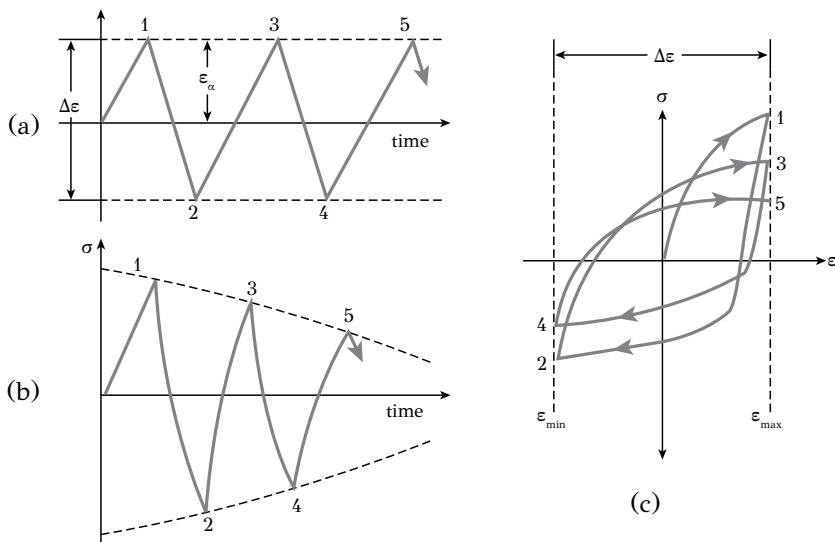


Figure 9.14: Cyclic softening: (a) cyclic loading with strain range $\Delta\epsilon$, (b) stress response, (c) hysteresis loops for cyclic softening.

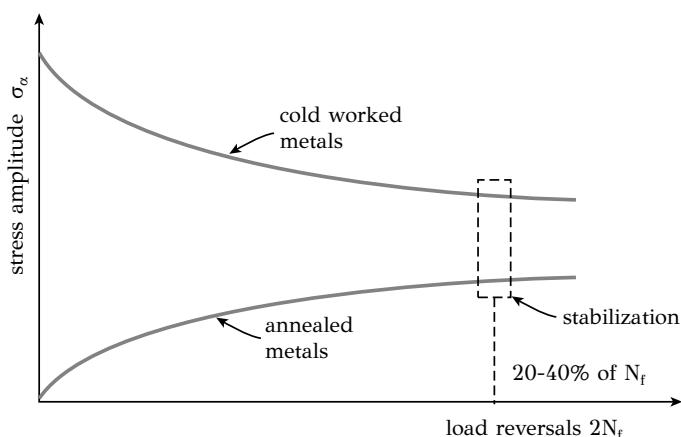


Figure 9.15: Stabilization of cyclic behavior after 20-40% of fatigue life.

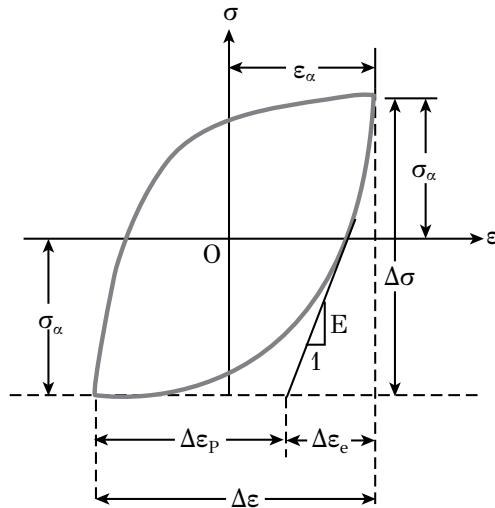


Figure 9.16: Stable hysteresis loop, indicating stable cyclic behavior.

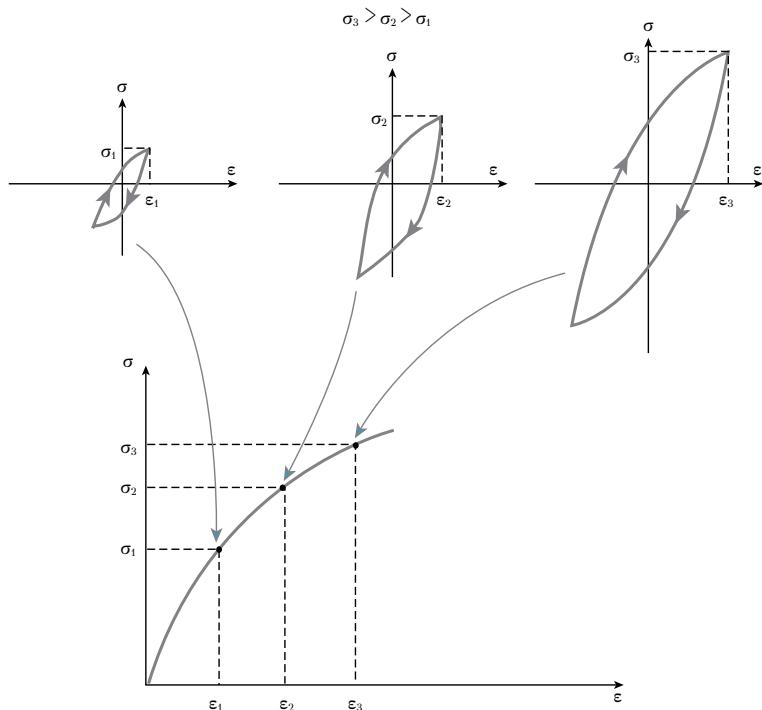


Figure 9.17: Construction of the cyclic stress-strain curve from the stable hysteresis loops.

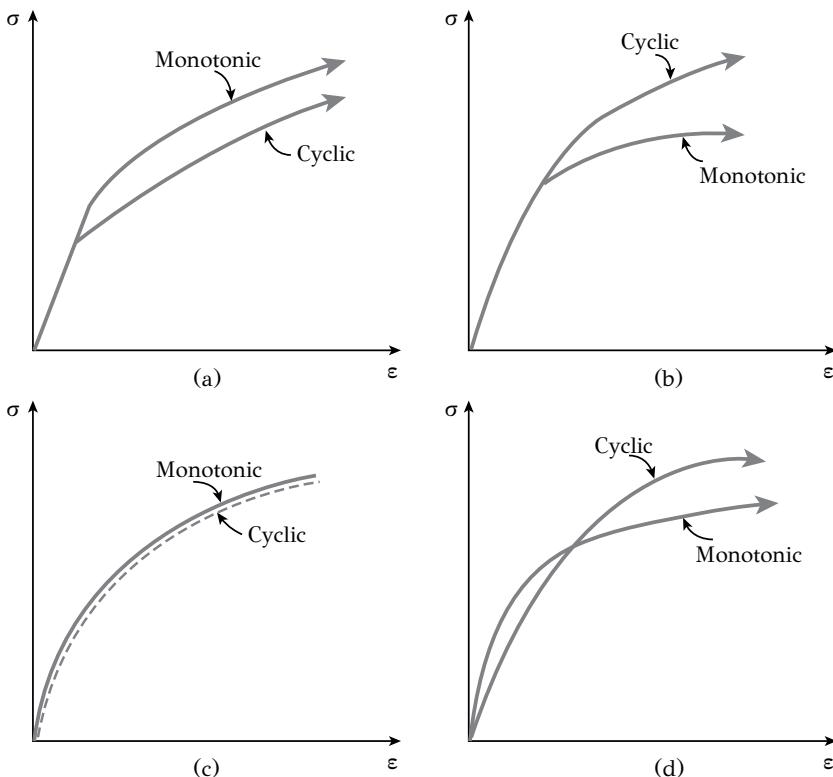


Figure 9.18: Comparison of monotonic and cyclic behavior: (a) cyclic softening, (b) cyclic hardening, (c) stable behavior, (d) mixed behavior.

ties, in an application involving cyclic loading, may underestimate the magnitude of plastic strain. The opposite is true for cyclic hardening, depicted in Figure 9.18b. A cyclic stable behavior is depicted in Figure 9.18c, while a mixed behavior is shown in Figure 9.18d. In general, metals with a high monotonic strain-hardening exponent ($n > 0.15$) exhibit cyclic hardening while metals with a low strain-hardening exponent ($n < 0.15$) exhibit cyclic softening. Metals, which are initially soft, exhibit cyclic hardening. In these metals the dislocation density increases rapidly with cyclic loading causing strain hardening. A stable dislocation configuration is formed after a certain number of load reversals, corresponding to the stabilization of the hysteresis loop, discussed above. The opposite is true for metals, which are initially hard. In those metals, the initial dislocation density is high and cyclic loading causes a dislocation rearrangement, which provides less resistance to dislocation glide and results in cyclic softening. Cyclic behavior is also influenced by cross slip. The presence of a close-packed direction in BCC metals makes cross slip easier. The frequent change of slip system, known as *wavy slip*, causes a more frequent interaction between dislo-

cations and results in a more intense cyclic hardening in annealed metals and cyclic softening in cold worked metals. On the other side, cross slip is not easy in FCC metals due to the dissociation of dislocations into partials. As discussed in Chapter 3, the dissociation is accompanied by the formation of a stacking fault. The structure of the fault is *HCP* and the fault is characterized by a stacking fault energy (*SFE*). In metals with a high *SFE*, such as copper, dislocation dissociation is difficult and cross slip is easy. As a result the cyclic loading of annealed copper leads to cyclic hardening while the cyclic loading of cold worked copper leads to cyclic softening. Alloying copper with aluminum reduces the *SFE* and cross slip becomes more difficult. As a result, dislocation interactions are reduced and cyclic hardening or cyclic softening is not so intense.

9.3.3 FATIGUE WITHOUT PRE-EXISTING CRACKS – STRAIN-LIFE APPROACH.

High cycle and low cycle fatigue. In machine elements, such as shafts, gears and bearings, the assumption is made that they do not contain any pre-existing cracks and design against fatigue is performed with the *strain-life approach*. Fatigue is then distinguished in *high-cycle fatigue (HCF)* and *low-cycle fatigue (LCF)*. In *HCF* the loads are low and the fatigue life is above $10^3 - 10^4$ cycles. In this case the cyclic stresses cause only elastic strains. In *LCF*, the loads are much higher and the fatigue life is short, usually below $10^3 - 10^4$ cycles. In this case the cyclic stresses cause significant plastic strains.

Cyclic loading in *HCF* is characterized by the stress range $\Delta\sigma$, the stress amplitude σ_α , the mean stress σ_m and the stress ratio R given by

$$\Delta\sigma = \sigma_{\max} - \sigma_{\min}$$

$$\sigma_\alpha = \frac{\Delta\sigma}{2} = \frac{\sigma_{\max} - \sigma_{\min}}{2}$$

$$\sigma_m = \frac{\sigma_{\max} + \sigma_{\min}}{2}$$

$$R = \frac{\sigma_{\min}}{\sigma_{\max}}$$

Several cases of cyclic loading, for different values of the stress ratio R , are shown in Figure 9.19. The dependence of fatigue life on stress amplitude is shown in Figure 9.20 for the case of $\sigma_m = 0$ with full load reversal, as a plot of σ_α vs. $\log N_f$. These plots are called *S – N* curves or *Wohler* curves and are very helpful in defining the *endurance limit* σ_e , which is the stress amplitude, below which the fatigue life is theoretically infinite. The endurance limit is easy to define in several steels. However in high-strength steels and non-ferrous alloys it is difficult to define a distinct endurance limit. In these cases, a conventional endurance limit is defined, e.g., for a life of $N_f = 10^7$ cycles. If the data of Figure 9.20 are placed in a double logarithmic scale of $\log\sigma_\alpha$ vs. $\log 2N_f$, where $2N_f$ represents the load reversals, then the relationship is linear, as depicted in Figure 9.21. It is described by

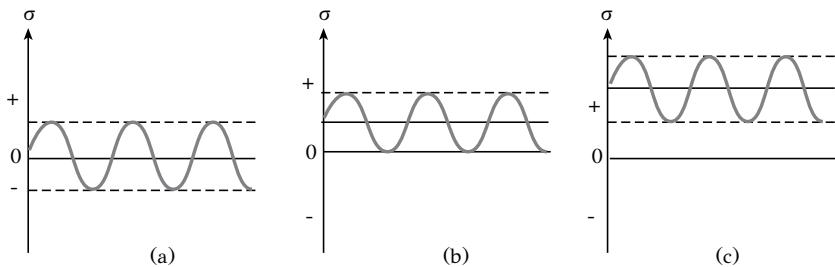


Figure 9.19: Cyclic loading : (a) $R = -1$, $\sigma_m = 0$, (b) $R = 0$, $\sigma_m > 0$, (c) $0 < R < 1$, $\sigma_m > 0$.

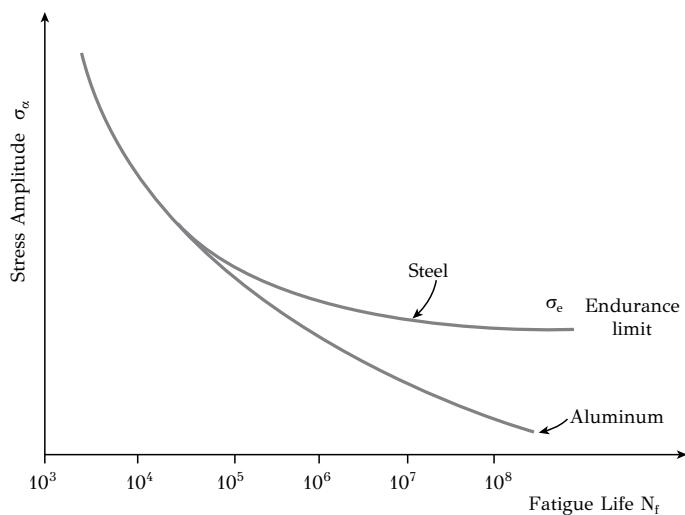


Figure 9.20: The S-N curve for steel and aluminum.

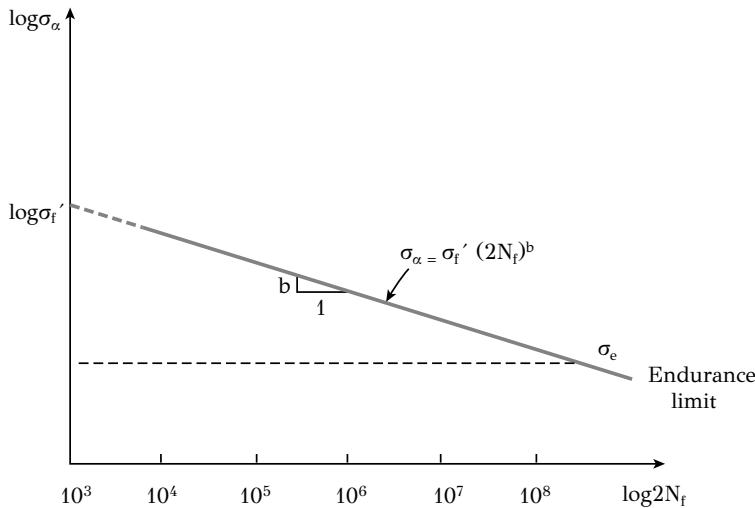


Figure 9.21: Stress amplitude vs. load reversals and the Basquin's law.

$$\sigma_\alpha = \sigma'_f (2N_f)^b \quad (9.13)$$

which is the *Basquin's law* for high cycle fatigue. σ'_f is the fatigue strength coefficient and b is the fatigue strength exponent or the Basquin exponent. The parameters σ'_f and b are material properties. In most cases $\sigma'_f = \sigma_f$ the stress at fracture in monotonic loading. The Basquin exponent b is in the range $-0.12 < b < -0.05$, with a typical value of $b = -0.1$.

Basquin's law fails to describe the low cycle fatigue (*LCF*) behavior, for lives less than $10^3 - 10^4$ cycles. In this case the fatigue life is correlated with the plastic strain amplitude

$$\frac{\Delta \varepsilon_p}{2} = \varepsilon'_f (2N_f)^c \quad (9.14)$$

which is known as the *Coffin-Manson law*, depicted in Figure 9.22. The parameter ε'_f is the fatigue ductility coefficient and c is the fatigue ductility exponent. Both are material properties. In most cases $\varepsilon'_f = \varepsilon_f$ the true strain at fracture. The exponent c is in the range $-0.7 < c < -0.5$ with a typical value $c = -0.6$.

The total strain amplitude is the sum of elastic and plastic contributions and is given by

$$\frac{\Delta \varepsilon}{2} = \varepsilon_\alpha = \frac{\sigma_\alpha}{E} + \frac{\Delta \varepsilon_p}{2}$$

Using the Basquin and Coffin-Manson relations it becomes

$$\frac{\Delta \varepsilon}{2} = \frac{\sigma'_f}{E} (2N_f)^b + \varepsilon'_f (2N_f)^c \quad (9.15)$$

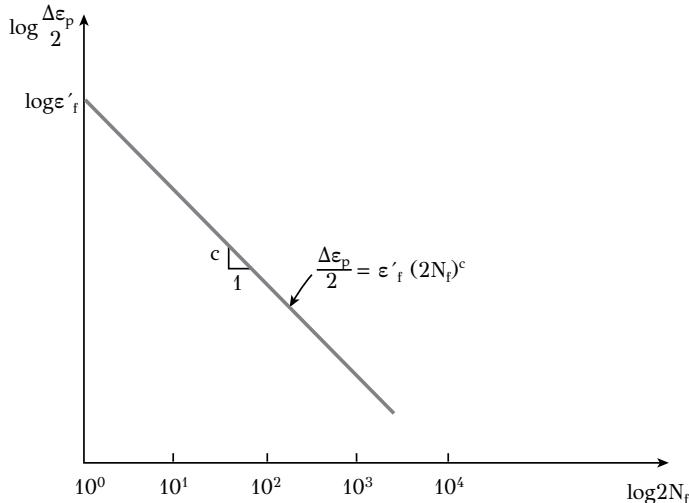


Figure 9.22: Plastic strain amplitude vs. load reversals and the Coffin-Manson law.

Relation (9.15) forms the basis of the *strain-life approach* and describes the fatigue behavior of components without pre-existing cracks, including both *HCF* and *LCF*. It is shown in Figure 9.23 as the curve, which is asymptotic to the two straight lines corresponding to *HCF* and *LCF* respectively. The point of intersection between the two lines defines the life $2N_t$, where the elastic and plastic components of strain are equal and is called the *transition fatigue life*. For lives less than $2N_t$, the deformation is mainly plastic and fatigue behavior is controlled by the ductility of the material. For lives greater than $2N_t$, the deformation is mainly elastic and fatigue behavior is controlled by the strength of the material. The balance between strength and ductility is depicted in Figure 9.24 for two steels, which exhibit different strength and ductility. The high-strength low-ductility steel performs better in *HCF* conditions. On the other side the soft, ductile steel, performs better in *LCF* conditions.

The effect of mean stress. Relation (9.15) correlates the fatigue life with strain range under conditions of zero mean stress and full load reversal ($\sigma_m = 0, R = -1$). The effect of mean stress can be significant though, especially in *HCF*, where the plastic deformations are small. It should be noted that in *LCF* the mean stress is relaxed rapidly through plastic deformation. The effect of mean stress in fatigue life can be estimated if we consider that the mean stress will reduce the fatigue strength coefficient of the Basquin Equation (9.13) so that

$$\sigma_\alpha = (\sigma'_f - \sigma_m)(2N_f)^b$$

The total strain amplitude then becomes

$$\frac{\Delta \varepsilon}{2} = \varepsilon_\alpha = \left(\frac{\sigma'_f - \sigma_m}{E} \right) (2N_f)^b + \varepsilon'_f (2N_f)^c$$

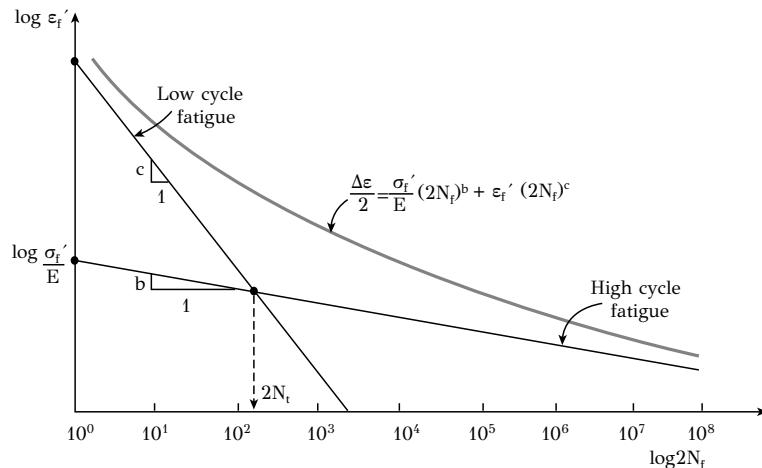


Figure 9.23: The total fatigue curve, including both Basquin's and Coffin-Manson laws corresponding to HCF and LCF.

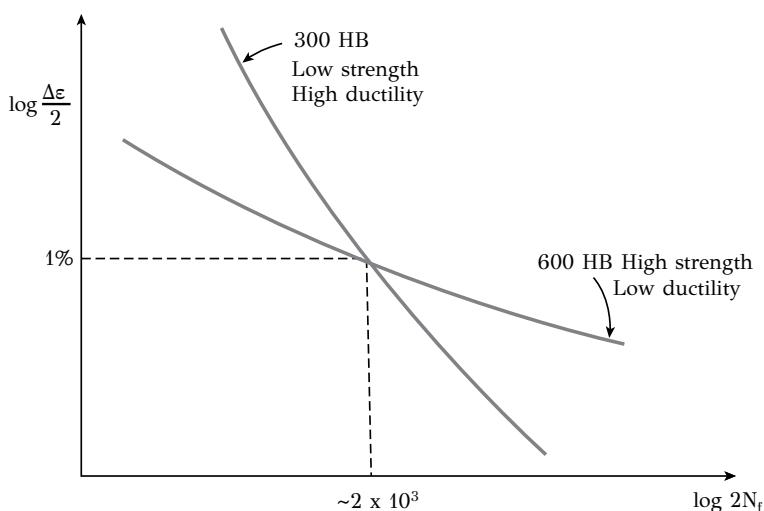


Figure 9.24: Fatigue curves of two steels with different hardness and ductility.

The effect of a tensile mean stress is depicted in Figure 9.25. Fatigue life is reduced relative to the condition $\sigma_m = 0$. On the other side, the fatigue life is increased when the mean stress is compressive. The new fatigue life is then

$$N_f = \left(1 - \frac{\sigma_m}{\sigma'_f}\right) N_{f0} \quad (9.16)$$

where N_{f0} is the fatigue life for $\sigma_m = 0$.

Variable loading amplitude - Miner's rule. The strain life equation holds for a constant strain amplitude for the entire fatigue life. However several engineering components are subjected to a loading spectrum, which involve variable stress or strain amplitude. Consider a spectrum consisting of two parts, in Figure 9.26.

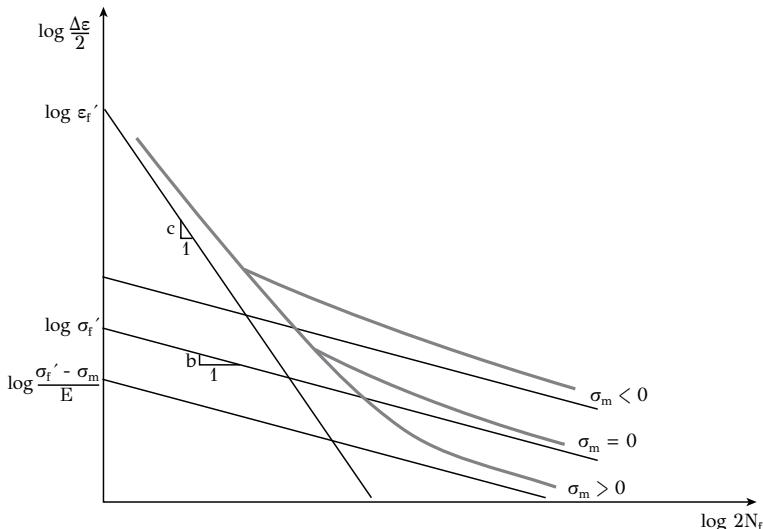


Figure 9.25: The effect of mean stress in high-cycle fatigue.

In the first part the material is loaded with a strain amplitude $\Delta\epsilon_1/2$ for n_1 cycles. In the second part the material is loaded with a higher strain amplitude $\Delta\epsilon_2/2$. We are interested in the number of cycles n_2 till failure. If N_{f1} is the fatigue life, that the material could exhibit if it was loaded only with the $\Delta\epsilon_1/2$ amplitude, then the parameter

$$d_1 = \frac{n_1}{N_{f1}}$$

expresses the damage accumulated in the material during the first part of the spectrum ($d_1 < 1$). The above equation can be generalized and to define the damage d_i accumulated in the material in part "i" of the spectrum with respective strain amplitude $\Delta\epsilon_i/2$ and life n_{fi} , then

$$d_i = \frac{n_i}{N_{fi}}, (0 \leq d_i \leq 1)$$

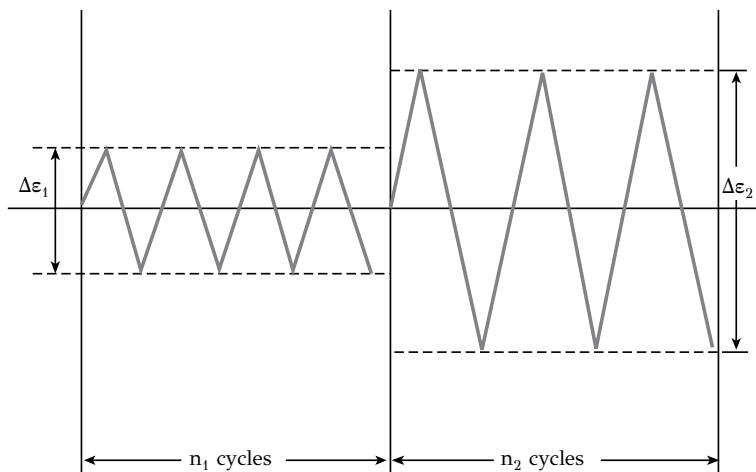


Figure 9.26: Cyclic loading with variable stress range.

The theory of damage accumulation states that failure from fatigue will take place when

$$\sum_i d_i = \sum_i \frac{n_i}{N_{fi}} = 1 \quad (9.17)$$

which is known as the *Miner's rule*. In the loading spectrum consisting of two loading parts, Equation (9.17) gives

$$\sum_{i=1}^2 d_i = \frac{n_1}{N_{f1}} + \frac{n_2}{N_{f2}} = 1$$

and solving for n_2

$$n_2 = N_{f2} \left(1 - \frac{n_1}{N_{f1}} \right)$$

which is the number of cycles in the second part until failure.

Though Miner's rule is a useful approximation, there are certain limitations, which should be considered in real loading cases. The most significant one is that the rule does not account for the loading sequence. For example if the higher strain amplitude is applied first, the damage accumulated could lead to crack initiation. Then the life spent in the second, lower amplitude part, would be higher than the prediction of Miner's rule. In addition, a higher load amplitude in the first part could cause the formation of a compressive residual stress that could retard fatigue crack growth. This would result in a higher fatigue life in the second part.

Crack initiation in fatigue. As has been discussed above, in engineering components, which do not contain any pre-existing cracks, most of fatigue life is consumed in crack initiation. Cracks usually form at discontinuities on the metal surface, at

inclusions or at scratches and notches, either produced from manufacturing or corrosion. In very smooth specimens, surface discontinuities are created in the surface of the metal, along zones of localized plastic deformation, called *persistent slip bands* (*PSB*), since they persist to form at the same location, even after their removal with electropolishing. The dislocation arrangement along the *PSBs* allows for easier dislocation glide along the bands than in the surrounding crystal lattice. Such a directional dislocation glide, generates slip steps at the surface, which, under cyclic slip, evolve to *intrusions and extrusions*, generating a localized surface roughening, as depicted in Figure 9.27. The fatigue cracks initiate either directly from these surface discontinuities or from the interaction of intrusions with inclusions and grain boundaries as depicted in Figure 9.28. Inclusions play a key role in industrial alloys while grain boundaries play a key role at high temperatures during creep fatigue.

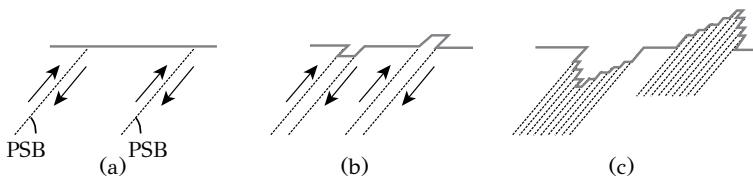


Figure 9.27: Localization of plastic deformation during fatigue: (a) Persistent slip bands (PSB), (b) slip bands, (c) intrusions and extrusions.

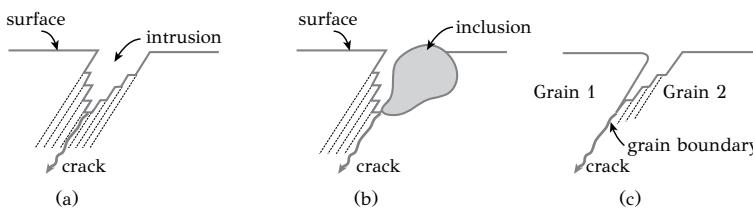


Figure 9.28: Mechanisms of fatigue crack initiation: (a) from an intrusion at the surface, (b) from an inclusion at the surface, (c) from a grain boundary.

In smooth specimens, most of fatigue life is consumed in crack initiation. Then the fatigue life depends on how easy *PSBs* form, i.e., the mechanism of *PSB* formation. In *FCC* metals with high stacking fault energy (*SFE*), such as *Cu*, *Al*, *Ni* and *Ag*, cross slip of dislocations is easier and those metals are subjected to a more intense cyclic hardening or softening in the first 20-40% of fatigue life. Under these conditions, the formation of *PSBs* is easier, leading to a shorter fatigue life. The same holds for low-strength steels, which are subjected to cyclic hardening. In *FCC* metals with low *SFE*, such as brasses (*Cu-Zn*), *Ni*-base superalloys (*Ni-Cr-Mo-Ti-Al-Fe*), austenitic stainless steels (*Fe-Cr-Ni*) as well as *HCP* metals, such as *Mg* and *Ti*, cross slip is more difficult. As a result *PSBs* are difficult to form, leading to a longer fatigue life.

General guidelines concerning the delay of fatigue crack initiation and the extension of fatigue life can now be considered. The condition of the surface of the alloy

should be emphasized. A smooth surface, exhibiting low roughness, contributes to a long fatigue life, since there are not too many discontinuities to cause crack initiation. For the same reason, corrosion or an improper manufacturing method would reduce fatigue life significantly. Increasing the flow stress at the surface delays plastic deformation and the associated *PSB* formation. Several surface modification technologies, such as shot peening, carburization, laser surface hardening, etc., either increase the surface hardness or create a compressive stress field, which delays crack initiation. In addition, a decrease in the level of inclusions, with “cleaner” metallurgical processing, would lead to a delay in crack initiation and a longer fatigue life.

9.3.4 FATIGUE WITH PRE-EXISTING CRACKS – DAMAGE TOLERANCE

Fatigue life. Fatigue life consists of two phases: crack initiation and crack propagation. There are certain engineering structures, where cracks or crack-like defects are created during the construction process. A methodology dealing with fatigue in those structures, is based on the concept of *damage tolerance*, i.e., it is accepted that the structure contains pre-existing crack-like defects. The fatigue life then corresponds to the number of cycles for the propagation of the pre-existing crack with size α_0 to critical size α_f , as depicted in Figure 9.29.

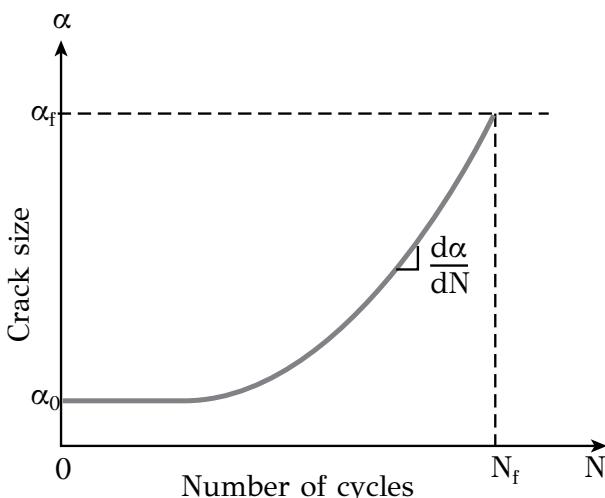


Figure 9.29: Fatigue crack growth: α_0 and α_f are the initial and final crack sizes respectively and da/dN the crack growth rate.

The initial crack size is determined either from non-destructive testing (*NDT*) techniques or is taken equal to the detection limit of the available *NDT* methodologies. The critical final crack size α_f , depends on the fracture toughness of the alloy, according to the fracture criterion. The crack growth rate is expressed with the increase of crack length per cycle, da/dN . If the crack growth rate is known then the

fatigue life can be estimated. The theory of *LEFM*, described in the previous section, can be used for this purpose, provided that small scale yielding conditions are obeyed. Cyclic loading in this case is expressed by the cyclic variation of the stress intensity factor ΔK as

$$\Delta K = K_{\max} - K_{\min}$$

Using Equation (9.7), the following relations can be derived

$$K_{\max} = Q\sigma_{\max}\sqrt{\pi\alpha}$$

$$K_{\min} = Q\sigma_{\min}\sqrt{\pi\alpha}$$

$$\Delta K = Q\Delta\sigma\sqrt{\pi\alpha}$$

$$\Delta\sigma = \sigma_{\max} - \sigma_{\min}$$

where K_{\max} and K_{\min} are the maximum and minimum stress intensity factors and σ_{\max} , σ_{\min} the maximum and minimum stresses of the loading cycle respectively. In most metallic alloys, the relationship between da/dN and ΔK is represented by a sigmoidal curve in a plot of $\log(da/dN)$ vs. $\log\Delta K$, as depicted in Figure 9.30. Three regions can be distinguished in Figure 9.30:

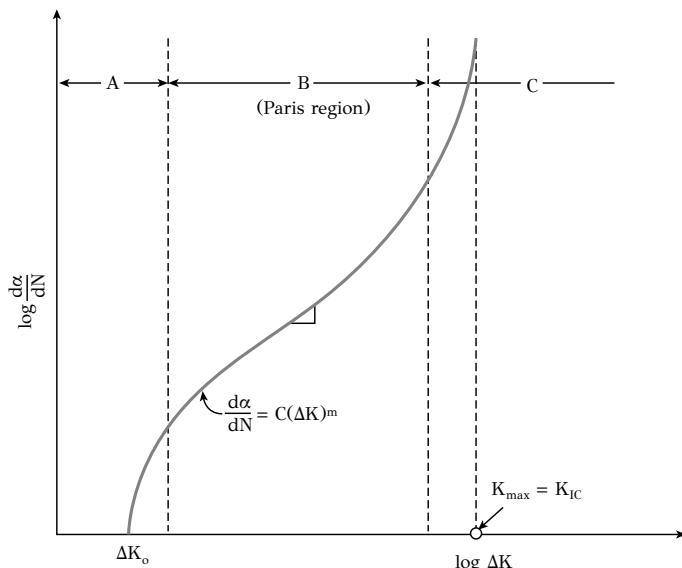


Figure 9.30: The curve expressing the crack growth rate vs. the range of stress intensity factor.

- *Region A:* For low values of ΔK , lower than a threshold ΔK_o , the crack growth rate is extremely small, lower than a lattice parameter per cycle. Essentially it could be considered that the crack does not grow below ΔK_o .
- *Region B:* There is stable crack growth under the action of ΔK and da/dN is correlated with ΔK by an empirical relation

$$\frac{d\alpha}{dN} = C(\Delta K)^m \quad (9.18)$$

which is known as the *Paris law*. The parameters C and m are material constants and depend on the microstructure, temperature, environment and stress ratio R . The exponent m varies between 2 and 4.

- *Region C:* Under large values of ΔK , the crack growth rate is large. The crack grows rapidly to the critical size for fracture.

The Paris law can be used to estimate the fatigue life of a component, which contains a pre-existing crack with initial size α_o . The fatigue life is the number of cycles required for the crack to grow from α_o to α_f . Under constant R and $\Delta\sigma$, Relation (9.18) becomes

$$\frac{d\alpha}{dN} = C(Q\Delta\sigma\sqrt{\pi\alpha})^m$$

Integrating the above relation

$$CQ^m(\Delta\sigma)^m\pi^{m/2} \int_0^{N_f} dN = \int_{\alpha_0}^{\alpha_f} \frac{d\alpha}{\alpha^{m/2}}$$

the fatigue life is derived as

$$N_f = \frac{2}{(m-2)CQ^m(\Delta\sigma)^m\pi^{m/2}} \left[\frac{1}{\alpha_0^{(m-2)/2}} - \frac{1}{\alpha_f^{(m-2)/2}} \right]$$

It should be noted that the critical crack size a_f is determined from the fracture criterion

$$K_{\max} = Q\sigma_{\max}\sqrt{\pi\alpha_f} = K_{IC} \Rightarrow \alpha_f = \frac{1}{\pi} \left(\frac{K_{IC}}{Q\sigma_{\max}} \right)^2$$

Mechanisms of crack growth during fatigue. Crack growth during fatigue is influenced by several factors including the plasticity of the metal, microstructure and the characteristics of cyclic loading. Similar to crack initiation, discussed in the previous section, crack growth involves localized plastic deformation at the crack tip. Two stages can be distinguished in fatigue crack growth: *stage I* and *stage II*. Stage I corresponds to region A of Figure 9.30. When the material is subjected to low ΔK , the crack length and the plastic zone size are comparable to the grain size. The crack grows by shear along the active slip system, as depicted in Figure 9.31a. The activation of a single slip system leads to a zig-zag fracture path, since the crack changes direction as it advances from grain to grain. In this case the fracture surface in stage I

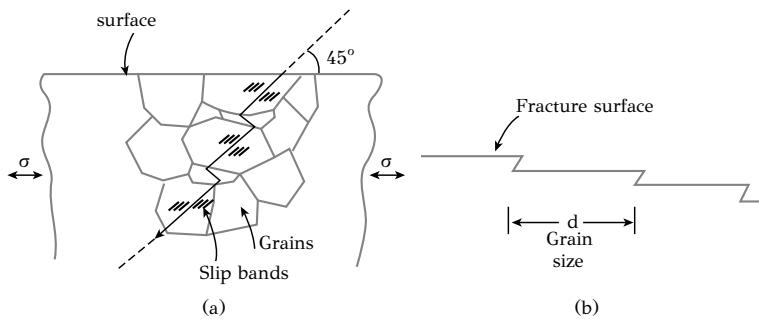


Figure 9.31: Fatigue crack growth during stage I: (a) crack growth with simple shear, (b) corresponding fracture surface.

contains steps, with the step size to correspond to the mean grain size (Figure 9.31b). As the principal slip system is parallel to the shear stress, the crack grows at an angle of about 45° relative to the applied cyclic tensile stress. At larger values of ΔK , crack growth enters stage II, which corresponds to region B of Figure 9.30 (Paris region). In this case the plastic zone is large enough to cover several grains ahead of the crack tip. In this case, at least two slip systems are activated and the crack grows along the intersection of the zones of localized deformation, without changing direction from grain to grain, as depicted in Figure 9.32. The direction of crack growth is perpen-

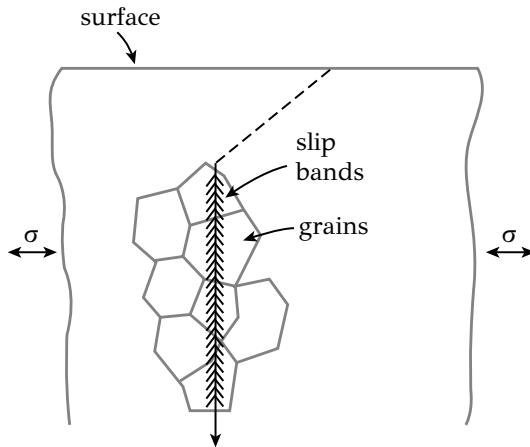


Figure 9.32: Fatigue crack growth during stage II along the intersection of slip bands.

dicular to the applied stress axis and the fracture surface exhibits the characteristic *striations*, shown in Figure 9.33. In several cases the striation spacing is correlated to the crack growth rate da/dN . The presence of striations is often used during failure analysis as a proof that the material failed by fatigue. However, it should be noted, that not all fatigue fractures exhibit striations on the fracture surface. Striations are

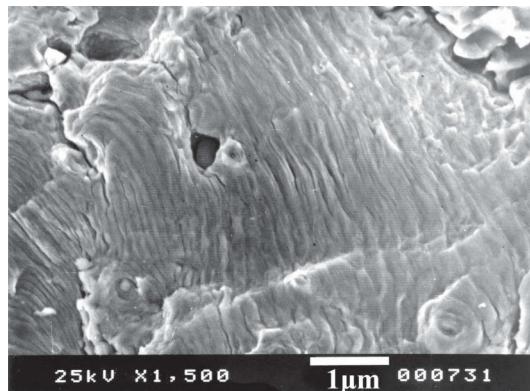


Figure 9.33: Fatigue fracture surface, depicting the characteristic fatigue striations.

more frequent in ductile metals, such as copper and aluminum while they are hard to observe in cold worked metals and steels. At high values of ΔK , the fracture surface contains, in addition to striations, microvoids, cleavage areas and even intergranular fracture, as depicted in Figure 9.34. In those cases, the correlation between striation

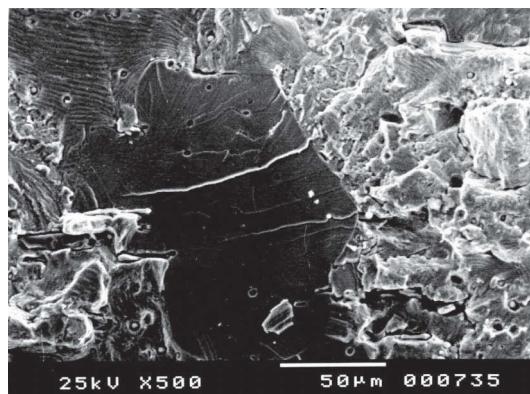


Figure 9.34: Fatigue fracture surface, at high ΔK values, exhibiting striations (top left), cleavage (center) and dimples (right).

spacing and crack growth rate is not accurate. From the above discussion it is clear that while the presence of striations is an evidence of fatigue, the absence of striations does not necessarily exclude fatigue from the possible fracture mechanisms.

The transition from stage I to stage II crack growth takes place when the plastic zone acquires a size comparable to the grain size. This simply means that the fatigue threshold as well as the crack growth rate in region A are influenced by the microstructure of the alloy. In general, ΔK_o scales with a characteristic microstructural length as $d^{1/2}$. This characteristic length could be the grain size in austenitic

steels or *HSLA* steels or the ferritic grain size in ferrite-pearlite steels. In cases where the grain is subdivided by packets of Widmanstaetten side plates or martensitic laths, the characteristic length corresponds to the packet size. The fatigue threshold and crack growth rate in region A, during stage I growth, are also influenced by the operative slip mechanism. Consider the heat-treatable aluminum alloys, which obtain their strength by precipitation hardening. Consider a 2024 alloy in two conditions, exhibiting the same yield strength. The first condition corresponds to underaging, where the θ'' precipitates are fully coherent with the matrix and can be penetrated by dislocations. In this case, dislocation glide can continue from the matrix into the particles. This requires the activation of one slip system and the crack grows according to stage I mechanism. The crack growth rate depends on grain size and is lower in coarse-grained materials. The second condition corresponds to overaging, where the θ precipitates are fully incoherent with the matrix. In this case the dislocations do not penetrate the particles and these obstacles are bypassed by the Orowan mechanism (see Chapter 8 for details). The fracture surface is flat and transgranular, while the crack growth rate is higher and independent of grain size. A contradiction appears in stage I growth, regarding the effect of grain size. For low ΔK values, an increase of grain size leads to higher ΔK_o and lower da/dN . At the same time the endurance limit in *HCF* increases with grain refinement due to the respective increase of yield strength. The contradiction can be resolved, since the first case concerns crack growth while the second case, crack initiation. If there are no pre-existing cracks, then grain refinement leads to an increase of fatigue life. However in structures with pre-existing cracks, which are subjected primarily to low ΔK loading, the use of coarse-grained alloys would lead to longer fatigue life.

In contrast with stage I, crack growth in stage II (Paris region) is not affected significantly by grain size. The plastic zone has the size of several grains and advances from grain to grain with no change in direction. In region C, the crack growth rate is high under large ΔK . Crack advance in this region has common characteristics with static fracture and is influenced by microstructure. Then the microstructure can play a significant role by affecting the fracture toughness K_{IC} . A higher fracture toughness leads, eventually, to a longer fatigue life.

9.4 CREEP

9.4.1 INTRODUCTION

When a metal or alloy is subjected to an applied stress, with magnitude above the yield strength, the material undergoes plastic deformation. As long as the stress remains constant, as depicted in Figure 9.35a, the plastic strain remains constant (Figure 9.35b). However when a constant stress is applied at a high temperature, with $T > 0.4 - 0.5T_m$, where T_m is the melting point, then the plastic strain increases with time. The time-dependent deformation under constant load is called *creep*. The resulting strain is called *creep strain*. Creep is important in the high-temperature service of materials. Creep can lead to unacceptable plastic strains or even fracture, called *creep fracture*. The understanding of the mechanisms of creep is important in

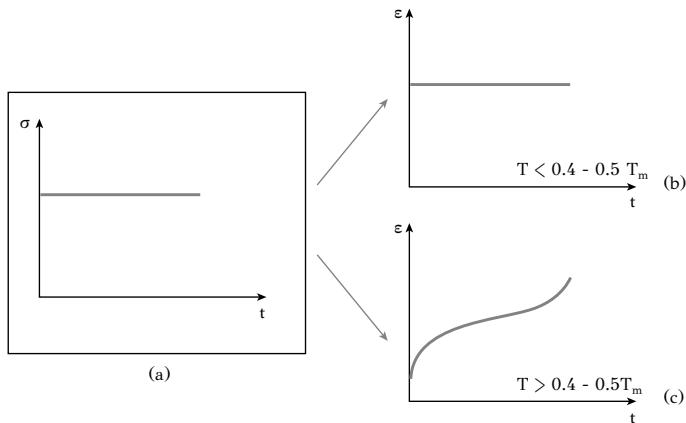


Figure 9.35: (a) Loading with a constant stress will cause: (b) constant deformation at low temperatures, (c) time-dependent deformation (creep) at high temperatures.

designing alloys, which exhibit creep resistance and can perform at high temperatures.

Plastic deformation in metals is produced by the glide of dislocations. The yield strength, which expresses the resistance of the material to plastic deformation, consists of several resistances, which correspond to the strengthening mechanisms studied in Chapter 8. Strengthening mechanisms involve obstacles to dislocation glide. At high temperatures, the thermal energy assists the dislocations to overcome those obstacles and dislocation glide can proceed under a lower stress. In addition, diffusion is activated at high temperatures. Diffusion can either assist dislocations to overcome obstacles and produce plastic strain or can activate additional deformation mechanisms, e.g., with mass flow in the material. Diffusion allows dislocations to acquire an additional degree of freedom. In addition to glide, the dislocations can climb, i.e., move in a direction perpendicular to the slip plane. With climb, the dislocations overcome obstacles that lie on the slip plane, as depicted in Figure 9.36. Since diffusion is a time-dependent process, so is the creep strain. It increases over time.

The time-dependence of the creep strain is depicted in Figure 9.37. The curve $\varepsilon - t$, which is called the *creep curve*, corresponds to the plastic strain produced under constant stress and temperature. Beyond the initial strain ε_0 , which is the instantaneous strain, caused by the application of stress, three stages are distinguished in creep. Stage I is called *primary or transient creep*, stage II is called *secondary or steady state creep* and stage III is called *tertiary creep*. An understanding of the shape of the creep curve can be obtained if it is considered that during creep there are two competitive processes operating: strain hardening and recovery. Strain hardening expresses the increase of the resistance to plastic deformation, and is caused by dislocation multiplication and dislocation interactions. On the other hand, recovery expresses the softening of the material, which is caused by a decrease in dislocation

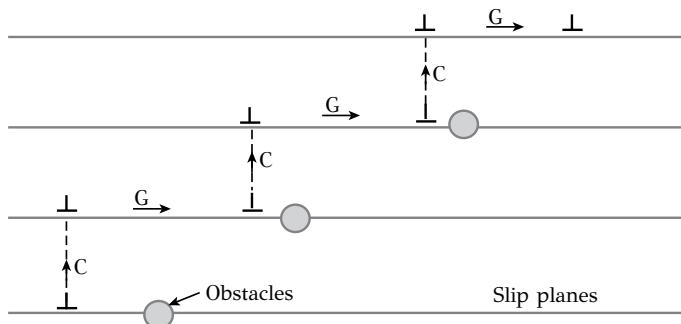


Figure 9.36: Production of plastic strain with combined glide (G) and climb (C) of dislocations.

density and dislocation rearrangements in lower-energy configurations, as has been shown in Chapter 8. An important contribution to recovery comes from the climb of dislocations.

During transient creep (stage I), the plastic deformation causes strain hardening, which dominates over recovery. The density of dislocations increases and dislocation glide is hindered. In this case the strain rate $\dot{\epsilon} = d\epsilon/dt$ decreases.

During steady-state creep (stage II), strain hardening is balanced by recovery. Both processes proceed with the same rate, resulting in a constant strain rate $\dot{\epsilon}$. Steady-state creep is the most important stage in creep, in terms of engineering design, since the creep strain increases with time, a process that could eventually lead to failure. The term failure does not necessarily mean fracture. Even plastic strains, above certain acceptable limits, e.g., 1% in 10 years, can render a component unsuitable for service.

During tertiary creep (stage III), creep is accelerated. The strain rate increases due to mechanisms of internal voiding and cracking at grain boundaries. In certain alloys, strengthening dispersions become unstable and particle coarsening can lead to tertiary creep. In all cases, tertiary creep ends up with the fracture of the material.

The two most important operational parameters, affecting creep strain rate, are stress and temperature. Their influence on the shape of the creep curve is shown schematically in Figure 9.38. An increase of stress raises the creep curve to higher deformation levels. A temperature rise increases the strain rate by accelerating diffusion and the associated recovery processes. Taking into account a large number of experimental data, gathered by the measurement of the steady-state creep strain rate, two creep behaviors could be distinguished. In the first case, the plastic deformation is caused exclusively by dislocation glide, which is thermally activated and assisted by climb. The creep in this case is termed *dislocation or power-law creep*. In the second case, plastic deformation is mainly caused by mass transfer in the material by diffusion. The creep in this case is termed *diffusional flow*.

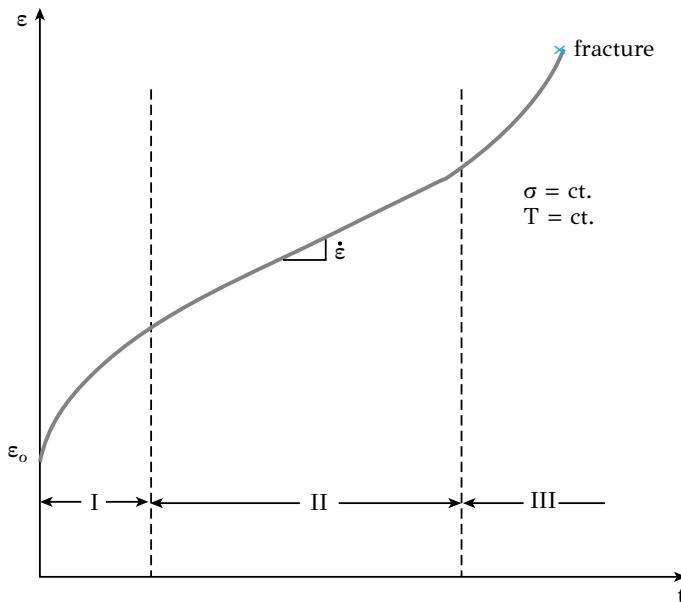


Figure 9.37: The creep curve with the stages of initial or transient creep, steady state creep and tertiary creep.

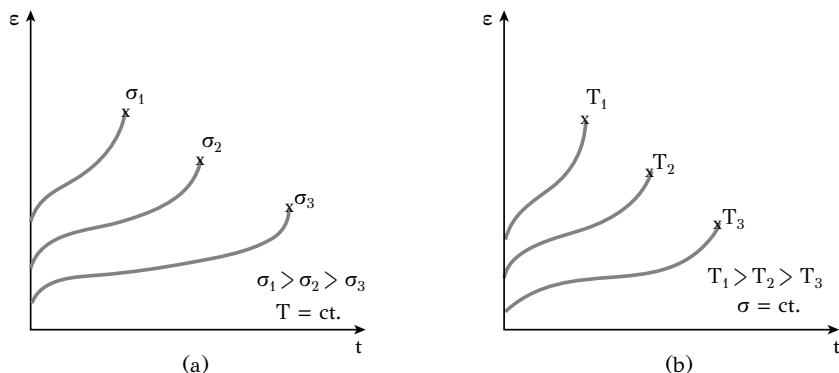


Figure 9.38: The effect of (a) stress and (b) temperature on the creep curve.

In dislocation creep or power-law creep, the strain rate is a function of stress

$$\dot{\varepsilon} \propto D\sigma^m \quad (9.19)$$

The exponent m depends on the alloy system and the micromechanisms of creep. It takes values in the range of 3-10. The temperature dependence of creep is taken into account through the diffusion coefficient D . In diffusional flow, dislocations contribute much less in producing creep strain and creep is caused almost exclusively by diffusion. Diffusional flow is activated at higher temperatures than dislocation creep. The strain rate for diffusional flow is

$$\dot{\varepsilon} \propto D\sigma \quad (9.20)$$

The temperature dependence is also taken into account through the diffusion coefficient. The mass transfer, during diffusional flow, causes a change in the shape of grains. In a polycrystalline material, the local adjustment of grain shape is performed by the relative movement of neighboring grains, a mechanism termed *grain boundary sliding*.

The creep mechanisms and the role of microstructure in creep will be discussed in the following sections.

9.4.2 DISLOCATION CREEP

In dislocation creep, the creep strain is produced exclusively by thermally activated dislocation glide. The operation of two competitive processes, strain hardening and recovery, at the same rate, result in a constant creep strain rate. Recovery removes the effects of strain hardening and allows dislocation glide to take place under constant stress and temperature. The micromechanisms for the production of plastic strain, during dislocation creep, are essentially mechanisms with which recovery takes place. Recovery allows dislocations to overcome obstacles. The specific mechanisms have been discussed in Chapters 7 and 8. A short discussion is presented here, by referring to Figure 9.39:

Overcoming the lattice resistance. Thermal activation assists the formation of kinks, which when formed, allow the stepwise movement of a dislocation on the slip plane. The dislocations obtain a higher mobility and overcome the lattice resistance, producing plastic deformation even under the application of a constant stress (Figure 9.39a).

Overcoming the Cottrell atmospheres. In most metals and alloys, Cottrell atmospheres, consisting of interstitial solutes, form around dislocations. At low temperatures, these atmospheres pin the dislocations and thus operate as an obstacle to dislocation glide. At high temperatures, diffusion allows the glide of the dislocation, since the solute atoms of the atmosphere can follow the dislocation (Figure 9.39b). The velocity of dislocation glide and, consequently, the creep strain rate depend on the rate of diffusion.

Dislocation climb. The most important role of climb during creep is that it facilitates dislocations to overcome obstacles on the slip plane (Figure 9.39c). Plastic

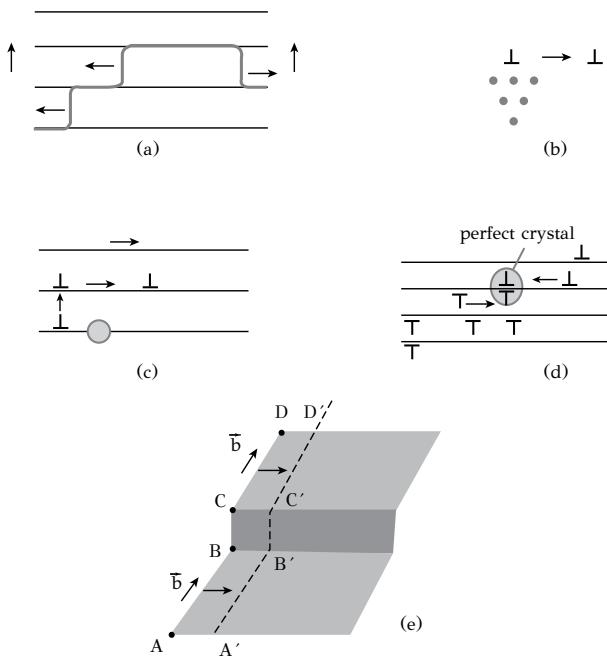


Figure 9.39: Micromechanisms of dislocation creep: (a) overcoming the lattice resistance, (b) overcoming Cottrell atmospheres, (c) dislocation climb, (d) annihilation of dislocations and (e) climb of dislocation jogs.

deformation is produced by dislocation glide, but this glide is controlled by climb. The dislocations rearrange into lower energy configurations, forming cells. The interior of the cells is essentially dislocation free, while the dislocations are concentrated at the cell boundaries, called subboundaries. The subboundaries are then transformed to low-angle grain boundaries. In addition, dislocation climb facilitates the annihilation of dislocations of opposite sign, contributing to the decrease of dislocation density (Figure 9.39d). So during creep, while new dislocations form with the activation of dislocation sources, on the other hand, dislocations annihilate with climb. In *FCC* metals, dislocation dissociation into partials makes climb difficult. This is particularly true for metals with a low *SFE*, in which the dissociation is easier. These low-*SFE* metals exhibit higher creep resistance, as depicted in Figure 9.40.

Climb of dislocation jogs. Jogs form during dislocation intersections. When two screw dislocations intersect, a jog with edge character forms. The glide of screw dislocations, which contain jogs, is blocked because the jog cannot glide on the slip plane of the screw dislocation. The only way for the screw dislocation to glide on its slip plane is when the jog is allowed to climb. For example, as depicted in Figure 9.39e, the screw dislocation $ABCD$ contains a jog BC , with edge character. The glide of the dislocation in the new position $A'B'C'D'$ requires the climb of the jog

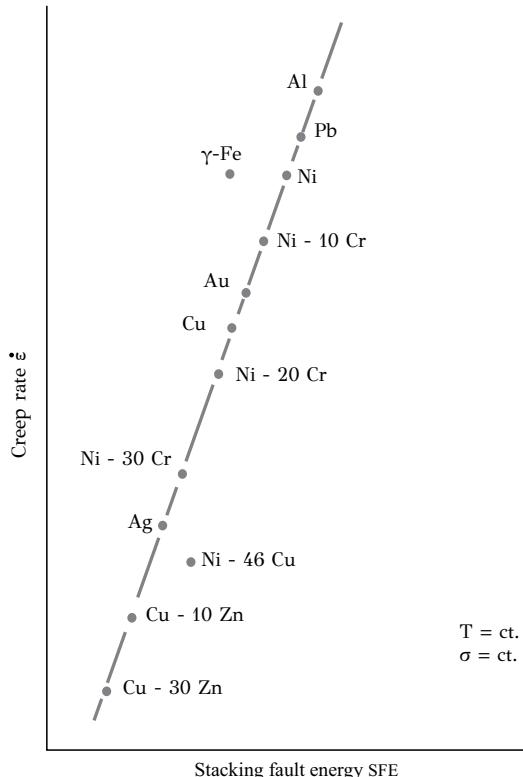


Figure 9.40: The effect of *SFE* on creep rate of various *FCC* metals and alloys.

to position $B'C'$, which is only possible at high temperatures, with the activation of diffusion.

Cross slip of partial dislocations. This mechanism refers mainly to *FCC* metals, in which perfect dislocations dissociate into partials. In this way cross slip of screw dislocations is difficult, resulting in significant strain hardening. However, at high temperatures, thermal activation allows the cross slip of partial dislocations, which contribute to the creep stain. This is important in *FCC* alloys with low *SFE*, where the dissociation into partials is easier.

Dislocation creep is characterized by a creep strain rate, which is a power-law function of stress. Relation (9.19) can be written as

$$\dot{\epsilon} = A \frac{\mu b}{kT} D \left(\frac{\sigma}{\mu} \right)^m \quad (9.21)$$

where A is a dimensionless constant and m is the creep exponent, μ is the shear modulus and k is the Boltzmann's constant. Values for A and m appear in Table 9.1. In pure metals, the exponent m takes a value of about 5. The addition of alloying elements for the formation of solid solutions, decreases m to a value around 3, for

Table 9.1**Values of the creep exponent m and constant A for dislocation creep.**

Material	m	A
FCC Metals	5	10^7
BCC Metals	5	10^9
HCP Metals	5	10^6
Alloys (Class I)	3	10^1
Alloys (Class II)	5	10^6

Reprinted from: S. Takeuchi, A.S. Argon, J. Mater. Sci. 11 (1976) 1542-1566, with permission from Springer.

class I alloys, while m remains around 5 for class II alloys. This differentiation is attributed to the interaction of dislocations with Cottrell atmospheres, which form in class I alloys. When the alloying content or the difference in atomic size increases, creep behavior approaches that of class I alloys with $m=3$. A typical way to increase creep resistance, is the formation of a stable precipitate dispersion, as in Ni-base superalloys, discussed in the following sections. In these alloys the creep exponent takes values between 7 and 12.

Summarizing the above discussion, the most important microstructural elements affecting the rate of dislocation creep are (a) the characteristics of the solid solution (atomic size, concentration of solute), (b) dislocation characteristics regarding climb and cross slip and (c) the *SFE* in *FCC* alloys. Dislocation creep is not influenced significantly by grain size.

9.4.3 DIFFUSIONAL FLOW

At low stresses, where dislocation glide is impossible, creep strain can be produced exclusively by diffusion at high temperatures, a mechanism termed diffusional flow. Consider a square grain of size d in a polycrystalline alloy subjected to a shear stress, as depicted in Figure 9.41a. The shear loading is equivalent to a set of normal stresses, a tensile stress acting on grain boundary (1) and a compressive stress acting on grain boundary (2). The stress gradient along the grain influences diffusion in a manner discussed in Chapter 5, section 5.9.2. The stress gradient influences the driving force for diffusion by altering the concentration of vacancies in boundaries (1) and (2). The equilibrium concentration of vacancies is

$$c_o = \exp\left(-\frac{E_v}{kT}\right)$$

where E_v is the vacancy energy of formation, discussed in Chapter 3. Vacancy formation is assisted by a tensile stress, since the energy of formation is reduced by the term $\sigma\Omega$, where Ω is the atomic volume. The opposite is true for a compressive

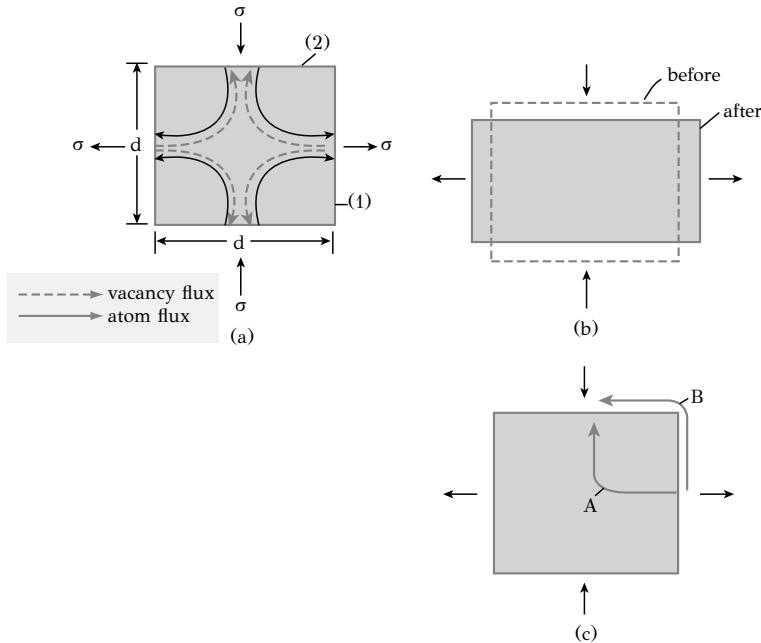


Figure 9.41: The mechanism of diffusional flow: (a) vacancy diffusion, (b) grain deformation, (c) bulk and grain boundary diffusion.

stress. The equilibrium vacancy concentration in boundaries (1) and (2) becomes

$$c_1 = \exp\left(-\frac{E_V - \sigma\Omega}{kT}\right)$$

$$c_2 = \exp\left(-\frac{E_V + \sigma\Omega}{kT}\right)$$

From the above relations it is clear that $c_1 > c_o$ and $c_2 < c_o$. The difference in vacancy concentrations between grain boundaries (1) and (2) sets up a diffusive flux of vacancies from boundary (1) to boundary (2) and a corresponding flux of atoms from boundary (2) to boundary (1). Mass is removed from the transverse boundaries and is added to the longitudinal boundaries. This process results in a macroscopic plastic deformation in a polycrystalline alloy, depicted in Figure 9.41b. The creep strain is not produced by dislocation glide but, instead, is produced by diffusional flow.

The diffusive flux of vacancies can take place either by bulk diffusion or by grain boundary diffusion, as depicted in Figure 9.41c. The first case is referred to as *Nabarro-Herring* creep while the second as *Coble* creep. The combination of bulk and grain boundary diffusion leads to a strain rate, for diffusional flow, given by

$$\dot{\epsilon} = B \left(\frac{\sigma\Omega}{kT} \right) \left(\frac{D_V}{d^2} \right) \left[1 + \frac{\pi}{2} \left(\frac{\delta}{d} \right) \frac{D_B}{D_V} \right] \quad (9.22)$$

where B is a constant, d is the grain size, δ is the grain boundary thickness (about 2-3 interatomic distances), while D_V and D_B are the bulk and grain boundary diffusivities respectively. Contrary to dislocation creep, the grain size plays an important role in diffusional flow, since the strain rate increases with a decrease in grain size. This is attributed to the shorter diffusion distances and the contribution of grain boundary diffusion as the grain size becomes shorter. The contribution of grain boundary diffusion becomes important when

$$\frac{\pi\delta}{2d} \frac{D_B}{D_V} > 1$$

Taking into account the temperature dependence of the diffusion coefficient

$$D = D_o \exp(-\Delta G^*/kT)$$

the above relation becomes

$$\frac{\pi\delta}{2d} \exp \frac{\Delta G_V^* - \Delta G_B^*}{kT} > 1$$

and indicates that diffusional flow will commence at lower temperatures in a material with a fine grain size. In addition, grain boundary diffusion is faster than bulk diffusion, since ΔG_B^* is about 60% of ΔG_V^* , as discussed in Chapter 5.

The creep mechanisms have been graphically assembled by Prof. Mike Ashby in the so-called deformation mechanism maps. An example is shown in Figure 9.42 for nickel with two grain sizes. The extent of diffusional flow over dislocation creep in the fine grain material is evident.

A phenomenon associated with diffusional flow in a polycrystalline material is the relative motion between neighboring grains, since the grains should adjust their shape. The motion takes place by *grain boundary sliding* and is depicted in Figure 9.43. The diffusional flow causes a change in the shape of grains, producing voids, as depicted in Figure 9.43a and b. Continuity is maintained by grain boundary sliding, as indicated by the relative motion of the markers in Figure 9.43c. Grain boundary sliding is, therefore, a consequence of diffusional flow.

9.4.4 CREEP FRACTURE AND CREEP LIFE

Creep is accelerated during the tertiary creep stage. The increase in strain rate is caused by the formation of voids and microcracks in the interior and mostly at the grain boundaries. Two types of intergranular cracking are observed. The first refers to wedge-type cracking, which forms on boundaries loaded with the highest shear stress. The cracks initiate at triple junctions and crack propagation is assisted by grain boundary sliding, as depicted in Figure 9.44a.

The second type of intergranular cracking is caused by the growth and coalescence of microvoids in boundaries, which are oriented perpendicular to the applied stress (Figure 9.44b). Void nucleation is assisted by precipitates located at grain boundaries, such as carbides in steels. Microvoids already form during steady-state creep (stage

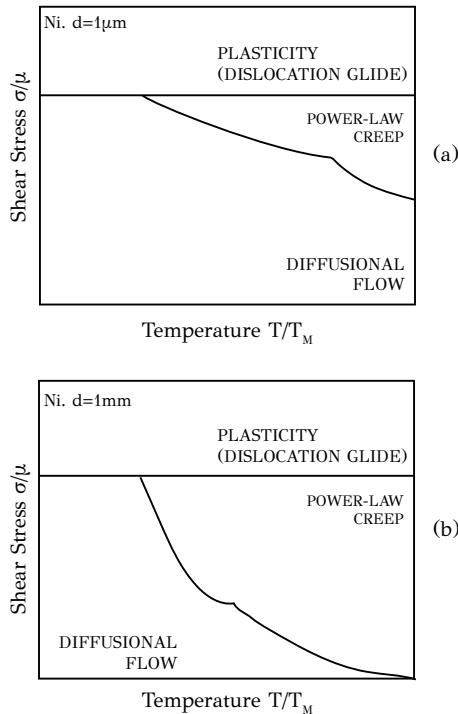


Figure 9.42: Deformation mechanism maps for nickel with (a) grain size of 1mm , (b) grain size of $1\mu\text{m}$ (schematic).

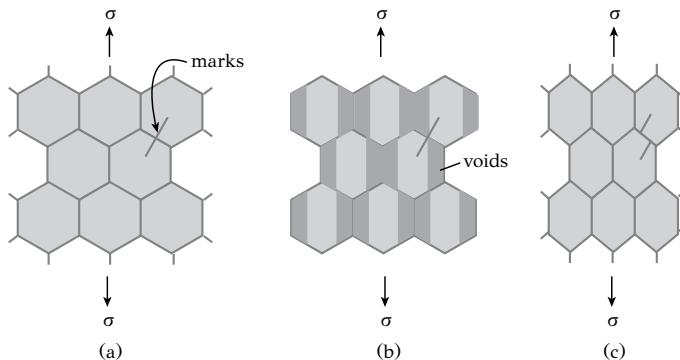


Figure 9.43: Schematic of grain boundary sliding during diffusional flow. Details in text.

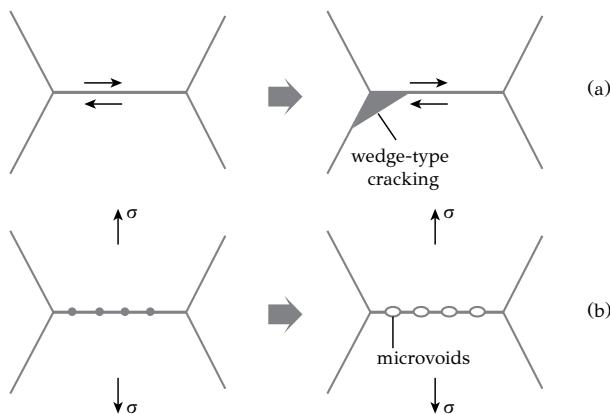


Figure 9.44: Intergranular cracking during creep: (a) wedge-type cracking, (b) formation and coalescence of cavities.

II) but they are isolated voids. With the evolution of creep, the microvoids grow and become cavities, which coalesce and form microcracks during tertiary creep.

The time to fracture or creep life, t_f , is the time necessary for the evolution of creep damage up to the point of fracture. The creep life, together with the steady state creep rate $\dot{\epsilon}$ express the creep resistance of an alloy. The time to fracture can be determined from creep tests and depends on stress and temperature, as depicted in Figure 9.45. During creep testing, the specimens are subjected to higher stresses than actual service stresses, so that the test can be completed in a reasonable time. Typical

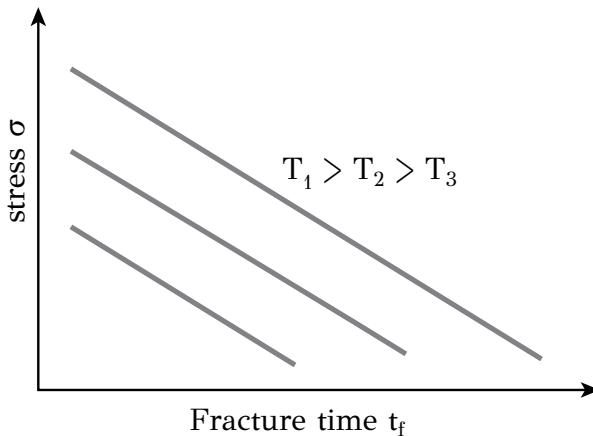


Figure 9.45: Time to fracture as a function of stress and temperature.

creep fracture times in laboratory experiments are between 1000 and 10000 hours. In actual service conditions creep life is expected to be of the order of 10000 hours

for aircraft turbine blades, while for stationary turbines the anticipated creep life is of the order of 100000 hours (about 10 years). In heat exchangers and boilers the creep life is of the order of 300000 hours (30 years). Extrapolation techniques for the estimation of creep life from creep tests have been developed. The most widely used one is the methodology developed by Larson and Miller¹, using the Larson-Miller parameter LM , which is defined by

$$LM = T(\log t_F + C) \quad (9.23)$$

where T is the absolute temperature, t_F the fracture time and C a constant, which depends on the material, with a value about 20. The LM parameter depends on applied stress, as depicted in Figure 9.46. Graphs like the one in Figure 9.46 are available in handbooks, for many high-temperature alloys. In this way it is possible to estimate the creep life for a certain stress and temperature combination or to estimate the allowable stress for an anticipated creep life. For example, the allowable stress for a creep life of, say, 100000 hours at a temperature T_1 is calculated by first finding the LM parameter from Equation (9.23) and then determining the stress from a corresponding graph, like the one of Figure 9.46. Care should be taken at low stress and very high temperatures, where creep could be dominated by diffusional flow. In this case the Larson-Miller technique overestimates the creep life.

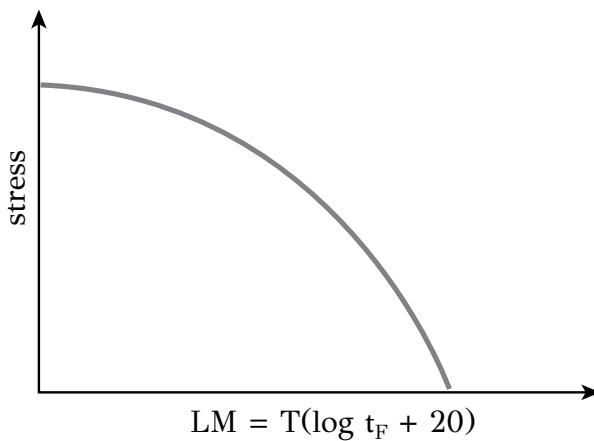


Figure 9.46: Larson-Miller plot correlating the LM parameter with the applied stress.

9.4.5 DESIGN OF CREEP RESISTANT ALLOYS

The design and development of creep resistant alloys is based on different criteria than for alloys used at low temperatures. The strengthening mechanisms, discussed

¹F.R. Larson, J. Miller, Trans. ASME, 74(1952)765.

in Chapter 8, do not necessarily operate at high temperatures. Grain size refinement is a good example. According to the Hall-Petch relation, grain size refinement leads to strengthening. However, at high temperatures, especially under diffusional flow conditions, creep resistance is increased with a coarse grain size. The basic criteria for the design of creep resistant alloys are the following:

Melting point. Since the diffusion coefficient decreases with an increase of melting point, it is reasonable that alloys, with a high melting point, are used in high-temperature applications. An example is the use of tungsten wire in light bulbs.

Crystal structure. *FCC* metals are more creep resistant than *BCC* metals for two reasons. The first is that *FCC* metals exhibit a lower diffusion coefficient, arising from close packing. The second is that in *FCC* metals, the dislocation dissociation into partials, which suppresses cross slip and climb of dislocations, is easier. Most high temperature alloys and superalloys have an *FCC* structure.

Grain size. This effect is more intense under conditions of diffusional flow, i.e., at low stresses and very high temperatures. Alloys are manufactured with a large grain size or as single crystals.

Processing. Microstructure is affected by processing. Consider turbine blades for example. The blade is subjected to a tensile stress along the axis of the blade during its rotation in the engine. Now consider the casting of the blade in Figure 9.47. Conventional casting results in a fine equiaxed microstructure (Figure 9.47a) with low creep resistance. Directional solidification results in the columnar structure of Figure 9.47b. This microstructure exhibits a lower number of grain boundaries, oriented perpendicular to the applied stress. As a result this microstructure exhibits a higher resistance to microvoid nucleation at the grain boundaries and, therefore, has a higher creep resistance. An even higher creep resistance is provided by the casting of single crystal blades (Figure 9.47c). The blade has no grain boundaries and, therefore, the nucleation of microvoids is prevented. The casting of single crystals is performed by the technique of investment casting or lost wax casting.

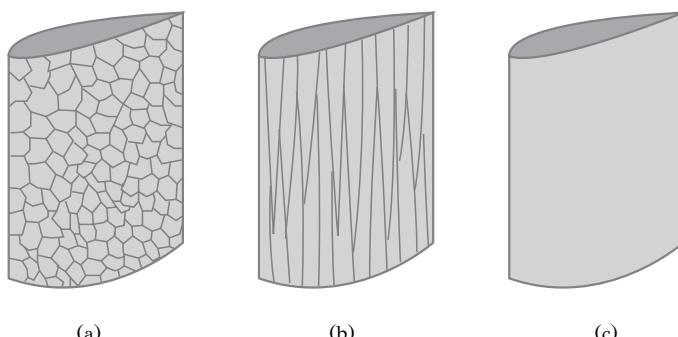


Figure 9.47: Three turbine blades with same chemical composition but different processing: (a) conventional casting with fine equiaxed grains, (b) directional solidification with columnar grains, (c) investment casting with single crystal blade.

Alloying. The most important role of alloying elements in creep resistance is that they suppress dislocation glide and climb. Alloying elements can be either in solid solution or form precipitates. Consider the case of nickel-base superalloys, which are nickel alloys with additions of *Cr*, *Al*, *Ti*, *Mo*, *Co* and *Fe*. The microstructure, depicted in Figure 9.48, consists of a fine dispersion of Ni_3Al (γ') phase in the nickel solid solution (γ) phase. These particles resist the glide of dislocations, even



Figure 9.48: Micrograph from the transmission electron microscope showing γ' (Ni_3Al) particle dispersion in an *FCC* matrix in an Inconel 718 superalloy.

at very high temperatures. This is attributed to the high thermodynamic stability of the γ' phase, which maintains coherency with the γ matrix and, therefore, exhibits low interfacial energy and low coarsening rate. The rest of the alloying elements are dissolved in the matrix. Their contribution to solid solution hardening is small, due to the small difference in atomic size with nickel. However they reduce the *SFE* of the matrix, allowing the easier dissociation of dislocations, which suppress the climb and cross slip of dislocations. Even greater thermodynamic stability is exhibited by inert oxide dispersions, like Al_2O_3 and ThO_2 . These oxides are mixed with metallic powders for the processing of engineering components by sintering. In heat-resistant steels, alloying leads to the formation of carbides such as VC , TiC , NbC , Mo_2C and $Cr_{23}C_6$. These carbides offer resistance to dislocation glide at moderate temperatures. Several creep resistant alloys are shown in Table 9.2.

Table 9.2
Several creep resistant alloys

Ferritic Steels	
1.25Cr – Mo	0.1C – 1.25Cr – 0.5Mo
5Cr – Mo	0.2C – 5Cr – 0.5Mo
Austenitic Steels	
316	0.08C – 18Cr – 8Ni – 2Mo
16 – 25 – 6	0.1C – 16Cr – 25Ni – 6Mo
Nickel Superalloys	
Inconel	<i>Ni</i> – 0.04C – 15Cr – 7Fe
Inconel 718	<i>Ni</i> – 0.04C – 19Cr – 3Mo – 5Nb – 0.8Ti – 0.6Al – 18Fe
Rene 41	<i>Ni</i> – 0.1C – 19Cr – 10Mo – 11Co – 3.2Ti – 1.6Al – 2Fe
Udimet 700	<i>Ni</i> – 0.15C – 15Cr – 5.2Mo – 18.5Co – 3.5Ti – 4Al – 1Fe
TD-Nickel	Ni-2% <i>ThO</i> ₂

9.5 SYNOPSIS

1. The term “mechanical behavior” expresses the response of a metallic material to applied loads. This includes static or dynamic loads, at low or high temperatures, while the material can be simultaneously exposed to corrosive environments. The material responds by either deformation or fracture.
2. When fracture takes place after significant plastic deformation, it is characterized as ductile fracture, otherwise it is termed brittle fracture.
3. Brittle fracture takes place by cleavage along specific crystallographic planes, leading to a transgranular mode of fracture. Brittle fracture can also take place as intergranular fracture, as a result of degradation processes at the grain boundaries.
4. Ductile fracture takes place by the nucleation, growth and coalescence of microvoids, a process associated with considerable plastic deformation. Void nucleation usually takes place at inclusions and second phase particle dispersions.
5. Structural steels, as well as many *BCC* metals, undergo a ductile-to-brittle transition with a drop in temperature. The ductile-to-brittle transition temperature (*DBTT*) is used as a measure of the notch toughness or impact strength of the material.
6. Brittle fracture is controlled, among others, by stress, pre-existing crack-like flaws and fracture toughness of the material. The correlation of these parameters in engineering design is the subject of fracture mechanics.

7. Fracture toughness is a material property and depends on chemical composition, microstructure and processing of the material.
8. When a metallic component is subjected to cyclic loading, it would fail at loads, which are much lower than those required for fracture under monotonic loading. In this case, failure is attributed to metal fatigue.
9. Fatigue involves the progressive accumulation of damage in the form of localized plastic deformation, the initiation and the propagation of cracks, until final fracture.
10. Two periods can be distinguished in fatigue life. The first period corresponds to the number of cycles N_i , which are required for the formation or initiation of a crack. The second period corresponds to the number of cycles N_p required for the propagation of the crack to the critical size.
11. In machine elements, such as shafts, gears and bearings, the assumption is made that they do not contain any pre-existing cracks. In this case, design against fatigue is performed with the strain-life approach. Fatigue is then distinguished in high-cycle fatigue (*HCF*) and low-cycle fatigue (*LCF*).
12. In *HCF* the loads are low and the fatigue life is above $10^3 - 10^4$ cycles. In this case, the cyclic stresses cause only elastic strains. In *LCF*, the loads are much higher and the fatigue life is short, usually below $10^3 - 10^4$ cycles. In this case, the cyclic stresses cause significant plastic strains.
13. The balance between strength and ductility is important in fatigue. A high-strength, low-ductility alloy, performs better in *HCF* conditions. On the other hand, a soft and ductile alloy, performs better in *LCF* conditions.
14. Crack initiation in smooth specimens, takes place at surface discontinuities, intrusions and extrusions, which are created at the surface of the metal, along zones of localized plastic deformation, called persistent slip bands (*PSB*). Fatigue cracks initiate either directly from these surface discontinuities or from the interaction of intrusions with inclusions and grain boundaries.
15. There are engineering structures, such as welded structures, where crack-like defects are created during the construction process. A methodology, dealing with fatigue in those structures, is based on the concept of damage tolerance, i.e., it is accepted that the structure contains pre-existing crack-like defects. The fatigue life then corresponds to the number of cycles for the propagation of the pre-existing crack to critical size.

16. The Paris law can be used to estimate the fatigue life of a component, which contains a pre-existing crack in the context of fracture mechanics.
17. The time-dependent deformation under constant load, at high temperatures, is called creep. Creep can lead to unacceptable plastic strains or even fracture.
18. In general, three stages are distinguished in creep: primary or transient creep, secondary or steady state creep and tertiary creep.
19. During creep there are two competitive processes operating: strain hardening and recovery. Strain hardening expresses the increase of the resistance to plastic deformation, and is caused by dislocation multiplication and dislocation interactions. Recovery expresses the softening of the material, which is caused by a decrease in dislocation density and dislocation rearrangements in lower-energy configurations.
20. Two creep behaviors can be distinguished. In the first case, the plastic deformation is caused exclusively by dislocation glide, which is thermally activated and assisted by climb. The creep in this case is termed dislocation or power-law creep. In the second case, plastic deformation is mainly caused by mass transfer in the material by diffusion. The creep in this case is termed diffusional flow.
21. The most important microstructural elements affecting the rate of dislocation creep are (a) the characteristics of the solid solution (atomic size, concentration of solute), (b) dislocation characteristics regarding climb and cross slip and (c) the SFE in FCC alloys. Dislocation creep is not influenced significantly by grain size.
22. Contrary to dislocation creep, the grain size plays an important role in diffusional flow, since the strain rate increases with a decrease in grain size.
23. Grain boundary sliding is the relative motion between neighboring grains and is a consequence of diffusional flow in a polycrystalline material.
24. Two types of intergranular cracking are observed in creep. The first refers to wedge-type cracking, which forms on boundaries loaded with the highest shear stress. The second is caused by the growth and coalescence of microvoids in boundaries, which are oriented perpendicular to the applied stress.
25. The basic criteria for the alloy design against creep are: melting point, crystal structure, grain size, alloying and processing.

9.6 REVIEW QUESTIONS

1. A large plate is going to be subjected to uniaxial tensile loading. The plate is to be made with one of the aluminum alloys, shown in the table below.

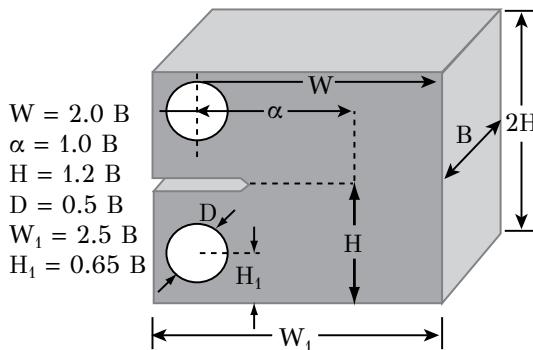
Alloy	Yield strength	Fracture toughness
	$\sigma_y(MPa)$	$K_{IC}(MPa\sqrt{m})$
2024-T4	330	34
7075-T651	503	27
7039-T651	338	32

You are asked to select the right alloy, which will allow you to maximize the design stress, with a safety factor of 2 against plastic deformation and safety factor of 2 against brittle fracture. Consider two cases regarding the detection limit of your ultrasonic device: (a) 20 mm, (b) 2 mm.

- Steel A533 is a pressure vessel steel, according to the *ASTM* specifications. It is going to be used for a low-temperature application. The dependence of fracture toughness and yield strength on temperature is shown in the table below. Consider a plate containing an edge crack ($K_I = 1.12\sigma\sqrt{\pi\alpha}$), subjected to a tensile stress $\sigma = \sigma_y/2$.

T (°C)	Fracture toughness	Yield strength
	$K_{IC}(MPa\sqrt{m})$	$\sigma_y(MPa)$
-160	32	730
-140	36	680
-120	39	640
-100	44	600
-80	50	570
-60	57	540
-40	69	520
-20	85	490

- Calculate the critical plastic zone size and the critical crack size as a function of temperature.
- Suppose you would like to determine the fracture toughness of the material, using a compact tension specimen, per *ASTM E399*, shown below. Calculate the required thickness and the associated weight of the specimen as a function of temperature, so that small-scale yielding conditions are satisfied. Comment on your results.
- A specimen made of aluminum alloy 2024-T3, which is going to be used for the construction of the fuselage of an aircraft, was subjected to fatigue testing under a sinusoidal stress with zero mean stress. With a stress amplitude of 140 MPa, the fatigue life was 10^5 cycles while when the stress amplitude decreased to 100



MPa, the fatigue life increased to 10^7 cycles. Calculate the expected fatigue life corresponding to a stress amplitude of 75 MPa. Consider now that a fuselage, constructed from this alloy, has been subjected to 4×10^8 cycles with a stress amplitude of 75 MPa. It is desired to extend the fatigue life for another 4×10^8 cycles. Calculate the required reduction in the stress amplitude, in order to achieve the desired life extension.

- Consider a steel plate with thickness 30 mm and width 300 mm. The plate is to be subjected to cyclic tensile loading between min and max loads of 1.8 MN and 2.7 MN respectively. The detection limit of the available NDT technique is 8.5 mm and the required fatigue life is at least 190000 cycles. The associated mechanical properties are $K_{IC} = 80 \text{ MPa m}^{1/2}$ and $da/dN = 4.624 \times 10^{-12} (\Delta K)^3$ (in m/cycle). Does the material satisfy the requirements? What options are there for life extension?
- A steel plate, with thickness 12.7 mm and width 300 mm, was subjected to cyclic tensile loading (in the length direction) between zero and an unknown maximum load. The plate fractured during service. Failure analysis indicated that fracture initiated at the edge of the plate. The striation spacing, determined by SEM, was $0.25 \mu\text{m}$ at a crack length of 25 mm. Calculate the maximum tensile load. Material properties: σ_y MPa, $K_{IC} = 41 \text{ MPa m}^{1/2}$ and $da/dN = 10^{-12} (\Delta K)^{3.25}$ (in m/cycle).
- Consider the design of three spherical pressure vessels with a diameter of 3.3 m, to be constructed from three different alloys, shown in the table below.

Material	σ_y (MPa)	K_{IC} ($\text{MPa m}^{1/2}$)	da/dN (m/cycle)	ρ (gr/cm^3)
Steel	1275	154	$5.6 \times 10^{-11} (\Delta K)^{2.2}$	7.61
Titanium alloy	896	83	$1.12 \times 10^{11} (\Delta K)^{3.08}$	4.35
Aluminum alloy	345	44	$3.63 \times 10^{-11} (\Delta K)^{3.34}$	2.72

Each vessel will be designed with a safety factor of 2 against plastic deformation,

however no safety factor is specified against fracture. The vessel is subjected to a cyclic internal pressure between 10.33 and 20.66 MPa. The detection limit of available *NDT* techniques is 7.6 mm. Assume a semi-elliptical crack ($Q = 1.073$) and determine the following for each vessel:

- a) the weight
- b) the failure type, i.e., fracture or leak-before-break
- c) the fatigue life

BIBLIOGRAPHY—FURTHER READING

1. Wells, J. M. & Landes, J. D. (1984). *Fracture: Interactions of Microstructure, Mechanisms, Mechanics*, A.I.M.E.
2. Barsom, J. M. & Rolfe, S. T. (1999). *Fracture and Fatigue Control in Structures*, 3rd edition, Butterworth Heinemann.
3. Ewalds, H. L. & Wanhill, R. J. H. (1984). *Fracture Mechanics*, Edward Arnold.
4. Broek, D. (1982). *Elementary Engineering Fracture Mechanics*, 3rd edition, Martinus Nijhoff Publishers, The Hague.
5. Broek, D. (1988). *The Practical Use of Fracture Mechanics*, Kluwer Academic Publishers, The Netherlands.
6. Kanninen, M. F. & Popelar, C. H. (1988). *Advanced Fracture Mechanics*, Oxford University Press.
7. Knott, J. F. (1979). *Fundamentals of Fracture Mechanics*, Butterworth, London.
8. Knott, J. F. & Withey, P. A. (1993). *Fracture Mechanics: Worked Examples*, The Institute of Materials, London.
9. Anderson, T. L. (1995). *Fracture Mechanics: Fundamentals and Applications*, CRC Press.
10. Tada, H., Paris, P.C., Irwin, G.R. (2000). *The Stress Analysis of Cracks Handbook*, 3rd edition, ASME, New York.
11. Hull, D. (1999). *Fractography*, Cambridge University Press.
12. Bannantine, J.A., Comer, J.J., Handrock, J.L. (1990). *Fundamentals of Metal Fatigue Analysis*, Prentice Hall.
13. Fuchs, H.O. & Stephens, R.I. (1980). *Metal Fatigue in Engineering*, John Wiley & Sons, New York.
14. Suresh, S. (1998). *Fatigue of Materials*, 2nd edition, Cambridge University Press.
15. Hertzberg, R. W. (1976). *Deformation and Fracture Mechanics of Engineering Materials*, 3rd edition, Wiley.
16. Lung, C. W. & March, N. H. (1999). *Mechanical Properties of Metals*, World Scientific.
17. Meyers, M. A. & Chawla, K. K. (1984). *Mechanical Metallurgy: Principles and Applications*, Prentice-Hall.
18. McClintock, F.A. & Argon, A.S. (1966). *Mechanical Behavior of Materials*, Addison Wesley.
19. Riedel, H. (1982). *Fracture at High Temperatures*, Springer-Verlag, Berlin.

10 Physical metallurgy of steels

10.1 INTRODUCTION

Steel is the most important engineering material used by man. Despite the advances in microelectronics, led by the development of silicon semiconductors, in terms of construction, mankind is still in the “steel age”. Steels constitute more than 80% of industrial alloys, used for the construction of buildings, bridges, cars, trains and ships. But steel is also used for making machine parts, such as shafts, gears and bearings as well as home appliances. The technological position of steels among other materials is attributed to several factors, including low cost, manufacturability and recyclability. With the implementation of cleaner steel making technologies, the CO_2 footprint has been reduced dramatically in the last 50 years.

Steels exhibit a vast diversity of microstructures and mechanical properties. Steel strength ranges between 200MPa and 5GPa. Additional reasons for the technological dominance of steel are:

- (a) The large reserves of steel ore, as the earth crust contains more than 4% iron. The ore is basically iron oxide, which can be easily reduced to metallic iron. This iron, in combination with the iron produced from the recycling of scrap, ensures an unlimited steel supply, keeping steel prices low.
- (b) The melting point of iron ($1539^\circ C$) allows the thermal activation of phase transformations during heat treatment and processing at temperatures, which can be easily achieved and controlled at an industrial scale.
- (c) The allotropy of iron and the associated phase transformations allow the formation of a large range of microstructures and properties.
- (d) Alloying elements in steels widen the range of steel applications: structural steels, weathering steels, automotive steels, stainless steels, piping steels, heat-resistant steels, tool steels, surgical steels, to name just a few.

Steel is the most complicated alloy, in terms of the interplay between alloying elements, processing, microstructure and properties. In the last 100 years, with the advent of new high-resolution characterization techniques there has been enormous progress in the understanding of steel. With the development of reliable models and computational tools, a new era is emerging: computational steel design. This knowledge-based methodology is going to boost steel development in the years to come. More about alloy design in Chapter 11.

In the following sections, the fundamental knowledge of thermodynamics, diffusion and phase transformations, developed in the previous chapters of this book, will be used for a structured study of steels. The phases in steels and the associated $Fe - C$

phase diagram will be presented first. The effect of alloying elements on phase stability will be discussed next. The core of the chapter concerns the phase transformations in steels, which take place under isothermal and continuous cooling conditions and shape the microstructure of steel. The development of desirable microstructures and properties through heat treatment will be discussed next. Several case studies, regarding the interplay between processing, microstructure and processing will also be presented.

10.2 PHASES IN STEELS

The various phases, which form in steels, are responsible for the large range of properties. The phases can be classified in solid solutions, intermetallic compounds (basically carbides and nitrides) and phase mixtures.

10.2.1 SOLID SOLUTIONS – FERRITE, AUSTENITE, AND MARTENSITE

Carbon is the most important alloying element in steels. It forms interstitial solid solutions with iron. Due to the allotropy of iron, two solid solutions are formed, ferrite and austenite. *Ferrite* is an interstitial solid solution of iron with carbon with a *BCC* structure. It appears as α -ferrite and δ -ferrite at low and high temperatures respectively. *Austenite* is an interstitial solid solution of iron with carbon, with an *FCC* structure. A third solid solution, *martensite*, is a supersaturated interstitial solid solution of iron with carbon, with a *BCT* structure. Martensite is a non-equilibrium phase. It forms by the rapid cooling of austenite, a process called *quenching*.

Solid solubility in ferrite and austenite is controlled by the size of interstitial sites in the two structures, *BCC* and *FCC*. As discussed in Chapter 2, the octahedral interstitial sites in *FCC* are larger than the respective sites in *BCC*. This is the main reason, austenite exhibits a higher maximum solid solubility for carbon, 2.1wt%, relative to only 0.021wt% for ferrite. This difference can be important in the processing of steel. For example, in *carburizing*, the steel is heated in the austenitic state, in a carbon-containing environment, where it can absorb more carbon in solution. Thus the surface of the steel is enriched in carbon. The steel is then cooled rapidly, forming martensite at the surface. The higher carbon content combined with the martensitic structure, provides the high surface hardening, responsible for the wear resistance of the carburized steel. The *case depth*, i.e., the thickness of the carburized layer, is controlled by the diffusion of carbon in austenite. A similar surface hardening technique is *nitriding*, during which the steel surface is enriched with nitrogen.

10.2.2 CARBIDES AND NITRIDES

In supersaturated solid solutions of iron with carbon or nitrogen ($Fe - C$ and $Fe - N$), the two elements precipitate forming metastable phases. These phases exhibit a lower energy barrier for nucleation than the thermodynamically stable counterparts. So if a steel, containing 0.025 wt% carbon, is heated at $720^{\circ}C$ and then is rapidly cooled to room temperature, the resulting ferrite will be supersaturated with carbon since

the maximum solid solubility at room temperature is only 0.00005 wt% carbon. If the steel is now heated at a low temperature, it will be subjected to *aging* and the excess carbon will precipitate in the form of metastable phases and not as graphite, which is the equilibrium phase. Aging at temperatures up to 200°C causes the formation of epsilon carbide (ε -carbide) with composition between Fe_{2.4}C and Fe₃C and hexagonal structure. Aging above 200°C causes the formation of cementite with composition Fe₃C and orthorhombic structure. The nucleation of cementite takes place at the interfaces between ferrite and ε -carbide, while the cementite grows at the expense of ε -carbide. The cementite grows as plates on the {110}_a planes in the <111>_a directions. The supersaturated Fe – N solid solutions exhibit similar precipitation sequence. Aging below 200°C causes the formation of the metastable Fe₁₆N₂ nitride (phase α'') with a *BCT* structure while with aging above 200°C the metastable nitride is substituted with the γ' nitride Fe₄N with an ordered *FCC* structure. In commercial steels, nitrogen is combined with aluminum, which is added for deoxidation purposes, to form AlN. So hardening in steels is caused primarily by carbon.

10.2.3 PHASE MIXTURES – PEARLITE AND BAINITE

In addition to solid solutions (ferrite and austenite) and precipitates (carbides and nitrides) there are two important phase mixtures, which form in steels, during cooling from the austenitic region. The first is *pearlite*, which is a mixture of ferrite and cementite and forms with the eutectoid reaction during the slow cooling of austenite. The morphology of pearlite is lamellar, with a stacking sequence of alternate lamellae of ferrite and cementite. The formation of pearlite will be discussed in detail in the following sections.

The second phase mixture is *bainite*, which is also a mixture of ferrite and cementite. Bainite forms during the fast cooling of austenite or by the isothermal transformation of austenite at temperatures lower than those required for the formation of pearlite. The details of bainite formation will also be discussed in detail in the following sections.

10.3 THE FE-C PHASE DIAGRAM

An initial discussion on the binary Fe – C phase diagram has been presented in Chapter 4. The diagram, depicted in Figure 10.1, represents the equilibrium between iron and cementite. Alloys with carbon content up to 2wt% are classified as *steels*, while alloys with carbon content greater than 2wt% are classified as *cast irons*. The difference in carbon solid solubility between ferrite and austenite is reflected in the extent of the respective phase fields. The maximum carbon solid solubility in austenite is 2.1wt% at 1149°C while the respective value in ferrite is just 0.021wt% at 723°C. The Fe – C phase diagram contains three invariant reactions: peritectic, eutectic and eutectoid. The eutectic reaction is important in cast irons, while the eutectoid reaction affects both steels and cast irons. A magnification of the lower left corner of the diagram, containing the eutectoid reaction, is depicted in Figure 10.2. The diagram

contains two critical temperatures. The first is A_1 , which is the temperature of the eutectoid reaction, $723^\circ C$, represented by line BED . The second is A_3 , the temperature above which the steel is fully austenitic, represented by line AE . Temperatures A_1 and A_3 are determined experimentally, by dilatometry or thermal analysis and depend on heating and cooling rates. Steels that do not contain major alloying elements, other than carbon, are called *carbon steels*.

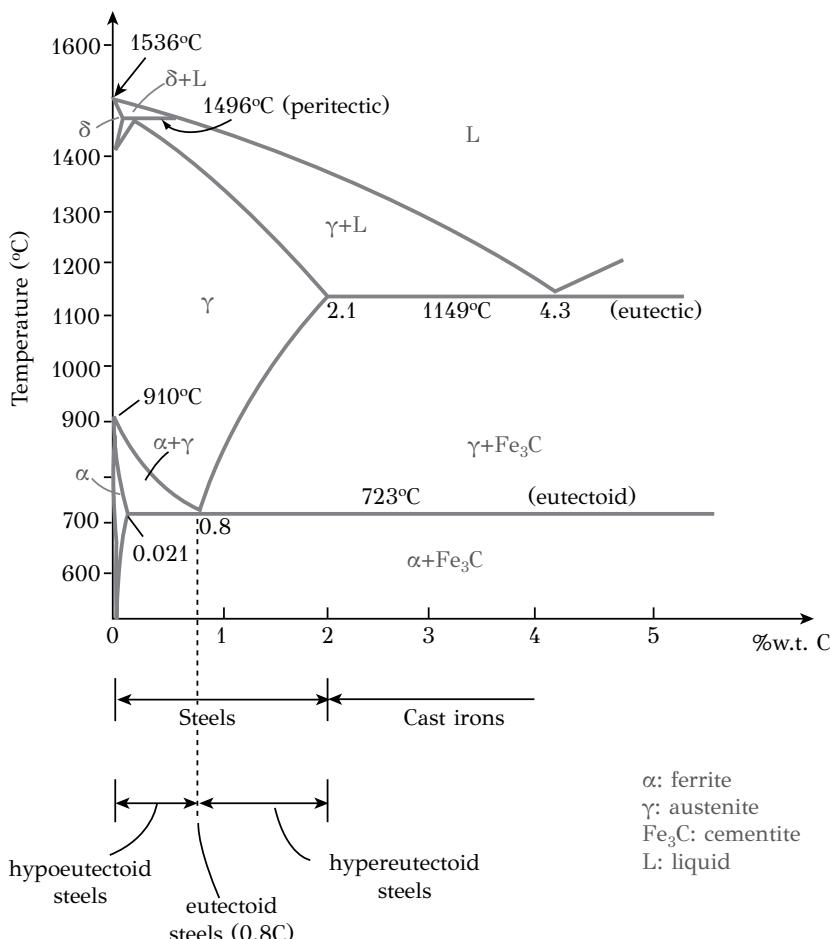


Figure 10.1: *Fe-C* equilibrium phase diagram.

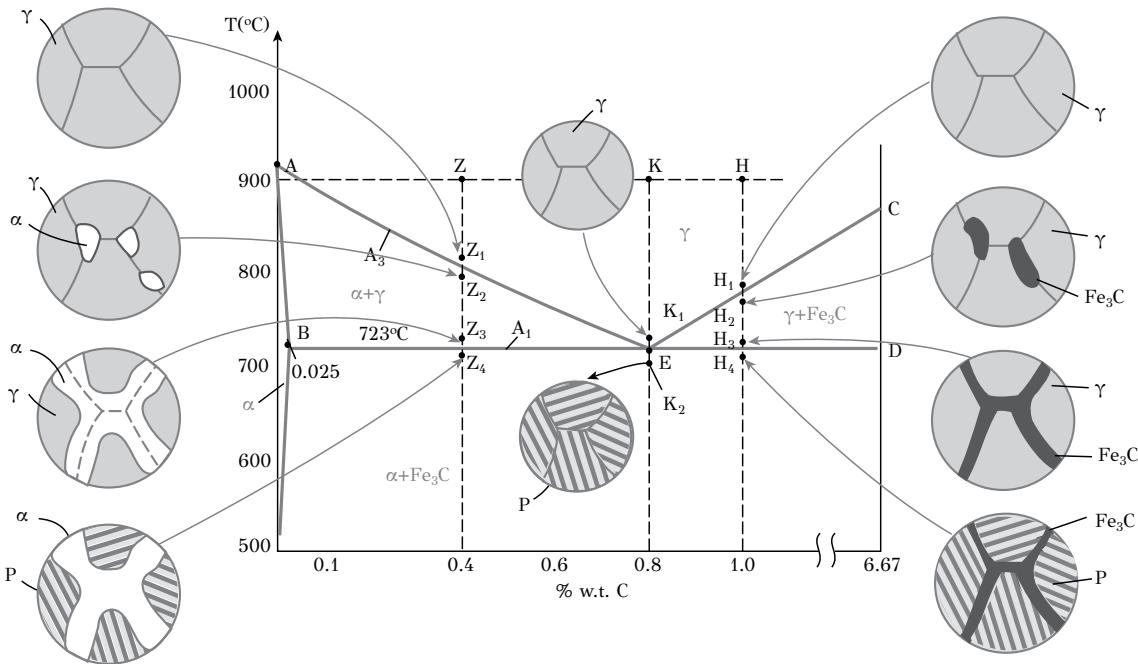
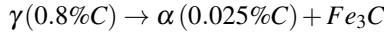


Figure 10.2: Section of the $\text{Fe}-\text{C}$ phase diagram with respective microstructures during equilibrium cooling of eutectoid, hypo-eutectoid and hyper-eutectoid steels.

Depending on the carbon content, these steels are distinguished in *eutectoid* (wt%*C*=0.8), *hypoeutectoid* (wt%*C*<0.8) and *hypereutectoid* (wt%*C*>0.8). When a eutectoid steel is heated in the austenitic region (point *K* in Figure 10.2) and then cooled slowly, it will be transformed to 100% pearlite according to the *eutectoid reaction*



The microstructure is depicted schematically at points *K*₁ and *K*₂ corresponding to temperatures *A*₁+θ and *A*₁-θ respectively. As discussed above, pearlite is not a phase but, instead, it is a phase mixture of ferrite and cementite. Pearlite forms by the cooperative growth of ferrite and cementite, resulting in a lamellar structure depicted in Figure 10.3a. The fractions of ferrite and cementite in pearlite can be determined

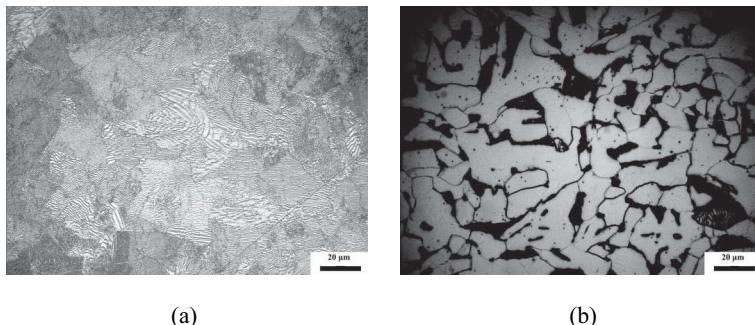


Figure 10.3: (a) Microstructure of pearlite in a eutectoid steel, (b) microstructure of proeutectoid ferrite and pearlite in a 0.4C steel.

by the lever rule at *A*₁:

$$f_a = \frac{6.67 - 0.8}{6.67 - 0.025} = 88.3\%$$

$$f_{Fe_3C} = \frac{0.8 - 0.025}{6.67 - 0.025} = 11.7\%$$

Consider now the slow cooling of a hypoeutectoid steel with 0.4wt% *C* from 900°C (point *Z* in Figure 10.2). At point *Z*₁ (*A*₃+θ) the steel is 100% austenite. On cooling below *A*₃, austenite transforms to ferrite, as depicted schematically at point *Z*₂ (*A*₃-θ). The nucleation of ferrite takes place at the prior austenite grain boundaries. During cooling from point *Z*₂ to point *Z*₃, inside the two-phase α+γ region, more ferrite forms, while the remaining austenite is enriched in carbon. The carbon enrichment of austenite follows the line *AE* of the phase diagram and is caused by the carbon rejection from the growing ferrite into austenite, due to the large difference in carbon solid solubility between the two phases. According to the lever rule, at temperature *A*₁+θ (point *Z*₃) the microstructure consists of ferrite and austenite with the following fractions

$$f_a = \frac{0.8 - 0.4}{0.8 - 0.025} = 51.6\%$$

$$f_{\gamma} = \frac{0.4 - 0.025}{0.8 - 0.025} = 48.4\%$$

The austenite has now obtained the eutectoid composition (0.8wt%*C*) and on cooling below the A_1 will be transformed to pearlite, according to the eutectoid reaction. Thus at temperature $A_1 - \theta$ (point Z_4) the steel contains 51.6% ferrite and 48.4% pearlite. This constitution is not altered significantly on further cooling to room temperature. The ferrite forming during cooling from the A_3 to the A_1 temperature is called *proeutectoid ferrite*, since it forms prior to the eutectoid reaction. As a result, hypoeutectoid steels consist of proeutectoid ferrite and pearlite, as depicted in Figure 10.3b. These steels are characterized by a ferrite-pearlite microstructure.

Consider now the slow cooling of a hypereutectoid steel with 1.0wt%*C* from 900°C (point *H* in Figure 10.2). At point *H*₁ the steel contains 100% austenite. During cooling below line *CE*, cementite precipitates from austenite. This cementite is called *proeutectoid cementite*. The nucleation of cementite takes place at the prior austenite grain boundaries as depicted in Figure 10.2 for point *H*₂. During cooling from *H*₂ to *H*₃ ($A_1 + \theta$) inside the two-phase $\gamma + Fe_3C$ region, more cementite forms, while the remaining austenite is depleted from carbon. The composition of carbon in austenite decreases, following line *CE* of the phase diagram. At temperature $A_1 + \theta$ the austenite has obtained the eutectoid composition and transforms to pearlite according to the eutectoid reaction. At temperature $A_1 - \theta$ (point *H*₄) the steel contains proeutectoid cementite and pearlite. As a result, hypereutectoid steels, containing between 0.8 and 2.0wt% carbon, consist of proeutectoid cementite and pearlite.

It is important to note that the redistribution of carbon between the phases during cooling takes place by diffusion. It is also important to note that slow cooling provides the necessary time for diffusion to take place and for thermodynamic equilibrium to be established. All phase transformations described above are diffusive phase transformations and take place by nucleation and growth, as discussed in detail in Chapter 6. Faster cooling limits diffusion and other competitive transformations, e.g., the bainitic or the martensitic transformation, take place. The phase transformations in steels will be discussed in the following sections of this chapter.

10.4 ALLOYING ELEMENTS IN STEELS

Although the most important alloying element in steel is carbon, other elements are intentionally added in steel for several reasons. The role of an alloying element is characterized by a certain degree of complexity. Alloying elements influence the stability of phases, the solid solubility of other elements, the rate of phase transformations, deformation and strengthening mechanisms as well as embrittlement phenomena.

A first attempt to discuss the role of alloying elements in steel, is to describe the basic requirements of alloying. These requirements and the respective alloying elements, which satisfy them, are listed below.

Deoxidation. It is accomplished by adding *Si* and/or *Al*. Regarding the deoxidation practice, steels are usually classified as *killed*, *semi-killed* and *rimmed*. In killed steels

deoxidation is complete and it is accomplished by adding both *Al* and *Si*. In semi-killed steels deoxidation is partial and is accomplished by adding *Si*. In rimmed steels deoxidation is limited, since only small additions of deoxidizing elements are made.

Desulfurization. The removal of sulfur is accomplished by adding sufficient amounts of *Mn*. Addition of *Mn* leads to the formation of *MnS*, which forms inclusions in the steel. However the effect of *MnS* inclusions is not so detrimental as the effect of free sulfur, which tends to segregate at the grain boundaries and cause embrittlement.

Stabilization of austenite. In certain cases it is desirable to stabilize the *FCC* austenitic structure at room temperature. This is accomplished by the addition of austenite stabilizers, such as *Ni* and *Mn*. The austenitic stainless steels and the high-*Mn* steels are good examples.

Carbide formers. The formation of carbides in steels causes strengthening. The most important carbide formers are *Cr*, *Mo*, *W*, *V*, *Nb*, *Ti* and *Ta*. Good examples are tool steels, which contain significant amounts of carbide forming elements.

Hardenability. The ability of steel to form martensite with lower cooling rates is termed hardenability. This property will be discussed in detail in later sections of this chapter. However it should be noted that the addition of alloying elements, such as *Mn*, *Si* or *Cr*, delays the eutectoid reaction and other diffusional transformations and promotes the formation of martensite at lower cooling rates.

Grain refinement. One of the major strengthening mechanisms is grain refinement or Hall-Petch strengthening, discussed in Chapter 8. Carbide or carbonitride forming elements are added. The precipitation of carbonitrides during the thermomechanical treatment of steel generates a pinning force, resisting recrystallization and grain growth. Good examples are HSLA steels, which contain microadditions of *Nb*, *Ti* or *V*.

Fracture toughness. Specific elements are added for high toughness. Nickel is the most important of these elements, as it lowers the ductile-to-brittle transition temperature. Good examples are cryogenic steels containing 9wt% *Ni*.

Oxidation and corrosion resistance. Oxidation resistance is promoted by elements which form protective oxides, such as *Cr* and *Al*. High-temperature ferritic heat-resisting steels contain 1-9 wt% *Cr*. Stainless steels contain at least 12wt% *Cr*. Several stainless steels contain also *Mo* for increased resistance to pitting corrosion.

According to the effect of the alloying elements, regarding the stabilization of either austenite or ferrite, alloying elements can be divided in two categories:

- *Austenite stabilizers:* These elements include *C*, *N*, *Ni*, *Mn*, *Co*, *Cu* and their addition expands the γ -field in the phase diagram.
- *Ferrite stabilizers:* These elements include *Si*, *Al*, *P*, *Cr*, *Mo*, *W*, *Nb*, *Sn*, *Sb* and their addition restricts the γ -field in the phase diagram.

According to the partitioning of the alloying elements during the transformation of austenite to ferrite and cementite on cooling, alloying elements are divided in two categories:

- *Partitioning in ferrite:* Ni, Cu, P and Si dissolve in ferrite.
- *Partitioning in cementite/carbides:* Mn, Cr, Mo, V, Ti, W, Nb. At low concentrations these elements dissolve in cementite (M_3C) and ferrite. At higher concentrations they form carbides, such as M_2C , $M_{23}C_6$, M_7C_3 , M_6C . Excess quantities of these elements dissolve in ferrite together with Ni and Si. Nitrogen forms nitrides or carbonitrides with Fe and other alloying elements in steel.

As stated above, alloying elements, together with the steel processing, play a key role in shaping the microstructure and properties of steels. However, it should be noted, that the influence of an element might be modified by the presence of another element. The interaction of alloying elements and their cooperative action is a complicated issue with implications on the thermodynamics and kinetics of phase transformations in steels. The description of the effect of alloying elements in steels, which follows below, is subjected to these restrictions.

Carbon. It is the most important element in steel with a large contribution to the strength of steel. Hardness, strength and hardenability increase with carbon, while at the same time toughness, ductility and weldability are reduced.

Manganese. It is present in most steels with concentrations above 0.2wt%. It is added mainly for desulfurization, forming MnS inclusions. The formation of MnS favors the workability and weldability. Mn increases the hardenability of steel. In larger concentrations Mn favors the stabilization of austenite.

Silicon. It is one of the two elements added for deoxidation. In killed steels the Si content is up to 0.6wt%, while in semi-killed steels it is lower. High Si contents degrade the workability of steel.

Aluminum. It is also added for deoxidation in steel and for grain refining. It forms AlN with nitrogen. It is added for oxidation resistance in ferritic heat-resisting steels. In larger concentrations it reduces the density of steels. A new generation of Al -containing light steels is actively investigated in the recent years.

Phosphorous. It segregates at grain boundaries and leads to a reduction of ductility and toughness, especially notch toughness in quenched and tempered steels. Phosphorous dissolves in ferrite increasing its strength. It is also beneficial in free-machining steels, where it improves the machinability.

Sulfur. It has detrimental effects on the ductility, toughness and weldability. It segregates at the grain boundaries causing embrittlement. It degrades the workability, especially hot working. Its presence increases the tendency for quench cracking in quenched and tempered steels. It has beneficial effects in free-machining steels, where it improves the machinability.

Nitrogen. It expands the γ -field and is an important austenite stabilizer, used in austenitic stainless steels. It is also used in microalloyed steels to form nitrides or carbonitrides, with carbon, during the thermomechanical processing for grain refinement.

Chromium. Is a strong carbide former, increasing the strength and wear resistance of tool steels. M_3C , M_7C_3 and $M_{23}C_6$ carbides form with increasing the chromium content. It increases the hardenability of quenched and tempered low alloy steels. Cr

promotes the oxidation resistance and is added in ferritic heat-resistant steels. *Cr* is also the major alloying element in stainless steels.

Nickel. Is a strong austenite stabilizer, non-carbide forming element in steels. It raises the ductility and toughness and reduces the ductile-to-brittle transition temperature in cryogenic steels. Important alloying element in austenitic stainless steels.

Molybdenum. Is a strong carbide former and increases the hardenability of steels. It forms carbides and causes secondary hardening during the tempering of tool steels. It increases the corrosion resistance, especially pitting resistance, of austenitic stainless steels.

Tungsten, Vanadium, Niobium, Titanium. They are carbide formers, causing an increase in strength and hardness. They cause grain refinement during working at high temperatures. They are added in tool steels for increased hardness at high temperatures. Microalloying additions of *Nb*, *V* and *Ti* are used for grain refinement in *HSLA* steels.

Cobalt. It decreases the hardenability of steel, except where *Cr* is present. It raises the M_s temperature and decreases the amount of retained austenite. It resists recovery and grain growth during tempering and promotes precipitation hardening, contributing to the strength of tool steels at high temperatures.

10.5 PHASE TRANSFORMATIONS IN STEELS

10.5.1 ISOTHERMAL AND CONTINUOUS COOLING TRANSFORMATIONS

As in all metallic alloys, the mechanical properties of steels are influenced by its microstructure and more importantly by the phase constitution of the steel. The phase constitution evolves by phase transformations, which take place during the heat treatment or the thermomechanical treatment of steel. Considering the $Fe - C$ phase diagram, it is important to say, from the beginning of this discussion, that all carbon steels, during cooling, enter the austenitic region. On further cooling, it is the austenite, which transforms to other phases, such as ferrite and pearlite, under equilibrium conditions. As will be discussed in the following sections, the austenite may transform to other phases or other phase mixtures, such as bainite or martensite, depending on the transformation temperature or the cooling rate. So when we consider phase transformations in steels, we mainly talk about the transformation of austenite, i.e., the austenite is the *parent phase*, which transforms to other phases or phase mixtures. In general we distinguish two types of phase transformations of austenite: isothermal and continuous cooling transformations. Take for example a eutectoid steel (0.8wt%*C*) in the austenitic region of the phase diagram. In the isothermal transformations, the steel is brought to a temperature below the A_1 temperature where the austenite transforms isothermally to pearlite or bainite, depending on the transformation temperature. In continuous cooling transformations, the steel is cooled continuously from the austenitic region and the steel transforms during cooling to pearlite, bainite or martensite, depending on the cooling rate. These transformations are depicted in the simplified diagram of Figure 10.4. The precise representation of these transformations is accomplished with the isothermal transformation (*IT*) or the

continuous cooling transformation (*CCT*) diagrams, to be discussed in the following sections. It is, however, important to discuss first, the main aspects of the phase transformations of austenite to ferrite, pearlite, bainite and martensite.

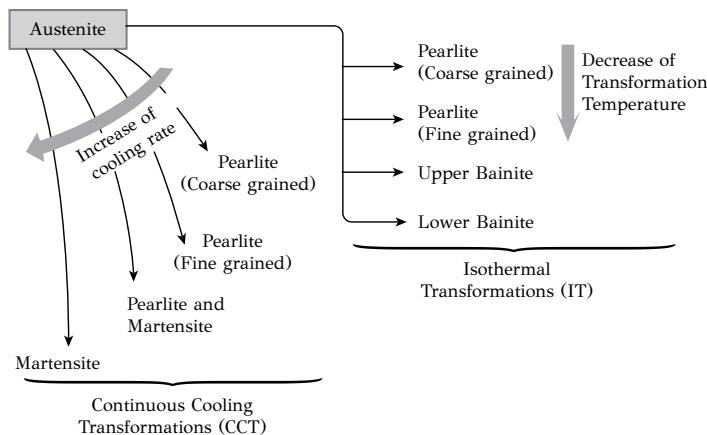


Figure 10.4: Simplified diagram of isothermal and continuous cooling transformations in a eutectoid steel.

10.5.2 THE TRANSFORMATION OF AUSTENITE TO FERRITE ($\gamma \rightarrow \alpha$)

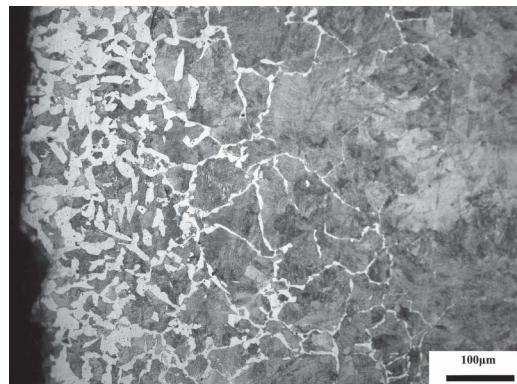
As discussed in section 10.3, during slow cooling of a hypoeutectoid steel (<0.8wt%*C*) between the A_3 and A_1 temperatures, the austenite transforms to proeutectoid ferrite. The same transformation takes place when the steel is instantaneously cooled to a temperature below the A_1 , say 600°C. The austenite will transform isothermally to ferrite. The transformation is diffusive and takes place by the nucleation and growth of ferrite. The nucleation of ferrite is heterogeneous and takes place at grain boundaries or other structural defects such as dislocations. The morphology of proeutectoid ferrite depends strongly on the transformation temperature. In general, with the reduction in transformation temperature, three morphologies are distinguished: (a) intergranular ferrite, (b) Widmanstätten ferrite and (c) transgranular ferrite.

Intergranular ferrite nucleates at higher transformation temperatures (e.g., 800°C) at the prior austenite grain boundaries. The ferrite grows both along the boundary as well as towards the grain interior, resulting in a morphology known as equiaxed ferrite, depicted in Figure 10.5a.

The orientation relationship with the parent austenitic phase follows the Kurdjumov-Sachs relation

$$\{111\}_\gamma // \{110\}_\alpha, <1\bar{1}0>_\gamma // <1\bar{1}1>_\alpha$$

According to the above relationship, the {110} planes of ferrite are parallel to the close-packed {111} planes of the austenite while the close-packed $<1\bar{1}1>$ di-



(a)



(b)



(c)

Figure 10.5: Morphology of proeutectoid ferrite in a 0.4C steel: (a) intergranular ferrite, (b) Widmanstätten ferrite, (c) transgranular ferrite.

rections of ferrite are parallel to the $<1\bar{1}0>$ close-packed directions of austenite. The ferrite/austenite interface is incoherent and exhibits high mobility at the high transformation temperatures.

Widmanstätten ferrite forms at lower transformation temperatures, as plates with a saw teeth or feather-like morphology, as depicted in Figure 10.5b. During growth, the Widmanstätten ferrite maintains the Kurdjumov-Sachs orientation relationship with austenite. The interfaces between Widmanstätten ferrite and austenite are coherent or semicoherent.

Transgranular ferrite forms at even lower temperatures in the interior of prior austenite grains (Figure 10.5c). The nucleation of ferrite in the grain interiors is caused by the large nucleation driving force, which activates heterogeneous nucleation sites with a lower potency than the grain boundaries.

In $Fe - C - X$ alloys, where X is a substitutional element, e.g., Mn , the $\gamma \rightarrow \alpha$ transformation exhibits the same morphological characteristics as in plain carbon steels. However the transformation kinetics is slower, since during transformation, in addition to carbon, other substitutional elements partition between the two phases. For example, in a $Fe - C - Mn$ steel, C and Mn partition to austenite during the growth of ferrite. Since carbon diffusion is faster than the manganese diffusion, the overall transformation rate is controlled by the manganese diffusion in austenite.

10.5.3 THE TRANSFORMATION OF AUSTENITE TO PEARLITE ($\gamma \rightarrow P$)

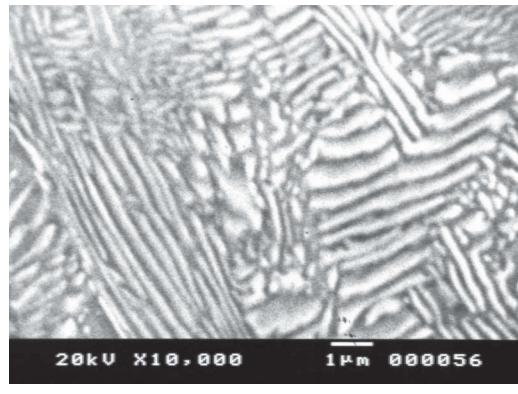
The austenite transforms below the eutectoid temperature, A_1 , to pearlite according to the eutectoid reaction, as discussed in section 10.3. Pearlite exhibits a lamellar morphology with sequential plates or lamellae of ferrite and cementite, as depicted in Figure 10.6. In these terms, pearlite forms by the cooperative growth of ferrite and cementite. In this structure, the ferrite is the continuous phase. The pearlitic structure exhibits, therefore, a periodicity, expressed by the interlamellar spacing λ_p . The fine structure of pearlite is also expressed by the magnitude of λ_p . The shorter the interlamellar spacing the finer the pearlite. This has a direct consequence on the mechanical properties of pearlite, and more importantly strength and hardness. Fine pearlite exhibits a significantly higher hardness than coarse pearlite, due to the higher density of ferrite/cementite interfaces, acting as obstacles to dislocation glide.

The interlamellar spacing depends on undercooling ΔT , below the equilibrium $T_e = A_1$ temperature. In a simplified treatment, if the strain energy associated with pearlite formation is neglected, then the driving force should at least be equal to the interfacial energy, associated with the formation of the new ferrite/cementite interfaces as

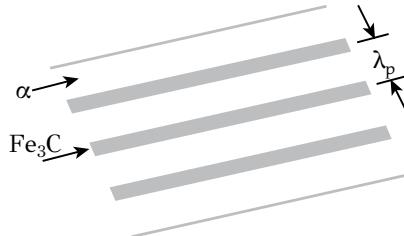
$$\Delta G = \frac{\Delta H \cdot \Delta T}{T_e} = \frac{2\gamma}{\lambda_p}$$

from which the interlamellar spacing can be deduced as

$$\lambda_p = \frac{2\gamma T_e}{\Delta H} \cdot \frac{1}{\Delta T}$$



(a)



(b)

Figure 10.6: (a) High magnification of pearlitic microstructure in a eutectoid steel, (b) definition of interlamellar spacing.

The interlamellar spacing is inversely proportional to undercooling ΔT . Thus when the transformation of austenite to pearlite takes place at low temperatures, the resulting pearlite is finer (shorter λ_p) and exhibits higher hardness than the pearlite formed at higher transformation temperatures. For example, the pearlite formed just below the A_1 temperature (e.g., 700°C) has $\lambda_p=10^{-3}\text{mm}$ and a hardness of 20 RC. The pearlite formed at a lower temperature, say 600°C , has $\lambda_p=10^{-4}\text{mm}$ and a hardness of 30 RC. It should however be noted that the fineness of the pearlitic microstructure is limited by the available driving force for the transformation.

The kinetics of the pearlite formation can be described with the methods presented in Chapter 6. The overall rate of the concurrent nucleation and growth of pearlite can be expressed by the Johnson-Mehl-Avrami equation, if the nucleation rate of pearlite is defined. In the case of pearlite, the nucleation rate is not constant, since the heterogeneous nucleation sites, such as grain boundaries, are consumed quickly with the transformation. In this case the pearlite fraction is given by equation 6.17,

which is reproduced here

$$f_p = 1 - \exp \left\{ -\frac{4}{3} \pi N_v^0 u^3 t^3 \right\}$$

where N_v^0 is the initial density of nucleation sites and u the growth rate of pearlite. The above relation results in the familiar sigmoidal transformation curve of Figure 10.7a. Regarding the effect of temperature, the kinetics of the pearlite formation follows a C-curve kinetics behavior, discussed in Chapter 6, section 6.5. At low undercooling, the diffusion is fast but the driving force is small. At high undercooling, the driving force is high but the diffusion is slow. In both cases the transformation rate is low. The maximum transformation rate is accomplished at intermediate transformation temperatures, leading to the familiar C-curve of transformation, depicted in Figure 10.7b. This curve forms the basis of the IT and CCT diagrams to be discussed in the following sections.

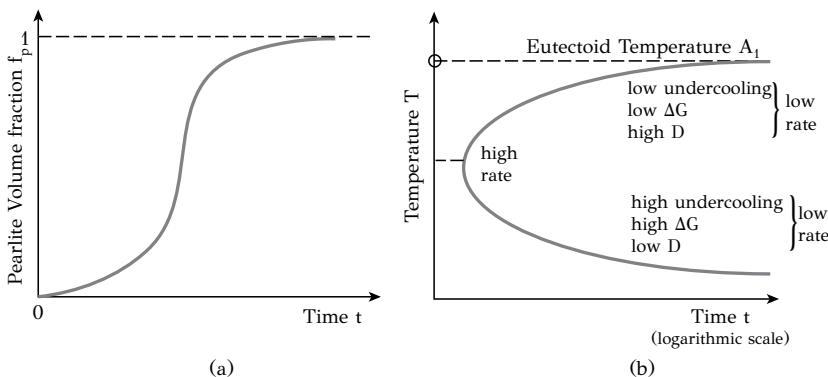


Figure 10.7: (a) Isothermal transformation of austenite to pearlite, (b) effect of temperature and C-curve kinetics.

The growth of pearlite in carbon steels takes place under diffusion control, while the rate-limiting step is the diffusion of carbon in austenite. Thus, in carbon steels, the formation of pearlite is fast. The addition of substitutional alloying elements, delays the pearlite formation for two reasons: (a) these elements should also diffuse and partition between phases and (b) some elements, such as B , segregate at the austenite grain boundaries and delay the nucleation of pearlite. As will be discussed in the following sections, a major role of alloying elements in steels is the delay of the pearlite reaction, during the cooling of steel from the austenitic region. The reason is that it is more desirable, that another competitive transformation takes place, the martensitic transformation, in order to form a harder phase, martensite.

Regarding the mechanical properties of pearlite, as discussed above, the strength and hardness increase with the reduction in interlamellar spacing. In hypoeutectoid steels the increase of the fraction of pearlite with carbon content, leads to an increase of both yield strength and ultimate tensile strength. However the presence of pearlite

leads to a decrease in ductility and fracture toughness, since the ferrite/cementite interfaces act as crack nucleation sites while the increased strain hardening potential of the pearlitic structure limits the plasticity ahead of the crack tip. As a result, in plain carbon steels, the ductile-to-brittle transition temperature (DBTT) increases with carbon content, as depicted in Figure 10.8.

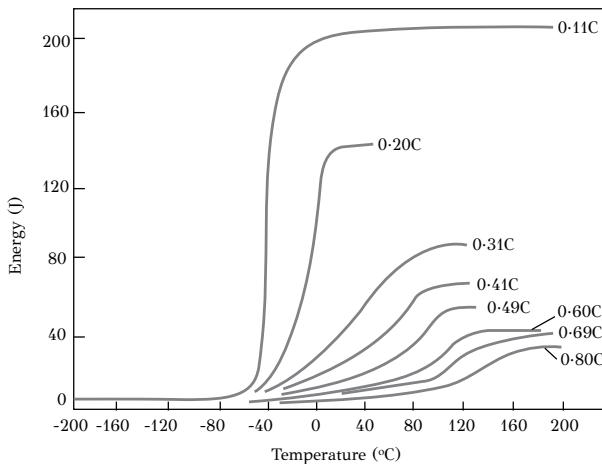


Figure 10.8: Effect of carbon (pearlite content) on the ductile-to-brittle transition temperature.

10.5.4 TRANSFORMATION OF AUSTENITE TO MARTENSITE ($\gamma \rightarrow M$)

The martensitic transformation is probably the most significant transformation in steels with a great technological impact. The characteristics of martensitic transformations including the phenomenological theory of the shape deformation, the nucleation and the morphology of martensite as well as the mechanical effects associated with martensite formation have been discussed in Chapter 6, section 6.8. The martensitic transformation is a displacive, diffusionless transformation. The strain energy of the transformation arises from shear-like displacements and influences the transformation kinetics and morphology of martensite. In steels, the martensitic transformation takes place during the rapid cooling of austenite, termed *quenching*. The cooling rate should be large enough so other competitive diffusional transformations, such as the formation of ferrite, pearlite or bainite, will not be activated. The critical cooling rate depends on chemical composition and the austenite grain size and characterizes the hardenability, which is a steel property to be discussed in the following sections.

The transformation is generally *athermal*, since the fraction of martensite increases with the drop in temperature. The nucleation of martensite is heterogeneous and takes place at structural defects as dislocation arrays, located at prior austenite grain boundaries. The martensitic transformation starts at the M_s temperature and ends at the M_f temperature, as depicted in Figure 10.9. At the M_s temperature the

driving force takes the critical value to overcome the energy barrier associated with the nucleation of the transformation.

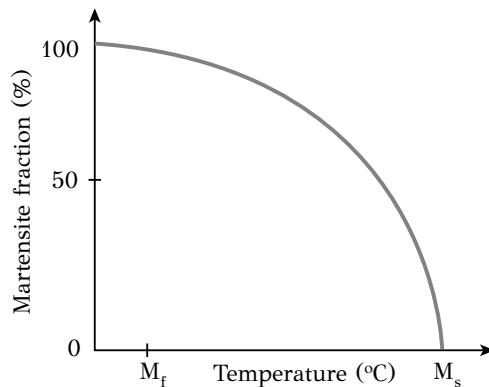


Figure 10.9: Athermal martensitic transformation: the martensite fraction increases on cooling between the M_s and M_f temperatures.

The M_s temperature exhibits a strong dependence from the alloying elements in the steel. The austenite stabilizers, such as C , N , Mn , Ni , reduce the M_s temperature, since a larger driving force is required for martensite nucleation. In addition, elements, such as Cr , Mo , Al , Si , W , and V , which are not austenite stabilizers, also reduce the M_s temperature. This is attributed to the solid solution hardening, hindering dislocation glide in the austenite, which is necessary for martensite nucleation. The effect of alloying elements on the M_s temperature is described by several empirical relationships, such as the Andrews equation:

$$M_s(C) = 539 - 463C - 30.4Mn - 17.7Ni - 12.1Cr - 7.5Mo$$

where the chemical symbols denote the alloying content in wt%.

The M_f temperature corresponds to the temperature where the martensitic transformation stops. However this may not correspond to the completion of the transformation. *Mechanical stabilization* of the austenite is one of the reasons, the transformation does not go into completion. It is attributed to the build up of strains in the austenitic lattice, which hinder the motion of glissile martensitic interfaces. The untransformed austenite, following a quench, is termed *retained austenite*.

The formation of martensite in carbon steels is associated with significant hardening, a fact that is not observed in martensitic transformations in other alloy systems. The high hardness of martensite is attributed to carbon in interstitial sites in the *BCT* martensite lattice. The tetragonality of the lattice is due to the occupation of octahedral interstitial sites and is expressed by the c/a ratio of the lattice parameters of the martensite unit cell as a function of carbon content:

$$c/a = 1 + 0.045 (\text{ %w.t. } C)$$

Taking into account that the martensitic transformation is diffusionless, carbon should occupy the same interstitial positions in austenite and martensite. The occupation of octahedral interstitial sites in austenite causes symmetric distortions, which lead to moderate strengthening. However, the occupation of the same sites in the *BCT* lattice of martensite, causes non-symmetric distortions and lead to significant solid solution strengthening, as discussed in Chapter 8. The difference in hardness between austenite and martensite, is depicted as a function of carbon, in Figure 10.10. In this particular case the austenite has been retained at room temperature with the

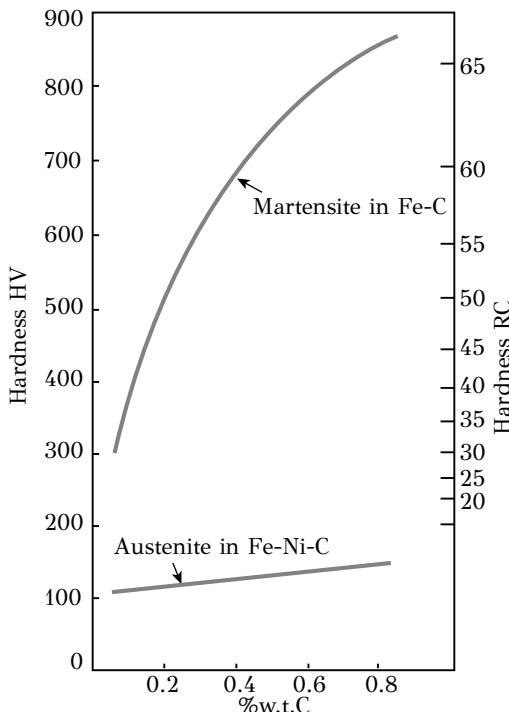


Figure 10.10: Effect of carbon content on the hardness of martensite and austenite.

addition of nickel. As discussed in Chapter 8, the solid solution strengthening due to carbon is described by the relation

$$\Delta\tau = k\zeta^{3/2} X_C^{1/2}$$

where ζ is the tetragonality factor

$$\zeta = \varepsilon_{11} - \varepsilon_{22}$$

which expresses the non-symmetric distortions caused by the carbon atoms. Strengthening mechanisms, other than solid solution hardening associated with carbon, are operational. These include strengthening arising from the presence of

boundaries and interfaces, leading to a Hall-Petch type strengthening, either from the prior austenite grain size or the mean size of martensite laths.

10.5.5 TRANSFORMATION OF AUSTENITE TO BAINITE ($\gamma \rightarrow B$)

Bainite is the second phase mixture, in addition to pearlite, which forms in carbon steels. Bainite forms by the decomposition of austenite at temperatures above the M_s but below that at which pearlite forms. While pearlite forms by the cooperative growth of ferrite and cementite, bainite forms by the growth of supersaturated ferrite followed by the precipitation of cementite as a result of carbon diffusion. The supersaturated ferrite forms by a displacive, shear-like transformation, which resembles the martensitic transformation, i.e., the associated shape strain is an invariant plane strain (IPS), discussed in Chapter 6. Depending on the transformation temperature, the cementite precipitates either between or inside the bainitic ferrite laths and the resulting structures correspond to *upper and lower bainite* respectively, as depicted in Figure 10.11.

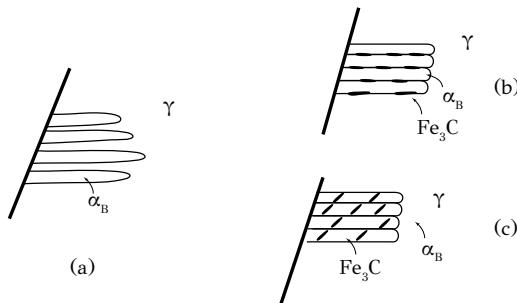


Figure 10.11: Formation of bainite: (a) formation of bainitic ferrite, (b) upper bainite, (c) lower bainite.

Upper bainite forms in two stages. In the first stage (Figure 10.11a) the bainitic ferrite forms by a displacive mechanism as supersaturated ferrite. In the second stage, cementite precipitates by diffusion from the supersaturated ferrite. Carbide precipitation takes place between the bainitic ferrite laths (Figure 10.11b). Upper bainite forms in *subunits* of parallel ferrite laths, called *bainite sheaves*. The accommodation of the transformation strain is accomplished by the plastic deformation of austenite. The associated increase of dislocation density in austenite suppresses the motion of the glissile interfaces and the growth of bainitic ferrite stops. As a result each bainite sheave has a size, which is shorter than the prior austenite grain size. The orientation relationship between the bainitic ferrite and austenite is according to Kurdjumov-Sachs

$$\{111\}_\gamma // \{110\}_{\alpha_B}, \quad <1\bar{1}0>_\gamma // <1\bar{1}1>_{\alpha_B}$$

while the orientation relationship between cementite and austenite is according to Pitch

$$[001]_{Fe_3C} // [\bar{2}\bar{2}5]_\gamma$$

$$[010]_{Fe_3C} // [\bar{1}\bar{1}0]_\gamma$$

$$[100]_{Fe_3C} // [5\bar{5}4]_\gamma$$

In alloy steels, iron and substitutional elements do not diffuse during the bainitic ferrite formation. The addition of *Si* or *Al* retard the cementite precipitation. In this case a *carbide-free bainite* is obtained, consisting of bainitic ferrite and retained austenite.

Lower bainite also forms in two stages. The first stage is similar to upper bainite. However in the second stage, cementite precipitates within the bainitic ferrite laths (Figure 10.11c). The transformation takes place at lower temperatures than those of upper bainite. The orientation relationship between cementite and bainitic ferrite is according to Bagaryatski:

$$[001]_{Fe_3C} // [\bar{1}01]_{\alpha_B}$$

$$[100]_{Fe_3C} // [111]_{\alpha_B}$$

$$[010]_{Fe_3C} // [\bar{1}2\bar{1}]_{\alpha_B}$$

The carbide dispersion in lower bainite is finer than in upper bainite and this in part explains the higher ductility of lower bainite.

Conventional bainitic steels do not exhibit the superior strength-toughness combination of tempered martensitic steels due to the relatively coarse carbide dispersion present in bainite. However, with the addition of *Si* or *Al*, the cementite precipitation can be suspended and a carbide-free bainite can form instead, as discussed above. This carbide-free bainite forms the basis of many modern steels, such as the TRIP steels used primarily in the automotive industry and piping steels for the oil and gas industry.

10.5.6 ISOTHERMAL TRANSFORMATION DIAGRAMS

As discussed in section 10.5.1, the isothermal transformations of austenite can be visualized with the Isothermal Transformation (*IT*) diagram, which is also usually denoted as the Time-Temperature-Transformation (*TTT*) diagram. The *IT* diagram for the eutectoid steel is depicted in Figure 10.12. The vertical axis is the temperature and the horizontal axis is time, in logarithmic scale. The two *C* curves correspond to the start and finish of the austenite decomposition. The intermediate dotted line corresponds to 50% of the transformation. The *C* curve for the start of the pearlite transformation approaches asymptotically, at large times, the horizontal dotted line, which corresponds to the A_1 eutectoid temperature. The *C* shape of the curves corresponds to the characteristic *C*-curve kinetics of nucleation and growth transformations discussed in Chapter 6. The two horizontal lines correspond to the temperatures M_s and M_f for the start and finish of the martensitic transformation. These lines are horizontal due to the athermal character of the martensitic reaction. The *IT* diagram

provides information on which phases form during the isothermal transformation of austenite, when the steel is cooled below the eutectoid temperature, where the austenite is unstable.

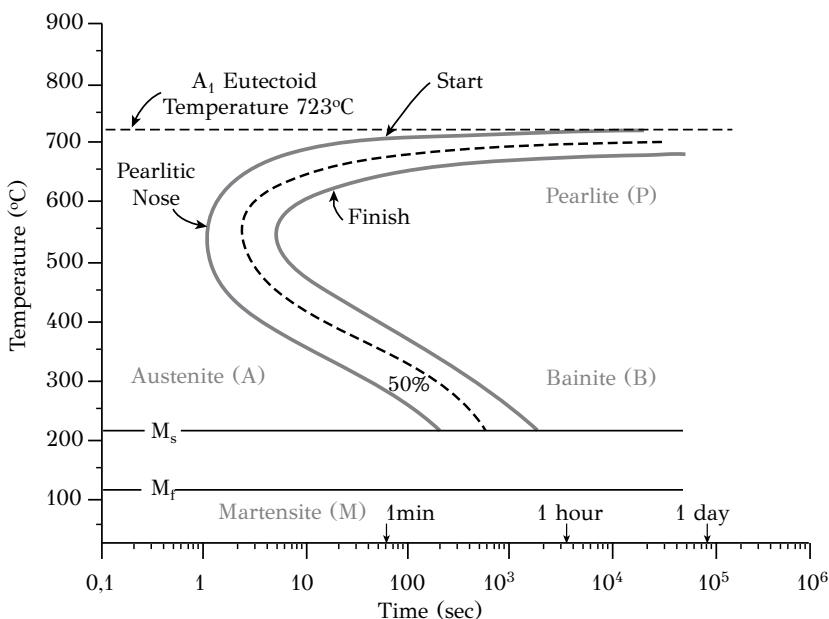


Figure 10.12: Isothermal transformation (IT) diagram for a eutectoid steel (0.8C).

In order to understand the *IT* diagram, consider five (5) specimens of eutectoid steel (0.8wt%*C*), which are austenitized at 800°C, as depicted in Figure 10.13a. The structure of these specimens is also represented by point O in the phase diagram of Figure 10.13b. The specimens are then rapidly cooled to specific temperatures where they remain for isothermal transformation. Specimen 1 is cooled to 650°C, a relatively high temperature, where it remains for 2000 sec (33 min) and then is quenched to room temperature. The horizontal line at 650°C crosses both the start and finish of the pearlite transformation, meaning that all austenite transforms to pearlite. Because the transformation temperature is high, the pearlite will be coarse, i.e., the interlamellar spacing, is large leading to low hardness. Specimen 2 is cooled to 550°C and remains for 700 sec (12 min). The pearlite transformation is completed in a shorter time than at 650°C. The fraction of pearlite vs. time curves for both 650 and 550°C are depicted in Figure 10.13c. The pearlite, which forms at 550°C is finer and exhibits higher hardness. Specimen 3 is cooled to 450°C and remains for 300 sec (5 min). The austenite is transformed completely to upper bainite. Specimen 4 is cooled to 350°C and remains for 200 sec (3.3 min). The austenite is transformed isothermally to lower bainite. The transformation curves for both 450 and 350°C are depicted in Figure 10.13d. Finally specimen 5 is quenched from 800°C to room temperature. The cooling curve of specimen 5 does not cross the C curves, however it crosses

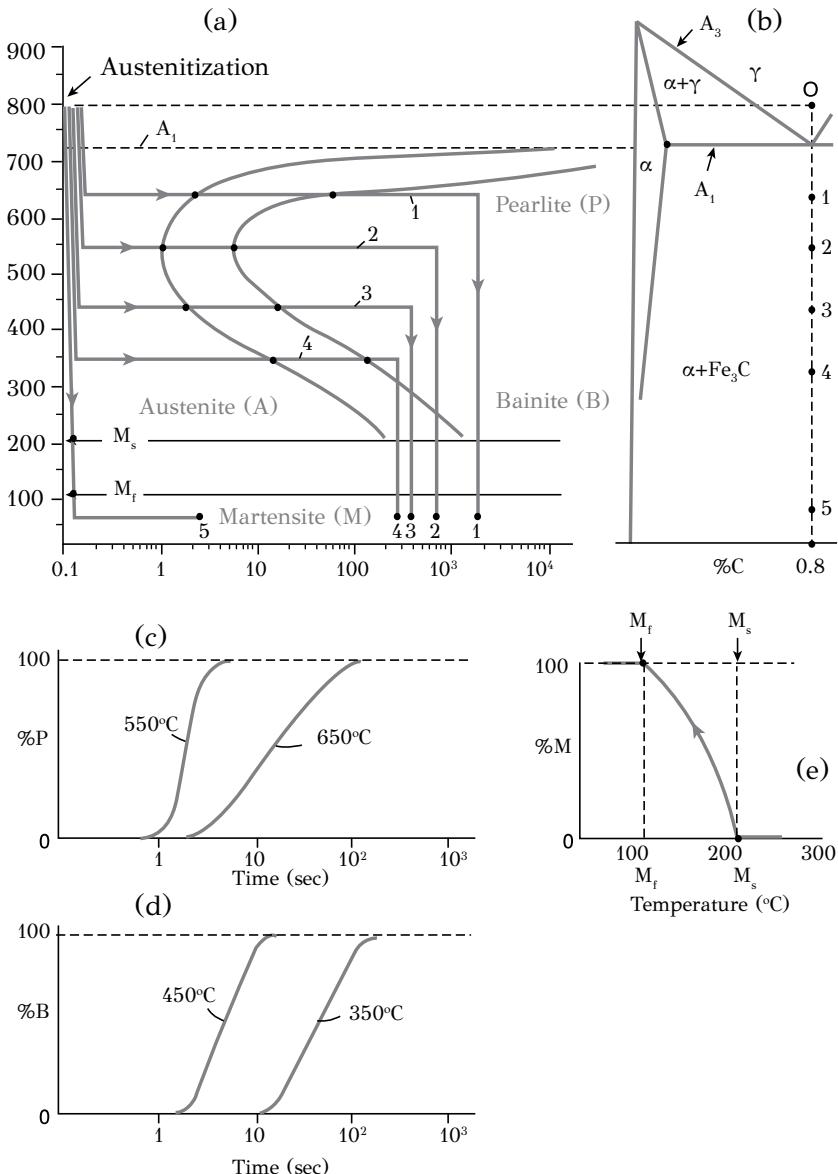


Figure 10.13: (a) IT diagram for a eutectoid steel, (b) section of the Fe-C phase diagram, (c) isothermal transformation of austenite to pearlite at 650 and 550°C, (d) isothermal transformation of austenite to bainite at 450 and 350°C, (e) martensite fraction as a function of temperature.

both the M_s and M_f lines and the austenite transforms to martensite. The martensite is not formed by isothermal transformation but by continuous cooling from the M_s to the M_f temperature. The transformation curve for martensite is depicted in Figure 10.13e.

An example, where the transformation is not allowed to proceed to completion is shown in Figure 10.14. Consider again the *IT* diagram of the same eutectoid steel

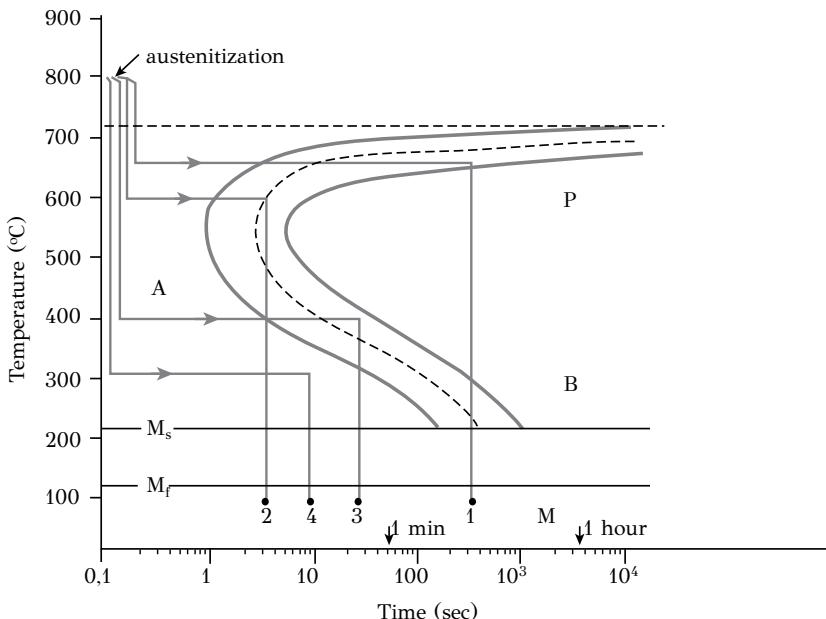


Figure 10.14: Isothermal transformation of four specimens of eutectoid steel.

and four (4) specimens austenitized at 800°C . Specimen 1 is cooled to 650°C where it remains for 300 sec and then it is quenched to room temperature. The pearlitic transformation proceeds up to the point where 90% pearlite has formed. With quenching from 650°C , the remaining austenite transforms to martensite. The final microstructure of specimen 1 will consist of 90% pearlite and 10% martensite. Specimen 2 is cooled to 600°C , where it remains for 3 sec and is quenched to room temperature. The final microstructure consists of 50% pearlite and 50% martensite. Specimen 3 is cooled to 400°C , where it remains for 30 sec and is quenched to room temperature. During the isothermal holding 80% bainite has formed. The final microstructure consists of 80% bainite and 20% martensite. Finally specimen 4 is cooled to 300°C , where it remains for 10 sec and is quenched to room temperature. The horizontal line does not cross the C curves and no pearlite or bainite forms. The specimen remains austenitic. Upon quenching, the austenite transforms to martensite. The final microstructure of specimen 4 is 100% martensite.

In all the above cycles, the cooling curves corresponding to the initial rapid cool-

ing to the isothermal holding temperature pass on the left of the *pearlitic nose*, so that no diffusional transformations are activated during cooling. It is obvious that in order to get a 100% martensitic structure, the cooling curve should pass on the left of the pearlitic nose and cross both the M_s and M_f lines.

Consider now the *IT* diagram of a hypoeutectoid steel, such as a 0.4wt% steel, depicted in Figure 10.15a. Relative to the *IT* diagram of the eutectoid steel, the *C*

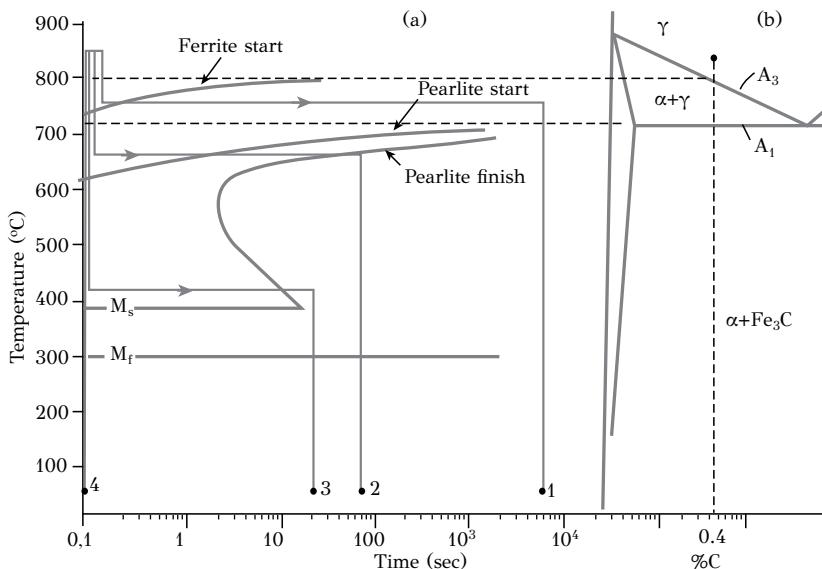


Figure 10.15: (a) Isothermal transformation (*IT*) diagram for a hypo-eutectoid steel (0.4C), (b) Section of the *Fe-C* phase diagram.

curves are displaced to the left, to shorter times, meaning that the diffusional transformations take place at a higher rate. In fact the pearlitic nose is crossed by the vertical axis. An additional difference is the presence of an additional *C* curve corresponding to the start of the formation of proeutectoid ferrite. This line extends asymptotically to the A_3 temperature of the steel, as depicted in the phase diagram of Figure 10.15b. Consider now four (4) specimens austenitized at $850^\circ C$. Specimen 1 is cooled rapidly to $750^\circ C$ (above the eutectoid temperature) where it remains for 7000 sec (about 2 hours). The horizontal line crosses the start of ferrite formation and the austenite transforms to proeutectoid ferrite. Due to the long holding at $750^\circ C$, thermodynamic equilibrium is established and the fraction of ferrite formed is the one predicted by the phase diagram. When the specimen is quenched to room temperature, the remaining austenite is transformed to martensite. The final microstructure of specimen 1 consists of ferrite and martensite. Specimen 2 is cooled to $650^\circ C$, where it remains for 100 sec. The pearlitic transformation is completed. However since the cooling line from 850 to $650^\circ C$ crosses the ferrite *C* curve, a small quantity of ferrite forms before the start of the pearlitic transformation. The final microstruc-

ture of specimen 2 consists of pearlite and a small quantity of ferrite. Specimen 3 is cooled rapidly to 400°C , where it remains for 30 sec. The bainitic transformation is completed. As before, the cooling curve now crosses both the ferrite and pearlite C curves, the final microstructure of specimen 3 consists of bainite and small quantities of ferrite and pearlite. Specimen 4 is quenched to room temperature and the cooling curve crosses both the M_s and M_f lines. As before, the final microstructure of specimen 4 will consist of martensite and small quantities of ferrite and pearlite. The high transformation rate, of austenite to ferrite or pearlite, does not allow the formation of a 100% martensitic structure in this steel.

10.5.7 CONTINUOUS COOLING TRANSFORMATION DIAGRAMS

The majority of heat treatments applied to steel include austenitizing followed by continuous cooling at different rates. In this case it is necessary to use the continuous cooling transformation diagrams (*CCT* diagrams). There are certain differences between *IT* and *CCT* diagrams for the same steel. The *CCT* (solid line) and *IT* (dotted line) diagrams are depicted in Figure 10.16 for the eutectoid steel. In the *CCT*

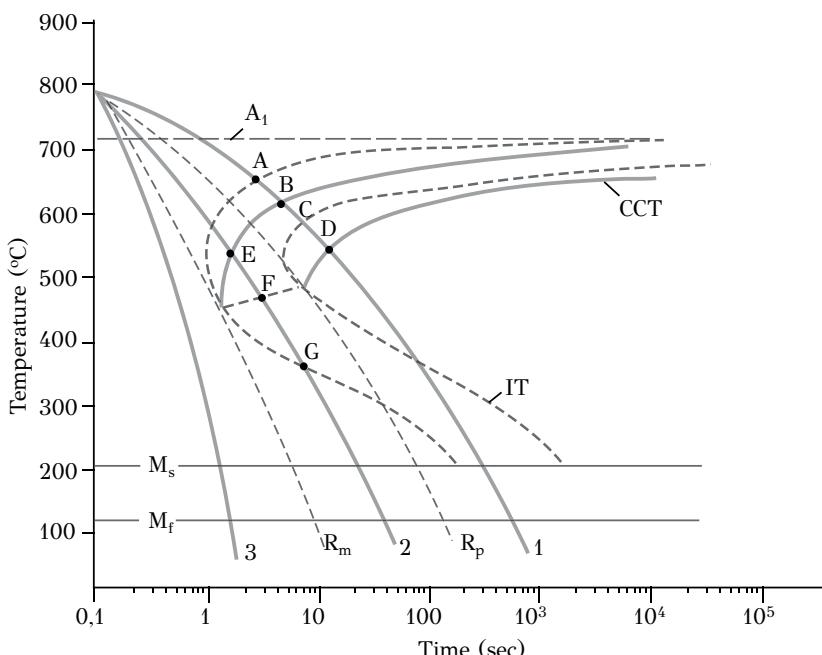


Figure 10.16: *CCT* (solid lines) and *IT* (dotted lines) diagrams for a eutectoid steel with critical cooling rates R_M and R_P .

diagram the C curves are displaced to longer times and lower temperatures relative to the *IT* diagram. This can be explained by considering the cooling curve of specimen 1, which is austenitized at 800°C . The cooling curve crosses the start line of the

IT diagram at point *A*. This point corresponds to time 2 sec, which is the time required for the start of the pearlite transformation if it was carried out isothermally at 650°C. However, during cooling to point *A*, the steel is at a higher temperature than 650°C, where more time is required for the start of the pearlite transformation. The longer time corresponds to a larger drop in temperature and, consequently, during continuous cooling, the transformation starts at a lower temperature, corresponding to point *B* on the *CCT* diagram. With the same arguments, the correspondence between points *C* and *D*, of the *IT* and *CCT* diagrams respectively, can be explained. An additional difference regarding the *CCT* diagram of the eutectoid steel is that it does not predict the formation of bainite. Actually the cooling curve of specimen 1 crosses both the start and finish curves for pearlite formation. This means that all austenite transforms on cooling to pearlite. Specimen 2 cools at a higher rate. The cooling curve crosses the start curve for pearlite but it does not cross the finish curve. This means that a quantity of austenite transforms to pearlite during cooling from point *E* to point *Z*. During cooling from *Z* to *H*, no bainite forms since the time spent at the region of bainite transformation is very short. The remaining austenite will transform to martensite. The final microstructure of specimen 2 will consist of pearlite and martensite and practically no bainite. Finally specimen 3 is cooled at a sufficiently high rate so that the cooling curve does not cross the pearlite curve. The austenite transforms to 100% martensite. Two critical cooling rates corresponding to curves R_P and R_M are indicated in the *CCT* diagram. In order to obtain 100% pearlite, the specimen should be cooled with a rate lower than R_P , as in specimen 1. In order to obtain 100% martensite, the specimen should be cooled at a rate higher than R_M , as in specimen 3.

For the application of the *CCT* diagram consider four specimens of eutectoid steel austenitized at 800°C and cooled at different rates, as depicted in Figure 10.17a. Specimen 1 is cooled inside the furnace, at a low rate. According to the *CCT* diagram of Figure 10.17b the time required for the specimen to reach room temperature is one day. The steel obtains a hardness of 15RC and elongation 30%. Specimen 2 is cooled in air and reaches room temperature in one hour. The hardness is now 25RC and elongation 20%. Specimen 3 is quenched in oil and reaches room temperature in one minute. The hardness is now 35RC and the elongation 10%. Finally specimen 4 is quenched in water and reaches room temperature in less than 2 sec. The hardness rises to 60RC but the elongation drops to the low value of 2%. Obviously, the different mechanical properties arise from the different microstructures obtained from the different cooling rates. As depicted in the *CCT* diagram, in specimen 1, the austenite transforms to coarse pearlite. This results in a low hardness (15RC) and high ductility, as indicated by the high elongation of 30%. In specimen 2, the austenite transforms to fine pearlite, since the transformation temperature is lower than before. This results in a higher hardness (25RC) and lower elongation (20%). In specimen 3, quenched in oil, the austenite transforms to pearlite and martensite. The hardness increases to 35RC. Finally in specimen 4, quenched in water, the austenite transforms to 100% martensite. The hardness climbs to 60RC, which is a very high value. As discussed in the previous sections, the high hardness of martensite

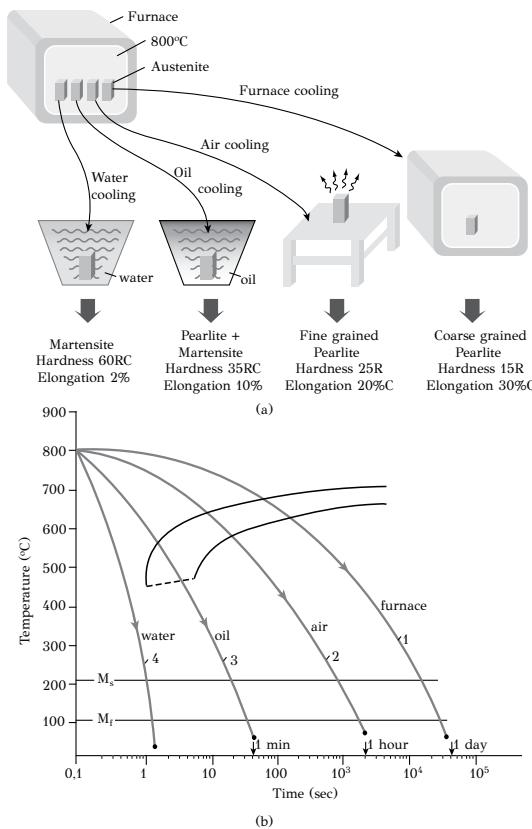


Figure 10.17: (a) Four specimens of eutectoid steel austenitized at 800°C and cooled at different rates, (b) CCT diagram of eutectoid steel indicating the different cooling rates.

arises from solid solution hardening, associated with the supersaturation with carbon at interstitial sites in the *BCT* lattice of martensite. However the high hardness is associated with a drop in ductility, as indicated by the low elongation of 2%. It is obvious that the steel should be given an extra heat treatment in order to restore ductility before it is put to service. This treatment is called *tempering* and will be discussed in a following section.

10.6 HARDENABILITY

Hardenability is an important property associated with the heat treatment of steels and in particular quenched and tempered steels. Consider two rods of eutectoid steel 1080 (0.8wt% C) with diameters 10 and 30 mm respectively. The two rods are austenitized at 800°C and then quenched in oil. The 10 mm rod is fully hardened, since

martensite has formed across the diameter of the rod. However the 30 mm rod exhibits a variation of hardness across the diameter and has formed martensite only at the surface. This is attributed to the different cooling rate between the surface and the interior of the rod. In fact every point across the diameter has a different cooling rate. Consider that the cross section is divided in three regions A, B and C, as depicted in Figure 10.18. Region A corresponds to the surface layers of the rod, region C corresponds to core of the rod and region B is intermediate between the other two regions. The corresponding cooling curves R_A , R_B and R_C are depicted in the CCT diagram of Figure 10.18.

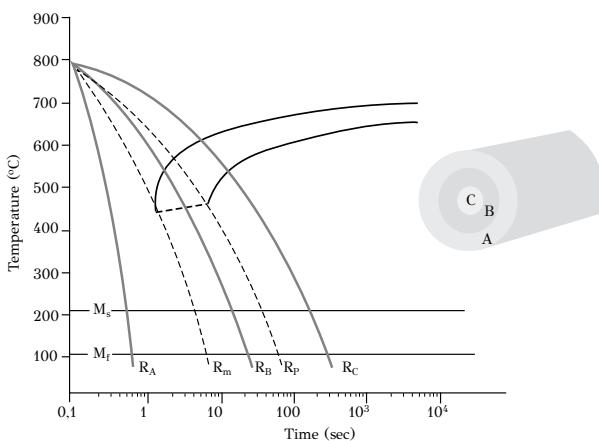


Figure 10.18: CCT diagram and cooling curves for three regions of the cross section of a rod made of eutectoid steel.

The cooling rate R_A is higher than the critical cooling rate R_M and, as a result, martensite will form at region A, at the surface of the rod. The cooling rate at the core of the rod R_C is lower than the critical rate R_P and, as a result, pearlite will form in region C. The cooling rate R_B is between R_P and R_M and, as a result, a mixture of martensite and pearlite will form in the intermediate region B. While in the small diameter rod it is possible to obtain 100% martensitic structure (full hard), the same is not possible in the larger diameter rod, where martensite forms only at the surface region. The rod is hard at the surface and soft in the interior. One way to obtain hardening in a larger portion of the rod cross section is to use a different *quenching medium* in order to achieve higher cooling rate. Consider a similar rod, 30 mm diameter, quenched in water instead of oil. As depicted in Figure 10.19a, regions A and B have transformed to martensite and the rod hardens in a larger depth from the surface. However the use of more severe quenching medium for the hardening of steels could result in two problems. The first is associated with the formation of *quench cracks*, formed due to the rapid cooling of the material. The second is *quench distortions* caused by the thermal stresses, which develop during quenching. In longitudinal elements, with

thin cross sections, the thermal stresses could cause significant bending distortions and render a part unsuitable for service. Both the above problems set severe limitations on the use of quenching mediums for the hardening of steels. The 1080 steel, discussed above, has a limited ability to fully harden by quenching. Hardenability can be defined as the ease with which a steel can form martensite. A steel, with a high hardenability forms martensite with a lower cooling rate than another steel, which requires higher cooling rate to obtain the same effect. A steel with a higher hardenability can form martensite at a larger depth from the surface than another steel with lower hardenability quenched in the same quenching medium.

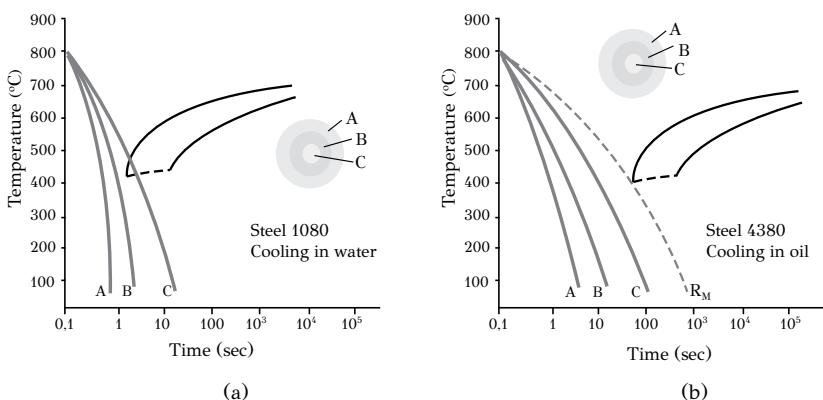


Figure 10.19: CCT diagram and cooling curves for (a) 1080 steel quenched in water and (b) 4380 steel quenched in oil.

Consider now a 30 mm diameter rod of 4380 steel. In addition to 0.8wt%*C*, the steel contains *Mn*, *Cr* and *Ni*. The rod is austenitized and then quenched in a milder quenching medium, such as oil. As depicted in the CCT diagram of Figure 10.19b, all cooling curves *R_A*, *R_B* and *R_C* are on the left of the critical *R_M* curve. This is because the *C* curves, corresponding to the pearlite transformation, moved to the right, to longer times, in relation with steel 1080. The 4380 steel has a higher hardenability than steel 1080. This means that the 4380 steel exhibits a higher potential to form martensite and harden in a larger depth by the use of milder quenching media than 1080 steel. In addition, the 4380 steel will not present the problems of quench cracks or quench distortions associated with the use of severe quenching media. Why does 4380 steel exhibit a higher hardenability? The answer is on the role of the alloying elements. *Mn*, *Cr* and *Ni* are substitutional alloying elements. In order for the eutectoid reaction to take place, in addition to carbon, the diffusion of *Cr*, *Mn* and *Ni* should take place. However the diffusion of these elements is slower than the diffusion of carbon. As a result, the pearlite formation is delayed and this is expressed by the displacement of the *C* curves to longer times in the CCT diagram.

The hardenability of steel can be expressed in various ways. One way is the maximum rod diameter, which can fully harden with martensitic transformation. Another

way is by the lowest cooling rate, which allows for complete martensitic transformation in the whole cross section. In this book, the depth of hardening (case depth) is adopted for the characterization of hardenability. The method is known as the *Jominy hardenability test* or end-quench test. According to this method, described in the ASTM A255 standard, after austenitizing, one end of the rod is quenched by a water spray, as depicted in Figure 10.20. Hardness measurements are then taken along the

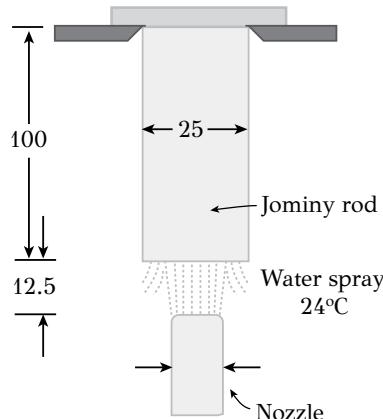


Figure 10.20: Jominy hardenability test.

length from the end of the rod and the hardenability curve is constructed, as depicted in Figure 10.21 for 4140 steel ($0.4C - 0.7Mn - 1Cr$). The hardness varies along the length of the rod due to the corresponding variation of cooling rate, as depicted by the cooling curves 1, 2 and 3 in Figure 10.21. Hardenability curves find practical application in comparing steels with different hardenability, as depicted in Figure 10.22. Steels 1040, 5140, 2340 and 4340 have identical carbon contents (0.4wt% C) but differ in the substitutional alloying elements. The addition of alloying elements improves the hardenability, allowing the steel to harden in a larger length from the quench end.

Hardenability is associated with the position of the pearlitic nose in the *CCT* diagram, i.e., the maximum rate for pearlite formation. Any factor, which delays the diffusive phase transformations, such as the formation of ferrite, pearlite or bainite, contributes to the increase of hardenability. The most important factors are: (a) carbon content and alloying elements and (b) the prior austenite grain size. We have already discussed the role of alloying elements on hardenability. Consider the effect of austenite grain size. The grain size depends on the temperature and time of the austenitizing treatment and is characterized, according to the ASTM E112 standard, by the number N , according to the relation

$$n = 2^{N-1}$$

where n is the number of grains per square inch in a metallographic specimen, which

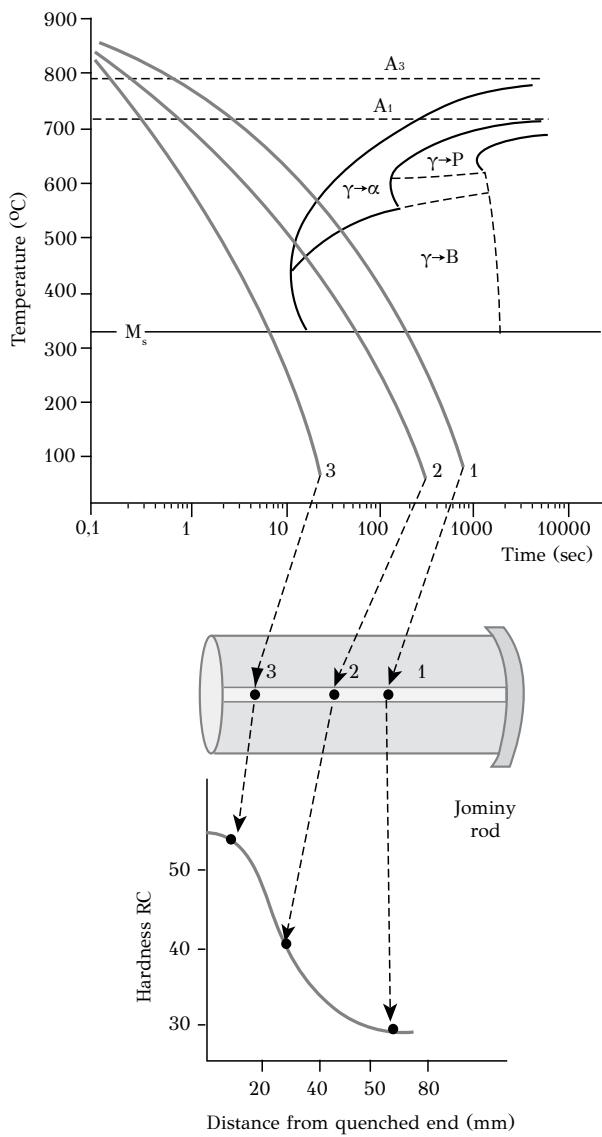


Figure 10.21: Correspondence between jominy hardenability curve and CCT diagram of 4140 steel.

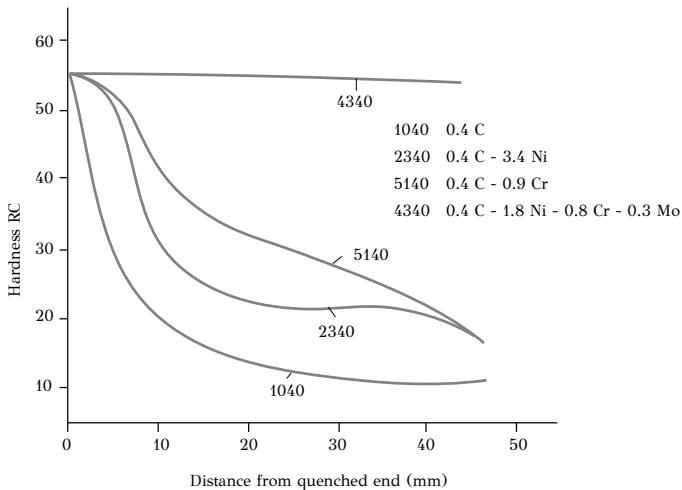


Figure 10.22: Hardenability curves for 1040, 2340, 5140 and 4340 steels with the same carbon content (0.4C).

is observed at a magnification 100X. The grain size number varies between 1 and 9. When $N = 4$ there are 8 grains/in², while when $N = 8$ there are 128 grains/in². A fine grained austenite with $N = 8$ contains much more grain boundaries than a coarse grained austenite with $N = 4$. Since the grain boundaries act as heterogeneous nucleation sites for the pearlite formation, the rate of the pearlitic transformation will be higher in the fine-grained steel. This simply means that the fine-grained steel will have lower hardenability. However, in practice, it is rare to seek a hardenability increase by a respective increase of the austenite grain size, since a larger grain size is associated with a lower ductility and fracture toughness of the steel. In any case, when a CCT diagram for certain steel is used from literature, the grain size of the austenite should always be noted.

10.7 TEMPERING OF MARTENSITE

The martensite in as-quenched steels is a very hard but at the same time a brittle phase. This makes the as-quenched steel unsuitable for the majority of industrial applications. In order to restore ductility, a heat treatment, termed *tempering*, follows the quenching of steel. With tempering, the ductility and fracture toughness are increased at a small expense in strength and hardness. The as-quenched martensite is a supersaturated $Fe - C$ solid solution with a BCT structure. The shape deformation and in particular the invariant-lattice deformation, which accompanies the martensitic transformation, raise the dislocation density to values up to 10^{12} cm^{-2} . During tempering, the carbon atoms as well as structural defects, such as vacancies and dislocations, acquire mobility and, as a result, several processes are activated:

- Relaxation of internal elastic stresses
- Rearrangement of dislocations into lower energy configurations
- Reduction of dislocation density
- Grain growth
- Carbide precipitation
- Transformation of retained austenite

The above processes reduce the hardness and raise the ductility of steel. In steels with a relatively high M_s temperature, some of the above processes are activated during quenching and cause the *autotempering* of steel.

In carbon steels, tempering includes four stages with increasing tempering temperature:

Stage 1. Precipitation of ε -carbide and reduction of the tetragonality of martensite (up to 250°C). The ε -carbide has *HCP* structure and exhibits an orientation relationship with martensite, according to Jack

$$(101)_{\alpha'} // (10\bar{1}1)_{\varepsilon}$$

$$(011)_{\alpha'} // (0001)_{\varepsilon}$$

$$[11\bar{1}]_{\alpha'} // [1\bar{2}10]_{\varepsilon}$$

The lattice misfit between planes $(101)\alpha$ and $(10\bar{1}1)\varepsilon$ is small, allowing the ε -carbide to maintain coherency in the initial stages of precipitation. The observed hardening between 50 and 100°C is caused by coherency hardening.

Stage 2. Transformation of retained austenite to ferrite and cementite (200 - 300°C). It should be noted that the retained austenite fraction is higher in high-carbon steels.

Stage 3. The ε -carbide is replaced by cementite (200 - 350°C). The precipitation of cementite takes place at the interfaces between ε -carbide and martensite, which loses its tetragonality and transforms to ferrite. Other heterogeneous nucleation sites are twin interfaces (in high-carbon steels), martensite lath interfaces and prior austenite grain boundaries. The cementite grows with Widmanstätten morphology and the orientation relationship with ferrite is according to Bagaryatski

$$(211)_{\alpha} // (001)_{F_3C}$$

$$[01\bar{1}]_{\alpha} // [100]_{F_3C}$$

$$[\bar{1}11]_{\alpha} // [010]_{Fe_3C}$$

Stage 4. Coarsening and spheroidization of cementite, recrystallization of ferrite (above 350°C). Coarsening is activated between 300 - 400°C . The driving force for coarsening is the reduction of the interfacial energy. With coarsening, the cementite loses the crystallographic relationship with ferrite and begins to spheroidize, a process, which takes place up to 700°C . At the same time the martensite lath boundaries

transform from low-angle to high-angle boundaries and become essentially the grain boundaries of ferrite, which adopts equiaxed grain morphology. The above processes include both recovery ($300\text{-}600^{\circ}\text{C}$) and recrystallization ($600\text{-}700^{\circ}\text{C}$). The final microstructure is equiaxed ferrite and a spheroidized dispersion of cementite.

Tempering is accompanied by a gradual decrease in hardness and tensile strength of the steel due to the loss of tetragonality of martensite and the escape of carbon from solid solution in order to form cementite. At higher tempering temperatures, a further decrease of hardness is caused by the coarsening of cementite and recrystallization of ferrite. The strength reduction is accompanied by an increase in ductility, as depicted in Figure 10.23. It is possible, with tempering, to achieve an optimum combination of strength and ductility. This is the basic principle of quenched and tempered (Q&T) steels, which exhibit excellent combinations of strength and fracture toughness.

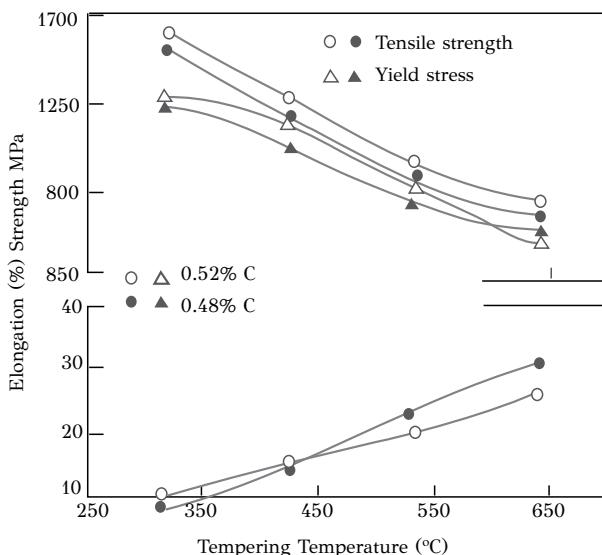


Figure 10.23: Effect of tempering temperature on the mechanical properties.

In alloy steels, the presence of alloying elements modifies several aspects of the tempering behavior, the most important of which are: (a) Si retards the transformation of ε -carbide to cementite up to 400°C , (b) Cr , Mo , W , Ti , Si , and V maintain the tetragonality of martensite to temperatures up to 500°C while Mn and Ni accelerate the loss of tetragonality, (c) Si , Cr , Mo and W delay the coarsening of cementite and allow cementite to act as a grain refiner for ferrite, (d) Co has been observed to retard the recovery processes during tempering and maintain high dislocation densities at high temperatures.

The presence of carbide formers (Cr , Mo , W , V and Ti) at sufficient amounts causes the precipitation of alloy carbides, which are more stable than cementite. These carbides form at higher tempering temperatures ($500\text{-}600^{\circ}\text{C}$), since enhanced

diffusivity is required. The alloy carbide dispersion replaces the cementite dispersion and is considerably finer. In addition it exhibits a higher resistance to coarsening. The increase in hardness associated with the precipitation of alloy carbides during tempering has been termed *secondary hardening*. The main characteristics of secondary hardening are similar to age hardening, which was discussed in Chapter 8. The tempering curve exhibits a peak, which is related with the transition from a coherency hardening to an Orowan hardening mechanism, as depicted in Figure 10.24a for Mo-containing steels. The Mo_2C carbide dispersion is shown for a $Fe - 4Mo - 0.2C$ steel in Figure 10.24b.

The nucleation of alloy carbides is heterogeneous. The major heterogeneous sites are cementite interfaces and dislocations. The precipitation sequence depends on the alloying elements. In steels containing Cr , the sequence is



In steels containing Mo or W , the sequence is



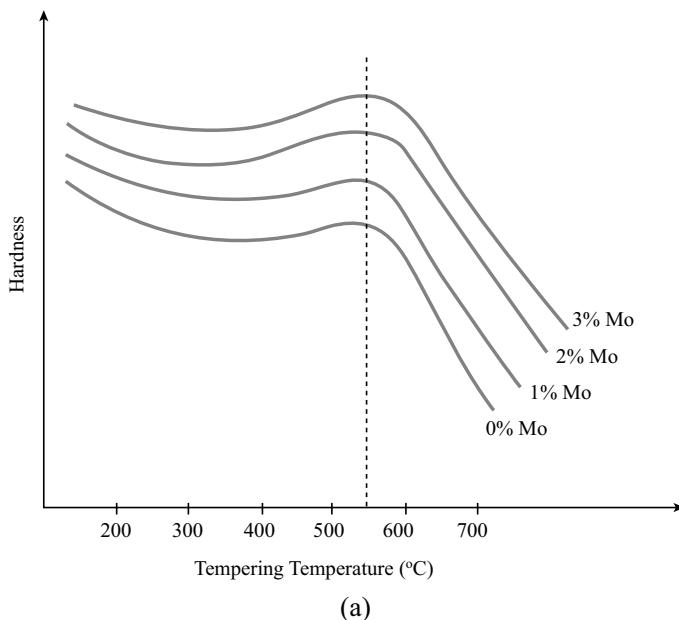
and in steels containing V , the carbide is VC .

A large advantage of secondary-hardening steels is that the alloy carbide dispersion forms at relatively high tempering temperatures ($500\text{-}600^\circ C$), meaning that the microstructure remains stable for service temperatures up to $500^\circ C$. As a result, these steels maintain their strength at high temperatures and can be used for high-temperature applications for creep resistance and as tool steels for hot working applications.

10.8 HEAT TREATMENT OF STEEL

The term heat treatment represents the controlled heating and cooling of steel in order to change its microstructure and obtain specific properties. Almost all machine elements made of steel, as shafts, gears and bearings, are subjected to some type of heat treatment either as an intermediate or as the final stage of production. Hundreds of heat treatment processes are available, some with a more general applicability while others are more specialized for certain types of steels. In this chapter, only the basic heat treatments will be described, those with a general application. Effort will be made to correlate these treatments with the relevant phase transformations discussed in the previous sections. Heat treatments are applied to steel for a number of reasons:

- To relieve the residual stresses, which arise from prior mechanical treatments or welding
- To homogenize the steel, in order to obtain a uniform distribution of alloying elements in the microstructure
- To soften the steel in order to be able to subject it to mechanical or deformation processing



(a)



(b)

Figure 10.24: Secondary hardening in a steel containing molybdenum: (a) hardness curves, (b) TEM micrograph showing the fine Mo_2C carbide dispersion.

- To increase the notch toughness for the steel to resist impact loading
- To increase strength and hardness for resistance to mechanical loads
- To modify the surface properties in order to increase the fatigue, wear and corrosion resistance.

Based on the above, heat treatments can be classified in three major categories: Annealing treatments, Strengthening treatments and Surface treatments.

Regarding the annealing treatments, the temperature regions where such treatments are performed appear in the phase diagram of Figure 10.25, while the respective thermal cycles appear in the *IT/CCT* diagram of the eutectoid steel, in Figure 10.26. Annealing was discussed in detail in Chapter 7, as a process, which removes the effects of strain hardening. In steels, the term annealing concerns those treatments, which are performed to soften the material or increase its ductility and fracture toughness.

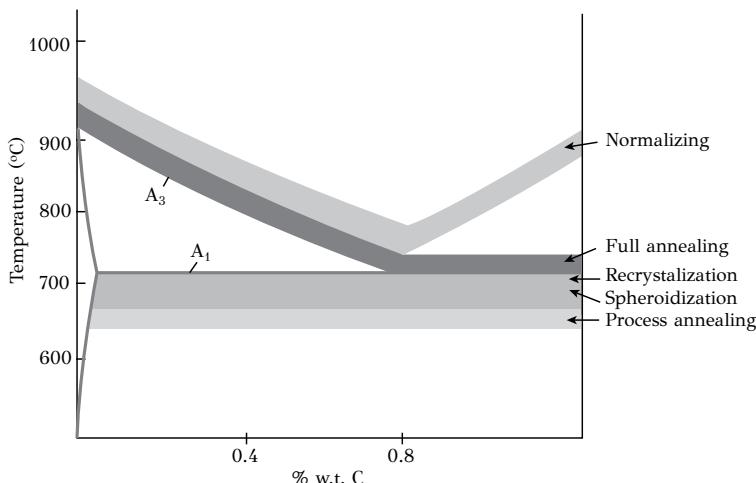


Figure 10.25: Temperature ranges for different types of annealing of steel.

Annealing involves heating the material to a specified temperature, holding at this temperature for a certain time, which depends on the size of the workpiece and cooling with a controlled rate at room temperature. Several types of annealing are the following:

Full annealing. It is performed in order to soften steel before any manufacturing process. Cooling is slow and is performed usually inside the furnace, as depicted by curve 1 in Figure 10.26. The final microstructure consists of coarse pearlite. Full annealing can also be performed isothermally, as depicted by curve 2 in Figure 10.26, termed isothermal annealing.

Normalizing. The main purpose is the homogenization of alloying elements in order to achieve a uniform transformation of austenite during cooling. It involves heating in the austenitic region, at temperatures higher than those of full annealing.

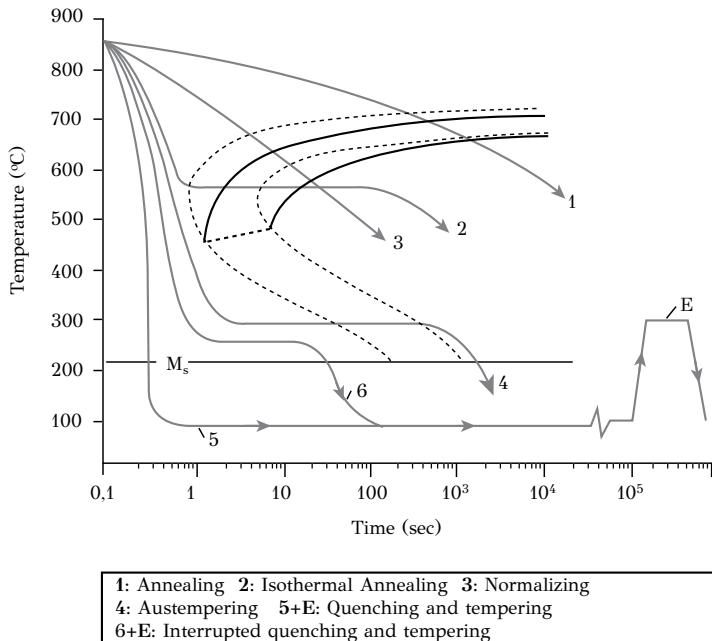


Figure 10.26: *IT/CCT* diagram of eutectoid steel depicting the basic heat treatments.

Cooling is performed in air, as depicted by curve 3 in Figure 10.26. The microstructure consists of fine-grained pearlite.

Process annealing. It concerns steel products produced by deformation processing, such as cold rolling or wire drawing. The purpose is to remove the effects of strain hardening. It is performed just below the eutectoid temperature.

Stress relief annealing. The main purpose is the relief of residual stresses arising from volumetric changes, associated with phase transformations in steels. It is applied when manufacturing processes need to be performed after quenching. In addition it is applied as a post weld heat treatment, in order to reduce the residual stresses caused by fusion welding processes.

Regarding strengthening heat treatments in steels, they are performed in order to form either martensite or bainite. In order to form martensite, the steel is austenitized at a certain temperature. The austenite grain size depends on the austenitizing temperature and holding time. Dissolution of carbides also takes place during austenitizing. After austenitization is complete, the steel is cooled rapidly in a quenching medium (air, oil, water, brine) depending on its hardenability. Quenching is then followed by a tempering treatment. The basic strengthening heat treatments for steel are:

Quenching and tempering. It is the traditional heat treatment for the hardening of steel, depicted by curve 5 in Figure 10.26. Important factors that should be taken into account are the hardenability of the steel in relation to the quenching medium

used and the size of the workpiece in order to avoid quench cracking and quenching distortions. Quenching is then followed by a tempering treatment, as depicted by curve E in Figure 10.26.

Interrupted quenching and tempering. Quenching is interrupted just above the M_s temperature and is continued at a lower cooling rate in the region of martensitic transformation, as depicted by curve 6 in Figure 10.26. A more uniform cooling between surface and interior of the workpiece is achieved in order to avoid quench cracking and quench distortions.

Austempering. The purpose of this heat treatment is to strengthen steel via the bainitic transformation. It involves austenitizing and quenching at a temperature above the M_s temperature. Holding at this temperature transforms the austenite to bainite, as depicted by curve 4 in Figure 10.26. Because of the isothermal nature of the treatment, there are no significant temperature gradients in the workpiece and cracking or distortions are avoided.

Regarding surface treatments in steels, they are performed in order to modify the surface of the workpiece and enhance the fatigue, wear or corrosion resistance. There are numerous surface treatments, which will not be considered here. The interested reader is advised to study the suggested reading list at the end of the chapter. The basic treatments as surface hardening, carburizing and nitriding will be discussed here.

Surface hardening. These treatments involve surface heating for the austenitization of a surface layer of the workpiece, while the interior remains cold. The steel is then quenched and the surface austenite layer transforms to martensite. In this way the workpiece obtains a high surface hardness while the interior remains soft and ductile. Tempering then follows quenching. The most usual techniques for surface hardening are *flame hardening* and *induction hardening*. In flame hardening, the heating of the workpiece is performed with an oxygen/acetylene burner. The case depth can reach 10 mm. In induction hardening, the heating is performed by electromagnetic induction. A coil, usually made of copper, surrounds the workpiece. The alternating current flowing in the coil generates an alternating magnetic field, which in turn generates eddy currents in the surface layers of the workpiece. Joule heating causes austenitization of the surface layers. With induction heating it is possible to control the case depth, which can reach 20 mm. In addition to flame hardening and induction hardening, other techniques, such as *laser surface hardening* have been developed. In laser hardening, the surface of the workpiece is austenitized by scanning with a laser beam, while the material is self-quenched with the transfer of heat to the interior of the material.

Carburizing. Carburizing as well as nitriding, which will be discussed below, are classified as *thermochemical treatments*, since the chemical composition is modified in a surface layer of the workpiece. In carburizing, the steel surface is enriched in carbon. Carbon is introduced by solid, liquid or gas carburizing media. In *pack carburizing*, the workpiece is covered by graphite powder and is heated at 900-950°C, where carbon diffusion is rapid. In *liquid carburizing*, the workpiece is placed in a bath of carburizing salts. In *gas carburizing*, the workpiece is placed in a furnace with

a controlled gas atmosphere, which supplies the necessary carbon potential. After certain time at the carburizing temperature the steel is quenched to form martensite in the carburized layer (case depth). The thickness of the carburized layer depends on carburizing temperature and time and can reach 2-3 mm. A carburized layer is depicted in Figure 10.27.

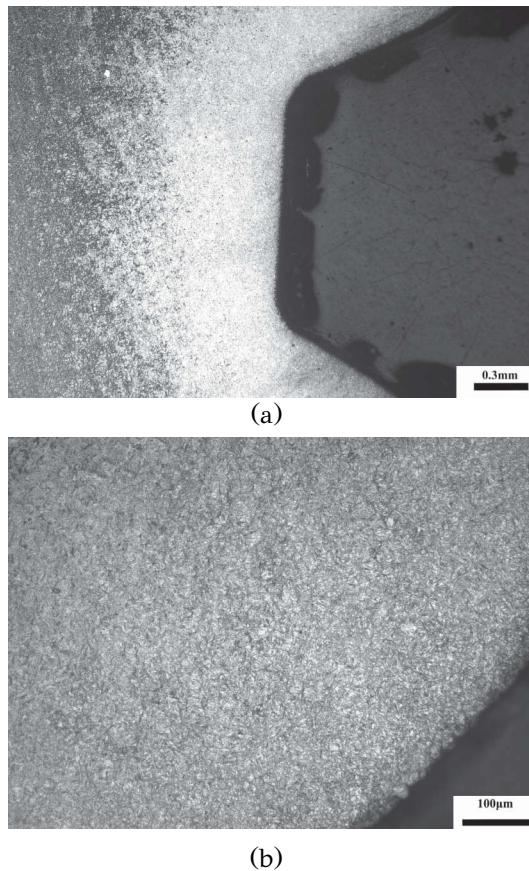


Figure 10.27: (a) micrograph of the carburized layer, (b) higher magnification depicting the martensitic structure in the carburized layer.

Nitriding. In nitriding the steel surface is enriched in nitrogen. Nitrogen diffuses in the steel and reacts with iron and the alloying elements to form nitrides, which raise the hardness of the nitrided layer. The nitriding layer typically consists of two zones. The white zone (because it etches white in metallography) or *compound zone* is the zone where nitrides form, such as the $\varepsilon - Fe_2N$ and $\gamma' - Fe_4N$. The *diffusion zone* consists mainly of a supersaturated solid solution of nitrogen in iron. In contrast to carburizing, the surface hardening is not achieved via a martensitic transformation. The hardness is due to the inherent hardness of the nitrides in the compound layer.

The compound layer hardness is responsible for the wear resistance, while the diffusion layer is responsible for the fatigue resistance of nitrided steels. The latter is due to the compressive residual stress profile, which develops during nitriding. Nitriding is performed between 500-550°C for 20-90 hours. Thus the nitriding temperature is much lower than the carburizing temperature and grain growth can be avoided.

10.9 CASE STUDIES IN STEELS

The main characteristics of some steels, in terms of structure-property relationships, will be presented in this section.

Pearlitic rail steels. The basic requirements for rail steels are: (a) high ultimate tensile strength, (b) wear resistance, (c) fatigue resistance and (d) weldability. The pearlitic microstructure provides a combination of properties, which address the above requirements, so pearlitic steels are widely used for the construction of rails. Rail steels are covered by international standards, such as UIC. An example is UIC-860 steel, Grade 900A with 0.8wt%*C* and UTS=880-1030 MPa. The steel has a pearlite microstructure with a fine interlamellar spacing, as depicted in Figure 10.28, providing a high hardness, which increases the wear resistance. However, due to the rolling contact between the wheel and the rail, another failure mechanism may appear in these steels, termed rolling contact fatigue (*RCF*). Small fatigue cracks initiate and propagate under the contact stresses, few mm below the rolling surface. An example is shown in Figure 10.29. In most cases, these cracks cause spalling of the rail. However, in certain cases, they would cause complete fracture of the rail. The mitigation of *RCF* in rails involves non-destructive testing using ultrasonic techniques for the detection of cracks as well as preventive grinding of rails for the removal of *RCF* cracks before they grow to critical size.

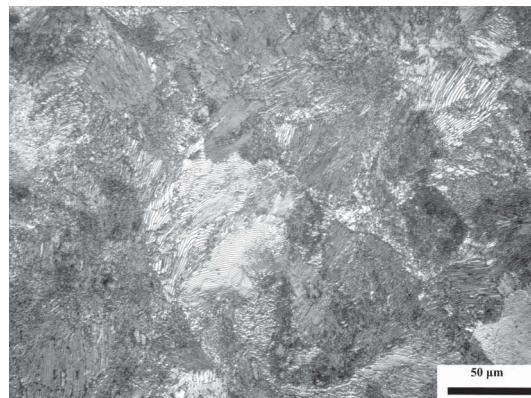


Figure 10.28: Microstructure of a pearlitic steel used in rails.

HSLA piping steels. The transport of oil and gas in large quantities has generated requirements for large diameter pipes. The need to keep the weight of the pipes in reasonable limits generates a requirement for high strength. Additional requirements

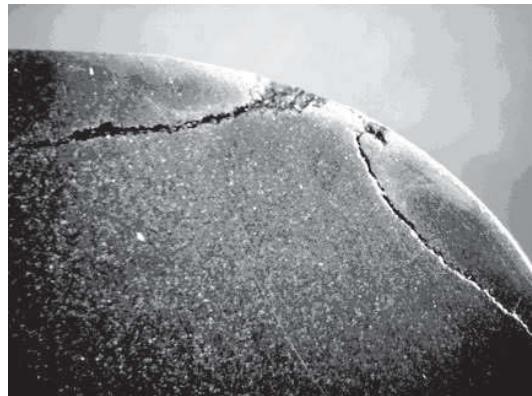


Figure 10.29: Formation of cracks on the surface of a rail by Rolling Contact Fatigue (RCF).

are high fracture toughness and weldability. High-strength low alloy steels (*HSLA*) address the above issues. These are steels with low carbon for enhanced weldability and microadditions of *Nb*, *V* or *Ti*, which form carbides, such as *NbC*, *VC* and *TiC*, or carbonitrides, such as *V(CN)*, *Nb(CN)*, *Ti(CN)*, *(Nb,Ti)(CN)*. During thermo-mechanical processing, these carbonitrides introduce a pinning effect on the grain boundaries and retard recrystallization and grain growth during controlled rolling of the austenite. The result is a fine ferrite grain size, of the order of $5\mu\text{m}$, which results in high strength and toughness.

Dual-Phase steels. *DP* steels were developed mainly for automotive applications. Their microstructure consists of soft ferrite and a dispersion of hard martensite. This combination provides a low initial yield strength but a very high strain hardening rate, leading to enhanced formability, higher than *HSLA* steels. The processing of *DP* steels involves intercritical annealing in the $\alpha + \gamma$ region, where carbon is rejected in austenite, increasing its hardenability. During cooling to room temperature, the austenite transforms to martensite. A typical microstructure of *DP* steel is shown in Figure 10.30.

TRIP steels. Transformation Induced Plasticity (*TRIP*) steels usually contain ferrite, bainite and retained austenite. The remarkable combination of strength and formability is a result of the operation of the *TRIP* effect, which was discussed in Chapter 6. Martensitic transformation can be activated under stress or plastic strain. Strain-induced martensitic transformation, taking place during the cold forming of these steels, can stabilize plastic flow and increase the formability limit of the steel. The basic requirement is to form a dispersion of retained austenite with adequate stability. The processing of *TRIP* steels is depicted in Figure 10.31 and involves two steps: (a) intercritical annealing to form ferrite and austenite followed by (b) isothermal bainitic transformation, during which the austenite is stabilized by carbon partitioning from the growing bainitic ferrite. An additional point here is that cementite precipitation during the bainitic transformation has been suspended by *Si*



Figure 10.30: Microstructure of a Dual-Phase steel. Dark color for martensite, white color for ferrite.

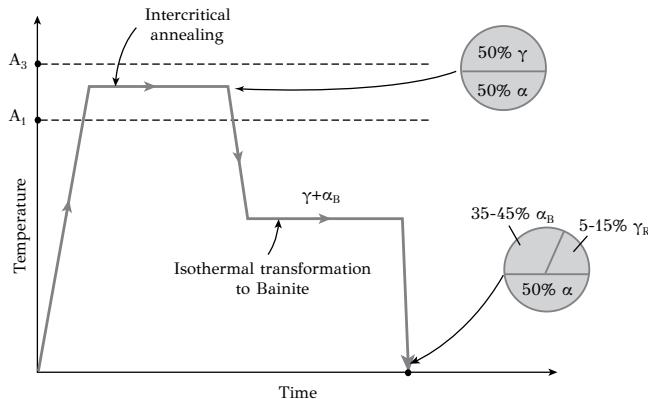


Figure 10.31: Annealing treatments for the production of low alloy TRIP steels.

or Al additions. In effect the bainite in TRIP steels is a carbide-free bainite. The microstructure of a 0.2C – 1Mn – 1Si TRIP steel is shown in Figure 10.32.

A comparison of HSLA, DP and TRIP steels is shown in Figure 10.33. The total elongation is plotted against maximum tensile strength. TRIP steels exhibit a better combination of strength and ductility than HSLA and DP steels, making them suitable for demanding applications in the automotive industry.

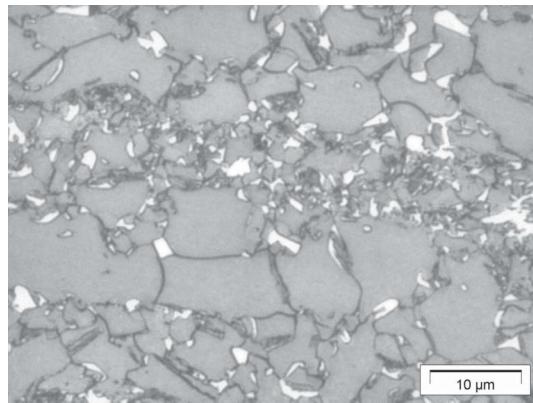


Figure 10.32: Microstructure of a *Fe-0.2C-1Mn-1Si TRIP* steel. Retained austenite appears white.

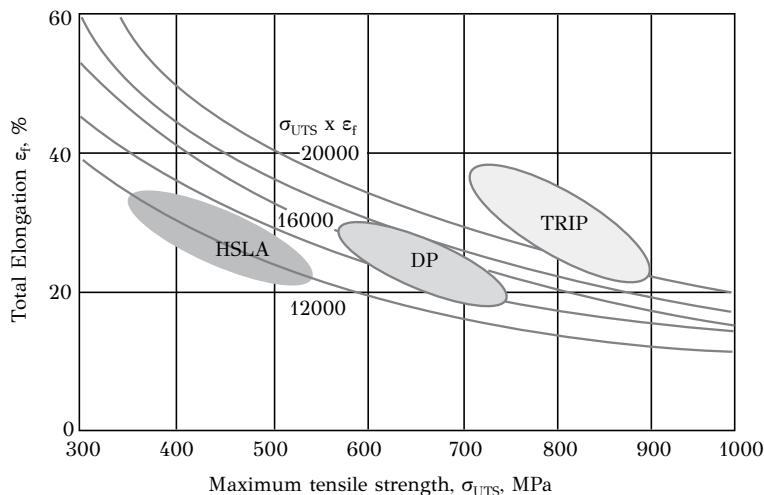


Figure 10.33: Comparison of *HSLA*, *DP* and *TRIP* steels used in the automotive industry.

10.10 SYNOPSIS

1. The technological position of steels among other materials is attributed to the diversity of properties, low cost, manufacturability and recyclability. Steels dominate over other metals and alloys due to the large reserves of steel ore, the melting point and allotropy of iron and finally the effect of alloying elements, which widen the range of steel applications.
2. The phases in steels are responsible for the large range of properties. The phases can be classified in solid solutions, intermetallic compounds (basically carbides and nitrides) and phase mixtures.
3. Solid solutions in steels are ferrite, austenite and martensite. Pearlite and bainite are phase mixtures consisting of ferrite and cementite.
4. Phase constitution in carbon steels can be studied with the $Fe - C$ phase diagram. A central point in the phase diagram is the eutectoid reaction. When a eutectoid steel is heated in the austenitic region and then cooled slowly, it will be transformed to 100% pearlite according to the eutectoid reaction. Hypoeutectoid steels consist of ferrite and pearlite while hypereutectoid steels consist of cementite and pearlite.
5. Alloying elements in steels influence the stability of phases, the solid solubility of other elements, the rate of phase transformations, deformation and strengthening as well as various embrittlement phenomena.
6. The main purposes of alloying elements are deoxidation, desulfurization, austenite stabilization, carbide formation, hardenability, grain refinement, increase of fracture toughness and oxidation resistance.
7. Ferrite in steels is the result of the decomposition of austenite. The transformation is diffusive and takes place by the nucleation and growth of ferrite. The morphology of proeutectoid ferrite depends strongly on the transformation temperature and is distinguished in: (a) intergranular ferrite, (b) Widmanstätten ferrite and (c) transgranular ferrite.
8. Pearlite in steels forms via the eutectoid reaction. During pearlite formation, ferrite and cementite grow cooperatively. Pearlite exhibits a lamellar morphology with sequential plates or lamellae of ferrite and cementite. Regarding the effect of temperature, the kinetics of the pearlite formation follows C -curve kinetics.
9. The growth of pearlite in carbon steels takes place under diffusion control, while the rate-limiting step is the diffusion of carbon in austenite. The addition of substitutional alloying elements, delays the pearlite formation for two reasons: (a) these elements should also diffuse and partition between phases and (b) some elements,

such as B, segregate at the austenite grain boundaries and delay the nucleation of pearlite.

10. The martensitic transformation is probably the most significant transformation in steels with a great technological impact. The martensitic transformation is a displacive, diffusionless transformation. The strain energy of the transformation arises from shear-like displacements and influences the transformation kinetics and morphology of martensite.
11. In steels, the martensitic transformation takes place during the rapid cooling of austenite, termed quenching. The cooling rate should be large enough so other competitive diffusional transformations, such as the formation of ferrite, pearlite or bainite, would not be activated. The critical cooling rate depends on chemical composition and the austenite grain size and characterizes the hardenability of steels.
12. The martensitic transformation is generally athermal, since the fraction of martensite increases with the drop in temperature. It starts at the M_s temperature and ends at the M_f temperature.
13. The M_s temperature exhibits a strong dependence on the alloying elements in the steel. The austenite stabilizers, such as C, N, Mn, Ni, reduce the M_s temperature, since a larger driving force is required for martensite nucleation. In addition, elements, such as Cr, Mo, Al, Si, W, and V, which are not austenite stabilizers, also reduce the M_s temperature. This is attributed to the solid solution hardening, hindering dislocation glide in the austenite, which is necessary for martensite nucleation.
14. The high hardness of martensite is attributed to carbon in interstitial sites in the *BCT* martensite lattice. Other mechanisms include strengthening arising from the presence of boundaries and interfaces, leading to a Hall-Petch type strengthening, either from the prior austenite grain size or the mean size of martensite laths.
15. Bainite forms by the decomposition of austenite at temperatures above the M_s but below that at which pearlite forms. While pearlite forms by the cooperative growth of ferrite and cementite, bainite forms by the growth of supersaturated ferrite followed by the precipitation of cementite as a result of carbon diffusion. The supersaturated ferrite forms by a displacive shear-like transformation, which resembles the martensitic transformation.
16. Depending on the transformation temperature, the cementite precipitates either between or inside the bainitic ferrite laths and the resulting structures correspond to upper and lower bainite respectively.

17. The precise representation of phase transformations in steels is accomplished with the isothermal transformation (*IT*) or the continuous cooling transformation (*CCT*) diagrams.
18. Hardenability is an important property of steels associated with the heat treatment of steels and in particular quenched and tempered steels. A steel, with a high hardenability forms martensite with a lower cooling rate than another steel, which requires higher cooling rate to obtain the same effect. A steel with a higher hardenability can form martensite at a larger depth from the surface than another steel with lower hardenability, quenched in the same quenching medium. Hardenability is influenced by alloying elements and the prior austenite grain size.
19. The martensite in as-quenched steels is a very hard but at the same time a brittle phase. In order to restore ductility, a heat treatment, termed tempering, follows the quenching of steel. With tempering, the ductility and fracture toughness are increased at a small expense in strength and hardness.
20. The heat treatment of steel involves the controlled heating and cooling of steel in order to change its microstructure and obtained specific properties. Heat treatments can be classified in three major categories: Annealing treatments, Strengthening treatments and Surface treatments.

10.11 REVIEW QUESTIONS

1. A hypoeutectoid steel with 0.4wt%*C* is heated to 900°C for austenitization. Then the steel is transferred to another furnace with temperature 750°C.
 - (a) Calculate the phase fractions at this temperature.
 - (b) The steel is transferred to another furnace at 700°C. How much pearlite forms?
 - (c) How much cementite forms?
 - (d) If the steel is left to cool slowly in the furnace at room temperature and then we turn on the furnace so it will reach a temperature of 750°C, how much austenite forms and what is its carbon content?
 - (e) Another steel contains 70% austenite at 750°C. What is its carbon content?
2. Dual phase steels usually consist of a mixture of ferrite and martensite and exhibit a good combination of strength and ductility.
 - (a) Describe the basic characteristics of the two phases (crystal structure, solubility limits).
 - (b) What are the similarities and the differences between the two phases?
 - (c) Describe a heat treatment for the production of a dual-phase steel containing 50% ferrite and 50% martensite, in which the carbon content of the martensite is 0.6wt%.
3. In the steels below calculate the carbon content and comment whether the steel is hypo- or hyper-eutectoid.
 - (a) Steel 1, containing 8% cementite and 92% ferrite at room temperature

- (b) Steel 2 containing 18% cementite and 82% ferrite at room temperature
 (c) Steel 3 containing 18% pearlite and 82% proeutectoid cementite at room temperature
 (d) Steel 4 containing 94% pearlite and 6% proeutectoid cementite at room temperature.
4. Consider the *IT* diagram of a eutectoid steel (Figure 10.12).
- What is the carbon content of the steel and what is the phase constitution (phases present) according to the $Fe - C$ phase diagram?
 - Four specimens are austenitized at $800^{\circ}C$ and cooled at different rates:
 - Specimen 1 is rapidly cooled to $650^{\circ}C$, where it remains for 100 sec and is water quenched
 - Specimen 2 is rapidly cooled to $600^{\circ}C$, where it remains for 2 sec and is water quenched
 - Specimen 3 is rapidly cooled to $400^{\circ}C$, where it remains for 1 hour and is water quenched
 - Specimen 4 is rapidly cooled to $300^{\circ}C$, where it remains for 1 min and is water quenched
 Describe the microstructure of the specimens after the end of the treatment.
 - Calculate the critical (minimum) cooling rate, in $^{\circ}C/\text{sec}$, which is required for cooling from $800^{\circ}C$ for the complete transformation of austenite to martensite.
 - The critical cooling rate, calculated above, was applied in a thin steel part and quench distortions were observed. What would you propose in order to obtain a fully martensitic structure and avoid the quench distortions?
5. A rod with diameter 25 mm is made of eutectoid steel. The rod is austenitized at $900^{\circ}C$ and quenched in oil. The metallographic investigation revealed that the structure consists of martensite, bainite and pearlite.
- Draw a schematic of the cross section of the rod and note where do you expect these phases to form.
 - Draw the expected hardness profile from the rod surface to the rod center.
 - The rod is subjected to a tempering treatment at $500^{\circ}C$ for 1 hour. Draw the new hardness profile and compare it with question (b) above.
 - What heat treatment could convert the rod to 100% pearlite?
 - Suppose that the rod, in addition to carbon, contains $2Mn$ and $1Cr$ (wt%). The rod is austenitized at $900^{\circ}C$ and is oil quenched. What basic difference will be observed from the metallographic investigation?
6. Consider the *IT* diagrams of the eutectoid steel and the 0.4wt%*C* hypoeutectoid steel (Figure 10.12 and 10.15 respectively). There are certain differences between the two diagrams.
- In the 0.4*C* steel the pearlitic nose is crossed by the vertical axis. Why?
 - In the 0.4*C* steel there is an additional *C* curve. What does it represent?
 - The M_s temperature is higher in the 0.4*C* steel. Why?
 - Which of the two steels exhibits a higher hardenability?

BIBLIOGRAPHY—FURTHER READING

1. Honeycombe, R. W. K. & Bhadeshia, H. K. D. H. (1995). *Steels: Microstructures and Properties*, Edward Arnold, UK.
2. Llewellyn, D. T. (1992). *Steels: Metallurgy and Applications*, Butterworth-Heinemann, UK.
3. Bhadeshia, H. K. D. H. (1992). *Bainite in Steels*, The Institute of Materials, UK.
4. Thelning, K. E. (1984). *Steel and its Heat Treatment*, Butterworths.
5. Hornbogen, E. (1983). *Physical Metallurgy of Steels, Ch.16 in Physical Metallurgy*, R.W. Cahn, P. Haasen (eds), Elsevier Science Publishers.
6. Krauss, G. (1994). *Steels: Heat Treatment and Processing Principles*, American Society of Metals, Materials Park, Ohio, USA.
7. Totten, G.E. (ed) (2007). *Steel Heat Treatment Handbook*, Taylor & Francis, NY.



Taylor & Francis
Taylor & Francis Group
<http://taylorandfrancis.com>

11 Alloy design

11.1 INTRODUCTION

The development of new metallic alloys is traditionally a very tedious process, which involves extensive experimental investigations by which composition and processing conditions are varied in a more or less erratic manner until the desired properties are achieved. The properties of metallic alloys depend on phases present, their composition and how they are geometrically arranged in the microstructure, involving structural elements on micro, nano and atomic scale. Therefore detailed information about temperature, composition and stability of the different phases as well as phase transformations and phase equilibria are indispensable not only for the modeling and simulation but, more importantly, *for the computational design of metallic alloys and their processing*.

Computational alloy design has its origins on the pioneering work of Prof. G.B. Olson at MIT and Northwestern University in the USA, where the concept of *materials by design* was first developed. This work led to the announcement, in 2011, by the White House, of the Materials Genome Initiative (*MGI*). This in turn initiated the global framework of Integrated Computational Materials Engineering (*ICME*), which encompasses computational materials modeling for engineering applications. Integrated Computational Materials Design (*ICMD*) is a framework based on *ICME* and *MGI* and aims at the accelerated knowledge-based design of new materials. The origins and the *ICMD* methodologies are described in a review paper by Xiong and Olson (see suggested reading at the end of the chapter).

For simple alloys, a large amount of constitutional information can be found in databases and reference books. If critically evaluated, such constitutional, thermodynamic and kinetic data develop *predictive power*, when appropriate thermodynamics and kinetics calculation software and model descriptions are applied. Several software packages, which allow calculation of phase equilibria as well as other thermodynamic quantities are now commercially available. Calculations considered a cutting edge a few years ago may now be performed on a routine basis. A more ambitious approach is to combine the thermodynamic calculations with kinetic models, e.g., diffusion calculations and thereby predict the rate of reactions and phase transformations. This approach is extremely powerful and may be used to simulate a wide range of different phenomena in alloys including precipitation, homogenization, diffusional interactions between substrate and coatings, welding, etc. With this approach it could be possible to treat multi-component alloys with realistic thermodynamic and kinetic properties and it could become possible to take anisotropy, elastic stresses etc. into account, as long as the computational work can be afforded. When computational alloy thermodynamics and kinetics tools are combined with other software packages in the area of mechanics of materials, such as Finite Element codes, inverse analysis programs, or more general multipurpose computational

tools such as Matlab, then it is possible to apply these tools towards the design of alloy compositions and processing.

11.2 THE ALLOYNEERING METHODOLOGY FOR ALLOY DESIGN

Alloyneering is a generic term, which combines the words *Alloy* and *Engineering* and describes a concise knowledge-based methodology leading to the development of Engineered Alloys. Alloyneering is based on the principles of ICMD and encompasses the application of computational alloy thermodynamics and kinetics, precipitation and phase-field models, for the simulation of microstructural evolution and the design of metallic alloys with tailored properties. The same approach can be used for the design of the required processing conditions to obtain the corresponding microstructure and thus to achieve the required properties. In its full development, Alloyneering enables the computer-aided design of both, alloy composition and processing. In the near future this would lead to a situation where the development of new alloys will take only a fraction of time and cost associated with the empirical (try-and-error) development. This constitutes a breakthrough step, which takes the alloy development process from an entirely empirical to a knowledge-based platform. According to the Alloyneering methodology, the alloy design process is performed in five stages depicted in Table. 11.1. These stages are

- Analysis
- Simulation
- Validation
- Mapping
- Optimization

Analysis. It is the first stage in alloy or process design. The aim is to set up the alloy design criteria, on which the design will be based. The alloy design criteria include property requirements, which can be translated into microstructural requirements. For example, in tool steels, the required hardness can be translated into specific microstructure requirements for the matrix and the carbide dispersions. The composition search space is defined with the range of alloying elements. In the tool steel example above, the alloying elements could be *C*, *Cr*, *Mo*, and *V*. In the case of process design, the process search space is defined. This would mean to define appropriate ranges of austenitizing and tempering temperatures and times.

Simulation. The aim of this stage is to perform a simulation of the microstructural evolution of the alloy and compute microstructure elements and properties. For example, in TRIP steels, the amount and composition of retained austenite can be calculated as a function of alloy composition and annealing process. In another example, the hardness of a heat treatable aluminum alloy can be calculated as a function of alloy composition and aging conditions. Simulations are performed with certain methods and tools, such as computational alloy thermodynamics and kinetics, precipitation and phase-field methods. These tools are described in section 11.3.

Validation. All simulations involve certain assumptions. It is, therefore, necessary to compare simulation results with available experimental data. For example, computed phase fractions and compositions should be compared with experimental data from quantitative metallography, XRD, EBSD, EDX or atom probe microanalysis. In this way, adjustments can be made regarding the simulation conditions, the input data or the geometric problem set-up. Only after the simulations are experimentally validated, the design can proceed to the next stage.

Mapping. The design involves the mapping of the composition or the process space. For example, in medium-Mn steels, the volume fraction of austenite and austenite composition are mapped on the composition search space, defined in the Analysis stage. The mapping generates a volume of data connecting the microstructural requirements to the alloy composition or process conditions. The alloy design criteria, established in the Analysis stage, are then applied to screen alloy compositions or process conditions. So only alloy compositions and process conditions, which satisfy the set alloy design criteria are selected for the optimization stage.

Optimization. In most alloy and process design cases, multiple alloy design criteria are defined. In this case a single optimum solution might not be feasible, since the improvement in one criterion might cause the deterioration of another criterion. As a result, a tradeoff between the design objectives is considered, leading to the identification of a number of optimal solutions. These problems can be modeled as multiobjective optimization (MOOP) problems, where an alloy composition or process condition can be considered optimal if it results in objectives that lay on the Pareto front. The Pareto optimal solutions are then ranked, using a heuristic approach based on a function that aggregates the design criteria into a single parameter. An example of this procedure is discussed in section 11.5 concerning the design of medium-*Mn* steels.

11.3 SIMULATION FRAMEWORK

As discussed above, the second stage in the alloy design process is simulation. The relevant simulation toolbox is described in Table 11.2. It involves computational thermodynamics, computational kinetics and phase-field modeling. A short description of these items is presented below.

11.3.1 COMPUTATIONAL ALLOY THERMODYNAMICS

Computational alloy thermodynamics, as presented in this book, is based on the CALPHAD approach, which involves the development of models of the free energy of the various phases in multiphase and multicomponent alloys. In this way thermodynamic properties in multicomponent systems can be calculated from those of binary or ternary systems. The approach involves critical assessment of thermochemical data and their incorporation into self-consistent databases. Only the fundamental equations describing the Gibbs free energy of substitutional solid solutions will be presented in this chapter. For a more rigorous treatment involving interstitial solid

Table 11.1**The five stages of the Alloyneering methodology for alloy and process design**

Design stages	Objectives
Analysis	Alloy design criteria Property objectives Microstructural objectives Composition search space Processing search space
Simulation	Thermodynamic equilibrium Phase transformations Microstructural evolution Properties (strength, toughness, processability etc.)
Validation	Experimental data Phase fractions Grain size Precipitate size distribution Phase compositions
Mapping	Composition search space Processing search space
Optimization	Multiobjective optimization Ranking of optimal solutions

Table 11.2**Simulation framework for alloy design**

Toolbox	Description
Computational thermodynamics	CALPHAD approach Thermodynamic databases Tools: Thermo-Calc
Computational kinetics	Multicomponent diffusion Mobility databases Tools: DICTRA
Phase-field modeling	Diffuse interface models Tools: MICRESS, Open Phase

solutions and intermetallic compounds, the reader is advised to see the suggested reading list at the end of the chapter.

Following the analysis of alloy thermodynamics, presented in Chapter 4, the free energy of a substitutional solid solution is

$$G = G^o + G^{id} + G^{xs} + G^{mag} \quad (11.1)$$

where G^o is the free energy of the pure components, G^{id} is the ideal or entropic term, G^{xs} is the excess term and G^{mag} is the magnetic term, which is important in systems exhibiting magnetic ordering. For a binary solid solution, the free energy is given by equation 4.13, which is repeated here

$$G = \sum_i X_i G_i^o + RT \sum_i X_i \ln X_i + \sum_i \sum_{j>i} X_i X_j \Omega_{ij} + G^{mag} \quad (11.2)$$

The terms of (11.2) correspond to the terms of equation (11.1). The composition dependence of the interaction parameters Ω_{ij} is taken into account by the introduction of the so-called Redlich-Kister polynomials and equation (11.2) becomes

$$G = \sum_i X_i G_i^o + RT \sum_i X_i \ln X_i + \sum_i \sum_{j>i} X_i X_j \sum_p \Omega_{ij}^{(p)} (X_i - X_j)^p + G^{mag} \quad (11.3)$$

where p denotes the order of the interaction. A further generalization into multicomponent systems requires the addition of suitable terms in equation (11.3) to describe the interaction between three or more elements. For three-element interactions these terms take the form

$$G_{ijk}^{xs} = X_i X_j X_k L_{ijk}$$

In general, using the symbol L for the interaction parameters in the CALPHAD approach, the excess term can be described with the Muggianu equation, which for three elements is

$$G^{xs} = X_A X_B \{L_{AB}^0 + L_{AB}^1 (X_A - X_B)\} + X_B X_C \{L_{BC}^0 + L_{BC}^1 (X_B - X_C)\} + X_A X_C \{L_{AC}^0 + L_{AC}^1 (X_A - X_C)\}$$

Therefore the expression for the Gibbs free energy in the CALPHAD approach is according to the Redlich-Kister-Muggianu (RKM) excess model:

$$G = \sum_i X_i G_i^o + RT \sum_i X_i \ln X_i + \sum_i \sum_{i \neq j} X_i X_j \sum_p L_{ij}^{(p)} (X_i - X_j)^p + \sum_i \sum_{i \neq j \neq k} X_i X_j X_k \sum_p L_{ijk}^{(p)} \quad (11.4)$$

In case of magnetic ordering, the magnetic term that should be added to equation (11.4) is

$$G^{mag} = RT \ln(\beta + 1) f(\tau) \quad (11.5)$$

where β is the magnetic moment, $\tau = T/T_c$ and T_c is the Curie temperature. These parameters are given as a function of composition by

$$\beta = \sum_i X_i \beta_i^o + \sum_i \sum_{i \neq j} X_i X_j \sum_p [\beta^{(p)} (X_i - X_j)^{(p)}] \quad (11.6)$$

and

$$Tc = \sum_i X_i Tc_i^o + \sum_i \sum_{i \neq j} X_i X_j \sum_p [T c^{(p)} (X_i - X_j)^{(p)}] \quad (11.7)$$

The above descriptions are implemented in the thermodynamics software Thermo-Calc, which is linked to relevant thermochemical databases to retrieve the required data for the calculations.

11.3.2 COMPUTATIONAL KINETICS

Computational kinetics is a computational framework for solving problems related to phase transformations in metals and alloys. The discussion in this chapter will be limited to diffusional transformations, where the transformation is entirely controlled by the diffusion and partitioning of the alloying elements between the transforming phases. The coupling between diffusion and thermodynamics, as discussed in Chapter 5, is presented first followed by multicomponent diffusion. Then the concept of the sharp interface for solving moving boundary problems will be presented. The relevant software, by which the various examples, presented below, have been performed, is DICTRA (Diffusion Controlled TRAnsformations).

Recall from Chapter 5, section 5.7.2 that the diffusive flux of an element is proportional to the chemical potential gradient

$$J_i = -M_i c_i \frac{\partial \mu_i}{\partial x}$$

where M_i is the mobility and c_i the concentration of component i . The above equation can be written as

$$J_i = -M_i c_i \frac{\partial \mu_i}{\partial c_i} \frac{\partial c_i}{\partial x}$$

Comparing with Fick's first law, the diffusion coefficient becomes

$$D_i = M_i c_i \frac{\partial \mu_i}{\partial c_i}$$

and following the discussion in section 5.7.2 it finally takes the form

$$D_i = M_i RT \left(1 + \frac{\partial \ln \gamma_i}{\partial \ln X_i} \right) \quad (11.8)$$

where γ_i is the activity coefficient and X_i the atomic fraction of component i . The term in parentheses is the thermodynamic factor. Equation (11.8) suggests that the diffusion coefficient is the product of a mobility parameter and the thermodynamic factor. In the special case of a binary solid solution $A - B$, the thermodynamic factor for component B takes the form

$$\left(1 + \frac{\partial \ln \gamma_B}{\partial \ln X_B} \right) = \frac{X_A X_B}{RT} \frac{\partial^2 G}{\partial X_B^2}$$

indicating that the thermodynamic factor is related to the curvature of the free energy-composition ($G - X$) curve. In the case of negative curvature, the thermodynamic factor and the associated diffusion coefficient are negative, resulting in uphill diffusion, as for example in spinodal decomposition. The main concentration dependence of the diffusion coefficient arises from the thermodynamic factor. However the mobilities are also concentration dependent. In the relevant mobility databases, used by DICTRA, the mobility of a component M_i in a given phase is described by a frequency factor M_i^o and activation energy ΔG_i^* as

$$M_i = \frac{M_i^o}{RT} \exp\left(-\frac{\Delta G_i^*}{RT}\right)$$

In the framework of the CALPHAD approach, in a multicomponent system, the composition dependence of the mobility parameters is expressed with Redlich-Kister-Muggianu polynomials, similar to the description of the associated thermodynamic quantities in the thermochemical databases.

Now consider a phase transformation, e.g., the growth of phase β in the matrix of phase α in a multicomponent system, as depicted in Figure 11.1. This is a moving-boundary problem and the growth rate of phase β is given by the velocity of the sharp interface, u , which can be calculated by the mass balance at the interface

$$J_i = u(c_\alpha - c_\beta) \quad (11.9)$$

where J_i is the diffusive flux of component i at the interface and c_α, c_β are the compositions of component i in phases α and β at the interface. Under the assumption of local thermodynamic equilibrium, the compositions c_α and c_β are calculated by computational alloy thermodynamics. The diffusive flux in a multicomponent alloy is given by the Fick-Onsager law (see section 5.9)

$$J_i = \sum_{j=1}^{n-1} D_{ij} \frac{\partial c_j}{\partial x} \quad (11.10)$$

Following (11.8), the diffusion matrix is the product of two matrices, one matrix containing the mobilities and one matrix with the associated thermodynamic factors. For the solution of the problem, $(n-1)(n-1)$ interdiffusivities are required. The evolution of the concentration profiles is provided by Fick's second law (mass conservation) as

$$\frac{\partial c_i}{\partial t} = -\nabla \cdot J_i \quad (11.11)$$

The solution of the system of equations (11.9), (11.10) and (11.11) provides the growth rate and the partitioning of alloying elements between phases α and β in this moving boundary problem. In DICTRA the solution of the diffusion equation (11.11) is performed by a finite element method. For the calculation of the interface compositions at local equilibrium and for the calculation of the thermodynamic factors, the program is linked to Thermo-Calc. The mobility parameters are stored in a separate mobility database. Simulation examples are presented in the following sections.

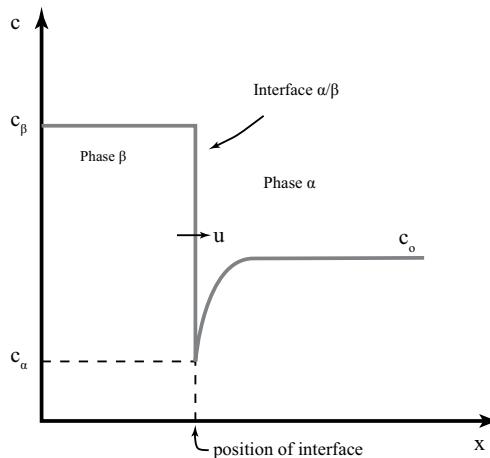


Figure 11.1: Concentration profile during the growth of phase β in the matrix phase α . The velocity of the sharp interface is calculated by a mass balance and corresponds to the growth rate.

11.3.3 PHASE-FIELD MODELING

The microstructure of metals is composed of grains and different phases, which are separated by interfaces, either grain boundaries or interphase boundaries. Microstructural evolution, during phase transformations, is accomplished with the migration of these interfaces. Phase-field modeling is a methodology, which allows the study of the thermodynamics and kinetics of interface migration and the associated microstructural evolution. Although phase-field modeling was first applied to the study of solidification, today the use of phase-field modeling has been extended to the study of solid-to-solid phase transformations, such as the decomposition of austenite in steels and the recrystallization and grain growth during hot rolling of HSLA steels. For a good review on the subject, the reader is advised to read the paper by Steinbach (see suggested reading at the end of this chapter). Due to space limitations, only the basic principles of phase-field modeling will be presented here.

Order parameter. Consider a precipitate within a matrix phase, as depicted in Figure 11.2a. The state of the entire microstructure can be represented by a simple variable, termed the *order parameter* ϕ , which inside the precipitate takes the value $\phi = 1$, in the matrix takes the value $\phi = 0$, while ϕ varies across the interface as $0 < \phi < 1$ (Figure 11.2b). Thus ϕ changes only in the region of the interface, i.e., $\nabla\phi \neq 0$ within the interface region, which is characterized by the interface thickness η . This corresponds to the concept of *diffuse interface*, discussed in section 6.7.1 in the framework of continuous transformations. The set of values of the order parameter $\phi(r)$ in the volume V is the *phase field* and characterizes the microstructure. Then the evolution of the microstructure is represented by the time derivative $\dot{\phi}(r)$. It is important to recognize that the free energy is a function of ϕ and $\nabla\phi$ so that $G =$

$G(\phi)$. In this case $G(\phi)$ is a functional. According to the principles of irreversible thermodynamics, discussed in section 5.9, a “flux” is proportional to a “force”, as in the following relation

$$\frac{\partial \phi}{\partial t} = M \frac{\delta G}{\delta \phi} \quad (11.12)$$

where $\delta G/\delta \phi$ is the variational derivative of G and M is the mobility. The above equation indicates that a gradient of G in ϕ causes a change of ϕ with time. Then $\partial \phi / \partial t$ describes the temporal evolution of the microstructure. The system evolves until a new set of ϕ values minimizes the total free energy.

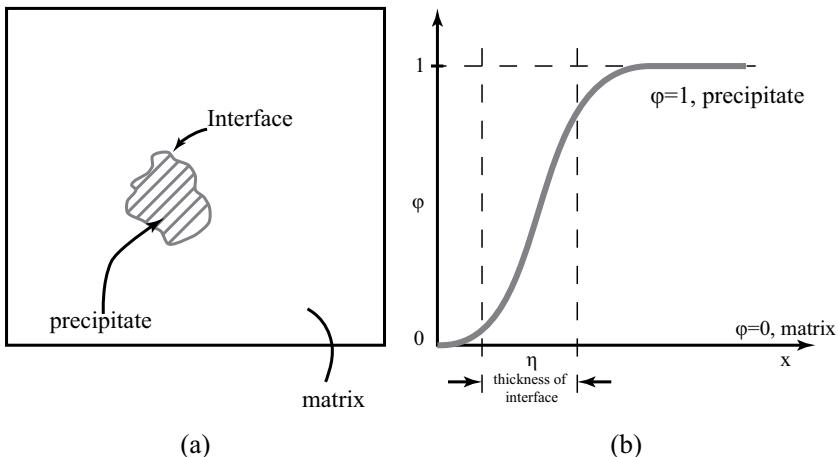


Figure 11.2: (a) Precipitate within a matrix phase, (b) variation of the order parameter. The order parameter is constant in the two phases and changes only in the interface region of thickness η .

Multiple order parameters. Any multiphase system can be described by multiple order parameters. As depicted in Figure 11.3a, three order parameters, ϕ_1 , ϕ_2 and ϕ_3 are assigned to the phases 1, 2 and 3 respectively. The following relations hold:

- region of phase 1: $\phi_1 = 1$ and $\phi_2 = \phi_3 = 0$
- region of phase 2: $\phi_2 = 1$ and $\phi_1 = \phi_3 = 0$
- region of phase 3: $\phi_3 = 1$ and $\phi_1 = \phi_2 = 0$
- interface 1/2: $\phi_1 + \phi_2 = 1$, $0 < \phi_1 < 1$, $0 < \phi_2 < 1$, $\phi_3 = 0$
- interface 1/3: $\phi_1 + \phi_3 = 1$, $0 < \phi_1 < 1$, $0 < \phi_3 < 1$, $\phi_2 = 0$
- interface 2/3: $\phi_2 + \phi_3 = 1$, $0 < \phi_2 < 1$, $0 < \phi_3 < 1$, $\phi_1 = 0$
- at a triple junction: $\phi_1 + \phi_2 + \phi_3 = 1$

In the general case at any location in the system

$$\sum \phi_i = 1$$

The variation of the order parameters at the 1/2 interface is shown in Figure 11.3b.

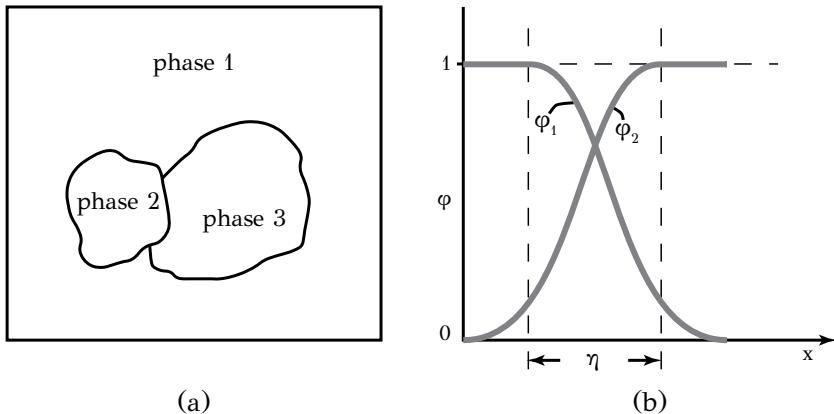


Figure 11.3: (a) Multiphase system consisting of three phases and characterized by three order parameters, (b) variation of the order parameters at the interface between phases 1 and 2.

Order parameters can be assigned to individual grains in a polycrystalline metal. The variation of the order parameters across line AB , which transverses several grains, is shown in Figure 11.4. As will be shown later, the phase field method is particularly appropriate to describe grain growth. Consider, for example, a region with ten grains, described by ϕ_i , $i=1,10$ (Figure 11.5a). Assume that grains 1, 4, 7 and 10 grow in expense of the others, as depicted in Figure 11.5b. In this case the order parameter is not conserved.

Free energy functional. Consider a region of volume V and two phases α and β , depicted in Figure 11.6a. The diffuse interface between the two phases is shown in Figure 11.6b, characterized by a smooth change of the order parameter. If the free energy density is $g(\phi)$, then the total free energy $G(\phi)$, if the system were homogeneous, would be

$$G(\phi) = \int g(\phi) dV$$

However, in heterogeneous systems, since the order parameter varies across the interface, the total free energy $G(\phi)$ is not only a function of ϕ but also of $\nabla\phi$. Since ϕ is a function of position, e.g., $\phi = \phi(x)$, then $G(\phi)$ is a functional. The dependence of G on $\nabla\phi$ can be considered by a Taylor series expansion of G

$$\begin{aligned} G = & g(\phi) + \frac{\partial g}{\partial \nabla \phi} \nabla \phi + \frac{1}{2} \frac{\partial^2 g}{\partial (\nabla \phi)^2} (\nabla \phi)^2 + \dots \\ & + \frac{\partial^2 g}{\partial \nabla^2 \phi} \nabla^2 \phi + \frac{1}{2} \frac{\partial^2 g}{\partial (\nabla^2 \phi)^2} (\nabla^2 \phi)^2 + \dots \end{aligned}$$

which after some math becomes

$$G = \int [g(\phi) + \epsilon^2 (\nabla \phi)^2] dV \quad (11.13)$$

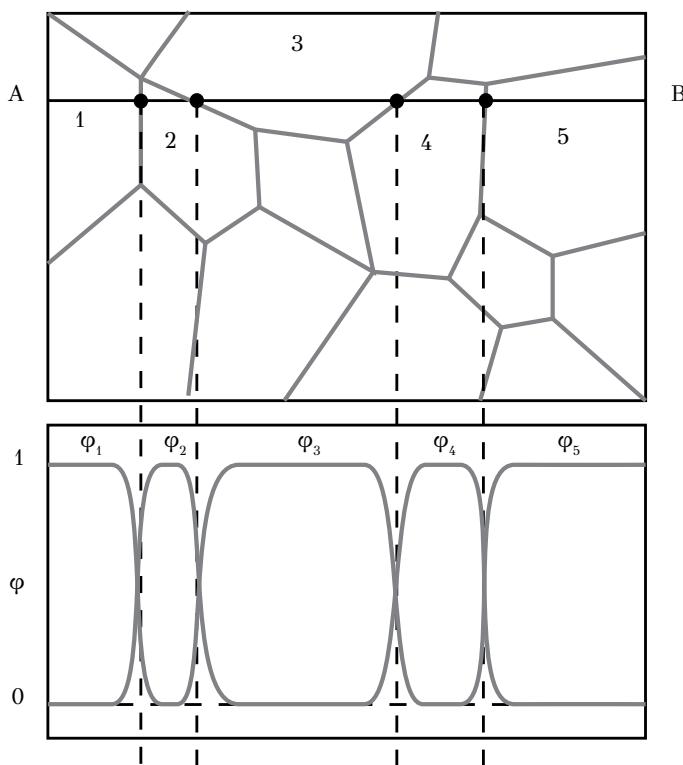


Figure 11.4: The variation of order parameters in a polycrystalline material.

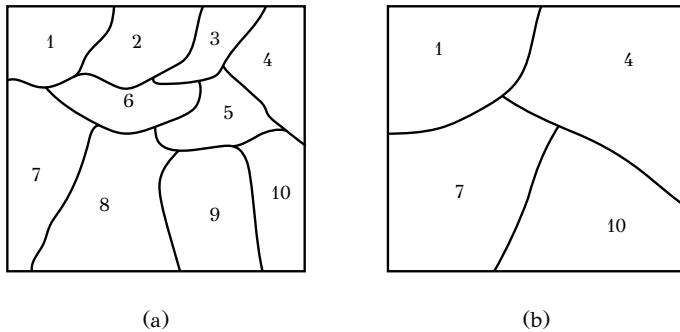


Figure 11.5: Non-conserved order parameter during grain growth: (a) initial microstructure, (b) final microstructure after grain growth.

The first term expresses the homogeneous part of free energy. The second term is termed the *gradient energy* and expresses the inhomogeneous part of the free energy. The parameter ε is the *gradient energy coefficient* and corresponds to the interfacial energy of the diffuse interface. It is given by

$$\varepsilon^2 = \frac{\partial^2 g}{\partial(\nabla\phi)^2} - 2 \frac{\partial}{\partial\phi} \left(\frac{\partial g}{\partial\nabla^2\phi} \right)$$

and scales with the interaction distance between atoms.

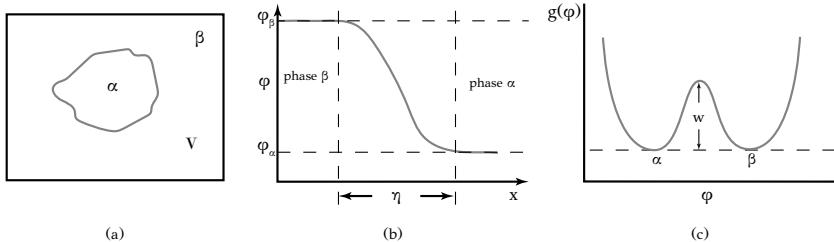


Figure 11.6: (a) Two phases α and β in a region of volume V , (b) variation of the order parameter in the diffuse interface, (c) free energy density $g(\phi)$ represented by a double-well potential.

Thermodynamics. The relation (11.13) above is the free energy functional and expresses the total free energy of an inhomogeneous system as a function of the order parameter and its gradients. At thermodynamic equilibrium the total free energy should be minimized. A function $\phi(x)$ is sought (in the one-dimensional case), which minimizes $G(\phi)$. By taking the variational derivative of G and setting it equal to zero, the following expression is obtained

$$\frac{\delta G}{\delta\phi} = \frac{\partial g}{\partial\phi} - 2\varepsilon^2\nabla^2\phi = 0 \quad (11.14)$$

An expression of $g(\phi)$ is required in order to proceed. A good example is to consider a double-well potential model for $g(\phi)$, depicted in Figure 11.6c and given by

$$g(\phi) = w\phi^2(1 - \phi^2)$$

where w is the height of the energy barrier between the two phases. Relation (11.14) becomes

$$2w\phi(1 - \phi)(1 - 2\phi) - 2\epsilon^2 \frac{d^2\phi}{dx^2} = 0$$

The function $\phi(x)$, which minimizes $G(\phi)$ is obtained by integration as

$$\phi(x) = \frac{1}{2}[1 + \tanh(\frac{x}{2\eta})]$$

where η is the *interface thickness*, given by

$$\eta = \sqrt{\frac{2}{w}}\epsilon$$

The interface thickness depends on the height of the energy barrier w and the gradient energy coefficient ϵ .

Kinetics. Two cases will be considered, (a) non-conserved and (b) conserved order parameter. The first case corresponds to problems where the change of the order parameter is due to gradients of the bulk free energy and no long-range diffusion is involved. Such problems are for example curvature-driven problems (e.g., grain growth), or order-disorder transformations. The second case corresponds to problems involving long-range diffusion and therefore, the mass conservation law should be obeyed. These problems include diffusive phase transformations and spinodal decomposition.

For the case of *non-conserved order parameter*, the evolution law of relation (11.12) becomes

$$\frac{\partial\phi}{\partial t} = -M\left(\frac{\partial g}{\partial\phi} - 2\epsilon^2\nabla^2\phi\right) \quad (11.15)$$

which is the Allen-Cahn equation.

For the case of *conserved order parameter*, the diffusive flux is proportional to the gradient of the diffusion potential, given by

$$J = -M\nabla\frac{\delta G}{\delta\phi} = -M\nabla\left(\frac{\partial g}{\partial\phi} - 2\epsilon^2\nabla^2\phi\right)$$

Conservation of mass is expressed by

$$\frac{\partial\phi}{\partial t} = -\nabla \cdot J$$

and finally the evolution law becomes

$$\frac{\partial\phi}{\partial t} = M\nabla^2\left(\frac{\partial g}{\partial\phi} - 2\epsilon^2\nabla^2\phi\right) \quad (11.16)$$

which is the Cahn-Hilliard equation. It is obvious that the Cahn-Hilliard equation, involves the Laplacian of the variational derivative of G , while the Allen-Cahn equation does not.

A special case of the Cahn-Hilliard equation, when the order parameter corresponds to the composition is

$$\frac{\partial c}{\partial t} = M \left(\frac{\partial^2 g}{\partial c^2} \frac{\partial^2 c}{\partial x^2} - 2\varepsilon^2 \frac{\partial^4 c}{\partial x^4} \right)$$

which involves the 4th derivative of $c(x)$. This equation corresponds to the modified Fick's second law for the treatment of spinodal decomposition and describes the evolution of compositional fluctuations, discussed in Chapter 6 (section 6.7.2).

The evolution laws expressed by equations (11.15) and (11.16) describe the temporal evolution of the order parameter as a response to gradients of G in ϕ and $\nabla\phi$. The system evolves to a new set of order parameters (new state of microstructure) in order to minimize the total free energy. The application of phase-field models in microstructural evolution is discussed later in this chapter with the example of recrystallization and grain growth during the hot rolling of austenite in steels.

11.4 SIMULATION EXAMPLES

Two examples regarding the simulation of diffusional phase transformations will be considered in this section. The first regards the formation of austenite and the associated solute partitioning in medium-*Mn* steels and the second the phase transformations during the homogenization of extrudable *Al* – *Mg* – *Si* alloys.

11.4.1 INTERCRITICAL ANNEALING OF A MEDIUM-MN STEEL

Problem statement. Medium-*Mn* steels belong to the 3rd generation of advanced high-strength steels where the *Mn* content is reduced relative to the high-*Mn* steels, to the level 5–12 wt%. The microstructure is a fine mixture of ferrite and austenite. The excellent combinations of strength and ductility of these steels is attributed to the transformation-induced plasticity (TRIP) effect of the retained austenite. One of the applied processing routes, in order to stabilize the austenitic phase, is intercritical annealing during which austenite forms and is enriched in stabilizing solutes. The retained austenite stability depends on the intercritical annealing temperature and time. The aim of the simulation is to map the austenite fraction and composition as a function of the annealing parameters in a *Fe* – 0.18*C* – 11*Mn* – 3.8*Al* (wt%) steel.

Simulation set-up. Austenite formation is a diffusive phase transformation, which takes place by solute partitioning. The transformation was simulated with the computational kinetics package DICTRA. Two databases were used. The first is a thermodynamic database for iron and steel, TCFE version 6. The second is a mobility database for steels, MOBFE version 2. As discussed in section 11.3.2, diffusivities in DICTRA are calculated by multiplying mobilities with thermodynamic factors. The thermodynamic factors and the interface compositions at local equilibrium are calculated via a link of DICTRA to the Thermo-Calc program. Austenite formation in

ferrite is a moving boundary problem, where the austenite/ferrite interface advances under a mass balance of alloying elements. Following the set-up of Figure 11.1, the moving boundary problem is depicted in Figure 11.7a. A thin austenite region, 50nm in size, is attached to the left of the ferrite region. The total simulation domain size is $1.55\mu\text{m}$. It is important to set the boundary and initial conditions correctly. As for the boundary conditions, the system is regarded as “closed” since no mass flow is allowed in or out. The initial conditions, concerning initial composition of phases, are depicted in Figure 11.7b. Although the compositions in the two phases are identical, the respective activities are not. Diffusional fluxes are generated between the two phases as a response to the activity gradients.

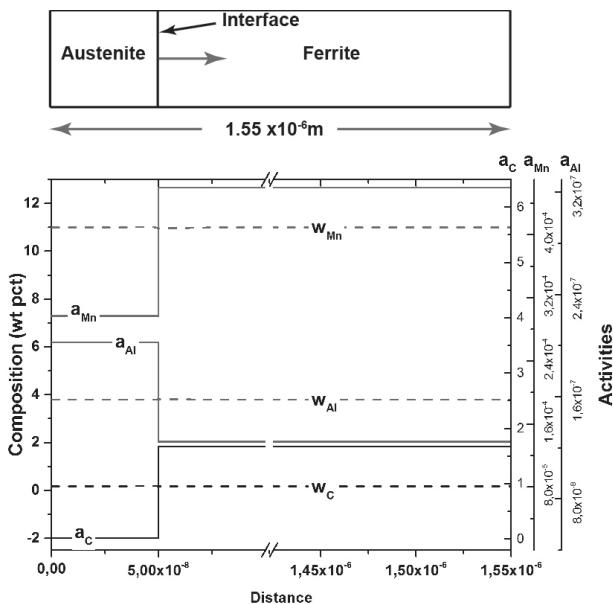


Figure 11.7: (a) Simulation domain for austenite growth in ferrite, (b) initial composition and activity profiles for Mn, Al and C.

Reprinted from: H. Kamoutsi, E. Gioti, G.N. Haidemenopoulos, Z. Cai and H. Ding, Metallurgical and Materials Transactions A, 46, No.11, pp. 4841-4846, 2015, with permission from Springer.

Results. The evolution of austenite volume fraction is depicted for several annealing temperatures in Figure 11.8. The simulation revealed three growth stages. In stage I, the initial rapid increase of austenite fraction is due to growth under no-partitioning local equilibrium conditions (NPLE), where austenite growth is controlled by carbon diffusion, while the substitutional alloying elements do not diffuse. In stage II, austenite growth is controlled by Mn diffusion in ferrite and takes place under partitioning local equilibrium (PLE) with partition of Mn and Al. In stage III, the final, very slow equilibration of austenite is controlled by Mn diffusion in austenite.

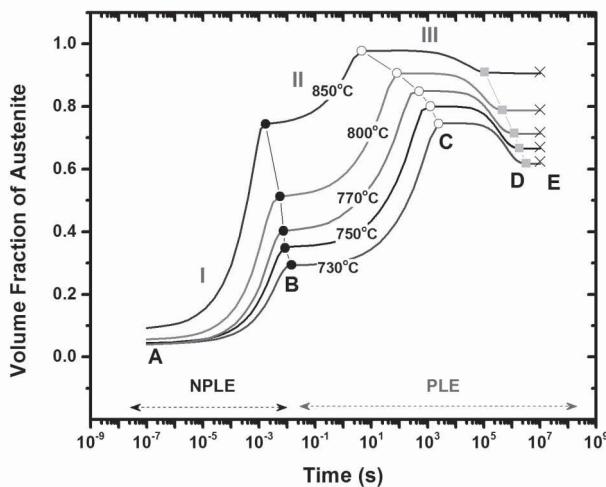


Figure 11.8: Evolution of austenite volume fraction with annealing time for several intercritical annealing temperatures, indicating the stages of NPLE and PLE growth.

Reprinted from: H. Kamoutsi, E. Gioti, G.N. Haidemenopoulos, Z. Cai and H. Ding, Metallurgical and Materials Transactions A, 46, No.11, pp. 4841-4846, 2015, with permission from Springer.

ite and is associated with a decrease of the volume fraction of austenite. Solute partitioning at 750°C, during PLE growth of stage *II* is shown in Figure 11.9. Austenite is enriched in *Mn* and *C*. The *Mn* enrichment is more intense at the austenite/ferrite interface. Due to the low diffusivity of *Mn* in austenite, the *Mn* is not homogenized in the austenitic phase. This takes place at much longer times. On the other hand *Al* partitions from the austenite to ferrite during the growth of austenite. More results and discussion about this simulation can be found in the paper by Kamoutsi et al.¹

11.4.2 SIMULATION OF HOMOGENIZATION IN AL-MG-SI ALLOYS

Problem statement. Aluminum extrusions find important applications in the building and transportation sectors. Large billets of *Al* – *Mg* – *Si* (6xxx series) alloys are produced by the direct-chill casting method. The as-cast material exhibits low extrudability due to the inhomogeneous microstructure formed during casting. Elemental and phase microsegregation as well as formation of *Fe*-containing intermetallics are among the major parameters degrading the extrudability of the material. A homogenization treatment is then applied in order to remove most of these inhomogeneities. The most important transformations during homogenization are: (a) the dissolution

¹H. Kamoutsi, E. Gioti., G.N. Haidemenopoulos, Z. Cai, H. Ding, Metall. Mater. Trans. A, 46A (2015) 4841.

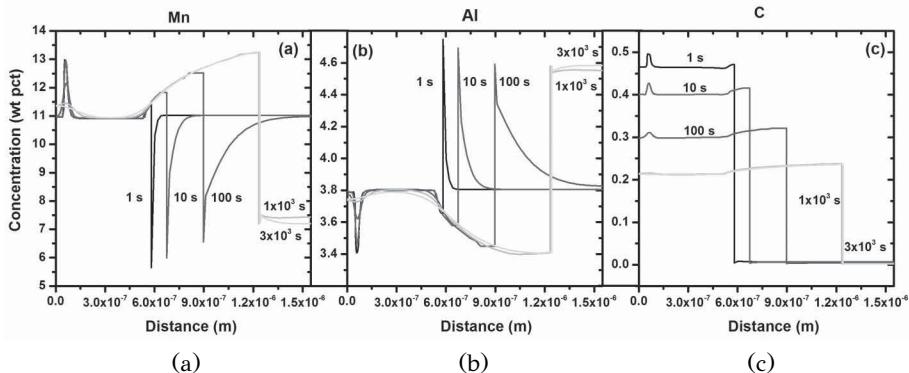


Figure 11.9: Profiles of (a) *Mn*, (b) *Al* and (c) *carbon* after annealing in the PLE region stage *II* for annealing temperature of 750°C and the annealing times indicated. The austenite and ferrite regions, for each annealing time, are on the left and right of the interface respectively.

Reprinted from: H. Kamoutsi, E. Gioti, G.N. Haidemenopoulos, Z. Cai and H. Ding, Metallurgical and Materials Transactions A, 46, No.11, pp. 4841-4846, 2015, with permission from Springer.

of Mg_2Si and reprecipitation during homogenization cooling and (b) the transformation of the $\beta - \text{AlFeSi}$ intermetallic, which is a hard intermetallic with sharp edges, to the more rounded $\alpha - \text{AlFeSi}$ intermetallic. The aim of the present example is the simulation of these transformations during the homogenization of a 6060 aluminum alloy.

Simulation set-up. The simulation is performed in two stages. The first stage deals with the simulation of microsegregation resulting from the solidification of the alloy. The cooling rate during direct-chill casting is fast enough to assume negligible diffusion of alloying elements in the growing solid. In this case the Scheil solidification model can be applied. The composition of the solid C_s at a fraction of solid f_s is given by the Scheil equation, which for a binary system is

$$C_s = C_0 k (1 - f_s)^{k-1}$$

where C_0 is the alloy nominal composition and k the partition coefficient. For multi-component alloy systems the Scheil-Gulliver model has been implemented in the so-called SCHEIL module of Thermo-Calc. Microsegregation profiles of elements and phases can be calculated. The second stage of the simulation deals with the homogenization treatment. The Mg_2Si dissolution/reprecipitation and the transformation of β to $\alpha - \text{AlFeSi}$ are simulated with DICTRA. The microsegregation profiles from the previous stage are taken as initial conditions for the simulation. Mg_2Si , β and $\alpha - \text{AlFeSi}$ phases are regarded as dispersed phases in the aluminum matrix. Due to variation in the grain size, a dual-grain model (DGM) was constructed.

Results. The microsegregation of phases and elements are shown in Figure 11.10. It is seen that the $\alpha - \text{AlFeSi}$ phase forms above 85 % solidification while all other

phases form above 95 % solidification. This means that the intermetallic phases form at the secondary dendrite arm boundaries towards the end of solidification. Accordingly the microsegregation of the alloying elements in the aluminum matrix follows a similar trend, i.e., increase towards the end of solidification. The drop in the concentration profile of Mg and Fe is attributed to the formation of Mg_2Si and iron intermetallics respectively. The spatial evolution of the $\beta - AlFeSi \rightarrow \alpha - AlFeSi$ transformation is depicted in Figure 11.11 for 50, 100 and 500 sec at 580°C. As discussed above, a dual grain model (DGM) was used. One grain of $90\mu m$ size attached to a grain of $20\mu m$ size. Two interesting features to be considered are: (a) the transformation starts at a distance away from the grain boundary while the transformation front moves towards the grain boundary and (b) there is an exact correspondence between the profiles of the $\alpha - AlFeSi$ and $\beta - AlFeSi$ phases, i.e., where there is a reduction in $\beta - AlFeSi$ there is a corresponding increase in $\alpha - AlFeSi$. This simulation is very helpful as it allows the calculation of the minimum homogenization times required for the transformation of the Fe -containing intermetallics. A more detailed treatment of this simulation can be found in the entry, by Haidemenopoulos and Sarafoglou¹, in the Encyclopedia of Aluminum and its Alloys.

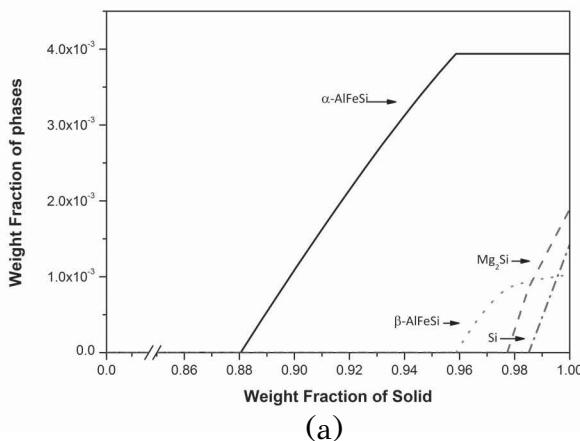
11.5 ALLOY DESIGN: MEDIUM MN STEELS

A computational alloy design example will be presented in this section. It is concerned with the design of alloy compositions for medium- Mn steels. The main characteristics of these steels have been discussed in section 11.4.1. The development of these steels has been, so far, based on empirical approaches. The aim of the design is to identify compositions, which exhibit process windows (PW), which in turn satisfy certain design objectives regarding the amount and stability of retained austenite. The Alloyneering methodology outlined in section 11.2 was followed.

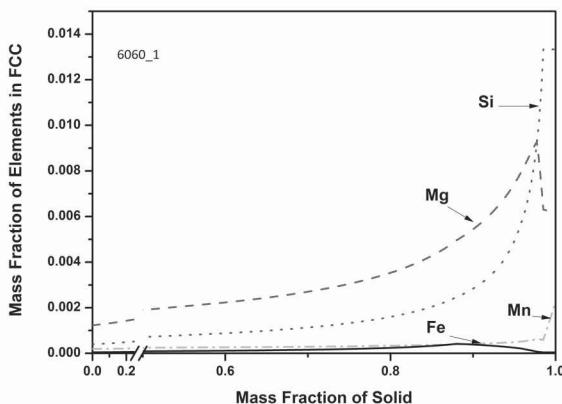
Analysis. The alloy design starts with the definition of the original composition space (OCS). This is actually the search space for the alloy compositions under consideration. In the $Fe - C - Mn - Si - Al - Ni$ system the search space is defined by the ranges of alloying elements. For *Carbon*: 0.1-0.3, for *Mn*: 2-10, for *Al*: 0-4, for *Ni*: 0-2, *Si* is fixed at 0.15. Considering a step increment in the search space for each alloying element, the total number of alloys in the OCS is 2835. The alloy design criteria involve the following parameters:

- Fraction of retained austenite set between 20 and 40%
- Stability of retained austenite, expressed by the M_s temperature, between -60 and $-20^{\circ}C$
- Width of process window larger than $10^{\circ}C$
- No cementite in microstructure

¹G.N. Haidemenopoulos, P.I. Sarafoglou, Extrudable Al-Si-Mg alloys: simulation of microsegregation and homogenization, in Encyclopedia of Aluminum and its Alloys, G. Totten, O. Kessler, M. Tiraykioglou (eds.), Taylor & Francis, USA, 2017.



(a)



(b)

Figure 11.10: (a) Microsegregation of phases and (b) Microsegregation of elements in a 6060 Al-alloy.

Reprinted from: G.N Haidemenopoulos and P.I. Sarafoglou, George Totten and Ollaf Kessler (editors), Taylor & Francis, New York, USA, 2016, with permission from Taylor & Francis.

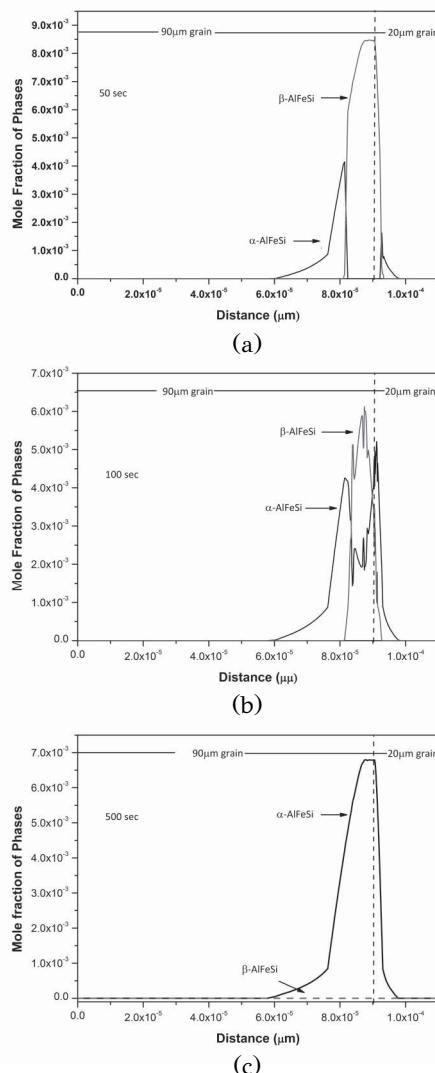


Figure 11.11: Spatial evolution of the mole fraction of $\alpha - AlFeSi$ and $\beta - AlFeSi$ for homogenization of the 6060 alloy at $580^{\circ}C$ for (a) 50, (b) 100 and (c) 500 sec.

Reprinted from: G.N Haidemenopoulos and P.I. Sarafoglou, Encyclopedia of Aluminium and Aluminum Alloys, George Totten and Ollaf Kessler (editors), Taylor & Francis, New York, USA, 2016, with permission from Taylor & Francis.

Simulation and mapping. Thermodynamic calculations were performed for all 2835 alloys of the OCS, in order to identify the compositions, which exhibit a PW and, therefore, satisfy the design criteria. The definition of a PW is shown for alloy $Fe - 0.15C - 8Mn - 2Al$ in Figure 11.12. The fraction of retained austenite and the M_s temperature are plotted against the intercritical annealing temperature. The fraction of retained austenite reaches a maximum value just above $650^{\circ}C$ and then drops since it transforms partially to martensite, as the M_s temperature is above room temperature. A_{CM} is the temperature above which cementite dissolves during intercritical annealing. The application of the design criteria defines the process window depicted in Figure 11.12 with a width from 625 to $636^{\circ}C$. If the specific alloy is intercritically annealed within this process window, then all design criteria, stated above, are satisfied.

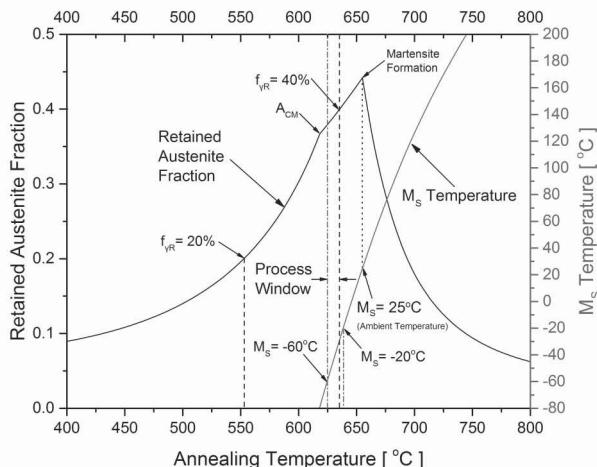


Figure 11.12: The definition of a process window (PW) for alloy $Fe - 0.15C - 8Mn - 2Al$ (wt%).

The Thermo-Calc program and the associated TCFE6 database were used for the calculations. Due to the large number of calculations a special interface operating on Thermo-Calc was created. The M_s temperature, in $^{\circ}C$, was calculated using the Andrews equation

$$M_s = 539 - 423W_C - 30.4W_{Mn} - 7.5W_{Si} - 17.7W_{Ni} + 30W_{Al}$$

where W_i are the alloying contents in wt%. The fraction f_m of austenite transformed to martensite was calculated by the Koistinen-Marburger model as

$$f_m(T) = 1 - \exp[-0.001(M_s(T) - T_R)]$$

where T_R is the ambient temperature. The volume fraction of retained austenite is

Table 11.3

Short list of medium-Mn steel compositions resulting from the optimization process

Rank	Alloy Composition (wt%)	Annealing Temperature (°C)	Fraction of Retained austenite (%)	M_s Temperature (°C)
1	0.15C-8Mn-2Al	632	39.4	-36.8
2	0.15C-6Mn-2Ni-1.5Al	632	38.8	-33.8
3	0.15C-8Mn-0.5Ni-2.5Al	640	39.5	-36.3
4	0.15C-5Mn-1.5Ni-2.5Al	654	28.8	-34.1
5	0.10C-8Mn-1Ni-2Al	622	39.3	-35.1

given by subtracting the martensite formed from the austenite formed at the intercritical annealing temperature

$$f_{\gamma R} = f_{\gamma}(T)[1 - f_m(T)]$$

From the 2835 alloys in the OCS only 305 alloys exhibit a PW. The next step is to apply an optimization procedure to identify the optimum compositions.

Optimization. The aim of the optimization process is to identify a short list of compositions of those found to exhibit a PW from the previous step. The selected design objectives are the volume fraction of retained austenite and the M_s temperature. In addition, a new parameter, termed the composition index (CI) is introduced to account for the total alloy content. The CI is used to favor less C and Mn and thus to improve weldability and to reduce raw material costs. Under these conditions, optimal compositions should maximize retained austenite and annealing temperature, while minimizing the M_s temperature and CI. Due to the multiple alloy design criteria, the problem can be modeled as a multi-objective optimization (MOOP) problem. In the framework of MOOP, an alloy composition can be considered optimal if it results in objectives lying on the Pareto front. A formal definition of Pareto optimality is the subset of solutions, for which it is not possible to improve an objective without simultaneously worsen at least one of the others. From the total 305 alloys exhibiting a PW, 173 alloys were identified as Pareto optimal solutions. The set of Pareto optimal solutions is extensive and, therefore, a method is required, to select a short list of solutions with overall better performance. In the present case the geometric mean of the objective parameters was calculated for each Pareto optimal solution. The five Pareto optimal solutions with the highest geometric mean values are the final result of the optimization process. These five compositions are shown in Table 11.3.

The alloy design example presented above is entirely based on computational alloy thermodynamics and therefore, equilibrium conditions were considered. It is anticipated that some of the annealing times required to achieve the required austenite fractions and associated stabilities might be long. It is, therefore, necessary to

follow-up this procedure with kinetic simulations, similar to the one presented in section 11.4.1 to define industrially feasible compositions. Nevertheless, the proposed methodology could be the first step towards the computational alloy design in this class of steels. More details regarding this example can be found in the paper by Aristeidakis and Haidemenopoulos¹.

11.6 PROCESS DESIGN: MULTI-PASS HOT ROLLING OF STEELS

The process design example presented in this section concerns the multi-pass hot rolling of steels. Hot rolling is the standard process, following casting, for the production of steel plates or strips. The plate enters the rolling mill, which consists of a series of rolls, at a high temperature. At each rolling pass the plate thickness is reduced by a certain amount by the deformation of austenite. During the interpass time t_{in} , recovery, recrystallization and grain growth processes take place. Due to these effects the grain size of austenite changes continuously throughout the hot rolling process. The aim of the simulation is to predict the final grain size of austenite after hot rolling. The simulation can then be used to design the multipass hot rolling process in terms of number of passes, thickness reduction per pass, interpass times and temperatures in order to achieve the desired grain size. Due to space limitations only the simulation stage will be presented in this chapter. The simulation presented here is part of a Ph.D. thesis of M.I.T. Tzini².

Analysis. A six-pass hot rolling of austenite will be simulated. The rolling set up is shown in Figure 11.13. The phase-field method is used for the simulation. As described in section 11.3.3, in recrystallization and grain growth problems, the order parameter is not conserved and the evolution of the order parameter is described by the Allen-Cahn equation, given by relation 11.15. Additional assumptions are as follows:

- Due to the short deformation and interpass times, the problem can be described as a discontinuous static recrystallization problem, taking place during the interpass time with nucleation of recrystallization at the roll exit.
- Each successive rolling pass takes place at a lower temperature. The grain boundary mobility is calculated as a function of temperature and is, therefore, lowered at each pass. An Arrhenius-type function is used: $\mu = \mu_0 \exp(-Q/RT)$, where Q is the activation energy for grain boundary migration.
- The stored energy and the nucleation density of recrystallized grains, at each rolling pass, are calculated as a function of strain rate and increase at each successive rolling pass.
- A constant interfacial grain boundary energy has been assumed for all rolling passes.

¹I.S. Aristeidakis, G.N. Haidemenopoulos, Metall. Mater. Trans. A, 48A (2017)2601

²M.I.T. Tzini, Computational process design of HSLA and Dual-Phase Steels, Doctoral Thesis, in progress, University of Thessaly, Greece.

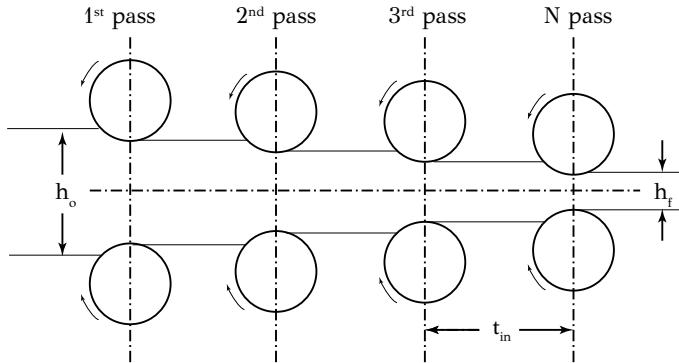


Figure 11.13: Schematic of the multipass hot rolling process.

The 2D simulation domain, representing the initial microstructure, is a 450×300 grid with cell size $1\mu m$. The thickness of the diffuse interface is selected as $\eta = 4$ cells. The region contains 18 grains, corresponding to an initial grain size of $100\mu m$. The problem was solved using the phase-field software MICRESS.

Results. The evolution of the grain size across the hot rolling process is depicted in Figure 11.14 as the variation of mean grain size vs. time. Starting at $100\mu m$ grain size, the plate enters the first pass. The interpass time between the first and second pass is 4 sec. The stored energy of the first pass causes the recrystallization and grain growth during the interpass time. The abrupt drop of the grain size from $100\mu m$ to $8\mu m$ is due to the nucleation of the new grains. The recrystallization and grain growth between the first and second passes are depicted in Figure 11.15a.

Recrystallization takes place from A to B. The respective recrystallization fraction is depicted, as a function of time, in Figure 11.15b. The shape of the curve is sigmoidal, as expected from the theory of recrystallization discussed in section 7.7.5 of Chapter 7. In the simulation of the remaining passes, the microstructure at the end of the first pass is entered as input at the beginning of the second pass and so on. In Figure 11.14, the points correspond to results from the work of Suehiro et al¹. It appears that the results of the phase-field model, described here, are in good agreement with the Suehiro et al. results. Typical microstructures at the end of the first and sixth passes are compared with the initial microstructure in Figure 11.16. The first rolling pass has the highest effect in refining the austenite grain size. The final grain size is of the order of $20\mu m$, which corresponds to a significant grain refinement of austenite during the hot rolling process. The simulation can be used for the design of the rolling process. Specific parameters of the rolling schedule, such as the number of rolling passes or the reduction per pass (strain per pass) can be identified in order to achieve the desired grain size at the end of the hot rolling process.

¹M. Suehiro, K. Sato, Y. Tsukano, H. Yada, T. Senuma, Y. Matsumura, Trans. ISIJ, 27 (1987) 439-445.

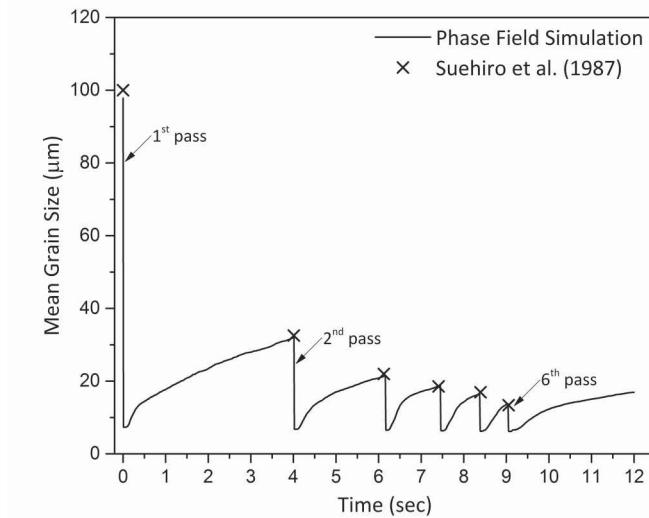


Figure 11.14: Evolution of the mean grain size during the multipass hot rolling process.

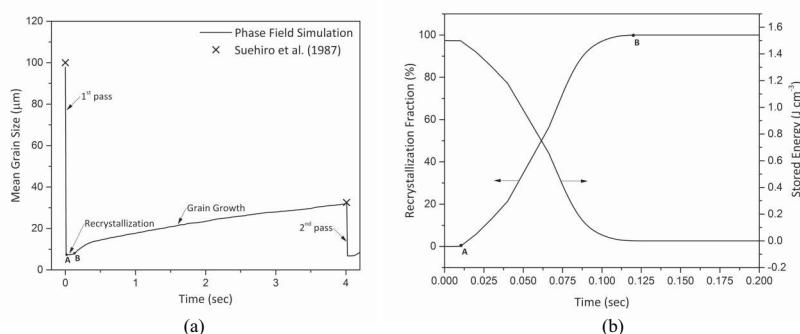


Figure 11.15: (a) Evolution of grain size during the interpass time after the first rolling pass, (b) recrystallization fraction as a function of time.

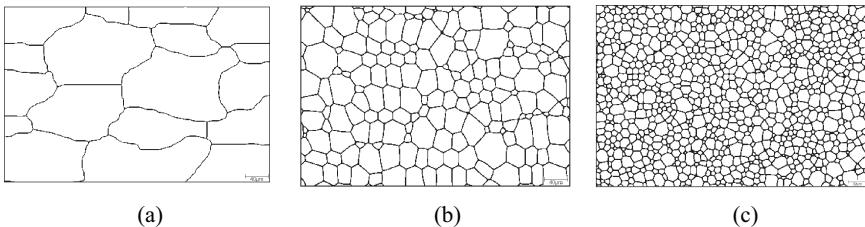


Figure 11.16: (a) Grain structures according to the simulation: (a) initial microstructure, (b) after the first pass, (c) after the sixth pass.

11.7 SYNOPSIS

1. *Alloyneering* is a generic term, which combines the words *Alloy* and *Engineering* and describes a concise knowledge-based methodology leading to the development of Engineered Alloys. Alloyneering is based on the principles of ICMD and encompasses the application of computational alloy thermodynamics and kinetics, precipitation and phase-field models, for the simulation of microstructural evolution and the design of metallic alloys with tailored properties. The same approach can be used for the design of the required processing conditions to obtain the corresponding microstructure and thus to achieve the required properties.
2. According to the Alloyneering methodology, the alloy design process is performed in five stages: Analysis, Simulation, Validation, Mapping and Optimization.
3. The simulation framework for alloy design includes computational alloy thermodynamics, computational kinetics and phase-field modeling.
4. Computational thermodynamics is based on the CALPHAD approach for the description of the Gibbs free energy of substitutional and interstitial solid solutions and intermetallic compounds in multicomponent systems. For the case of substitutional solutions, the excess term is modeled through a Redlich-Kister-Muggianu polynomial involving composition terms and interaction parameters. The relevant software for computational alloy thermodynamics is Thermo-Calc.
5. Computational kinetics is a computational framework for solving problems related to diffusional phase transformations in metals and alloys. Multicomponent diffusion is taken into account by the incorporation of composition and temperature dependent mobilities, while the associated thermodynamic factors are calculated by computational alloy thermodynamics. The numerical solution of the diffusion equation and the mass balance equation at the interface provides the solution to moving boundary problems. The relevant software for computational kinetics is DICTRA.

6. Microstructural evolution, during phase transformations, is accomplished with the migration of interfaces. Phase-field modeling is a methodology, which allows the study of the thermodynamics and kinetics of interface migration and the associated microstructural evolution. The state of the entire microstructure can be represented by the order parameter ϕ . The set of values of the order parameter $\phi(r)$ in the volume V is the phase field and characterizes the microstructure. Then the evolution of the microstructure is represented by the time derivative $\dot{\phi}(r)$. The system evolves until a new set of ϕ values minimizes the total free energy of the system.

BIBLIOGRAPHY—FURTHER READING

1. Xiong, W., Olson, G.B. (2015). Integrated computational materials design for high-performance alloys, *MRS Bulletin*, 40, 1035-1043.
2. Kaufman, L., Bernstein, H. (1970). *Computer Calculation of Phase Diagrams*, Academic Press, NY.
3. Saunders, N., Miodownik, P. (1998). *Calphad*, Pergamon materials series, Vol.1 (ed. R.W. Cahn).
4. Lukas, H. L., Fries, S.G., Sundman, B. (2007). *Computational Thermodynamics and the Calphad Method*, Cambridge University Press.
5. Thermo-Calc and DICTRA: www.thermocalc.com
6. Steinbach, I. (2009). Phase-field models in materials science, *Modelling Simul. Mater. Sci. Eng.* 17, 073001 (review from solidification to phase-field method)
7. Schmitz, G. J., Prahl, U. (2012). *Integrative Computational Materials Engineering*, Wiley-VCH, Germany.
8. Micress: <http://web.access.rwth-aachen.de/MICRESS/>



Taylor & Francis
Taylor & Francis Group
<http://taylorandfrancis.com>

Index

- accommodation factor, 203
activation energy, 35, 304
activation energy for nucleation, 191
activation volume, 304
activity, 102
activity coefficient, 102
activity gradients, 453
age hardening, 323
aging, 199, 323, 391
aging curve, 327
Al₂Cu intermetallic phase, 75
Al-Cu alloys, 75, 199
Allen-Cahn equation, 451
allotropic forms, 16
allotropic transformations, 95
allotropy, 16
alloy composition, 92
alloying elements in steel, 395
Alloyneering, 440
alternating stress, 349
amorphous materials, 5
Andrews equation, 405, 459
anelasticity, 21
anisotropy, 296
anisotropy of crystals, 6
annealing, 283, 425
annealing treatments, 425
annihilation of dislocations, 285
anticoherency dislocations, 241
antiphase boundaries, 229
Arrhenius kinetics, 37
Arrhenius law, 37
athermal, 404
atomic arrangement, 5
atomic configuration, 5
atomic fraction, 92
atomic mobility, 144
atomic packing factor, 14
austempering, 427
austenite, 390
austenite stabilizers, 396
autotempering, 421
average concentration, 174
Bagaryatski, 408
bainite, 391, 407
bainite sheaves, 407
bainitic ferrite, 407
bake hardening, 316
basal slip, 266
Bending force, 72
bivariant equilibrium, 107
body-centered cubic, 11
Boltzmann equation, 31
Boltzmann equation for entropy, 33
boundary conditions, 453
Bravais lattices, 6
brittle fracture, 336
bulk diffusion, 136, 376
Burgers circuit, 55
Burgers vector, 55
C-curve kinetics, 403, 408
Cahn-Hilliard equation, 452
CALPHAD approach, 441
carbide dispersion, 332
carbide formers, 396
carbide-free bainite, 408
carbon steels, 392
carbonitrides, 321, 396
carburized layer, 428
carburizing, 176, 390, 427
case depth, 390, 428
cells, 285
cementite, 391
chemical equilibrium, 27
chemical potential, 91, 101, 149
chemical potential gradient, 147
chemical spinodal, 224
cleavage, 338
climb, 369
close-packed directions, 11, 265
close-packed planes, 13, 265

- clusters, 18
coarsening, 213
coarsening rate, 215
Coble creep, 376
Coffin-Manson law, 357
coherency dislocations, 241
coherency hardening, 327
coherency loss, 198
coherency strains, 196
coherent, 26, 81
coherent interfacial energy, 81
coherent spinodal, 224
cold work, 283
cold working, 44, 272, 283
common tangent, 104
compatibility criterion, 271
complete solid solubility, 22
compound zone, 428
computational alloy design, 439, 456
computational kinetics, 444
concentration gradient, 147
concurrent nucleation and growth, 207
configurational entropy, 34, 42
conserved order parameter, 451
continuous cooling transformation diagrams, 413
continuous transformations, 188, 215
controlled rolling, 430
cooperative growth, 401
coordination number, 11
core diffusion, 137
corrosion fatigue, 350
Cottrell atmospheres, 166, 316
crack growth rate, 363
crack initiation, 361, 363
crack propagation, 363
crack-tip opening displacement, 349
creep, 336, 368
creep curve, 369
creep fracture, 368
creep life, 379
creep strain, 368
creep-fatigue, 350
critical radius for nucleation, 191
critical resolved shear stress, 46, 264
cross slip, 264, 354, 362, 374
CRSS, 264
cryogenic steels, 396
crystal planes, 6
crystal rotation, 270
crystal structure, 5
crystallographic directions, 6
Crystallography, 5
Curved grain boundaries, 291
cyclic hardening, 351, 354
cyclic loading, 349, 354
cyclic softening, 351
cyclic stress-strain curve, 351
damage, 349
damage tolerance, 346, 363
Darken equation, 157
Debye frequency, 203
deformation mechanism maps, 377
deformation texture, 295
Deoxidation, 395
destabilization, 279
desulfurization, 396
diffuse interface, 216, 446
diffusion, 115, 135, 369
diffusion coefficient, 138
diffusion control, 200, 403
diffusion equation, 167
diffusion matrix, 445
Diffusion mechanisms, 135
diffusion problems, 168
diffusion zone, 428
diffusion-controlled growth, 205
diffusional flow, 166, 370, 375
diffusionless transformation, 230
diffusive phase transformations, 186
dimples, 340
Directional solidification, 381
discontinuous plastic flow, 317
dislocation, 26
dislocation climb, 372
dislocation creep, 372
dislocation density, 60, 273, 285, 330
dislocation dissociation, 275, 373

- dislocation energy E , 69
dislocation glide, 53
dislocation interactions, 272
dislocation intersections, 275
dislocation jogs, 65, 373
dislocation kinks, 65
dislocation loops, 58
dislocation multiplication, 272
dislocation width, 309
Dislocations, 49
dissociation, 63
dissociation into partials, 63
double-well potential model, 451
driving force, 35, 146
driving force for coarsening, 213
driving force for nucleation, 190
driving force for recrystallization, 287
driving force for solidification, 95
ductile fracture, 336
ductile-to-brittle transition, 338
ductile-to-brittle transition temperature, 342
ductility, 279
dynamic recovery, 274, 284
dynamic recrystallization, 284
dynamic strain aging, 317
earing, 296
easy glide, 274
edge dislocation, 51
elastic accommodation, 249
elastic energy of a dislocation, 60
elastic miscibility gap, 224
elastic modulus, 335
Elastic-Plastic Fracture Mechanics, 349
elements of microstructure, 330
embrittlement, 397
end-quench test, 418
endurance limit, 355
energy barrier ΔG^* , 35
enthalpy, 29, 93
enthalpy of mixing, 96
entropic term, 35
entropy, 29, 94
entropy of mixing, 31, 96
equiaxed grain morphology, 422
equilibrium shape, 77
equivalent crystal planes, 9
error function, 172
eutectic alloy, 113
eutectic reaction, 108, 112, 391
eutectic temperature, 111
eutectoid reaction, 108, 394, 401
eutectoid steel, 394
excess term, 98
extended volume, 210
extra half plane, 49, 53
extrudability, 454
extrusions, 362
face-centered cubic, 11
failure analysis, 340
fatigue, 349
fatigue life, 350
Fe-1C-4Mo steel, 193
Fe-C phase diagram, 391
ferrite, 390
ferrite stabilizers, 396
ferrite-pearlite microstructure, 395
fibrous texture, 295
Fick's first law of diffusion, 138
Fick's second law, 141, 142
Fick-Onsager law, 445
flame hardening, 427
forest hardening, 277
formability, 279
fraction of interstitials, 44
fraction of vacancies, 44
fractography, 340
fracture criterion, 345
fracture mechanics, 337, 344
fracture toughness, 396
Frank's rule, 62, 276
Frank-Reed source, 273, 319
free energy, 94
free energy curves, 189
free energy functional, 448
free energy of mixing, 96
free surfaces, 76, 77
free-machining steels, 397

- fretting fatigue, 350
 full annealing, 425
 gas carburizing, 427
 Gibbs free energy, 28, 34
 Gibbs phase rule, 91, 105
 Gibbs-Thomson effect, 163, 213
 glass, 5
 glass transition temperature T_g , 5
 Glide force, 71
 glissile interfaces, 200, 241
 GP zones, 323
 gradient energy, 218, 450
 gradient energy coefficient, 450
 gradient energy term, 221
 grain boundaries, 24, 76, 77
 grain boundary diffusion, 136, 376
 grain boundary migration, 292
 grain boundary sliding, 372, 377
 grain boundary strengthening, 307, 319
 grain growth, 283, 289, 461
 grain refinement, 294, 321, 396
 grain size, 25, 330, 377
 grains, 24
 growth, 189, 200
 growth processes, 200
 growth rate, 204
 habit plane, 233, 239
 Hall-Petch strengthening, 319
 hard-sphere model, 11
 hardenability, 396, 415
 heat treatment, 185, 423
 Helmholtz free energy F , 28
 heterogeneous nucleation, 192
 heterogeneous nucleation at grain boundaries, 194
 hexagonal close-packed, 11
 high-angle boundaries, 78
 high-cycle fatigue, 355
 high-diffusivity paths, 137
 high-strength low alloy steels, 430
 homogeneous nucleation, 190
 homogeneous shear, 280
 homogenization, 167
 homogenization treatment, 454
 hot working, 283
 HSLA piping steels, 429
 HSLA steels, 321
 Hume-Rothery rules, 22
 hydrogen outgassing, 173
 hypereutectic alloys, 118
 hypereutectoid steel, 395
 hypoeutectic alloys, 114
 hypoeutectoid steel, 394
 ideal c/a , 13
 ideal shear strength, 47
 ideal solution, 98
 impact strength, 342
 impenetrable particles, 327
 incoherent, 26
 incoherent boundaries, 81
 incoherent interfacial energy, 82
 independent slip system, 271
 induction hardening, 427
 instantaneous source, 168
 intercritical annealing, 430, 452
 interdiffusion, 153
 interdiffusion coefficient, 157, 220
 interface control, 200
 interface migration, 446
 interface thickness, 446, 451
 interface-controlled growth, 202
 Interfaces, 75
 interfacial energy, 26, 77, 216
 interfacial mobility, 204
 interfacial mobility of grain boundaries, 292
 intergranular cracking, 377
 intergranular ferrite, 400
 intergranular fracture, 338
 interlamellar spacing, 401
 intermediate phases, 24
 intermetallic compounds, 23
 internal energy, 28
 internal friction, 21
 interparticle spacing, 326
 interphase boundaries, 76, 81

- interrupted quenching and tempering, 427
interstitial diffusion, 135
interstitial sites, 19
interstitial solid solutions, 18
interstitials, 41
intrinsic diffusion coefficients, 154
intrusions, 362
invariant equilibrium, 107
invariant reactions, 108
invariant-lattice deformation, 238
invariant-line strain, 236
invariant-plane strain, 233
inverse Hall-Petch effect, 322
irreversible thermodynamics, 158
Isothermal Transformation (*IT*) diagram, 408
isotropic materials, 5

J-integral, 349
Johnson-Mehl equation, 210
Johnson-Mehl-Avrami equation, 402
Johnson-Mehl-Avrami-Kolmogorov equation, 212
Jominy hardenability test, 418

killed steels, 395
kinetics of coarsening, 214
kink formation, 310
Kirkendall effect, 154
Koistinen-Marburger model, 459
Kurdjumov-Sachs, 399, 407

lamellar structure, 394
Laplace transform, 175
laser surface hardening, 427
latent heat of fusion, 94
lattice, 6
lattice diffusion, 136
lattice misfit, 196
lattice parameter, 11
lattice resistance, 306, 307
lattice walk, 139
ledeburite, 122
left handed screw dislocation, 58
lever rule, 105
limited solid solubility, 18, 22
Line tension, 72
Linear Elastic Fracture Mechanics, 349
linear imperfections, 49
linear strain hardening, 274
liquid carburizing, 427
liquidus, 110, 112
load drop, 316
local thermodynamic equilibrium, 445
localization, 279
Lomer-Cottrell lock, 276
long diffusion times, 173
long-range obstacles, 302
low-angle boundaries, 78
low-cycle fatigue, 355
lower bainite, 408
lower yield point, 316
LSW coarsening theory, 214
Lueders band, 316

M_f temperature, 405
M_s temperature, 405
machinability, 397
magnetic ordering, 443
martensite, 230, 390
martensite morphologies, 239
martensitic embryo, 243
martensitic interfaces, 241
martensitic transformation, 230, 263, 404
Materials Genome Initiative, 439
maximum transformation rate, 208
Maxwell-Boltzmann energy distribution, 35
mean grain size, 292
mechanical behavior, 335
mechanical driving force, 248
Mechanical equilibrium, 27
mechanical fatigue, 350
mechanical properties, 335
mechanical stabilization, 405
mechanical testing, 335
mechanical twinning, 263, 279
medium-*Mn* steels, 452, 456

- metal fatigue, 349
 metallic bonding, 11
 metallic glasses, 5
 metastable equilibrium, 27
 metatectic reaction, 108
 miscosegregation, 454
 microstructural evolution, 446
 microstructure, 335
 microvoid formation, 340
 Miller indices, 9
 Miner's rule, 361
 miscibility gap, 110, 224
 misfit dislocations, 81
 mixed control, 200
 mixed dislocation, 56
 MnS inclusions, 396
 Mo₂C carbides, 193, 215
 mobility databases, 445
 mobility of a component, 445
 mole fraction, 92
 monotectic reaction, 108
 monotonic loading, 350
 moving-boundary problem, 176, 445
 multi-objective optimization, 460
 multi-pass hot rolling of steels, 461
 multicomponent diffusion, 160
 multiple order parameters, 447
 multiple slip, 270
 Nabarro-Herring creep, 376
 nanocrystal, 322
 negative edge dislocation, 58
 net diffusion distance, 140
 Ni-base superalloys, 215
 Ni-Cu Monel alloy, 78
 Ni₃Al particles, 215
 nitriding, 390, 428
 no-partitioning local equilibrium, 453
 non-conserved order parameter, 451
 non-glissile interfaces, 200
 non-symmetric distortions, 406
 non-symmetric strain field, 313
 normalizing, 425
 notch toughness, 342
 nucleation, 188
 nucleation and growth transformations, 186
 nucleation of martensite, 243
 nucleation rate, 192
 nucleation site density, 192
 octahedral sites, 19
 Olson and Cohen theory of martensitic nucleation, 243
 Onsager symmetry principle, 160
 order parameter, 228, 446
 order-disorder transformations, 226
 ordered domains, 229
 orientation relationship, 407
 original composition space, 456
 Orowan by-passing, 327
 Orowan hardening, 327
 overaging, 326
 overall transformation rate, 207
 oxidation, 37
 pack carburizing, 427
 Pareto front, 460
 Pareto optimal solutions, 460
 Paris law, 365
 Partial dislocations, 61
 particle shape, 197
 particle size distribution, 214
 partitioning, 116
 partitioning in cementite/carbides, 397
 partitioning in ferrite, 397
 partitioning local equilibrium, 453
 partitioning of alloying elements, 445
 pearlite, 122, 391, 394, 401
 pearlitic nose, 412
 pearlitic rail steels, 429
 Peierls-Nabarro stress, 309
 penetrable particles, 327
 perfect dislocation, 60
 periodicity, 5
 peritectic point, 118
 peritectic reaction, 108
 peritectoid reaction, 108
 persistent slip bands, 362
 phase, 27, 92

- Phase composition, 92
phase diagram, 91, 108
phase transformations, 185
phase-field method, 461
phase-field modeling, 446
Pitch, 407
plane strain condition, 348
plane stress conditions, 348
plastic constraint, 342
plastic deformation, 262
plastic zone, 347
Point defects, 41
polycrystalline aggregates, 270
polycrystalline material, 24
polygonization, 285
polymorphism, 16
Portevin Le Chatelier effect, 317
positive edge dislocation, 58
power-law creep, 370
pre-existing cracks, 338
precipitation, 116, 322
precipitation hardening, 185
precipitation sequence, 199, 323, 423
precipitation strengthening, 307, 322
primary creep, 369
prismatic slip, 266
process annealing, 426
process design, 461
process windows, 456
proeutectic, 117
proeutectoid cementite, 395
proeutectoid ferrite, 395

quench cracks, 416
quench distortions, 416
quenched-in vacancies, 44
quenching, 390
quenching and tempering, 426
quenching medium, 416

random walk, 139
rate-limiting step, 38, 403
recovery, 283, 285, 369, 422
recrystallization, 283, 287, 422, 461
recrystallization fraction, 287

recrystallization kinetics, 287
recrystallization temperature, 283, 287
recrystallization texture, 295
redistribution of carbon, 395
Redlich-Kister polynomials, 443
reference state, 93
regular solution model, 96
reorientation of the crystal, 267
retained austenite, 405
right-handed screw dislocation, 58
rimmed steels, 396
rolling contact fatigue, 350, 429
rolling texture, 295

S-N curves, 355
Scheil solidification model, 455
Scheil-Gulliver model, 455
Schmid factor, 264
screw dislocation, 51
secondary creep, 369
secondary hardening, 423
segregation, 167
self-diffusion, 136, 152
semi-killed steels, 396
semicoherent, 81
semicoherent interfacial energy, 82
sessile dislocations, 275
shape deformation, 232
shape memory effect, 253
shape of the new phase, 197
sharp interfaces, 216
shear deformation, 263
shear lips, 348
Shockley partial dislocations, 62
short-range obstacles, 302
simple slip, 265
simulation domain, 453
single crystal plasticity, 261
single crystals, 263, 270
sliding contact fatigue, 350
slip, 263
slip as a thermally-activated process, 302
slip directions, 26
slip lines, 45
slip plane, 26, 53

- slip system, 46, 265
 small scale yielding, 348
 Snoek effect, 21
 solid solubility, 390
 solid solution strengthening, 306, 310
 solidus, 110
 solidus lines, 112
 solute drag, 293
 solution treatment, 323
 solvus lines, 112
 specific obstacle strength, 302
 spheroidization, 421
 spinodal decomposition, 111, 220, 452
 stabilization of austenite, 396
 stable element reference, 93
 stable equilibrium, 27
 stable hysteresis loop, 351
 stacking fault, 63, 275
 stacking fault energy, 63, 275, 355, 362
 Stacking faults, 76
 stacking sequence, 14
 steady state creep, 369
 stored energy, 283, 461
 strain aging, 316
 strain energy, 26
 strain hardening, 272, 306, 369
 strain-life approach, 355, 358
 strength of martensite, 314
 strengthening heat treatments, 426
 Strengthening mechanisms, 75
 strengthening mechanisms, 301, 330
 stress corrosion cracking, 346
 stress fields of dislocations, 69
 stress intensity factor, 344
 stress relief annealing, 426
 striations, 366
 Structural imperfections, 41
 substitutional diffusion, 45, 135
 substitutional solid solutions, 18
 superlattices, 18
 supersaturated ferrite, 407
 supersaturated solid solution, 322
 supersaturation, 206
 supersaturation ratio, 190
 surface hardening, 427
 surface tension requirement, 289
 surface treatments, 427
 symmetric distortions, 406
 syntectic reaction, 108
 tangent unit vector, 55
 Taylor criterion, 271
 Taylor factor, 272
 temperature dependence of the flow stress, 302
 tempering, 415, 420
 terminal solid solutions, 119
 tertiary creep, 369, 377
 tetragonality factor, 313, 406
 tetrahedral sites, 19
 tetrakaidecahedral grains, 289
 texture, 295
 The dissociation of a perfect dislocation, 61
 thermal activation, 302
 thermal equilibrium, 27
 thermo-mechanical fatigue, 350
 thermochemical treatments, 427
 thermodynamic equilibrium, 27, 34, 91
 thermodynamic factor, 150, 444
 thermoelastic martensitic transformation, 249
 thermoelasticity, 249
 thermomechanical processing, 430
 tilt boundary, 77
 Time-Temperature-Transformation (TTT) diagram, 408
 topological requirement, 289
 total elongation, 279
 Transformation Induced Plasticity, 249, 430
 transgranular ferrite, 399
 transgranular fracture, 338
 transient creep, 369
 transition fatigue life, 358
 transition phases, 324
 TRIP steels, 249, 430
 twinned crystal, 279
 twinning direction, 280

- twinning plane, 280
- twinning shear, 281
- twist boundary, 78
- ultimate tensile strength, 279
- ultrahigh-strength steel, 331
- undercooling, 95
- uniform elongation, 279
- unit cell, 5, 6
- univariant equilibrium, 107
- uphill diffusion, 147, 220
- upper bainite, 407
- upper yield point, 316
- vacancies, 41
- vacancy diffusion, 136
- vacancy discs, 315
- vacancy energy of formation, 42
- vacancy mechanism of diffusion, 45
- velocity of the interface, 206
- vibrational entropy, 34, 43
- virtual leak test, 172
- Volterra cylinders, 49
- volume diffusion, 136
- volume fraction, 330
- volumetric strains, 196
- weight fraction, 92
- white zone, 428
- Whoeler curves, 355
- Widmanstätten ferrite, 399
- work hardening, 272
- workability of steel, 397
- Wulff shape, 77
- X-ray diffraction, 5
- yield criterion, 345
- yield drop, 316
- yield point, 316
- yield strength, 279, 335
- Zener pinning, 293